JVLC

Journal of Visual Language and Computing

Volume 2020, Number 1

DOI: 10.18293/JVLC2020-N1

ISSN: 2573-7147 (online)

Journal preparation, editing and printing are sponsored by KSI Research Inc.

Journal of
Visual Language and Computing

An Open Access Journal published by
KSI Research Inc.

156 Park Square Lane, Pittsburgh, PA 15238 USA

Journal of

Visual Language and Computing

Volume 2020, Number 1

July 2020

Table of Contents

Regular Papers

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

A Framework for Intrusion Detection Targeted at Non-Expert Users

Bernardo Breve[a,*], Stefano Cirillo[a,*] and Vincenzo Deufemia[a,*]

[a] University of Salerno, via Giovanni Paolo II, 132 84084 Fisciano (SA), Italy

ARTICLE INFO

Article History:
Submitted 4.20.2020
Revised 6.21.2020
Second Revision 7.20.2020
Accepted 7.26.2020

Keywords:
Intrusion detection system
Voice assistant
Computer networks
Human-computer interfaces

ABSTRACT

The wide spreading of the Internet leads to the born of a whole interconnected world. Among all these devices, smart voice assistants are gaining particular attention thanks to their ease of use, allowing users to comfortably deploy commands for controlling other devices. The simplicity of use of voice assistants allowed non-expert to interact with complex systems, leading to that category of users with limited knowledge, to interact with s without being perfectly aware of the risks they are exposed to. For example, common network monitoring systems are so useful as they are complex to use for non-expert users. This paper presents a framework for intrusion detection specifically designed to be used by any category of users, using visual interfaces for simplifying the user interaction with the framework, allowing him/her to properly configure and run an Intrusion Detection System (IDS). The implementation of voice assistants as a communication channel will further improve the overall user experience.

1. Introduction

The spread of the Internet in all socio-economic sectors has led to the need of educating people on the use of this tool. The main goal is to create a society able to exploit the power of Internet for improving daily life. The Internet of Things (IoT) has become one of the most important technologies of this century, allowing users to connect each type of object, e.g., kitchen appliances, cars, thermostats, baby monitors, to the internet in order to establish continuous communication between people, processes, and things.

A large number of organizations benefit from the use of these types of devices in their business processes. In sectors as automotive [6, 23], public sector [34] and healthcare [3, 14], IoT has led to a real revolution. The use of intelligent systems that take advantage of IoT devices has improved safety in cars, has speeded up the time for the rescue of a person, or simply increased the productivity of the public administration. However, this has led to the birth of new security issues, since it is necessary to ensure that no one can interfere with their operations. Thus, the field of information security has become vitally important to the safety and economic well-being. The personal information of each person and what is connected to has enormous value and therefore must be preserved. To this end, new secure and safe information systems have been provided, by using firewalls, intrusion detection and prevention systems, authentication, and other hardware and software solutions.

The reasons that may induce an attack to an information system can be grouped into three main categories: access information, alter information, or render a system unusable [4]. These have led to the birth of intrusion detection systems (IDS), which represent tools for monitoring suspicious activities on the network. IDS can be defined as an alarm that monitors the network and reports intrusion to the users. Over the years, a large number of IDSs [8] have been developed, which were later extended through the use of data mining tools [25], data relationships [11, 37, 10], and machine learning approaches [17]. In general, we can consider several desirable characteristics for an IDS. In particular, an IDS should be run continuously without human supervision, and be fault-tolerant and survivable. Moreover, it should impose minimal overhead and be easily adapted to a specific network to observe the anomaly in network traffic.

Although there are a large number of IDSs, one of the main problems is to install and configure them in order to

* Corresponding author
** Principal corresponding author
bbreve@unisa.it (B. Breve); scirillo@unisa.it (S. Cirillo); deufemia@unisa.it (V. Deufemia)
ORCID(s):

DOI reference number: 10.18293/JVLC2020-N1-014

monitor a specific network. In fact, most of the existing IDSs are used by domain experts who are able to carry out complex configurations and installations. However, most of the people subject to these types of attacks do not have skills for configuring these systems. Furthermore, being the IDSs similar to alarms, it is required to customize the devices and the notification methods of these systems. Although several visual languages and visualization techniques have been proposed to support the management of security issues in the context of Web applications [9, 12], it is necessary to use technologies that are familiar to a large part of users.

Voice assistants are increasingly popular and functional, and they have become a routine part of everyday life for many people. Initially, these assistants did not bring big news. But their developers knew they still had a bright future because, like any other technology, voice recognition needed some more time to evolve. In fact, over time, a large number of features have been developed that take advantage of artificial intelligence (AI) and machine learning for allowing users to use complex tools through their voice. For these reasons, in this paper, we propose a new framework that allows non-expert users to install and configure an IDS on the network. In particular, we propose a new modular architecture with an easy-to-use user interface to customize an IDS. Moreover, we propose an innovative module for interacting with Alexa aiming to execute and monitor the status of an IDS via voice commands.

The paper is organized as follows. Section 2 describes recent work concerning IDSs and tools to support non-expert users in the use of systems that require deep domain knowledge. Section 3 provides an overview of the different types of IDSs. Section 4 presents the architecture of the proposed framework by describing the underlying components. Section 5 briefly discusses the most crucial aspects for non-expert users interacting with an IDS and how we are focusing our efforts to meet their needs. Section 6 concludes the paper by presenting our conclusions and future directions.

2. Related Works

The goal of the proposed solution is to allow a large number of users to use IDS systems despite their inexperience. In fact, the recently proposed IDSs do not consider how they can be used by non-expert users. As an example, the IDS proposed in [19] takes advantage of a deep learning approach based on the self-taught learning technique (STL), but the authors explicitly declare that this tool is targeted at network administrators, and not at common users.

One of the most relevant work has been presented in [21]. Here the authors proposed an innovative network IDS to combat increasingly sophisticated network attacks. It takes advantage of a Hidden Naïve Bayes multiclass classifier to create an effective IDS that outperforms one of the most used IDS based on SVM [2]. The goal of both researches were to create efficient tools without considering if non-expert users are capable to use them or not.

In this work we introduce an innovative framework to support non-expert users in the use of different types of IDSs. Through this framework we can increase the user's awareness of what is happening on their network. This topic has been widely discussed by researchers, who have created several tools and user interfaces to increase interaction between users and systems.

Recently, one of the studies that addressed the problem described above is [15]. The authors developed a visual interface for non-expert users, in order to increase awareness of what happens on the network during daily browsing sessions. Indeed, they have shown that most users are unaware of the type of information are exchanged during the browsing sessions and need specific tools to solve this problem. The proposal has been deeply evaluated and analyzed from the point of view of the user experience in [9].

In [7], authors have compared 13 different visualization tools for network analysis aiming to outline their pros and cons. They have used qualitative coding as part of their research design in order to select several metrics to evaluate the advantages and disadvantages of the analyzed tools. Their primary purpose is to increase the security analyst's situational awareness without considering the final users.

In literature, few tools have been proposed to facilitate the use of IDSs by non-expert users. In [29], authors have defined a simplified sound-assistant that mitigates the sound in order to uniquely notify network attacks. In particular, they exploit distinctive sounds for each attack scenario so that the users easily identify the type of attack. The proposed tool could be integrated within network IDSs.

Other research on human-computer interfaces for supporting IDS has focused on bimodal applications, visual and sound, to notify network intrusions. For example, in [26] the authors introduce immersive spatial audio representations of network events that exploit 3D visual representations for interactive auto-stereoscopic.

3. Overview of IDSs

In this section, we provide a general overview of Intrusion Detection Systems (IDSs). The latter can be classified as Network-based IDSs (NIDSs) and Host-based IDSs (HIDSs) [36].

A NIDS is designed to observe the passing traffic on the entire subnet, detecting attacks that involve all the devices on the network [33]. A HIDS, instead, runs on an independent device of the network and monitor the incoming and outcoming packets from the device, looking for the presence of any malicious activity occurring to the system the HIDS is attached to. Alongside these two categories of IDSs, there also exist hybrid solutions, which combine the information provided by both the network and single devices' feedback to develop a complete view over the network system [35].

IDSs can also be classified based on the methodology used for the identification of intrusions. In this case, they can be classified in two main categories: Signature-based detection (SD), Anomaly-based detection (AD) [5, 20].

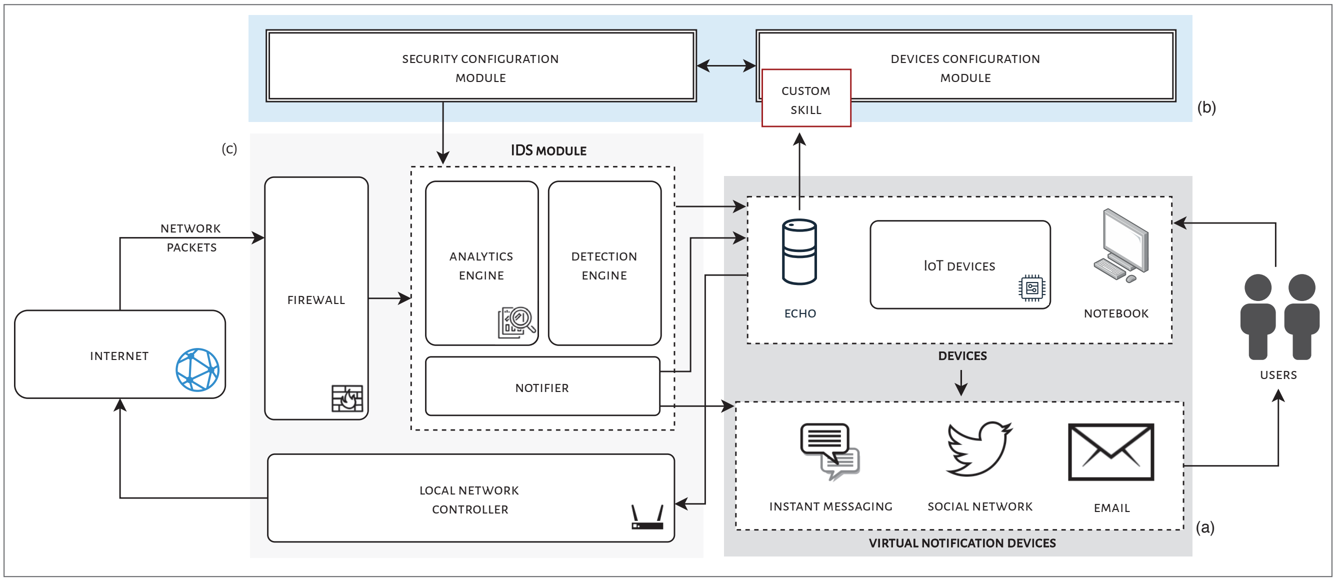


Figure 1: Architecture of the proposed framework.

3.1. Signature-based detection

Signature-based detection systems rely on a set of specific patterns (or strings), called *signatures*, representing the network traffic trend during a certain type of known malicious attack. Following the analysis of the network traffic performed by the system, the extracted data is compared with the stored signatures. When a correspondence is found, this would immediately lead to a report of an attack in progress, and it would provide details about the type of attack and its characteristics. The comparison with the stored signatures can be performed with different techniques such as data mining [31], design patterns [16, 27], or involving both centralized and distributed components [18]. The main advantage of this intrusion detection methodology relies on the simplicity of the identification process, which mainly involves an extrapolation process followed by a comparison [24]. This method is ideal for identifying known attacks and obtaining simultaneously all the details about them. On the contrary, identifying an attack based on a restricted set of patterns limits the number of intrusions that could be recognized. Also, the knowledge base requires to be kept continuously updated, which is often a difficult and time-consuming process [32].

3.2. Anomaly-based detection

The anomaly-based detection methodology relies on the application of machine learning techniques for building a trustful activity model [30]. This methodology looks for any deviation from the known behavior derived from monitoring the system activities over a certain period. Indeed, after the system has been trained long enough to generate a model of what activities, hosts, or even users affect the system, any incoming and outcoming anomaly traffic will be compared and declared dangerous if some of its characteristics cannot be found in the model.

The main advantages are the high dynamicity and extensibility of the model since they capable of identifying new and unforeseen anomalies afflicting the system. On the contrary, its weak point relies on the low accuracy of acquiring attacks' information since the methodology does not use a proper knowledge base, as done by signature-based approaches since it is strictly connected to the information acquired from the observed events. Moreover, the system cannot be operative straight away after its installation, but requires a certain amount of time for training the model, and adapt its analysis in response to the usual behavior of the network (or host) activities.

4. Framework

The proposed framework has been designed to improve the user's awareness of the network traffic and to simplify the IDS configuration process for receiving alerts when an attack is identified. The main idea is to create a modular framework that adapts to the different types of IDSs. This framework allows the user to manage their network through a visual interface and voice commands. The goal is to facilitate the installation and configuration of an IDS while ensuring the correct operations. The architecture of the proposed framework and its phases are described in the following sections.

4.1. Overall framework architecture

Normally, people use different types of devices without knowing what really happens during each usage and what the risks are. Therefore, it is difficult for them to configure network security tools. For these reasons, we have analyzed all the communication phases, starting from the interaction between users and devices to design different components for the architecture of the proposed framework. The architecture involves components designed to ensure high modularity, adaptability, and ease of configuration.

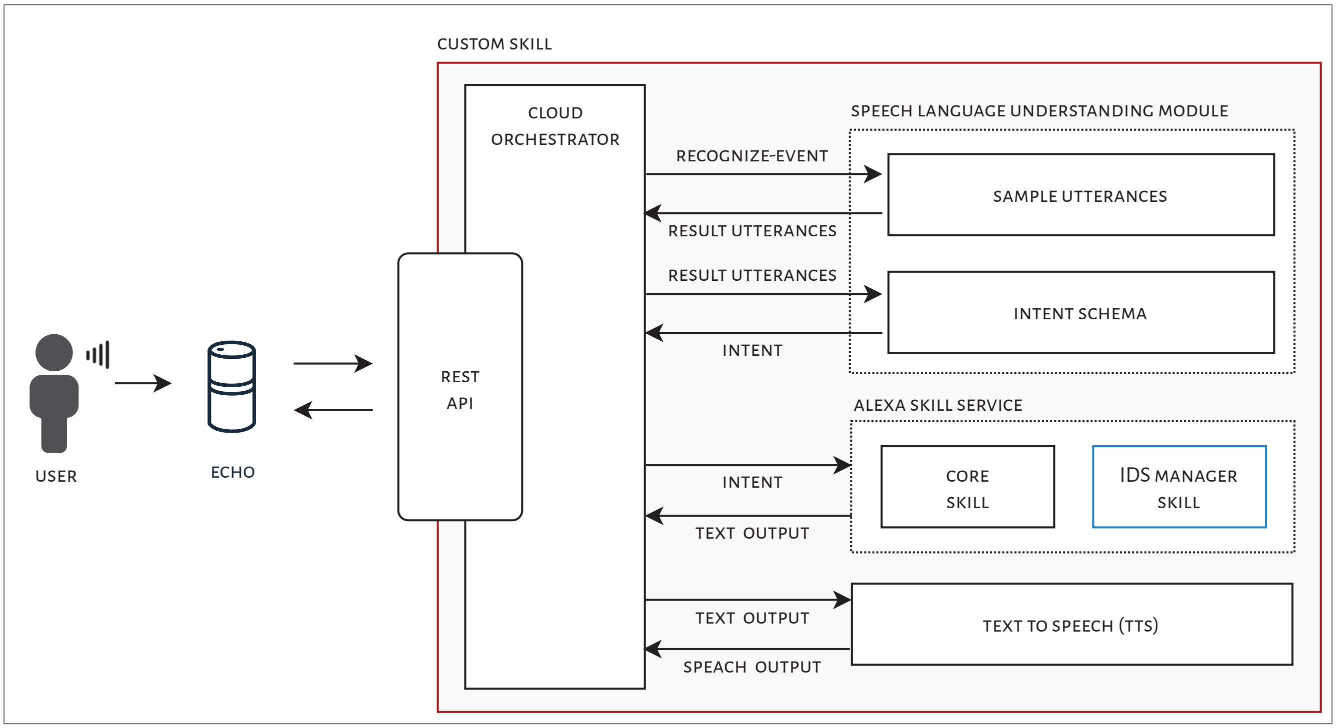


Figure 2: Architecture of the proposed skill.

The first challenge was to define an easily configurable and usable module by any type of device (Figure 1(a)). Thus, we have divided the architecture into three different layers that contain all the main components involving during network packets exchange. More specifically, our purpose is to monitor the outgoing and incoming traffic from the network during the connection between the devices and the local network controller. The first layer (Figure 1(b)) is divided into two distinct modules: devices and security configuration modules. The first is connected to Alexa so that it listens to voice messages and extracts the intents of the users. Through this module, the user defines the IDS configuration parameters. The security configuration module communicates with the third layer (Figure 1(c)). Using the parameters defined by the user, it automatically configures and executes the IDS manager on the network, in order to monitor web traffic. *IDS Manager* is a stand-alone component that can be easily replaced or updated as long as the framework configuration phase is repeated. Moreover, one of the main components of IDS is the *Notifier*. It is one of the main elements of interaction with the user. In fact, the Notifier communicates directly with devices for sending reports of attacks to physical and virtual devices. The interaction modules with instant messaging apps and some of the most well-known social networks will be integrated into the framework. Users will be able to customize the notification devices by using the device configuration module. The framework allows each user, with experience or not, to configure an IDS for their network and customize any device for receiving any alert.

4.2. Alexa custom skill architecture

One of the main proposals concerning this paper relies on the usage of voice assistant capabilities to ease the overall user experience with the system. Interaction with domestic voice assistants has been gaining prominence in the last period, becoming a very useful and simple communication channel [28]. Voice assistants, such as Google Home or Alexa, provide SDKs for the implementation of customized functionalities allowing for the definition of both the interaction model with the user and the logic to deploy the commands on other devices. For this reason, we are planning the implementation of a customized functionality for the Amazon Alexa voice assistant, called "skill" [1], through which the user can vocally interact with the IDS modules.

Figure 2 shows in detail how the vocal requests of the users are transformed into the corresponding commands that are deployed to the IDS modules. In particular, the user can launch the skill by pronouncing its name preceded by keywords like: “Alexa run” or “Alexa start”. This starts the skill and enables the process of communication between the user and the framework through the voice assistant. Any pronounced command deployed by the user to the skill is received and passed through the API at the cloud orchestrator. It has the goal of communicating and synchronizing the actions of all the other modules in the Amazon cloud. The first involved module is the Speech Language Understanding (SLU) whose task is trying to match the specific request with the action. Indeed, all the actions a skill can execute, called *intents*, are associated with several utterances the user can pronounce to trigger that intent. When a match is found, the corresponding intent is passed back to the cloud orches-

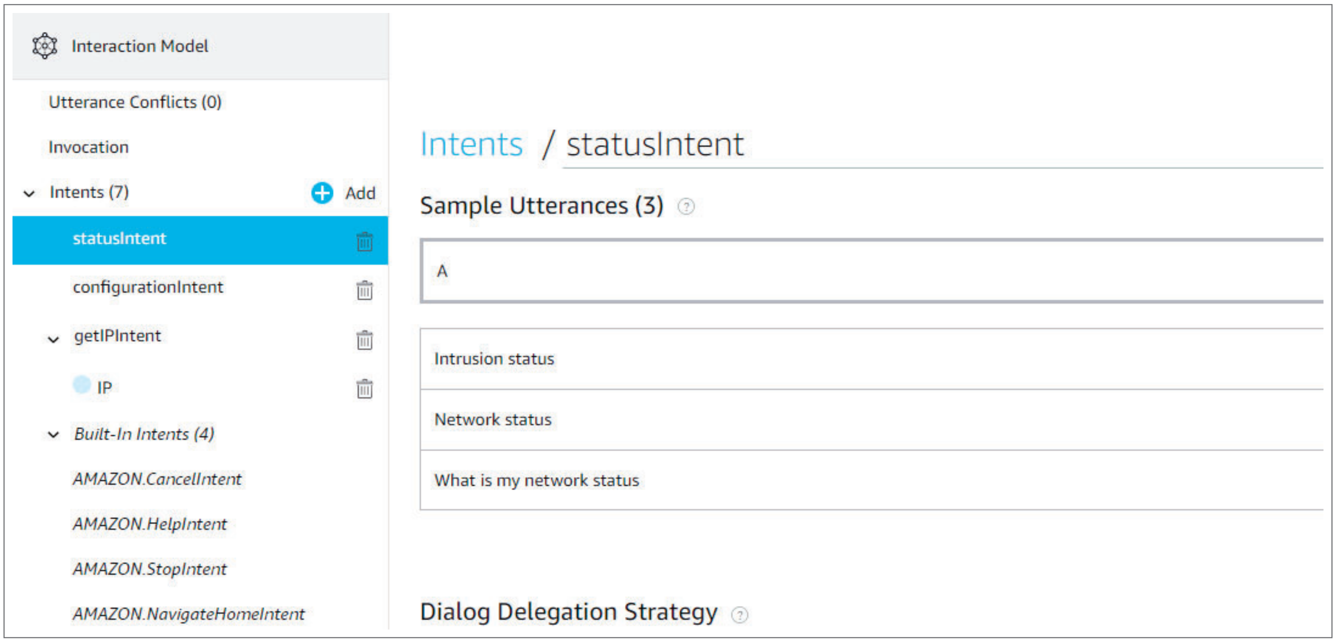


Figure 3: Interaction model of the custom skill

trator, which asks the Alexa Skill Service to perform that intent. The custom skill we are planning to design will at this point contact the IDS to fulfill the user's request. After that, the system will return back to the skill with responses like the system status and notification about any intrusion occurring. In the last phase, the response received from the system is sent from the cloud orchestrator to the Text to Speech (TTS) module, which is responsible to translate the textual content into the voice that will be played by the Alexa device.

Technically, the implementation process of a custom Alexa skill relies on two main components:

- the front-end (or *Interaction Model*);
- the back-end handled through an *AWS Lambda function*.

4.2.1. Interaction model

The Interaction model allows for the definition of the different Intents that the skill should allow. For each intent, a list of sample utterances needs to be defined, which will help the Alexa Skill Service to associate a user vocal command to the corresponding action.

Figure 3 shows a prototype of the interaction model we're designing for the Alexa custom skill. On the left-hand side, a list of all the intent we defined is proposed. In particular, the figure shows a total of seven intents, three of which are the intents we added, and the remaining ones are the so called *Built-In Intents*, which are added by the Alexa Skill Service by default. These intents are inserted for providing the basic functionalities such as the cancellation of the previous spelled token, the possibility to ask Alexa for help about the available commands for the skill, and the capability to both to exit the skill or asking Alexa to go back to the homepage for the skill, i.e., the first interaction process happening between the user and the skill. However, the utterances for enabling the intents and the logic for executing such procedures need to be specified by the developer. By inserting these intents by default and especially making them not removable, the Alexa Skill Services ensures that the developer provides these basic functionalities that are considered mandatory.

The three custom intents we defined so far allow users to query the current state of the network environment and begin the process of configuring the IDS. The *statusIntent*, whose details are shown in Figure 3, handle all the requests coming from the user concerning the status of the network environment, specifically, if any type of intrusions has been detected in the recent period. For this intent, we provided some basic utterances such as "What is my network status" or simply "Network status". Luckily, the AI-driven Natural Language Processing module offered by Alexa, expands autonomously the set of possible utterances said by the user, so that we do not have to define specifically all the utterances the user could say for triggering the intent.

The *configurationIntent* is responsible for starting the configuration process. After the user pronounces the command "start configuration" the back-end handler of the intent sends the request to the *Device Configuration Module*, which replies with the parameters required to correctly configure the IDS. Among the parameters, there is the IP address of the machine on which the IDS is installed. The acquisition process of this parameter is handled by the *getIPIntent*. Figure 4 shows how we designed the utterance for correctly interpreting the IP address said by the user. To this end, we applied another useful feature offered by the Alexa Development Skill Kit: the intent slots. Since we do not know in advance what numbers the user will provide as IP address, we use the intent slot to define a pattern of four groups of numbers separated by the

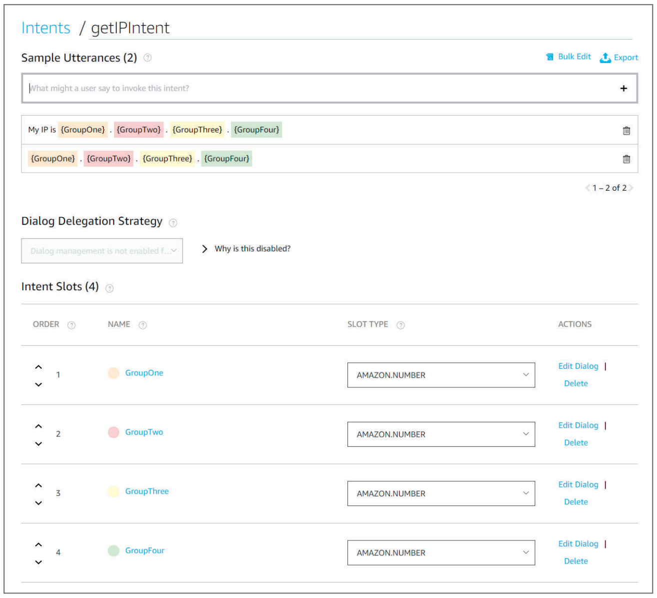


Figure 4: The interaction model of getIPIntent

```
sb = SkillBuilder()

sb.add_request_handler(LaunchRequestHandler())
sb.add_request_handler(StatusIntentHandler())
sb.add_request_handler(StartConfigurationHandler())
sb.add_request_handler(GetIPHandler())
sb.add_request_handler(HelpIntentHandler())
sb.add_request_handler(CancelOrStopIntentHandler())
sb.add_request_handler(SessionEndedRequestHandler())
sb.add_request_handler(IntentReflectorHandler())

sb.add_exception_handler(CatchAllExceptionHandler())

lambda_handler = sb.lambda_handler()
```

Figure 5: The list of handlers in response to user's inputs

dot sign. Each group of digits can be later associated in the back-end with four different variables, which are then combined to produce an IP address string (e.g., 172.16.254.12) that is sent to the Device Configuration Module. In order to avoid abnormal values, each group of digits has been provided with a *validator*, which sets boundaries of expected values for each intent slot. In our case, we expect each group of digits to have a value between 0 and 255.

4.2.2. AWS lambda function

We are implementing the back-end side of our skill through the AWS Lambda Function offered by the Amazon AWS Cloud service. The whole function is written in Python and is responsible for providing the computation logic in response to the user's commands. Figure 5 shows the list of the handlers we are working on at the time this paper is written. All the handlers define a class composed of two crucial methods: *can_handle* and *handle*. The former is always called at any time an input is received by the interaction model, and has the task of verifying if that particular handler is the one in charge of managing the user's request. Instead, the latter is the method that performs the computational operations and formats the response string that will be said by Alexa.

```
class LaunchRequestHandler(AbstractRequestHandler):
    """Handler for Skill Launch."""
    def can_handle(self, handler_input):
        # type: (HandlerInput) -> bool

        return ask_utils.is_request_type("LaunchRequest")(handler_input)

    def handle(self, handler_input):

        speak_output = "Welcome, to the network monitoring skill! Say \"Start configuration\" to start"

        return (
            handler_input.response_builder
                .speak(speak_output)
                .ask(speak_output)
                .response
        )
```

Figure 6: An example of Intent Handler

Figure 6 shows an example of request handler. In particular, this handler is the one responding to the launch command of the skill, i.e., the user asking "Alexa, start Network Monitoring". The *can_handle* method do not apply any specific verification other than ascertaining that the intent involved in the command is *LaunchRequest*. This type of verification is automatically performed by the Alexa Skill API thanks to the method *is_request_type*, which returns true if the matching between the user's command and the sample utterances is positive. The *handle* method, in this case only format the response message that has to be presented to the user. Moreover, other than only welcoming the user, the method also asks the user to pronounce the utterance "start communication" which we will be intercepted by the aforesaid *configurationIntent*. The *.ask* method is responsible for keeping the Alexa on listen, waiting for the next utterance to be pronounced. Whether that method is removed, this will force Alexa to stop any listening process until the wake word, i.e., "Alexa", is pronounced by the user.

4.3. Intrusion Detection System architecture

The intrusion detection system architecture will also rely on two-layer architecture comprehensive of a front-end for communicating results about the monitor process to both services and Alexa, and a back-end for monitoring the network and detecting possible attack pattern.

4.3.1. Front-end technologies

The graphic user interface is based on the bootstrap library in order to provide a fluid interaction, a modern look, and also the possibility to have a responsive interface capable of adapting the different components according to the device where the interface is displayed, such as a smartphone, a tablet, or a personal computer.

Another important aspect that needs to be addressed by the front-end layer is the social account setup. Indeed, the user may want to receive feedback about the changes of status in the network with a message on Telegram, or by publishing a tweet through his Twitter account. In order to achieve this, the user has to grant authorization for allowing the IDS to automatically operate on the user's accounts. The process of displaying authorization messages is required to be performed through proprietary interfaces, which also sometimes are provided with *security captcha* for verifying that the one who is granting authorization is an actual human being. Unfortunately, this process cannot be performed through the vocal interaction of the Alexa skill.

4.3.2. Back-end technologies

The back-end of the IDS defines the logic for detecting specific anomalies in the network environment, identifying the attack patterns, and possibly obtain details about the attacker/s.

The algorithm for monitoring the network offers the detection of common attacks based on probing the network environment. In these types of attacks, the attacker sends modified packets to the user network monitoring the automatic response the system provides; by doing so the attacker has the possibility to quickly verify whether there is a possible flaw s/he can exploit for penetrating the system. For example, a *FIN scan* is a type of attack characterized by sending anomalous TCP packets to the victim's ports, having only the FIN flag active. By technical specification, the network that receives a packet with only FIN flag active on a closed, secure port, has to response with another packet. On the contrary, if this packet is sent on an open and so vulnerable port, the network has to ignore the received packet. Thus, by monitoring whether a contacted host response to the anomalous packet or not, the attackers can understand if there are any flaws in his/her victim's network. Other than detecting a *FIN scan* attacks, the IDS also provides detection over *TCP ACK&Window*, *UDP*, *ICMP*, *XMAS* and *NULL* scans attacks.

The IDS provides also detection over Denial of Service (DoS) attacks, which represent a particularly delicate issue, especially in the IoT domain [22]. DoS attacks aim at exhausting the resources of a computer system, this can be achieved by saturating with packets the network environment in such a rate that the system is not able to process all incoming requests, causing it to fail. These attacks are particularly popular in the IoT domain, since the low computational power of IoT device, makes them prone to fail under a conspicuous amount of requests. Other than detecting DoS attacks afflicting the network, the IDS is also capable of recognizing some other type of attacks aiming at making the network fail, or even penetrating the system by granting unauthorized access, e.g., *DHCP Exhaustion*, *Man in The Middle* (MiTM), *SYN flood attack*, *Fake access point*, *Ping of Death attack*.

Finally, the IDS also provides the possibility of obtaining partial information about who performed the attack. Indeed, if the system achieves to obtain the IP (or the IPs) of the attackers, it is possible to exploit external services which are able to geolocalize the position of an IP address all around the world, together with other information, such as: is that IP associated with a website domain? What host has to sell the domain? Who registered the domain? It is worth mentioning, however, that really often attackers make use of tools and techniques for hiding their true IP address by, for example, making sure that a request is passed through a server located in a remote area, i.e., a proxy, before attacking the victim. Hence, the IP the victim intercepts is the IP address of the server and it becomes very difficult to go back to the original sender, especially if the proxy step is repeated several times.

5. Discussion

In this section, we will go through some of the most crucial and difficult aspects that a non-expert user needs to tackle down to correctly set up an IDS. We will also provide a general discussion about the solutions we are planning to implement for making all these steps possible.

We identified four main phases required for correctly using an IDS and we will walk through each of them describing our contribution plans.

5.1. Installation

The installation process is the first step a non-expert user needs to face when approaching an IDS. The main problem with this phase relies on the conspicuous amount of prerequisites the user has to deal with before actually proceed with the installation. Furthermore, most of the commands need to be deployed through a console command line, which represents an uncomfortable tool to interact with. To ease this issue, we are planning to include all the installation steps in an installer, a GUI-based software commonly found in the Microsoft Windows OS domain. An installer is composed of several windows describing the necessary steps to pursue the installation of the software.

Hence, through a minimal interface user will have the possibility to specify the paths where all the required files will be saved and granting the mandatory authorizations for the correct execution of the IDS.

5.2. Configuration

After the installation phase, another crucial step is configuring the IDS. Indeed, it is necessary for the user to provide some essential parameters to obtain a correct network traffic monitoring together with the identification of intrusions. For example, it is fundamental for the user to provide the name of his/her network interface, i.e. the physical inbound and outbound connection port connecting the computer on which the IDS has been installed to the router and so the Internet.

To ease this type of process, we have planned to rely on a specifically designed visual interface, which will implement several visual metaphors designed to be suitable for the knowledge level of the non-expert users approaching it. The introduction of this new level of abstraction will help the users to complete a correct system configuration without getting lost into the details of technical terms.

5.3. Usage

Being a monitoring system whose main task is to silently monitor and evaluate in the background the quality and type of network packets being exchanged, the active contribution by the non-expert users is reduced to the necessity of starting the IDS. However, this operation needs to be performed through a command launched from bash, which as mentioned above represents a particularly complicated step for inexperienced users. For this reason, we have planned the introduction of a series of automatism allowing the system to start without the user having to forcefully act on the system.

Alongside this choice, a useful alternative in this scenario would have been provided by the interaction capabilities offered by voice assistants. The Alexa custom skill we described earlier would allow the user to easily interact with the whole system, requesting to start the IDS and asking for information regarding its state of running.

5.4. Notification

Finally, the last step is the one that involves how to notify the user of the presence of an intrusion within the system. Even at this juncture, the use of a voice assistant providing immediate notification of the system's security status seems to be a suitable choice with respect to the knowledge level of the non-expert users. For this implementation phase, the challenge will be to program the type of message that the voice assistant will have to pronounce, avoiding phrases that can mislead the user and make the seriousness of the danger unclear. For example, the use of a phrase such as: "The system is under DDoS attack" is totally incomprehensible to a user who is unable to understand the seriousness of the danger of a DDoS attack which, in the IoT context, was the cause of one of the most devastating hacker attacks, the Mirai Botnet [22].

Therefore, it will be essential to find the right formulation to prevent the user from underestimating (or overestimating) the severity of the intrusion.

5.5. User involvement

Being a framework specifically designed for end-users, the overall involvement of them in the realization and testing phases plays a fundamental role in the achievement of a simple, functional, and effective system. For this reason, the development of the user interface will see the collaboration of some users, to whom we will submit some surveys to test their preferences. By doing so it is possible to direct the system towards the development of an interface more akin to user needs.

Another important phase will be the evaluation of the quality of the user experience. Thus, this type of evaluation will be planned with the involvement of a large group of users, which we'll seek among who has little or no knowledge of computer technologies. Moreover, we will ask them to fill in different surveys in order to evaluate the usability and effectiveness of the framework.

6. Conclusion

Intrusion detection systems (IDS) have been defined as an essential security measure in any type of network. They are an important component that permits to identify network attacks by analyzing network traffic. This paper presents a framework that allows non-expert users to monitor their network and identify any attacks. In particular, we have defined two different modules connected to the main components of an IDS. Through this approach, it is possible to adapt our framework in different IDS systems. Moreover, an innovative skill for Alexa has been proposed, in order to allow users to run the IDS through voice commands.

In the future, we would like to continue implementing the framework, integrating it with different IDSs. Moreover, we intend to perform supervised tests by involving people with different qualifications and knowledge skills. Each test will consist of several configuration tasks that allow the users to simulate the installation phases of a custom IDS through the proposed framework. The tests will focus on highlighting the difficulties encountered by each user during the interaction with the framework. At the beginning of each test, users will have to fill a background survey in order to state the prior domain knowledge. After completing the supervised test, we will provide users with a final evaluation survey which will allow us to verify the effectiveness of our framework.

Finally, we plan to extend the approaches proposed to capture the user navigation intents for improving the intent understanding task [13].

References

[1] Amazon, 2020. Alexa Skill Kit. https://developer.amazon.com/it-IT/docs/alexa/sdk/alexa-skills-kit-sdks.html/. [Online; accessed 10-April-2020].

[2] Ambwani, T., 2003. Multi class support vector machine implementation to intrusion detection, in: Proceedings of the International Joint Conference on Neural Networks, 2003, IEEE. pp. 2300–2305.

[3] Amendola, S., Lodato, R., Manzari, S., Occhiuzzi, C., Marrocco, G., 2014. RFID technology for IoT-based personal healthcare in smart spaces. IEEE Internet of things journal 1, 144–152.

[4] Anderson, J.P., 1980. Computer security threat monitoring and surveillance. Technical Report, James P. Anderson Company .

[5] Anjum, F., Subhadrabandhu, D., Sarkar, S., 2003. Signature based intrusion detection for wireless ad-hoc networks: A comparative study of various routing protocols, in: Proceedings of 2003 IEEE 58th Vehicular Technology Conference, IEEE. pp. 2152–2156.

[6] Aris, I.B., Sahbusdin, R.K.Z., Amin, A.F.M., 2015. Impacts of IoT and big data to automotive industry, in: Proceedings of 2015 10th Asian Control Conference (ASCC), IEEE. pp. 1–5.

[7] Attipoe, A.E., Yan, J., Turner, C., Richards, D., 2016. Visualization tools for network security. Electronic Imaging 2016, 1–8.

[8] Axelsson, S., 2000. Intrusion detection systems: A survey and taxonomy. Technical Report. Chalmers University of Technology.

[9] Breve, B., Caruccio, L., Cirillo, S., Desiato, D., Deufemia, V., Polese, G., 2020. Enhancing user awareness during internet browsing, in: Proceedings of the Fourth Italian Conference on Cyber Security, CEUR Workshop Proceedings 2597. pp. 71–81.

[10] Caruccio, L., Cirillo, S., . Incremental discovery of imprecise functional dependencies. Journal of Data and Information Quality (JDIQ) .

[11] Caruccio, L., Cirillo, S., Deufemia, V., Polese, G., 2019. Incremental discovery of functional dependencies with a bit-vector algorithm, in: Mecella, M., Amato, G., Gennaro, C. (Eds.), Proceedings of the 27th Italian Symposium on Advanced Database Systems, CEUR-WS.org. pp. 1–12.

[12] Caruccio, L., Deufemia, V., D'Souza, C., Ginige, A., Polese, G., 2015a. A tool supporting end-user development of access control in

web applications. International Journal of Software Engineering and Knowledge Engineering 25, 307–331.

[13] Caruccio, L., Deufemia, V., Polese, G., 2015b. Understanding user intent on the web through interaction mining. J. Vis. Lang. Comput. 31, 230–236.

[14] Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M.L., Tarricone, L., 2015. An IoT-aware architecture for smart healthcare systems. IEEE Internet of Things Journal 2, 515–526.

[15] Cirillo, S., Desiato, D., Breve, B., 2019. Chravat-chronology awareness visual analytic tool, in: Proceedings of 23rd International Conference Information Visualisation (IV), IEEE. pp. 255–260.

[16] De Lucia, A., Deufemia, V., Gravino, C., Risi, M., 2018. Detecting the behavior of design patterns through model checking and dynamic analysis. ACM Trans. Softw. Eng. Methodol. 26, 13:1–13:41.

[17] Haq, N.F., Onik, A.R., Hridoy, M.A.K., Rafni, M., Shah, F.M., Farid, D.M., 2015. Application of machine learning approaches in intrusion detection system: a survey. International Journal of Advanced Research in Artificial Intelligence 4, 9–18.

[18] Ioulianou, P., Vasilakis, V., Moscholios, I., Logothetis, M., 2018. A signature-based intrusion detection system for the internet of things, in: Proceedings of IEICE Information and Communication Technology Form, pp. 1–6.

[19] Javaid, A., Niyaz, Q., Sun, W., Alam, M., 2016. A deep learning approach for network intrusion detection system, in: Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), pp. 21–26.

[20] Jyothsna, V., Prasad, V.R., Prasad, K.M., 2011. A review of anomaly based intrusion detection systems. International Journal of Computer Applications 28, 26–35.

[21] Koc, L., Mazzuchi, T.A., Sarkani, S., 2012. A network intrusion detection system based on a hidden naïve bayes multiclass classifier. Expert Systems with Applications 39, 13492–13500.

[22] Kolias, C., Kambourakis, G., Stavrou, A., Voas, J., 2017. DDoS in the IoT: Mirai and other botnets. Computer 50, 80–84.

[23] Krasniqi, X., Hajrizi, E., 2016. Use of IoT technology to drive the automotive industry from connected to full autonomous vehicles. IFAC-PapersOnLine 49, 269–274.

[24] Liao, H.J., Lin, C.H.R., Lin, Y.C., Tung, K.Y., 2013. Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications 36, 16–24.

[25] Nadiammai, G., Hemalatha, M., 2014. Effective approach toward intrusion detection system using data mining techniques. Egyptian Informatics Journal 15, 37–50.

[26] Papadopoulos, C., Kyriakakis, C., Sawchuk, A., He, X., 2004. Cyberseer: 3D audio-visual immersion for network security and management, in: Proceedings of ACM Workshop on Visualization and Data Mining for Computer Security, pp. 90–98.

[27] Patil, R.M., Patil, M.R., Ramakrishnan, K.V., Manjunath, T., 2010. Iddp: Novel development of an intrusion detection system through design patterns. International Journal of Computer Applications 7, 22–29.

[28] Polyakov, E., Mazhanov, M., Rolich, A., Voskov, L., Kachalova, M., Polyakov, S., 2018. Investigation and development of the intelligent voice assistant for the internet of things using machine learning, in: Proceedings of Moscow Workshop on Electronic and Networking Technologies (MWENT), IEEE. pp. 1–5.

[29] Qi, L., Martin, M.V., Kapralos, B., Green, M., García-Ruiz, M., 2007. Toward sound-assisted intrusion detection systems, in: Proceedings of OTM Confederated International Conferences On the Move to Meaningful Internet Systems, Springer. pp. 1634–1645.

[30] Sangkatsanee, P., Wattanapongsakorn, N., Charnsripinyo, C., 2011. Practical real-time intrusion detection using machine learning approaches. Computer Communications 34, 2227–2235.

[31] Singh, V., Puthran, S., 2016. Intrusion detection system using data mining a review, in: Proceedings of 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), pp. 587–592.

[32] Uddin, M., Rahman, A.A., 2010. Dynamic multi layer signature based intrusion detection system using mobile agents. International Journal of Network Security & Its Applications (IJNSA), Vol.2, No.4 .

[33] Vigna, G., Kemmerer, R.A., 1999. Netstat: A network-based intrusion detection system. Journal of Computer Security 7, 37–71.

[34] Wirtz, B.W., Weyerer, J.C., Schichtel, F.T., 2019. An integrative public IoT framework for smart government. Government Information Quarterly 36, 333–345.

[35] Zaraska, K., 2003. Prelude IDS: current state and development perspectives. URL http://www.prelude-ids.org/download/misc/pingwinaria/2003/paper.pdf .

[36] Zarpelão, B.B., Miani, R.S., Kawakani, C.T., de Alvarenga, S.C., 2017. A survey of intrusion detection in internet of things. Journal of Network and Computer Applications 84, 25–37.

[37] Zuo, Y., Panda, B., 2004. Fuzzy dependency and its applications in damage assessment and recovery, in: Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004., IEEE. pp. 350–357.

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

Self-training algorithm combining density peak and cut edge weight

Yang Liu[a,*]

[a]Mathematics and Statistics, China, and Chongqing University, Chongqing

ARTICLE INFO

Article History:
Submitted 3.23.2020
Revised 6.19.2020
Second Revision 7.07.2020
Accepted 7.08.2020

Keywords:
self-training
Density
cut edge weight
hypothesis testing

ABSTRACT

In view of the influence of mislabeled samples on the performance of self-training algorithm in the process of iteration, a self-training algorithm based on density peak and cut edge weight is proposed. Firstly, the representative unlabeled samples are selected for labels prediction by space structure, which is discovered by clustering method based on density of data. Secondly, cut edge weight is used as statistics to make hypothesis testing. This technique is for identifying whether samples are labeled correctly. And then the set of labeled data is gradually enlarged until all unlabeled samples are labeled. The proposed method not only makes full use of space structure information, but also solves the problem that some data may be classified incorrectly. Thus, the classification accuracy of algorithm is improved in a great measure. Extensive experiments on real datasets clearly illustrate the effectiveness of proposed method.

1. Introduction

Data classification is a very active research direction in the field of machine learning. In order to train an effective classifier, traditional supervised classification methods often require a large number of labeled samples. However, in practical applications, the acquisition of labeled samples requires a large price and is not easy to obtain, and the acquisition of unlabeled samples is relatively easy. Therefore, when the number of labeled samples is small, supervised classification methods are difficult to train an effective classifier[5,9].In this case, the semi-supervised classification method, which requires only a small number of labeled samples and makes full use of a large number of unlabeled samples, has attracted more and more attention. [4,7] Self-training is one of the commonly used methods in semi-supervised classification. First, an initial classifier is trained with a small number of labeled samples, and the unlabeled samples are classified. Then, select unlabeled samples with higher confidence and their predicted labels, expand the labeled sample set, and update the classifier. These two processes continue to iterate until the algorithm converges.

[1,3,7]Self-training methods do not require any specific assumptions, are simple and effective, and have been widely used in many fields such as text classification, face recognition, biomedicine, and so on. But self-training classification algorithms also have some drawbacks, such as the classification performance is limited by the size of the initial labeled data set and their distribution across the entire data set. Aiming at the shortcomings of the self-training method, [11] Considering the spatial distribution of the data set, a semi-supervised fuzzy c-means clustering method is proposed to optimize the self-training algorithm (ST-FCM). This method integrates the semi-supervised clustering technology as an auxiliary strategy into the self-training process. The semi-supervised clustering technology can effectively mine the internal data spatial structure information contained in the unlabeled samples and better train the classifier. However, the fuzzy c-means clustering method cannot find the spatial structure of non-Gaussian distributed data sets well. [2,5,8] proposed Self-training based on density peak of data (ST-DP). In the ST-DP algorithm, the spatial structure of the data is found using density peak clustering. Although the method based on density peak clustering can make effective use of the spatial structure of various data distributions, the ST-DP classification of some datasets with more overlapping samples after visualization Ineffective. Subsequently, [11,14] used Differential evolution (DE) to improve

DOI reference number: 10.18293/JVLC2020-N1-004

the self-training algorithm, and proposed a self-training algorithm based on differential evolution (ST-DE). [15] This method uses DE algorithm to optimize the newly added labeled samples during self-training. Although the ST-DE algorithm solves the problem of overlapping samples, the optimization algorithm brings too many complicated operations to a certain extent. This method does not fundamentally solve the shortcomings of the ST-DP algorithm. The main reason is that in the self-training labeling process, those overlapping samples after visualization are extremely easy to be labeled. The ST-DP algorithm uses these mislabeled samples directly for subsequent iterative labeling, which ultimately reduces the performance of the trained classifier.

Based on the ST-DP algorithm, this paper proposes a Self-training method based on density peak and cut edge weight (ST-DP-CEW). This method not only selects unlabeled samples, uses the density clustering-based method to discover the underlying spatial structure of the data set, and selects representative samples for label prediction. Further, the correctness of the predicted labels can be identified by using the statistical method of cutting edge weights. Cutting edge weights and density peak clustering make full use of the sample spatial structure and unlabeled sample information, solve the problem of some samples being labeled incorrectly, reduce the accumulation of errors during iteration, and can effectively improve the performance of the classifier.

.

2. Algorithm construction

In this paper, we improve the classification accuracy of the self-trained semi-supervised classification algorithm by starting with the wrongly labeled samples during the self-training process. Based on ST-DP, the ST-DP-CEW algorithm is proposed. First, the spatial structure of the data set is discovered by density clustering method, and representative samples can be preferentially selected for label prediction during each iteration. Then, we use the statistical method of cutting edge weights to judge whether the samples are correctly labeled, and update the labeled set with the correctly labeled samples. The above process is iterated until all unlabeled samples are completely labeled.

1.Spatial structure of data

Clustering is a typical unsupervised learning method. The process of clustering can discover the spatial structure of data. The method based on density clustering can find the spatial structure of non-Gaussian distributed data sets and can automatically determine the number of clusters.

In this paper, let $L=\{(x_i, y_i)\}$ be the labeled sample set, where x_i is the training sample, and y_i is its label. $y_{i1} \in \{\omega_1, \omega_2, \cdots, \omega_s\}$, $i=1,2,\cdots,m$. S is the number of categories. $U=\{x_{m+1}, x_{m+2}, \cdots, x_n)$ is the unlabeled sample set. The local density of sample x_i is defined as follows:

$$\rho_i = \sum \chi\left(d_{ij} - d_c\right)$$

Among them:

$$\chi(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases}$$

d_{ij} is the Euclidean distance between samples x_i and x_i , and d_c is called the truncation distance. It is a constant that has no fixed value and is related to the data set itself(Wang & Xu, 2017). After calculating the ρ_i value of each sample x_i , find the sample x_j that is closest to sample x_i and has a greater local density, point x_i to x_j , and find the spatial structure of the data set.

2. Statistical method of cutting edge weights

[7]Trim weighting is a method to identify and process mislabeled samples. First, in order to illustrate the similarity of the samples, a relative adjacency graph is established on the data set. The two samples x_i and x_j are connected side by side, if the following conditions are met: $d(x_i, x_j) \leq \max\left(d(x_i, x_m), d(x_j, x_m)\right), \forall m \neq i, j$,Where $d(x_i, x_j)$ is the distance between samples x_i and x_j . In an adjacency graph, if two samples with edges connected by different labels, this edge is called a cut edge. In an adjacency graph, if two samples with edges connected by different labels, this edge is called a cut edge. If x_i has many cut edges, that is, most of the samples in the neighborhood have labels that are different from those of x_i , it is considered that it may be labeled incorrectly. Therefore, cut edges play an important role in identifying mislabeled samples. For different samples, they may have the same number of cutting edges, but the importance of each cutting edge is different, so each edge in the adjacent graph is given a weight. Let w_{ij} be the weight of the edges connecting samples x_i and x_j .

. Finally, the hypothesis test was used to identify whether sample x_i was labeled incorrectly. The sum of the trimming weights J_i of sample x_i is defined as follows:

$$J_i = \sum_{j=1}^{n_i} w_{ij} I_i(j)$$

Among them,

$$I_i(j) = \begin{cases} 1, & y_i \neq y_j \\ 0, & y_i = y_j \end{cases}$$

$\mathbf{n}_i$ is the number of samples with edges connected to sample x_i , and y_i is the label of sample x_i . If the J_i value of the sample x_i to be tested is large, it is considered that the sample may be labeled incorrectly. For hypothesis testing, the null hypothesis is defined as follows:

H_0 : All samples in the adjacent graph are labeled independently of each other according to the same probability distribution pro_y . pro_y represents the probability that the sample label is y .

In order to do a bilateral test, you must first analyze the distribution of J_i under H_0 . Under the null hypothesis, $I_i(j)$ is an independent identically distributed random variable subject to a Boolean parameter of $1-pro_{y_i}$. So the expected μ_0 and variance σ^2 of J_i under H_0 are:

$$\mu_0 = \left(1-\mathrm{pro}_{y_i}\right)\sum_{j=1}^{n_i} w_{ij}$$

$$\sigma^2 = \mathrm{pro}_{y_i}\left(1-\mathrm{pro}_{y_i}\right)\sum_{j=1}^{n_i} w_{ij}^2$$

J_i follows the normal distribution $J_i \sim N\left(\mu_0, \sigma^2\right)$ under the original hypothesis H_0 , so the selected test statistic is

$$u = \frac{J_i - \mu_0}{\sigma}$$

Given a significance level of α , the rejection domain is:

$$W = \{|u| \geq u_{1-\alpha/2}\}$$

The rejection domain that gets the sum of the trimming weights is

$$W = \left[-\infty, \mu_0 - \sigma \cdot u_{1-\alpha/2}\right] \cup \left[\mu_0 + \sigma \cdot u_{1-\alpha/2}, +\infty\right]$$

For sample x_i to be tested, if the value of J_i is significantly lower than the expected value under H_0 , that is, the rejection field on the left, the sample is marked correctly, otherwise it may be marked incorrectly. The main steps of the algorithm for identifying wrongly labeled samples using the edge-cut weights statistical method are as follows:

Step1. Create a relative adjacency graph for the sample set, and initialize the correctly labeled sample set $T=\{\varnothing\}$ and the incorrectly labeled sample set $T'=\{\varnothing\}$.

Step2. To assign weights to each edge in the adjacent graph, calculate the cut-edge weights of each sample and the expected and variance under the original hypothesis.

Step3. Given the significance level, calculate the rejection domain.

Step4. If the value of J_i is in the rejection field on the left, the label is correct, and the correct label set is updated; if it is not in the rejection field on the left, look at its neighbor samples. If the neighbor samples are all within T, then relabel with most label markers Otherwise, x_i mark errors, update the error mark set.

Step5. Repeat the above steps until all samples are tested.

3. Weight selection

The weight of each edge plays an important role in the statistical method of the edge weight. In this paper, the weight is first used to normalize the other nearest neighbor distances in the neighborhood by using the maximum nearest neighbor distance of each sample. Then calculate the probability that the sample has the same label as each neighboring sample, which is the weight of the edge.

Let sample set $\{x|_{i,1}, x_{i,2}, \cdots, x_{i,k}\}$ be the k adjacent samples of sample $\left(x_i, y_i\right)$, that is, they are connected to x_i with edges. x_i is the training sample, y_i is the label of x_i , and the distance between each adjacent sample and x_i satisfies the condition: $d\left(x_{i,1}, x_i\right) \leq d\left(x_{i,2}, x_i\right) \leq \cdots \leq d\left(x_{i,k}, x_i\right)$. Use the k -th nearest neighbor sample distance of x_i to normalize the distance from the first $k-1$ adjacent samples to x_i , then the normalized distance is:

$$D\left(x_{i,j}, x_i\right) = \frac{d\left(x_{i,j}, x_i\right)}{d\left(x_{i,k}, x_i\right)}, \quad j = 1, 2, \cdots, k$$

The weight of each edge in the adjacency graph is:

$$w_{ij} = P\left(x_{i,j} \mid x_i\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{D\left(x_{i,j}, x_i\right)}{2}\right)$$

4. Self-training algorithm based on density and trimming weights

Since the self-training algorithm tends to mark unlabeled samples at each iteration, these errors will be involved in the next iteration, which will affect the training of the classifier and reduce the performance of the algorithm. Therefore, in the process of self-training, identifying wrong labeled samples plays an important role in the performance of the algorithm. There are many methods for identifying sample labels, the common ones are filtering methods based on classifiers and data editing techniques based on nearest neighbor rules.

The classifier-based filtering method mainly divides the existing labeled sample set into n subsets during each iteration training, and uses the same learning algorithm such as C4.5 to train n in all possible n-1 subsets to get n Different classifiers. Then use n classifiers to classify the unlabeled samples, and select the labels of the samples according to the principle of consensus or majority voting. The data editing technique based on the nearest neighbor rule mainly relies on distance, and judges whether the label of the sample to be predicted is correct according to the labels of k nearest neighbor samples.

Classifier-based methods have extremely high requirements for the partitioning of sample sets and the selection of learning algorithms. The selection of distance metrics and values based on the nearest neighbor method need to be set in advance. If it is not selected properly in advance, it will cause a judgment error and affect the final classification effect. In addition, neither of these two methods uses a lot of valuable information carried by unlabeled samples in the recognition process, which reduces the accuracy of recognition. The method of cutting edge weight statistics to identify wrongly labeled samples does not need to set any parameters in advance, and it can also make full use of the information of unlabeled samples. Therefore, in order to improve the classification accuracy of the self-training algorithm, this paper incorporates the method of cutting edge weights to statistically identify the wrong label samples into the

ST-DP algorithm, and proposes the ST-DP-CEW algorithm. The algorithm first uses the density clustering method to discover the spatial structure of the data set, and uses the spatial result information to preferentially select representative unlabeled samples for label prediction during the iteration process, which improves the accuracy of predicting labels. Then use the method of cutting edge weight statistics to judge whether the prediction label is correct. Use the correctly labeled samples for the next training. The specific steps of the algorithm are described as follows:

Step1. Use the density clustering method to find the true space structure of the entire data set.

Step2. (a) Use KNN or SVM as the base classifier, and train an initial classifier with the initial labeled sample set;

(b) label prediction on the "next" unlabeled sample of all samples in;

(c) identify whether the "next" sample is correctly labeled by using the method of trimming edge weights to obtain a correctly labeled sample;

(d) Repeat (a) through (c) until all "next" samples of have been marked.

Step3. (a) Perform label prediction on the "previous" unlabeled samples of all the updated samples;

(b) Identify the "previous" sample using the edge-cut weighting statistical method to obtain the correct labeled sample, and then update the classifier;

(c) Repeat (a) and (b) until all "previous" samples of have been marked.

Obviously, Step3 is similar to Step2, except that the "next" in Step2 is replaced with "previous".

3. Experimental results and analysis

In order to illustrate the effectiveness of the algorithm, the proposed algorithm is compared with existing self-training algorithms on 8 real data sets. The datasets are derived from the KEEL database[6]. Samples with missing values are deleted from the Cleveland and Dermatology datasets, and the rest of the datasets are not processed. Related information is shown in Table 1.

Table 1 Experimental data set

data set	size	dimension	category
Bupa	345	6	2
Cleveland	297	13	5
Dermatology	358	33	6
Glass	214	9	7
Haberman	306	3	2
Ionosphere	351	34	2
pima	768	8	2
yeast	1484	8	10

The comparison algorithms used are: traditional self-training algorithms using KNN and SVM as classifiers, self-training classification algorithms based on fuzzy c-means clustering (ST-FCM), density-based self-training classification algorithms (ST-DP), and Self-training classification algorithm (ST-DE) based on differential evolution. The specific parameter settings are shown in Table 2.

Table 2 Parameter settings of related algorithms in the experiment

algorithm	parameter
KNN	K=3
SVM	Same settings as Literature(Chih-Chung & Chih-Jen, 2011)
ST-FCM	$\varepsilon_1 = 1$
ST-DP	$P_a = 2$
ST-DE	$P_a = 2$; $DE-POAC\left(L^{'}, L\right)$ Same settings as Literature(Chih-Chung & Chih-Jen, 2011)
ST-DP-CEW	$P_a = 2$;Significance level: $\alpha = 0.05$

1. Implementation of the experiment

A ten-fold cross-validation strategy was used to perform experiments on the dataset using KNN and SVM as base classifiers. Take one fold as the test set and the remaining nine fold as the training set. In each experiment, 10% of the samples in the training set are randomly selected as the initial labeled sample set, and the rest are unlabeled sets. In order to ensure the accuracy of the experiment, the ten-fold cross-validation experiment was repeated ten times, and the average value of the ten experiments was finally selected as the final experimental result. Accuracy rate (AR), Mean accuracy rate (MAR), and Standard deviation (SD-AR) are used as comparison criteria for the classification performance of the algorithm. Calculated as follows:

$$AR = \frac{1}{N_{T_s}} \sum_{i=1}^{N_{T_S}} \psi\left(\omega, f\left(x_i\right)\right)$$

$$MAR = \frac{1}{n} \sum_{k=1}^{n} AR_k$$

$$SD-AR = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(AR_k - MAR\right)^2}$$

$f\left(x_i\right)$ is the predicted label of the sample, N_{T_s} is the size of the test set, n is the number of times the experiment is repeated, MAR represents the classification performance of the algorithm, and SD-AR represents the robustness of the algorithm. MAR ± SD-AR is selected as the basis for judging the performance of the algorithm.

Tables 3 and 4 show the experimental results of the data set with KNN and SVM as the base classifier, respectively. The bold data indicates that the algorithm performs better in classification. As shown in Tables 1 and 2, when the initial labeled sample is 10%, the average classification accuracy of ST-DP-CEW on multiple data sets is significantly better than other comparison algorithms. However, when the algorithm is based on the SVM classifier, the classification accuracy of ST-DP-CEW on the dataset Cleveland has basically not improved. This is mainly because the values of most attributes in the dataset are close to 0. For the same attribute, The differences between the samples are small, resulting in a small difference between the samples as a whole, and the discrimination of each category is reduced, which affects the final classification effect.

Table 3 Experimental results when the base classifier is KNN (MAR ± SD-AR, %)

data set	Classifier: KNN				
	KNN only	ST-FCM	ST-DP	ST-DE	ST-DP-CEW
Bupa	54.48± 7.99	56.91± 9.34	58.88± 8.79	59.13± 8.43	**62.27± 6.21**
Cleveland	46.79± 6.70	46.47± 7.46	48.16± 8.65	49.15± 8.54	**52.17± 7.84**
Dermatology	53.60± 8.10	56.18± 7.58	70.94± 8.18	73.98± 7.21	**78.19± 6.64**
Glass	50.54± 7.59	5L58± 7.67	55.26 M.84	57.40± 8.35	**61.65± 6.83**
Haberman	67.59± 9.28	67.92± 9.52	69.31± 6.91	68.91± 8.29	**72.19± 7.11**
Ionosphere	74.35± 8.00	72.35± 8.33	80.61± 4.05	81.20± 5.44	**83.45± 7.78**
pi ma	67.72± 5.32	64.98± 4.56	66.40± 2.54	66.93± 4.57	**70.05± 2.70**
yeast	45.96± 5.83	48.32± 3.22	49.19± 3.28	50.74± 4.71	**53.10± 3.62**

Table 4 Experimental results when the base classifier is SVM (MAR ± SD-AR, %)

data set	Classifier: SVM				
	KNN only	ST-FCM	ST-DP	ST-DE	ST-DP-CEW
Bupa	60.86± 7.33	62.57± 7.70	65.50± 7.56	65.80± 6.30	**67.01± 8.20**
Cleveland	**53.84± 8.33**	**53.84± 4.29**	53.82± 8.76	53.82± 7.39	**53.84± 9.32**
Dermatology	56.41± 9.64	57.28± 9.65	68.14± 6.54	72.36± 9.72	**78.25± 9.28**
Glass	44.81± 9.87	46.34± 8.07	49.46± 9.10	51.36± 7.99	**54.72± 7.75**
Haberman	70.59± 7.06	71.61± 4.10	71.85± 5.56	72.24± 7.62	**74.62± 5.71**
Ionosphere	78.33± 4.16	79.75± 8.16	80.92± 6.10	82.34± 5.22	**84.92± 6.82**
pi ma	71.75± 6.13	72.53± 6.37	75.12± 4.72	75.78± 2.40	**77.23± 3.16**
yeast	31.54± 2.29	30.76± 3.68	31.21± 3.34	32.43± 4.25	**35.81± 2.63**

4. Conclusion

In this paper, based on the ST-DP algorithm, a self-training algorithm based on density peaks and edge trimming weights is proposed based on the samples that may be mislabeled during the self-training iteration process. That is, the method of statistically identifying cut-off weights to identify incorrectly labeled samples is integrated into the ST-DP algorithm. It not only considers the spatial structure of the data set, but also solves the problem that the samples are incorrectly labeled. In addition, the calculation of the weights in the adjacency graph also makes better use of the spatial structure of the data set and the information carried by the unlabeled samples. The effectiveness of the ST-DP-CEW algorithm is fully analyzed on the real data set. Especially when the proportion of initially labeled samples is low, the proposed algorithm has greatly improved performance compared to existing algorithms. In the subsequent work, we will discuss how to better construct the adjacency graph, and introduce a function that measures the probability of label error in the recognition process to make label recognition more accurate.

References

[1]Alcala-Fdez, J., Fernández, A., Luengo, J., Derrac,J., & García, S. (2011). KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. *Multiple-Valued Logic and Soft Computing*, *17*, 255–287.

[2]Chih-Chung, C., & Chih-Jen, L. (2011). Libsvm: a library for support vector machines. *ACM Trans. Intell. Syst. Technol., Vol. 2(3)*, 1–27.

[3]Gan, H., Sang, N., Huang, R., Tong, X., & Dan, Z.(2013). Using clustering analysis to improve semi-supervised classification. *Neurocomputing*, *101*, 290–298. https://doi.org/https://doi.org/10.1016/j.neucom.2012.08.020

[4] Liu, J., Gong, M., & He, H. (2019). Deep associative neural network for associative memory based on unsupervised representation learning. *Neural Networks : The Official Journal of the International Neural Network Society*, *113*, 41–53. https://doi.org/10.1016/j.neunet.2019.01.004

[5]Pavlinek, M., & Podgorelec, V. (2017). Text classification method based on self-training and LDA topic models. *Expert Systems with Applications*, *80*, 83–93. https://doi.org/https://doi.org/10.1016/j.eswa.2017.03.020

[6]fast search and find of density peaks. *Science*, *344*(6191), 1492–1496. https://doi.org/10.1126/science.1242072

[7]Triguero, I., Sáez, J. A., Luengo, J., García, S., & Herrera, F. (2014). On the characterization of noise filters for self-training semi-supervised in nearest neighbor classification. *Neurocomputing*, *132*, 30–41. https://doi.org/https://doi.org/10.1016/j.neucom.2013.05.055

[8]Vijayan, A., Kareem, S., & Kizhakkethottam, J. J. (2016). Face Recognition Across Gender Transformation Using SVM Classifier. *Procedia Technology*, *24*, 1366–1373. https://doi.org/https://doi.org/10.1016/j.protcy.2016.05.150

[9]Wang, X.-F., & Xu, Y. (2017). Fast clustering using adaptive density peak detection. *Statistical Methods in Medical Research*, *26*(6), 2800–2811. https://doi.org/10.1177/0962280215609948

[10]Wu, D, Luo, X., Wang, G., Shang, M., Yuan, Y., & Yan, H. (2018). A Highly Accurate Framework for Self-Labeled Semisupervised Classification in Industrial Applications. *IEEE Transactions on Industrial Informatics*, *14*(3), 909–920. https://doi.org/10.1109/TII.2017.2737827

[11]Wu, Di, Shang, M., Luo, X., Xu, J., Yan, H., Deng, W., & Wang, G. (2018). Self-training semi-supervised classification based on density peaks of data. *Neurocomputing*, *275*, 180–191. https://doi.org/https://doi.org/10.1016/j.neucom.2017.05.072

[12]Wu, Di, Shang, M. S., Wang, G., & Li, L. (2018). *A self-training semi-supervised classification algorithm based on density peaks of data and differential evolution*. 1–6. https://doi.org/10.1109/ICNSC.2018.8361359

[13]Xu, G., Zhang, M., Zhu, H., & Xu, J. (2017). A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene*, *604*, 33–40. https://doi.org/https://doi.org/10.1016/j.gene.2016.12.016

[14]Alcala-Fdez, J., Fernández, A., Luengo, J., Derrac, J., & García, S. (2011). KEEL Data-Mining Software Tool: Data Set

Repository, Integration of Algorithms and Experimental Analysis Framework. Multiple-Valued Logic and Soft Computing, 17, 255–287.
[15]Chih-Chung, C., & Chih-Jen, L. (2011). Libsvm: a library for support vector machines. ACM Trans. Intell. Syst. Technol., Vol. 2(3), 1–27.
[16]Liu, J., Gong, M., & He, H. (2019). Deep associative neural network for associative memory based on unsupervised representation learning. Neural Networks : The Official Journal of the International Neural Network Society, 113, 41–53. https://doi.org/10.1016/j.neunet.2019.01.004.

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

An Innovative System for Supporting Acquisition and Reproduction of Gestures in Storytelling Humanoid Robots

Agnese Augello[a], Angelo Ciulla[b], Alfredo Cuzzocrea[c], Salvatore Gaglio[a,b], Giovanni Pilato[a] and Filippo Vella[a]

[a]Institute for High Performance Computing and Networking National Research Council of Italy (ICAR-CNR), Palermo, Italy
[b]Dipartimento di Ingegneria, Università degli Studi di Palermo, Palermo, Italy
[c]iDEA Lab, University of Calabria, Rende, Italy

ARTICLE INFO

Article History:
Submitted 4.26.2020
Revised 7.14.2020
Accepted 7.27.2020

Keywords:
Storyteller
humanoid robot
gesture
automatic mimic

ABSTRACT

The work describes a module that has been implemented for being included in a social humanoid robot architecture, in particular a storyteller robot, named NarRob. This module gives a humanoid robot the capability of mimicking and acquiring the motion of a human user in real-time. This allows the robot to increase the population of his dataset of gestures. The module relies on a Kinect based acquisition setup. The gestures are acquired by observing the typical gesture displayed by humans. The movements are then annotated by several evaluators according to their particular meaning, and they are organized considering a specific typology in the knowledge base of the robot. The properly annotated gestures are then used to enrich the narration of the stories. During the narration, the robot semantically analyses the textual content of the story in order to detect meaningful terms in the sentences and emotions that can be expressed. This analysis drives the choice of the gesture that accompanies the sentences when the story is read.

1. Introduction

Nowadays, robots collaborate more and more with human beings, helping them to achieve different goals. In this context, the recognition and subsequent reproduction of gestures becomes extremely important.
Considering that communication between human beings is based not only on verbal interaction but also through nonverbal cues, a humanoid robot capable of interacting with people combining both speech and gestures would improve the effectiveness of social robots. On the other hand, other studies like [28, 22, 10] consider *knowledge management techniques* (e.g., [10]) to improve this phase. Sometimes, it is possible to enrich or even replace one's narrative exposure with gestures. This ability can be particularly relevant in a storytelling context, and, in general, in assistive robotics. In this context, a gestural expressiveness is indeed desirable to strengthen the meaning conveyed by the words spoken by a robot.
Processing the semantic content of a text and adding the execution of proper gestures while the robot is telling something, or generally, while it is interacting with human beings, is essential for improving the effectiveness of communication, avoiding trivial and boring situations. The capability to exhibit emotions and to show expressiveness during storytelling is fundamental to obtain an effective engagement [32] [13] [3]. As reported in [35], the use of bodily expressions in robots facilitates a mood induction process of the story and improves the storytelling experience.

In our Lab we are working on *NarRob*, a social robot

*Giovanni Pilato
**Alfredo Cuzzocrea
agnese.augello@icar.cnr.it (A. Augello); angelo.ciulla@community.unipa.it (A. Ciulla); alfredo.cuzzocrea@unical.it (A. Cuzzocrea); salvatore.gaglio@unipa.it (S. Gaglio); giovanni.pilato@cnr.it (G. Pilato); f.vella@pa.icar.cnr.it (F. Vella)
https://www.icar.cnr.it/persone/augello (A. Augello); http://si.deis.unical.it/cuzzocrea (A. Cuzzocrea); https://www.unipa.it/persone/docenti/g/salvatore.gaglio (S. Gaglio); https://www.icar.cnr.it/persone/pilato (G. Pilato); https://www.icar.cnr.it/persone/vella (F. Vella)
ORCID(s):

DOI reference number: 10-18293/JVLC20 20-N1-017

which plays the role of a storyteller [4] [3]. It is embodied in the Softbank robotic platforms Pepper and NAO, and it can show a body expressiveness to emphasize the semantic and the emotions arising from the text. In an early version, NarRob performs an analysis of the text in a story, and it then annotates each sentence by *a)* the emotion expressed in the text and *b)* the terms that can be associated with gestures, which are then performed while the robot is telling the story. In the specific, NarRob exploits: 1) a gesture module to acquire gestures and store them in a knowledge base 2) a chatbot module integrating an OWL ontology, used by the robot to have a conversation with users, before and after the narration, to deepen specific concepts involved in the narration; 3) an annotation module that makes it possible for the robot to detect meaningful terms as well as an emotion expressed in the text and to annotate the story with specific tags 4) an expressivity module that is used by the robot to convert the tags in gestures and expressions.

NarRob was equipped with a repository of manually annotated gestures. The gestures mainly came from the Pepper gesture dataset, and some other samples have been added using the timeline utility from the SoftBank Choregraphe suite. In this case, the gestures are acquired by using the timeline of Choregraphe, recording the sequence of postures obtained by using the robot like a puppet. Another modality consists of using an acquisition device that allows for tracking of the skeleton of a human performing the posture. This paper reports the evolution of NarRob with the implementation of a module designed to acquire and reproduce the gestures performed by a human during an interactive session.

In particular, we illustrate the development of a module of NarRob, named *Gestures Module*, aimed at acquiring in real-time gestures from human users, recording them, and executing them both as they are being recorded, and in a mirrored version. The module allows the robot to reproduce also simple poses instead of complete gestures. We have defined a mapping algorithm to allow a *SoftBank Pepper* robot to reproduce the tracked gestures as close as possible to the original ones.

The development of this module makes it possible for the robot to learn more and more gestures in real-time by using a low-cost RGBD camera. This allows enriching in a faster manner the gestures dataset of the robot.

In our approach, we exploited a *Microsoft Kinect* sensor to capture the motion data from a user. The Microsoft Kinect is a popular choice for any research that involves body motion capture. It is an affordable and low-cost device that can be used for noninvasive, marker-less tracking of body gestures. In particular, we used the *OpenNi* driver for the *Kinect*, the NiTE 2.2 libraries for detecting the user skeleton, and the Kinetic version of ROS with the module pepper_dcm to provide package exchange and bridging between the computer and the robot and Ubuntu 16.04. We focused on the movements of the arms and the head, laying the basis for the extension of the same approach to the remaining parts.

In what follows, after a section where related works are reported, we describe the NarRob components, focusing on the module aimed at the acquisition and reproduction of gestures.

2. Related Works

Using motion capture to control or teach a robot is a concept that have gained more and more attention in recent years, due to the many applications that it can have, from industrial work to social interactions. A device often used for this purpose is Microsoft Kinect, due to the cost and complexity of a standard motion capture setup. For example, Baron et al. [6] controlled a Mindstorm NXT artificial arm with sensor Kinect, employing gesture recognition to regulate arm movement. Chang et al. [8] developed a Kinect-based gesture command control method for driving a humanoid robot to learn human actions, using a Kinect sensor and three different recognition mechanisms: dynamic time wrapping (DTW), Hidden Markov model (HMM) and principal component analysis (PCA).

Sylvain Filiatrault and Ana-Maria Cretu [18] used sensor Kinect to mimic the motion of a human arm to a NAO humanoid robot. In their case, the software architecture is based on three modules: Kinect Manager, Interaction Manager, and NAO manager. The Kinect Manager deals with the events and data captured by the Kinect. The class Kinect Transformer is used to get the Euler angles of the desired joints. The Interaction Manager is the intermediary between the Kinect and the robot and contains the repository for the joints used by the other two modules. The use of a joint repository of all articulations allows reducing the data to be processed as some joints are not needed. Finally, the NAO manager contains the static and dynamic constraints to apply to each one of the articulations, as well as the methods that allow the control of the robot movements. To be sure that the robot has enough time to execute the gesture, a delay of 200 ms between one cycle and the next has been introduced.

Augello et al. [1] modelled a computational creativity behavior in a dancing robot using deep learning . The dataset used as a base was collected using a Kinect and together with a Variational Autoencoder that allows mapping input patterns in a latent space it allowed the robot to create new dancing moves in real time as it listened to music. Itauma et al. [19] used a Kinect to teach an NAO robot some basic Sign Language gestures. The aim was teaching Sign Language to impaired children by employing different machine learning techniques in the process. Shohin et al. [24] used three different methods to make a robot NAO imitate human motion: direct angle mapping, inverse kinematics using fuzzy logic and iterative Jacobian.

Miguel et al. [25] used a Kinect sensor and a Convolutional Neural Network (CNN) trained with the MSRC-12 dataset [33] to capture and classify gestures of a user and send related commands to a mobile robot. The used dataset was created by Microsft and had 6244 gesture instances of 12 actions. To have gestures of the same length, without losing relevant information, the system used a Fast Dynamic Time Warping algorithm (FastDTW) to find the optimal match between sequences by non linearly warping them along the time axis. This resulted in all gestures normalized to sequences of 667 frames, with each frame having 80 variables, corresponding to the x,y,z values for each of the 20 joints, plus a separation value for each joint. The resulting 667x80 matrix is used as the input of the CNN, which classifies it in one of the 12 possible gestures. The CNN was trained using two strategies, combined training consisting of a single CNN to recognize all 12 gestures and individual training with 12 different CNN, each capable of recognizing only one gesture. The accuracy rates were 72.08% for combined training and 81.25% for the individual training.

Moreover, Unai et al. [36] developed a natural talking gesture generation behavior for *Pepper* by feeding a Generative Adversarial Network (GAN) with human talking gestures recorded by a Kinect. Their approach in mapping the movements detected by Kinect on the robot is very similar to what we used, but while they feed the resulting values to a neural network (a GAN), we use the (filtered) values directly.

3. The Humanoid Storyteller Modules

In what follows we mainly describe the new *Gestures Module* that has been implemented in NarRob and its interaction with the other storytelling modules of NarRob, i.e. the *Annotation Module*, the *Chatbot Module*, and the *Expressivity Module*.

3.1. Gestures Module

The Gesture module deals with collecting acquired gestures, saving them in a knowledge base. In a previous implementation of this module [4] the dataset was mainly based on the animations available in the NAO robot animations package, and enriched with a limited set created by recording and annotating a sequence of postures of the robot through Softbank Choregraphe Suite.

The gesture module has been improved to enrich the dataset with new gestures acquired by capturing the movements of human people with a low cost RGB-D camera.

The module is structured in a set of components, to increase its versatility for future projects and to simplify extensions of the current project.

The first module is named *Viewer*, and it extracts data frames from the Kinect camera (consisting of nine float values: three values for a joint position in 3D space, four values for quaternion from the origin and reliability values for both) and sends them in a pipe. The module also provides the feed of the Kinect camera with the overlay of the tracked user's skeleton. The frame data is long 8640 bytes. It is composed of 15 joints, with 9 values each and a representation of 64 bytes for each value. The data is sent to the *Gesture_Brain* module.

The *Gesture_Brain* module works both as a gateway for the ROS system [34] and as the module where actual data processing takes place. The gathered data cannot be used directly: a mapping is required to correctly associate each joint user position to the equivalent one in the Pepper robot. For this reason, the data is parsed and structured in a 15×9

Figure 1: Structure of a single row of the array sent by the Viewer module

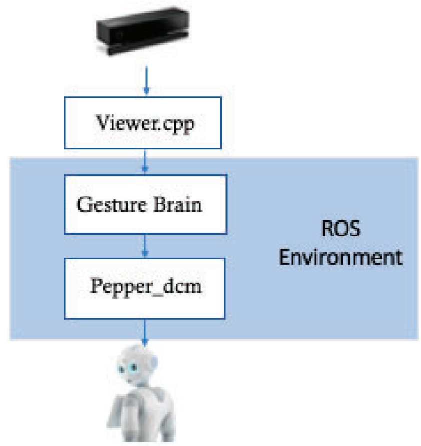


Figure 2: Structure of the gesture acquisition module

float matrix, which is then split into three matrices: one for coordinates, one for quaternions, and one for reliability values. In our approach, we exploit only the first matrix considering always high the reliability. We, therefore, assume at this stage that the captured joint data is accurate enough for our purpose, as the Kinect already discards joints whose reliability values are too low.

The joint position data is then used to estimate Pepper joint angles, specifically shoulder pitch, shoulder roll, elbow roll and elbow yaw for both arms and head yaw for the head.

The additional left and right wrist yaw and head pitch joint angles have been set to 0, since the Kinect camera is too imprecise to give a good estimate of that points.

When all the required values have been collected, the joint angles are sent to the robot using the ROS threads provided by the *pepper_dcm* bridge. These threads consist of multiple joint angles divided into groups, each group representing a body part. As we are interested only in the movement of arms and head, we use three of them: head, left arm, and right arm.

The bridge reads the sent values and the time stamp between each capture to dynamically compute in real-time the gesture trajectory. The spatio-temporal information allow the system to be as accurate as possible so that the gesture can be properly executed. The bridge itself is endowed with the in-built Self Collision Avoidance (part of the NaoQi library) while the *wait* and *breathe* animations have been deactivated, since they interfere with the commands sent by the pepper_dcm.

3.1.1. Mapping user Gestures into Pepper Movements

The *Pepper* robot has five degrees of freedom for each arm (each one associated with a joint), unlike human beings who have seven. A mapping is thus required between the detected user gestures and their reproduction by the robot.

From the *Kinect* camera the Cartesian coordinates for each joint, the quaternion for each segment (both referenced globally), and a reliability value for both are extracted. The bridge *pepper_dcm* uses the Euler angles to communicate to the robot the new position of its joint angles. We preferred to use the 3D space coordinates since the extracted quaternions do not represent the rotation from the previous position, but rather the rotation from a reference quaternion. This could lead to excessive inaccuracies when the values are converted in Euler angles.
Let $\overline{x}$, $\overline{y}$ and $\overline{z}$ be the unit vectors for each axis, that is:

$$\underline{x} = (1, 0, 0)$$
$$\underline{y} = (0, 1, 0)$$
$$z = (0, 0, 1)$$

Let S_L , E_L and W_L be the coordinates of the shoulder, the elbow and the wrist of the left arm respectively, $\overline{S_L E_L}$ and $\overline{E_L W_L}$ are defined as:

$$\overline{\underline{S_L E_L}} = E_L - S_L$$
$$E_L W_L = W_L - E_L$$

SR_L is the supplementary to the angle between $\overline{S_L E_L}$ and $-x$ axis:

$$SR_L = \frac{\pi}{2} - arcos(\overline{S_L E_L} \cdot -x) \tag{1}$$

SP_L is the angle between the projection of $\overline{S_L E_L}$ on zy plane and z axis, shifted in range to avoid the jump discontinuity at π and $-\pi$:

$$SP_L = \pi - mod_{2\pi}(\frac{3}{2}\pi + arctan(\overline{S_L E_{Lz}}, \overline{S_L E_{Ly}}) \tag{2}$$

For values of SR_L close to $\frac{\pi}{2}$, SP_L become unstable. As such, in the algorithm is assigned a value of 0 for $SR_L > 1.3$. ER_L is the angle between $\overline{E_L W_L}$ and $\overline{S_L E_L}$, shifted by $\frac{\pi}{2}$:

$$ER_L = \frac{\pi}{2} + arcos(\overline{E_L W_L} \cdot \overline{S_L E_L}) \tag{3}$$

EY_L is the angle between the projection of $\overline{E_L W_L}$ on zy plane and z axis, shifted in range for stability reasons, plus $-SP_L$:

$$EY_L = \pi - mod_{2\pi}(\frac{3}{2}\pi + \\ + arctan(E_L W_{Lz}, E_L W_{Ly}) - SP_L \tag{4}$$

The right arm is almost the same as the left arm, the only difference is that some angles have the opposite sign.

Let $\overline{HN}$ be the difference between the coordinates of the joints H (head) and N:

$$\overline{HN} = H - N$$

The head yaw HY is the angle between the projection of $\overline{HN}$ on the xz plane and the z axis:

$$HY = -arctan(\overline{HN}_z, \overline{HN}_y) - \frac{\pi}{2} \tag{5}$$

Smoothing movements through a Line of Best Fit: The *Kinect* joint detection is based on the shape of the user, which is redrawn at every frame. While calibrating the sensor helps to reduce the resulting jerkiness but a significant amount of noise can still remain. This noise can be approximately classified in two categories:

- a constant gaussian noise caused by small alteration on the detected shape and
- large "spikes" when the *Kinect* fail to guess the position of one or more joints (especially common when part of the limb is outside the frame or when two or more joints overlap).

A simple way to compensate part of this noise is to use a *line of best fit*.

In particular, given k points in (x, y) coordinates system, we must find the values c_0 and c_1 in the equation

$$p(x) = c_0 x + c_1$$

that define the straight line minimizing the squared error

$$E = \sum_{j=0}^{k} |p(x_j) - y_j|^2$$

in the equations

$$x_0 c_0 + c_1 = y_0$$
$$x_1 c_0 + c_1 = y_1$$
$$...$$
$$x_k c_0 + c_1 = y_k$$

The result is a smoother movement, especially when the Kinect camera is not able to detect the exact coordinates of a given joint. This is because, given a disturbing signal, the line of best fit can be seen as an approximation of the tangent line that the signal would have at that point if all the noise were removed. This is not always true, especially when the signal changes rapidly, but it's close enough in most cases to give a generally cleaner movement.

3.1.2. Modes of Operation

Besides acquiring the user movements in order to increase the gestures knowledge base of the robot, the Gesture module has also some additional features implemented to increase the breadth of experiments that can be performed with the system or to help with future projects. In particular, the module can act into three modalities: 1) Mimic mode, 2) pose mode, and 3) Playback mode.

The behavior of the module is managed by a set of input parameters, named *mode, mirror_flag, json_file_name, LAjpos, RAjpos, Hjpos*. The first parameter determines which one of the three different modes of operation will be used (default 0), the second one determines if the mirror mode is activated or not (default false), the third defines the name of the text file used to record (in mode 0 and 1) or read (mode 2) the gestures (the default value is NULL, that is no recording) and set a flag (record_flag) to 1, the fourth, fifth and sixth ones are used to determine the pose to use in mode

1 (as default, the robot will spread its arms parallel to the ground, in a pose that in animation is known as "T-pose"). More in details, the modes of operation of the main program are:

- Mode 0, or "Mimic Mode", is the default mode and it allows the robot mimicking the movements of the user. The record flag is used to store the data, so the output is not just sent to the ROS publishers, but recorded in a JSON(JavaScript Object Notation) file. This file is then linked to a dataset of gestures that can be stored and annotated in order to be reproduced later. If the *mirror* flag is active, each movement is mirrored during the execution. In case both the record and mirror flags are active, the mirrored movement will be recorded and saved in a specific file to be linked to the gesture dataset.
- Mode 1, or "Pose Mode", make the robot execute a pose (defined at the beginning by the value of the given parameters) already stored in the gestures ontology. We considered a scenario where the Pepper robot showed a specific pose and the user had to replicate it as closely as possible. The scenario involves the use of both the normal mode and the mirror mode. A proper distance algorithm calculates how close is the pose of the user to that one performed by the robot. The distance is separately evaluated for the head, the right upper arm, the left upper arm, the right forearm, and the left forearm. If the user pose keeps all body parts below their respective distance thresholds (defined separately for each body part), the pose is assumed to be correctly replicated. If the record flag is activated the returned distance values are saved. When the mirror flag is set to "on" the user should try to replicate the mirrored version of the shown pose.
- Mode 2, or "Playback Mode", consists of reproducing a previously recorded gesture. The mirror flag, even if selected, doesn't have any effect on this procedure. The name, necessary to activate the record flag, is used as the name of the gesture linked to the gesture ontology. In this case the goal is twofold: just emulating a gesture or asking a user to emulate the gesture itself. This could be done either for learning activities (e.g. teach pupils some gestures) or for therapeutic ones (e.g. physical exercises for elderly).

3.1.3. Auxiliary Algorithms

In the following we describe a set of auxiliary algorithms that we have included in the Gesture Module to implement its tasks.

Distance measurement To measure the distance between the robot pose and the detected pose, we used a distance measurement algorithm. For each arm, the related joint angle group (consisting of Elbow roll, Elbow yaw, Shoulder pitch,

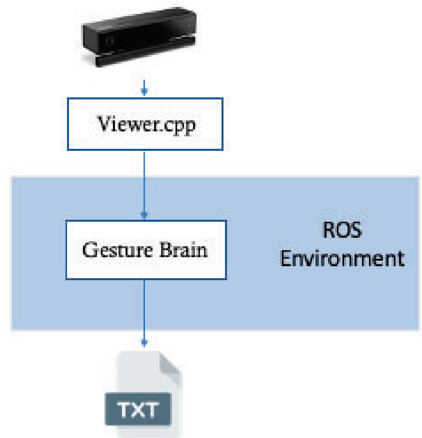


Figure 3: Structure of the algorithm when recording (the text file can be either the recorded gesture in mode 0 or the record of distance values in mode 1)

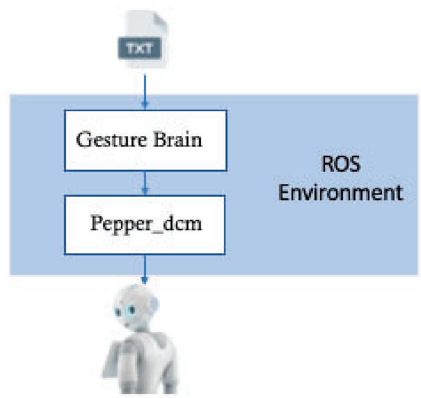


Figure 4: Structure of the algorithm in mode 2 (the txt file in this case is the coding of a previously recorded gesture)

Shoulder roll, and Wrist yaw) is split into two vectors, representing the upper arm (Shoulder pitch and Shoulder roll) and the forearm (Elbow roll, Elbow yaw, and Wrist yaw). Including the joint angle group for the head (Head yaw and Head pitch). The position is represented with five vectors. To avoid that joint angles with a larger range of values could excessively affect the final result, each angle is normalized dividing the value by its maximum value. This process is applied both to the robot pose and the detected pose of the user, producing five pairs of vectors. Finally, a Mean Squared Error algorithm is applied at each pair of vectors, resulting in five distances defined in the interval [0,1]. With this separated evaluation we found that the final measure is more precise and reliable allowing to define separate thresholds if a pass/fail system is implemented (like in our case).

Mirroring: The activation of the mirror flag activates the mirror algorithm for any detected movement. This means that in mode 0, the robot will mirror the user (and if the program is recording, it will record the mirrored movement), while in mode 1, the user will have to mirror the shown pose. The mirroring mode consists in switching the detected angles between the left and the right arm, changing the sign of every angle except shoulder pitch. For the head yaw angle, a simple change in the sign is enough.

Recording and Playback: If the recording flag is active, the program will write a text file in *JSON* syntax with the joint angles sent to the robot (mode 0) or the distance from the given pose (mode 1) for each frame.
In mode 2, the records created in mode 0 can be reproduced by the robot.
In mode 0, the fields for each frame are:

- AbsTime: Date and time when the frame was captured, to the microsecond.
- Left_Arm: Vector containing the joint angles of the left arm joint group.
- Right_Arm: Vector containing the joint angles of the right arm joint group.
- Head: Vector containing the joint angles of the joint head group.

In mode 1, the fields for each frame are:

- AbsTime: Date and time when the frame was captured, to the microsecond.
- Error_Left_Arm_Shoulder: the measured distance between the given position and the detected position of the left upper arm.
- Error_Left_Arm_Elbow: the measured distance between the given position and the detected position of the left forearm.
- Error_Right_Arm_Shoulder: the measured distance between the given position and the detected position of the right upper arm.
- Error_Right_Arm_Elbow: the measured distance between the given position and the detected position of the right forearm.
- Error_Head: the measured distance between the given position and the detected position of the head.

4. From Words to Gestures and Expressions

The gestures acquired by the kinect are used by the robot in order to improving its communication abilities by adding expressiveness to the narration, by including gesticulation to emphasize what is being said by the robot.

The operation of the module that draws the gestures from the dataset has two main phase; during the first phase, it parses the sentence that has to be said, looking for the terms that directly correspond to a given gesture; if a match is found, the term is annotated for being associated with a gesture; otherwise, the focus of the parser in oriented at finding the verbs in a sentence. A POS is then executed and the verbs present in the sentence are analyzed by using Wordnet: in particular, a lemmatization task is run in order to find the lemma of the verb under exam; once the lemma has been found, a score $s(lemma\_verb1, lemma\_kb) \in [0, 1]$ is computed between the lemma of the verb in a sentence and the lemmas that are associated, in the knowledge base of the annotated gestures. The score is determined by calculating the shortest path linking two senses in WordNet by considering the "is-a" relationship [23]. In this first version of the system, the score computation takes into account only the most frequent sense for each lemma, speeding up the analysis. A similarity threshold $T_{sim} \in [0, 1]$ is experimentally fixed. If $s(lemma_{verb1}, lemma_{kb})$ g T_{sim}), the action corresponding to the $lemma_{kb}$ that gives the highest value of $s(lemma_{verb1}, lemma_{kb})$ between the lemma associated to the verb detected in the sentence and the lemma associated to the action stored in the list of gestures that can be performed by the robot. If the value of $s(lemma\_verb1, lemma\_kb)$ is below the value of T_{sim}), the verb is ignored and no specific gesture is performed by the robot.

Furthermore, an emotion detection module, presented in [26], has been exploited to find if there is a basic emotion that could be associated to that particular sentence that will be said. The emotion detection module acts on the communicative channels of the robot that could be associated by a human observer to some emotions, e.g., the color of its LEDs, the speed of its speech, and the inclination of the head [17][20] [5].

The emotion that can be associated to a given sentence is one of the well known Ekman basic emotions: *anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise*. If no emotion is detected, the *neutral* label is used. The module exploits a literature lexicon derived from the Word-Net Affect Lexicon [30] [31] and the methodology, based on the Latent Semantic Analysis (LSA)[21] paradigm, that has been presented in [26].

That approach assumes that a sentence d can be encoded as a point in a Data Driven "conceptual" space, by calculating a vector $\mathbf{d}$ whose i-th component is the number of times the i-th word of the vocabulary is present in d. The vector $\mathbf{d}$ is subsequently mapped into a reduced-dimensionality "conceptual" space induced by LSA.
At the same time, a set of vectors acting as emotional "beacons" have been used to map a text from the conceptual space to an emotional space. In particular, six sets E_{anger} , $E_{disgust}$, $\cdots$, $E_{surprise}$ of vectors constituting the sub-symbolic coding of each "beacon" identifying a basic emotion have been used. The generic vector belonging to one of the sets is encoded in the same "conceptual" space together with the sentence s of the story that has to be said. Once the sentence s is mapped into the "conceptual" space, it is possible to compute its emotional fingerprint following the methodology illustrated in [26], which consists in exploiting seman-

tic similarity with the vectors that are associated to each one of the six E_i sets. The emotional space is build as a six-dimensional hypersphere where all the sentences are encodes. Each region of this hypersphere is associated to a set of emotional expression of the robot. The element associated to the highest value of emotion determines the main emotion expressed by the sentence. A minimum value of threshold $Th\_e \in [0, 1]$ has been experimentally established in order to label sentences that do not carry any emotion as "neutral".

Thanks to this encoding, the system is provided with a module that makes it possible to generate an expressive behavior, according to the story annotations. In particular a gesture tag triggers a proper movement according to what is is stored in the gestures dataset, and to produce an emotional expression.

For the emotions, we have encoded six possible emotional expressions, corresponding to the Ekman categories (i.e., anger, disgust, fear, joy, sadness, surprise), that has been defined by setting some of the robot communicative channels that can be correlated by a human observer to some emotions, such as the color of its LEDs, the pitch and speech rate, and the head inclination. [17] [20] [5] [29].

4.1. Annotation Module

This module of NarRob is aimed at the semantic annotation of possible actions and the emotional labelling of sentences.

The gestures acquired with the Gestures Module are annotated in an interactive manner according to their specific meaning and labelled into the gesture ontology.

The robot, before reading a story, analyzes its composing sentences, extracts the actions to perform and the emotions to manifest as well.

Each story is then offline analysed through a dependency-parsing by using the Stanford CoreNLP tool. This analysis is performed to obtain the dependencies graphs of the sentences. Each graph is then analyzed starting from the root, where each node is compared with the annotations of the gestures in the KB. Each time a match is found, the tag $^{s}tart(gesturename)$ is added to the text to give an indication of the gesture that robot should perform in that specific point of the story. Figure 5 shows an example for the sentence "When they were gone, Cinderella, whose heart was very sad, cried bitterly", extracted from the well known story of Cinderella [1]. In the specific case there are two words corresponding to the annotations in the gestures dataset. However, the priority is given according to the order of the dependencies, in this case one of the words (the verb *cry*), is the root of the sentence and will be chosen to add the gestures tag.

The annotations related to the emotional content of the story are inserted according the six Ekman categories. These categories have been widely used also in other contexts [27] [16] [14]. Here we exploit the approach described in section 4 When a specific emotion is detected, the tag *mood("EmotionName", EmotionWeight)* is added in the text, where the weight specify how intense is the emotion according to the emotion detection algorithm.

[1]The text is extracted from https://www.storyberries.com/fairy-tales-cinderella-or-the-little-glass-slipper/

```
root ( ROOT-0 , cried-14 )
advmod ( gone-4 , When-1 )
nsubjpass ( gone-4 , they-2 )
auxpass ( gone-4 , were-3 )
advcl ( cried-14 , gone-4 )
nsubj ( cried-14 , Cinderella-6 )
nmod:poss ( heart-9 , whose-8 )
nsubj ( sad-12 , heart-9 )
cop ( sad-12 , was-10 )
advmod ( sad-12 , very-11 )
acl:relcl ( Cinderella-6 , sad-12 )
advmod ( cried-14 , bitterly-15 )
```

Figure 5: The dependencies of the analyzed sentence

This text passage is an example of annotated story, taken from " The Real Princess"[2] by Hans Christian Andersen:

> *mood("SADNESS") gesture(ThunderGesture)*
> *One evening a fearful tempest arose, it thundered and lightened, and the rain poured down from the sky in torrents:*
> *mood("SADNESS")*
> *besides, it was as dark as pitch.*
> *mood("ANGER") gesture("HearGesture")*
> *All at once there was heard a violent knocking at the door, and the old King, the Prince's father, went out himself to open it.*

4.2. The Chatbot Module

This module relies on a chat-bot engine based on the SALVE architecture [2], and a conversational Knowledge Base described through a set of question-answers modules, named categories, and written in an enhanced AIML markup language, named S-AIML. This rule engine manages the interaction according to typical practices of interaction.

The chatbot exploits also an OWL[3] ontology, formalizing the main concepts related to the specific domain for which the robot has been suited[4].

Moreover, the Chatbot is not only aimed at interacting with the user during the storytelling activity. It is exploited also to label the recently acquired gestures with proper descriptions through an interactive session. An example of dialogue is reported below:

U: NarRob, please pay attention to this new gesture
R: Ok. I am waiting and recording
The user makes a gesture that is acquired by the Kinect
U:End of gesture
R: I have acquired a new gesture. I can repeat it
Narrob executes the PlayBack mode of the recently acquired gesture
R: Has the gesture been correctly reproduced?
U: Yes, it has
R: What label should I assign to the gesture?
U: The gesture can be labelled with "come here"

[2]http://www.gutenberg.org/cache/epub/1597/pg1597.txt
[3]W3C Web Ontology Language (OWL)

R: Ok. Done. Gesture acquired, labelled and saved for future storytelling.

5. Conclusions and Developments

We have presented in detail the realisation of a specific module for our NarRob Architecture aimed at acquiring gestures from an user by using a low-cost RGB-D Camera, such as the Kinect.

NarRob exploits an annotated corpus of text and an ontology of gestures to associate body signs and emotional expressions to the narration, by selecting the best appropriate gestures fitting the meaning of a sentence or part of a sentence. The goal is to obtain a not trivial, not pre-defined, and a more involving interaction with a user. Gestures have been acquired by a Gesture Module, which is capable of detecting, storing and reproducing the user poses with a Kinect camera with sufficient accuracy.

The first experiments show that the reproduced movements are quite accurate and smooth; the recording and execution of the gestures are very close to the real-time movements. However, sometimes, certain positions cannot be reliably detected, due to imprecise behavior of the Kinect output when joints overlap each other, and to excessive reliance on the silhouette to detect the human body and the lack of joints in key points of the detected skeleton (like the hands). There is also an environmental factor, like lightning and positioning, that can make accurate user detection problem.

The robot can acquire gestures form an human user and reproduce them in an adequate manner by using also a proper mapping procedure that allows to approximate the gestures of human beings. Furthermore, the robot, thanks to the Gesture Module, is potentially autonomously capable of acting both as an instructor and a learner by exploiting the gesture mirroring feature.

In future works, a neural network will be also trained to recognize and classify the gestures to give a proper answer, creating a more realistic verbal communication between humans and robots. Possible improvements should include a more effective detection algorithm for the Kinect, more efficient ways to execute the mapping, the use of all points detected (not just limbs and head), the use of the official Microsoft SDK to have even more points detected (provided it's possible to retrieve all the necessary libraries) and a more general user-friendly experience (like the ability to set a timer for recording). Moreover, we plan to improve our framework as to deal with novel and emerging *big data trends* including performance (e.g., [7, 12, 9]), and privacy and security (e.g., [11, 15]).

References

[1] Augello, A., Cipolla, E., Infantino, I., Manfre, A., Pilato, G., Vella, F., 2017. Creative robot dance with variational encoder. arXiv preprint arXiv:1707.01489 .

[2] Augello, A., Gentile, M., Weideveld, L., Dignum, F., 2016. Dialogues as social practices for serious games, in: Proceedings of the Twenty-second European Conference on Artificial Intelligence, pp. 1732–1733.

[3] Augello, A., Infantino, I., Maniscalco, U., Pilato, G., Vella, F., 2018a. Introducing narrob, a robotic storyteller, in: International Conference on Games and Learning Alliance, Springer. pp. 387–396.

[4] Augello, A., Infantino, I., Maniscalco, U., Pilato, G., Vella, F., 2018b. Narrob: A humanoid social storyteller with emotional expression capabilities, in: Biologically Inspired Cognitive Architectures Meeting, Springer. pp. 9–15.

[5] Bänziger, T., Scherer, K.R., 2005. The role of intonation in emotional expressions. Speech communication 46, 252–267.

[6] Baron, G., Czekalski, P., Malicki, D., Tokarz, K., 2013. Remote control of the artificial arm model using 3d hand tracking, in: 2013 International Symposium on Electrodynamic and Mechatronic Systems (SELM), IEEE. pp. 9–10.

[7] Bonifati, A., Cuzzocrea, A., 2006. Storing and retrieving xpath fragments in structured P2P networks. Data Knowl. Eng. 59, 247–269.

[8] Chang, C.w., He, C.j., et al., 2014. A kinect-based gesture command control method for human action imitations of humanoid robots, in: 2014 International Conference on Fuzzy Theory and Its Applications (iFUZZY2014), IEEE. pp. 208–211.

[9] Chatzimilioudis, G., Cuzzocrea, A., Gunopulos, D., Mamoulis, N., 2013. A novel distributed framework for optimizing query routing trees in wireless sensor networks via optimal operator placement. J. Comput. Syst. Sci. 79, 349–368.

[10] Cuzzocrea, A., 2006. Combining multidimensional user models and knowledge representation and management techniques for making web services knowledge-aware. Web Intelligence and Agent Systems 4, 289–312.

[11] Cuzzocrea, A., Bertino, E., 2011. Privacy preserving OLAP over distributed XML data: A theoretically-sound secure-multiparty-computation approach. J. Comput. Syst. Sci. 77, 965–987.

[12] Cuzzocrea, A., Moussa, R., Xu, G., 2013. Olap*: Effectively and efficiently supporting parallel OLAP over big data, in: Model and Data Engineering - Third International Conference, MEDI 2013, Amantea, Italy, September 25-27, 2013. Proceedings, pp. 38–49.

[13] Cuzzocrea, A., Pilato, G., 2018. Taxonomy-based detection of user emotions for advanced artificial intelligent applications, in: International Conference on Hybrid Artificial Intelligence Systems, Springer. pp. 573–585.

[14] Cuzzocrea, A., Pilato, G., 2019. An innovative user-attentive framework for supporting real-time detection and mining of streaming microblog posts. Soft Computing , 1–20.

[15] Cuzzocrea, A., Russo, V., 2009. Privacy preserving OLAP and OLAP security, in: Encyclopedia of Data Warehousing and Mining, Second Edition (4 Volumes), pp. 1575–1581.

[16] D'Avanzo, E., Pilato, G., Lytras, M., 2017. Using twitter sentiment and emotions analysis of google trends for decisions making. Program: electronic library and information systems 51, 322–350.

[17] Feldmaier, J., Marmat, T., Kuhn, J., Diepold, K., 2016. Evaluation of a RGB-LED-based emotion display for affective agents. arXiv preprint arXiv:1612.07303 .

[18] Filiatrault, S., Cretu, A.M., 2014. Human arm motion imitation by a humanoid robot, in: 2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings, IEEE. pp. 31–36.

[19] Itauma, I.I., Kivrak, H., Kose, H., 2012. Gesture imitation using machine learning techniques, in: 2012 20th Signal Processing and Communications Applications Conference (SIU), IEEE. pp. 1–4.

[20] Johnson, D.O., Cuijpers, R.H., van der Pol, D., 2013. Imitating human emotions with artificial facial expressions. International Journal of Social Robotics 5, 503–513.

[21] Landauer, T.K., Foltz, P.W., Laham, D., 1998. An introduction to latent semantic analysis. Discourse Processes 25, 259–284. doi:10.1080/01638539809545028, arXiv:https://doi.org/10.1080/01638539809545028.

[22] Lau, M.C., Anderson, J., Baltes, J., 2019. A sketch drawing humanoid robot using image-based visual servoing. Knowledge Eng. Review 34, e18.

[23] Lingling, M., Runqing, H., Junzhong, G., 2013. A review of semantic similarity measures in wordnet, in: International Journal of Hybrid Information Technology, p. Vol 6 No 1.

[24] Mukherjee, S., Paramkusam, D., Dwivedy, S.K., 2015. Inverse kinematics of a nao humanoid robot using kinect to track and imitate human motion, in: 2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE), IEEE. pp. 1–7.

[25] Pfitscher, M., Welfer, D., Cuadros, M.A.d.S.L., Gamarra, D.F.T., 2018. Activity gesture recognition on kinect sensor using convolutional neural networks and fastdtw for the msrc-12 dataset, in: International Conference on Intelligent Systems Design and Applications, Springer. pp. 230–239.

[26] Pilato, G., D'Avanzo, E., 2018a. Data-driven social mood analysis through the conceptualization of emotional fingerprints. Procedia Computer Science 123, 360 – 365. doi:https://doi.org/10.1016/j.procs.2018.01.056. 8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017 (Eighth Annual Meeting of the BICA Society), held August 1-6, 2017 in Moscow, Russia.

[27] Pilato, G., D'Avanzo, E., 2018b. Data-driven social mood analysis through the conceptualization of emotional fingerprints. Procedia computer science 123, 360–365.

[28] Regier, P., Milioto, A., Karkowski, P., Stachniss, C., Bennewitz, M., 2018. Classifying obstacles and exploiting knowledge about classes for efficient humanoid navigation, in: 18th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2018, Beijing, China, November 6-9, 2018, pp. 820–826.

[29] Rodriguez, I., Manfre, A., Vella, F., Infantino, I., Lazkano, E., 2018. Talking with sentiment: Adaptive expression generation behavior for social robots, in: The 19th International Workshop of Physical Agents (WAF).

[30] Strapparava, C., Mihalcea, R., 2007. Semeval-2007 task 14: Affective text, in: Proceedings of the 4th International Workshop on Semantic Evaluations, Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 70–74.

[31] Strapparava, C., Mihalcea, R., 2008. Learning to identify emotions in text, in: Proceedings of the 2008 ACM Symposium on Applied Computing, ACM, New York, NY, USA. pp. 1556–1560. doi:10.1145/1363686.1364052.

[32] Striepe, H., Lugrin, B., 2017. There once was a robot storyteller: Measuring the effects of emotion and non-verbal behaviour, in: International Conference on Social Robotics, Springer. pp. 126–136.

[33] Url, a. Msrc-12 dataset; https://www.microsoft.com/en-us/ download/ details.aspx? id=52283.

[34] Url, b. Ros kinetic; http://wiki.ros.org/kinetic.

[35] Xu, J., Broekens, J., Hindriks, K., Neerincx, M.A., 2015. Effects of a robotic storyteller's moody gestures on storytelling perception, in: Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on, IEEE. pp. 449–455.

[36] Zabala, U., Rodriguez, I., Martínez-Otzeta, J.M., Lazkano, E., 2019. Learning to gesticulate by observation using a deep generative approach. arXiv preprint arXiv:1909.01768 .

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc/

A Visualization Analysis of China's Leisure Sports Research⋆

Sun Ming[a], Zeng Ji[a,*] and Hu Tian[b]

[a]Hubei Leisure Sport Development Research Center, School of Physical Education, Hubei University, China
[b]Center for Studies of Education and Psychology of Ethnic Minorities In Southwest China, Southwest University, China

ARTICLE INFO

Article History:
Submitted 3.25.2020
Revised 6.19.2020
Accepted 7.10.2020

Keywords:
Chinese leisure sports
Leisure research hotspots
Leisure sports research
Visualization analysis

ABSTRACT

The social economy's continual development has produced gradual increases in disposable income, which have enriched leisure activities. In order to engage progress in the research of leisure sports in China over the past 20 years and contribute a new angle and direction to the research of leisure sports in China, this paper summarizes the internal and external characteristics of the country's leisure sports research. The data visualization software CiteSpace5.0 is applied to provide a macro analysis of the development of China's leisure sports research over the past 20 years. This method has not been applied to research of this kind before, and this reiterates the innovative character of this paper. It achieves this by extracting the 1998-2017 leisure sports literature from the CSSCI database before applying bibliometric Visual Analysis Software CiteSpace 5.0 to this material. It finds that the momentum of the development of the county's leisure sports is positive, and observes the initial establishment of a cooperation network. However, it also highlights a number of problems, which include the insufficient depth of disciplinary research, the lack of interdisciplinary research results (as shown by the adoption of a single research paradigm) and the neglect of frontier issues.

1. Introduction

China's leisure sports research mainly focuses on the history, development status and problems encountered by leisure sports research. Although this research had a late start (Kong Chuihui, 2009), its development has been rapid and it has provided enriching perspectives and generated methods that can be applied (Xiaoyu and Hai, 2016).

But there are a number of general problems in the study of China's leisure sports that need to be resolved. These are as follows: 1) The definition of the leisure sports concept is ambiguous and is not sufficiently comprehensive (Wu Jiangang, 2003; Li Xiangru, 2015); (2) Leisure sports research has limited depth and its results are not sufficiently influential (Xiaoyu and Hai, 2016); (3) Most of the research has an international focus, and does not sufficiently refer to China (Xiaoyu and Hai, 2016); (4) a single research method tends to be adopted and cooperative and interdisciplinary research is limited; (ibid) (5) research groups are not sufficiently diverse and do not sufficiently acknowledge engagement in leisure sports activities across different social classes (Jiangang, 2003; Chuihui, 2009).

There are grounds for believing that the

⋆ This work was supported by Hubei Leisure Sports Development Research Center 2020 Open Fund Project Grant Number 2020Y019 and by Hubei Leisure Sports Development Research Center 2018 Open Fund Project Grant Number 2018Z002

*Corresponding author
Email address: sunny_hung@foxmail.com
ORCID: 0000-0003-4394-9104

DOI reference number: 10.18293/JVLC2020-N1-005

development of leisure sports will make an important contribution to the rapid transformation of Chinese society and its future modernization. China therefore provides an ideal opportunity to study the relationship between leisure sports research and the stage of national development. Here it should be remembered that most of the research samples of leisure sports have been taken from Western contexts. The study of the development of China's leisure sports research over the past 20 years will provide insight into its future direction.

This paper will explore dynamic changes in China's leisure sports hotspots during the period 1998-2018; it does so with the intention of identifying key characteristics of China's leisure sports research, predicting future research trends and highlighting problems that can be engaged by Chinese researchers. Although development trends in China's leisure sports research have been extensively engaged (Jiangang, 2003; Limin, 2007; Xiaoming, 2008; Chuihui, 2009; Xin, 2012; Xiaoyu, 2016; Xiuyu, 2016), there are still research shortcomings that need to be directly addressed. This paper seeks to contribute to this process by providing an overview of all literature related to leisure sports topics that were published by CSSCI journals and incorporated into the China Knowledge Network (CNKI) during the period 1998-2018. It will apply Citespace with the intention of gaining insight into the literature's features, strengths and deficiencies, along with its dynamic development.

2. Literature Review

China is the world's second largest economy. In 2018, its gross domestic product (GDP) reached \$90.03 trillion (USD) and its population expanded to 1.395 billion, accounting for about one-fifth of the world's population. (World Bank, 2018) Since the 1980s, China's annual economic growth rate has, on average, exceeded 9 percent, which far surpasses the annual growth rate of 2.3 percent recorded for developed economies during the same period. (World Bank, 2011). Rapid economic development has greatly improved Chinese living standards (Li et al., 2012) and residents of mainland China now have the time and financial resources to participate in leisure activities. (Liang & Walker, 2011).

In recent times, China's booming economy and overall national strength has created huge dividends for its citizens. Personal income has also increased as a result of more efficient resource allocation and more incentives for private investment (Easterlin et al., 2012). During this transition period, both the leisure time (Yin, 2005) and leisure consumption of Chinese residents have increased significantly (Zhai & Xiao, 2004).

2.1 The conceptualization of 'leisure sports'

Chinese scholars have not yet reached an agreement on what 'leisure sports' are, and this has interrupted research into conceptualization. The most renowned experts in this area are Tian Hui, Ma Huizhen, Zhou Aiguang, Lu Feng and Xiao Huanqi. Ma Huizhen (2008) and Zhou Aiguang (2009). They define 'leisure sports' in similar terms, and observe they are a sports activity engaged in for leisure, which relieves stress, eliminates fatigue and benefits body and mind. Tian Hui (2006), Xiao Huanqi (2010) and Yu Kehong (2003) define them as sports activities that are engaged in during leisure time, which assist economic, environmental, personal and social development. Lu Feng (2004) suggests they are a collective term for various sports activities engaged under relatively free conditions.

The classification and definition of 'leisure sports' by scholars shows that China lacks metacognition of this concept . It is part of a foreign vocabulary, and has no Chinese counterpart. The word needs to be traced back to its root in order to explore its original meaning.

2.2 Leisure Sports and individual and collective development

A substantial amount of research has engaged with the developmental contribution of leisure sports. Since the beginning of the 21st century, population and economic aggregates have both increased and the individual and collective influence of leisure sports has attracted growing academic interest. Lu Feng (2004) suggests that the development of leisure sports has personal, developmental, social, social fashion (communication), social group (organization) and social symbolic functions. Zhang Rui (2014) observes that leisure sports do not just enhance physical health but also improve mental health, promote social interaction, create a relaxed and comfortable environment and contribute to an improved quality of life.

It is only by understanding and promoting leisure sports, in addition to grasping their essential purpose, that it will be possible to promote a healthier body and mind. Lu Gaofeng (2014) suggests that leisure sports provide a basis for various capital conversions. By engaging in leisure sports, higher social classes can accumulate cultural capital, expand social capital, and maintain or develop physical capital by investing in economic capital. Lower social classes can also transform physical capital into economic capital by engaging in leisure sports or can instead transform it into a form of cultural or social capital.

Qiu Yajun (2014) finds that women who do not participate in leisure sports are mainly held back by their own limitations and structural constraints. Female participants in recreational sports activities frequently encounter perceptual and experience limitations. Xiong Huan (2014) suggests that sport dissolves individual and micro-level limitations of women's leisure, and therefore promotes their freedom of choice and empowerment – this, however, is conditional on a more equal, free and reasonable social system. Only when the cultural environment is present will it be possible to effectively achieve this.

Jinyin Day (2015) finds that the leisure sports activities of Shanghai residents are characterized by activity space circle, activity demand differentiation, wide-area characteristics of travel space and the regularization of activity time. Wu Xiaoyang's (2015) research of the literature demonstrates that leisure sports are one way through which rural citizens integrate into an urban environment. The number of people involved in leisure sports therefore provides insight into the extent of urbanization.

Ye Xin (2015) proposes that gender order is the most important part of women's leisure sports behavior, and suggests it establishes a basis for consciousness while putting in place a material foundation that underpins women's leisure sports behavior. Yan Ke (2016) suggests that the healthy development of leisure sports relies on a benign interaction between individual self-development and social norms. This entails that the psychology and behavior of social individuals will produce their active participation in leisure sports and will encourage their use of social laws, regulations, norms and ethics that control, regulate and stimulate society. Guo Xiujin (2016) observes that the best popular leisure sports involve nature and civilization, and she therefore presents leisure sports as a kind of 'green living style' that helps to make the world a more harmonious place. Yan Ke (2017) notes that individuals make the nation's leisure and sports life possible, and observes the development of national leisure sports promotes an accelerated individualization. Fashion and market tendencies are the social characteristics that correspond to the development of national leisure sports in an individualized era.

Chen Dexu (2017) notes how leisure sports have promoted material development by stimulating sports consumption and economic growth, strengthening social governance, enhancing the stability of political civilization, improving the quality of the population and promoting the healthy development of utility. Sun Fenglin (2018) applies the theory of the ecological food chain network to a case study, and observes that the park can meet the multi-level and multi-type leisure needs of the elderly. Wang Min (2018) suggests that leisure sports are the most active, effective and economical way to deal with an aging society. They can enhance the physical condition, mental health and life satisfaction of the elderly, while reducing their risk of illness. They can alleviate decreased participation in the labour market caused by an aging society, and can also reduce healthcare expenditures and the social burden.

Li Hui (2018) observes that women participate in leisure sport activities, and notes that this does not just challenge the traditional gender order, but also provides a new approach to the construction of this order in the new era. It also enables a healthy China to be constructed along both vertical and horizontal axes. In the first sense, the participation of women in leisure sports activities are based on a challenge to the traditional gender order, the realization of gender role self-identification, an empowered gender freedom, the expansion of gender space and the promotion of women's individuality on terms that encompass the whole life-cycle.

Chinese scholars have a lot of empirical research on the development of human and social development in leisure sports, and there are few theoretical innovations. The research results mainly reflect the subsidiary meaning of leisure sports, and the interpretation of its essential functions is less, and further research is needed.

2.3 Research into leisure sports majors

Since the start of the century, China's comprehensive national strength has continually grown, along with related sport activities. The concept of 'leisure sports' has become increasingly widely accepted. Colleges and universities have also established leisure sports-related majors that help to meet growing demand for professional leisure sports in China.Because the major is newly developing, it is not amateur enough.The construction of China's leisure sports major has gradually become a research 'hot spot'.

Peng Guoqiang (2014) compares the construction of leisure sports majors in American colleges and universities, and proposes that China's leisure sports majors should meet the needs of the social market, cultivate composite talents that combine sports and leisure, broaden the coverage of leisure sports courses and reflect the curriculum.

The dynamic development of the system, the reform and construction of the leisure sports curriculum and the professional setting reflect diversified social needs, the establishment of a third-party specialized agency and the need for a leisure sports professional audit system that is adapted to China. Wang Xiaoyun (2017) observes that current talent training in the construction of leisure sports is not closely integrated with social requirements, neglects the improvement of comprehensive quality, and also gives insufficient attention to professional practice and the low standardization of training programs. He suggests the country should be market-oriented, should analyze demand for leisure sports professional ability, should clarify the training specifications of leisure sports professionals and should build a competency-oriented and standard curriculum system that helps to achieve the overall optimization of leisure sports professionals. Xu Dapeng (2017) conducts a comparative study of leisure sports majors and social sports majors in the Capital Institute of Physical Education, which concludes that the professional setting of leisure sports should focus on innovation and entrepreneurship education, increase the specific sport choices that are available and also inspire students' innovative spirit.

Due to the lack of a correct understanding of the term "leisure sports", Chinese scholars have limited research on leisure sports majors, and the relevant results are not rich, and there is no research result that reflects the connotation of leisure sports.

2.4. The industrialization of leisure sports

The sports and leisure industry has become indispensable to China's national economic development and its construction of a harmonious society. The question of how to align China's leisure sports industry with national conditions has come to preoccupy many scholars. Luo Lin (2006) draws on the perspective of cultural chemistry to put forward three principles for the development of the leisure sports industry, specifically an orientation towards people, the accommodation of multicultural values and the development of a national traditional sports culture and valued leisure industry. Wang Xianliang (2015) finds that China's leisure sports industry is characterized by consumption and production simultaneity, industrial integration and sports characteristics. The development of the leisure sports industry should clarify the industrial value chain, optimize the value chain layout and give regional advantages in order to optimize regional layout. Ye Xiaoyu (2016) finds the research of China's leisure sports industry is developing in accordance with contemporary trends, and adopts a macro-level perspective of the layout of the leisure sports industry, development strategy and policy design. Yang Yukai (2017), after studying the development of the leisure sports industry in developed countries, concludes that China's leisure sports industry needs to acknowledge the leading role of leisure sports culture in the healthy development of leisure sports, grasp the brand effect and integrate the industrial chain; this, he observes, will promote its own development.

Zhao Lefa (2017) uses the 'diamond model' of strategic management to analyze factors that restrict the competitiveness of China's leisure sports industry. He cites the lag of the mass consumption concept, the small scale of the leisure sports industry, limited integration with other industries and a lack of professional talent. In response, he suggests it is necessary to deepen the health concept, strengthen macro guidance, integrate cross-border industry integration and broaden talent training channels. Yang Lei's (2017) comparative research finds that, while the market of China's leisure sports industry has great potential, it has not yet shown short-term economic effects. Lei suggests that the sound development of China's leisure sports industry should be based on an understanding of its basic laws of development and distinctive characteristics. Both should be combined with innovation, coordinated development and scientific management.

The research on China's leisure sports industry mainly focuses on quantitative research. The results of qualitative research are limited. It is necessary to change the research paradigm and produce more theoretically in-depth research results.

When studying China's leisure sports, it is important to explore their function and significance from different angles. The past 20 years have only revealed a few research hotspots and frontier changes in China's leisure sports research, and there is still an ongoing need for further research.

3. Methods

3.1 Study Setting and Data Collection

The Chinese Social Science Citation Index (CSSCI) database was selected and documents from the period 1998-2017 were searched by using the keywords "leisure sports" and "leisure". After screening, 371 documents that satisfied the requirements were selected as this study's dataset.

3.2 Quantitative analysis

This study uses Visual analysis tools and CiteSpace 5.0 to study research hotspots and leisure sports trends in China that have emerged since the reform and 'opening up'. Their application will avoid subjective judgement and will better demonstrate the process and distribution of leisure sports research. The Citespace 5.0 is installed and run on JAVA platform; it is then processed, before the document data is downloaded by CSSCI. The year slice is set to one year and the threshold value is set to 50. Keyword, author and dispatch agency are generated in the corresponding maps and are then used to analyze the corresponding high frequency vocabulary, core authors and research institutes.

Bibliographic information statistical analysis tools, including SATI3.0 and Microsoft Office Excel 2007, are applied. In addition, the number of papers, organizations, authors and other information in Citespace 5.0 software are summarized and counted for purposes of further analysis.

3.3 Qualitative analysis

The high citation-rate literature is cut out from a knowledge map, read in full text and then analyzed through a visual map. The high citation-rate literature cut out in the knowledge map is read in full text, and combined with the visualization map and related papers for analysis. Comment on the context, characteristics and future trends of leisure sports development in China, in order to provide reference for new researchers to discover new research points.

4. Data Analysis

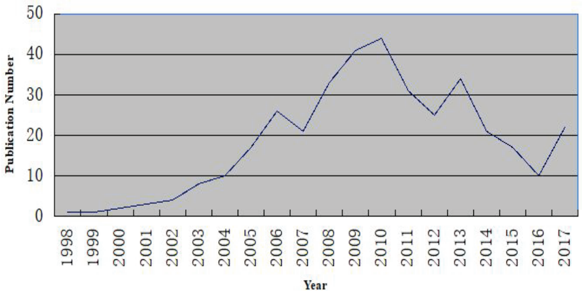


Figure 1: Distribution of leisure sports publications in China: 1998-2017

The time distribution of knowledge domain research can reflect the overall progress and development of this research field. The number of publications during the period 1998-2017 (see Fig. 1) show that, from the end of the 1990s onwards, Chinese scholars began to increasingly engage with the study of leisure sports. The concept of ‘leisure’ was introduced to China for the first time by the scholar Yu Guangyuan[1], who observes that ‘ leisure is an important goal of productivity development. The length of leisure time is parallel to the development of humans[2]’.

After it was gradually acknowledged that sports activities were part of leisure activities, research papers on leisure sports began to appear. Since the beginning of the 21st century, the amount of publications gradually increased, peaking in 2010, and then evidencing a spiral upward distribution trend. Although the volume of papers has declined, the annual number of published papers has remained above 10.

Changes in the number of articles published on leisure sports were closely related to the development of China's economy and society. The World Leisure Organization observes that per capita GDP of $2000 (USD) was the threshold for the rapid growth of leisure demand, and notes that leisure activities began to diversify beyond this point. (Li Xiangru, Ling Ping, Lu Feng,2011) When per capita GDP exceeds $3000 (USD), leisure demand generally arises[3]. In 2006 and 2008, China's per capita GDP respectively exceeded $2000 and $3000 (both USD)[4]. Rapid increases in leisure demand led to the development of related research, and reiterated that leisure sports research is a contemporary concern.

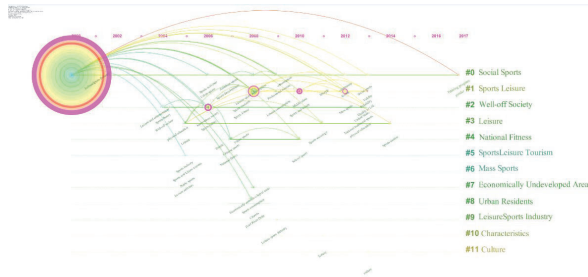


Figure 2: Timeline Analysis of Key Words

Figure3: Emergence Information Analysis: Keywords

Leisure sports was a dynamic research field that was closely related to economic and social development in China. Material progress was a precondition for the development of leisure sports, which impacted related research fields.

In order to present the dynamic changes of the theme of leisure sports research in China, this study uses CiteSpace 5.0 to generate Visualization. After running on smoothly, choosing Cluster View, adjusting Threshold、FontSize、Node Size, and then, choosing Timeline to get the figure No.2. In the figure, nodes in the outside of the annual ring are the key nodes, what they show the most core keywords among the connection of cited literatures. With these keywords, we can conclude the direction and trend of Chinese leisure sports research, in addition, also get the major the source of theory and structure. As we can see in the figure, the area of the circles on different nodes are different, and the composition of the color of the annual ring on different nodes are also different. Over the past 20 years, Chinese scholars have explored the topic of local leisure sports. They have some popular keywords, including Recreation, leisure activities, urban sports, leisure sports consumption, middle class sports consumption, lifestyle, mass sports, social sports, quality of life, training programs, etc. The Time Zone Distribution shows two periods: the first include for parts, one is the conception, theory about leisure sports; one is the deep study about leisure sports to the consumption of leisure sports; one is the analysis the relationship of leisure sports and urbanization, globalization, school sports; one is the discussion about the specific development direction of leisure sports, and the fusion path between leisure sports and different sports forms. The other one is constructing the theory system of leisure sports which includes construction of college sports leisure majors and cultivation of talents.

In the CiteSpace, calculating the burst literature information of ‘leisure sports’ research by clicking Burstness to get the keyword chart (Fig. 3). It can be concluded that different keywords in different periods are different, reflecting the dynamic changes of Chinese leisure sports research hotspots in different time periods. For instance, the keyword ‘well-off’ society, grids 2004-2006 are marked red that correspond to the time of Begin-End on the left, it means there are lots of published papers in this field. But after 2006, the research topic shift to ‘sports culture’ ‘harmonious society’ and ‘sports development’ and so on. During the period of 2010-2013, the research between sports and leisure has been focused, and many research results of leisure sports combining sports theory and leisure theory have been published, reflecting the development of China's leisure sports research toward standardization, discipline, and specialization.

The Time Zone Distribution Chart (Fig. 2) and the keyword chart (Fig. 3) make it possible to identify dynamic changes in research hotspots that have occurred since reform and ‘opening up’. The combination of graphic illustrations and interviews with authoritative experts in the field makes it possible to divide research from the last 20 years into three periods:

4.1 Germination Period (1990s-2000)

Research of leisure sports in China began in the early 1990s[5], when scholars began to focus on sports and leisure. Some of the most influential contributions were Leisure Sports Theory (Cheng Zhili, 1990) and Sports and the Leisure Life of Chinese Urban Residents (Liu Depei, 1990). These papers provided theoretical and empirical insight, and highlighted the relationship between social development and the diversification of leisure lifestyle. Due to conceptual underdevelopment, leisure sports was only considered within ‘sports and health’ and ‘social sports’, and was even viewed as ‘tourism and entertainment’, which further underlined its struggle to become an independent discipline. In 1995, China issued an outline of the National Fitness Program, an outline of the Development of the Sports Industry and the Sports Law of the People's Republic of China.

Although research into leisure sports continued to develop, progress was hindered by the under-development of concepts and the inability of the field to distinguish itself from its predecessors. This meant that research results for leisure sports remained limited[6]. At this time, China’s per capita income was not sufficient to sustain leisure sports, and so related research remained under-developed.

4.2 Initial period (2001-2005)

As China became an increasingly developed country in the 21st century, the research on leisure sports began to progress. Some of the most well-known contributions were: Cultural Significance of the Rise of Leisure Sports (Chen Rong, 2002), Investigation and Development Strategy of Urban Leisure Sports Consumption (Hu Chunwang, 2003), and Sports Leisure Science (Lu Feng, 2005). These contributions considered the significance and value of leisure sports by drawing on cultural, economics, leisure and sports perspectives. At this stage, leisure sports mainly focused on "leisure and entertainment", "mass sports", "well-off society", "entertainment and leisure" and "sports and leisure tourism", in addition to other contents.

4.3 Development period (2006-2017)

During the period 2006-2017, the research hotspots of leisure sports mainly focused on "sports culture", "national fitness", "harmonious society", "American sports development" and "leisure sports consumption". The research results focused on "leisure sports industry" increased substantially. The content involved the basic theory, economic value, educational value, cultural connotation, planning and design of leisure sports. The most influential contributions included Introduction to Leisure Sports (Xu Zongxiang, 2007), On Sports Leisure (Hu Xiaoming, 2008), Thoughts on the Cultivation of Leisure Sports Professionals (Chen Qi, 2008), Thoughts on the Construction of Leisure Sports Specialty in China from the Perspective of Leisure (Li Xiangru, 2009), Formation and Development of Leisure Sports Discipline (Liang Limin, 2010). Introduction to Leisure Sports (Li Xiangru, 2011), Perspective on China's Leisure Sports (Li Xiangru, 2012), Theory and Thinking of China's Leisure Sports Development (Zhong Bingshu, 2015) and Research on China's Leisure Sports Practice (Li Xiangru, 2016).

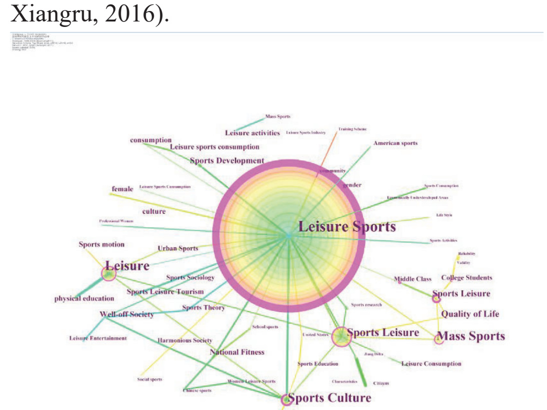


Figure 4: Cluster Co-occurrence Map: Keywords

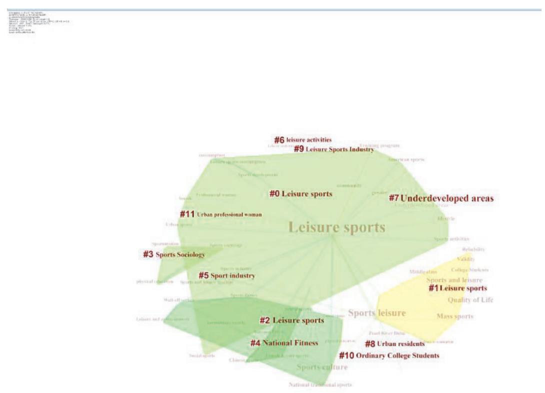


Figure 5: Clustering Common View: Keywords

The keywords in the literature are the core words extracted from the articles. They are core components of an article, and are referenced in the summary and conclusion. Keywords that occur frequency are used to identify hot issues in a research field. The most frequently occurring words collected from CSSCI data are 'leisure sports', 'sports leisure', 'leisure', 'sports culture' and 'sports leisure' (Fig. 4). This shows that, since the reform and 'opening up', scholars in the field of leisure sports have mainly focused on these contents, with the consequence that the research scope of leisure sports in China has generally progressed. Foreign experience, meanwhile, helped to stimulate domestic demand, increase employment and improve national happiness. But extended research into leisure sports in China was still insufficient. There were too many theoretical macro-studies and too few empirical micro-papers. Some leisure sports research directly copied from foreign research, and the results were divorced from China's national conditions.

Leisure sports was a compound subject produced by the combination of 'Leisure Science' and 'Sports Science'. The current tendency to integrate disciplines made it clear that leisure sports cannot exist as an isolated island, and therefore needs to be combined with different disciplines to produce new academic breakthroughs. The research scope of leisure sports is currently mainly combined with the contents of 'sports industry', 'national fitness' and 'gender' (Fig. 5[①]). However, there are too many quantitative studies, too few qualitative studies, insufficient research paradigms, no significant interdisciplinary research and insufficient integration with sociology, pedagogy, economics, anthropology and other disciplines.

For example, an exploration of the social significance of leisure sports can draw on sociology, while the behavioral motivations of leisure sports participants can be engaged from the perspective of Social Action Theory. The symbolic meaning of leisure sports can be explored from the perspective of Symbolic Interaction Theory, and the interactive and communicative value of leisure sports can be explored from the perspective of Daily Life Theory. Leisure sports can also be analyzed from the perspective of pedagogy. The interactive mode of leisure sports can be constructed with the theory of experiential learning, and the relationship between leisure sports and human status acquisition and identity can be analyzed by drawing on the theory of educational stratification and conflict. Leisure sports can also be analyzed by drawing on the theory of economics, while the economic value of leisure sports can be explored through theories of human capital and social reproduction. Anthropological perspectives can be drawn on to assess the cultural significance of leisure sports.

Figure 6: Co-occurrence Map: Research Institutions

Leisure sports have emerged as a 'hot spot' of sports research in recent years. Analysis of the common view spectrum of leisure sports research institutions can provide insight into the core institutions that promoted leisure sports research, and this can help researchers to better understand related research fields from the perspective of research institutions.

In the citespace5.0, selecting time interval '1998-2017', node type "institution", setting the threshold to top 50 per slice, selecting MST (minimum spanning tree) algorithm to simplify the network, and finally get 10 nodes, 0 links, density 0 research Institutional map(Fig.6). The nodes are ring-shaped. The larger the nodes, the larger the font of the research institution, indicating that the overall frequency of the institution is higher. The thickness of the annual ring is proportional to the frequency of institutional words in that year. Some rings have a purple outer ring, which means that research institutions have a greater degree of centrality. The connection between institutions is zero, indicating that there is no cooperation between two or more institutions.

Leisure sports research in China is concentrated in different universities (Fig.6), of which sports and comprehensive universities account for the highest proportion (five and three respectively). Sports colleges and universities tend to focus more on leisure

[①] The reason why there are two 'leisure sports' in the fig.5 is because the Chinese meaning of them are the same 'leisure spots' and 'leisure and sports'.

sports research. Comprehensive colleges and universities, meanwhile, engage on an interdisciplinary basis, adopt diverse research perspectives and have a clear advantage with regard to the quality of research. China's leisure sports research institutions tend to be based in economically developed cities, and this reflects both economic thresholds and the objective law of disciplinary development.

While leisure sports research in China has given rise to different views, the total number of views has not been high; furthermore, the nodes of institutions have been relatively isolated, with no connection between them. Leisure sports research institutions have fewer cross-regional alliances, low efficiency of resource integration and greater disciplinary limitations, none of which are conducive to cross-disciplinary integration of leisure sports research. Research institutions should accordingly strengthen cooperation with the intention of achieving complementary research resources and should work to promote the development of leisure sports research.

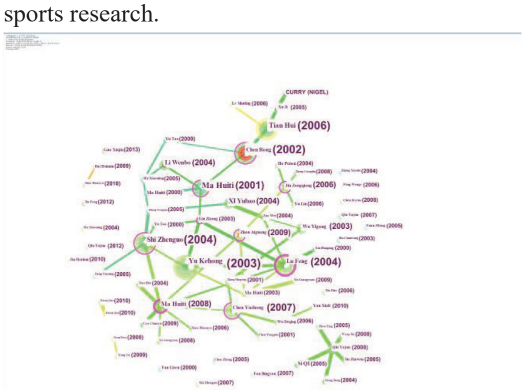
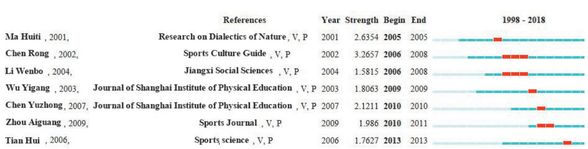


Figure 7: Co-occurrence map: highly-cited literature



	References	Year	Strength	Begin	End	1998 - 2018
Ma Huiti, 2001,	Research on Dialectics of Nature, V, P	2001	2.6354	**2005**	2005	
Chen Rong, 2002,	Sports Culture Guide, V, P	2002	3.2657	**2006**	2008	
Li Wenbo, 2004,	Jiangxi Social Sciences, V, P	2004	1.5815	**2006**	2008	
Wu Yigang, 2003,	Journal of Shanghai Institute of Physical Education, V, P	2003	1.8063	**2009**	2009	
Chen Yuzhong, 2007,	Journal of Shanghai Institute of Physical Education, V, P	2007	2.1211	**2010**	2010	
Zhou Aiguang, 2009,	Sports Journal, V, P	2009	1.986	**2010**	2011	
Tian Hui, 2006,	Sports science, V, P	2006	1.7627	**2013**	2013	

Figure 8: Co-occurrence information analysis

Frequently cited contributions to the literature can reveal the knowledge base of leisure sports research (Fig. 7). The most frequently cited paper is Yu Kehong's 'On Leisure Sports from the Definition of Leisure', which was published in China Sports Science and Technology in 2003. It mainly compares and summarizes the definitions of 'leisure' and 'leisure sports' put forward by Chinese and foreign scholars, with the intention of developing a more objective, comprehensive and clear definition[7]. The second paper is Tian Hui's 'Leisure, Leisure Sports and Its Development Trend', which was published in Sports Science in 2006. It mainly analyzes and interprets the meaning of leisure and leisure sports by tracing the historical development of leisure sports and discusses its content[8]. The third is 'Cultural Significance of the Rise of Leisure Sports' by Chen Rong, which was published in 2002 in Sports Culture Guide. This paper adopts a cultural perspective to discuss the cultural characteristics and significance of leisure sports[9]. Shi Zhenguo published 'Leisure, Leisure and Leisure Sports' in Sports Culture Guide in 2004. It analyzes the historical origins of leisure and leisure sports, and also engages the concept of leisuresportsactivities[10]. In 2004, Lu Feng's 'Discrimination of Leisure Sports Concepts' was published in the Journal of Chengdu Institute of Physical Education. It defines the concept of leisure sports and incorporates the author's personal opinions to establish three constructive dimensions of leisure sports[11].

Forward-looking research has also received extensive attention in the research community (Fig. 8). The authors' understanding of the corresponding time period in Figure 8 is consistent with that in Figure 3.This figure mainly reflects the time period in which the academic achievements of the highly cited authors of China's leisure sports research are heavily cited. It is not difficult to see that many research results are forward-looking, and they have received a lot of attention and citations many years after the paper was published.

Ma Huidi's '21st Century and Leisure Economy, Leisure Industry and Leisure Culture' was published in Dialectics of Nature in 2001. It anticipates a stronger interrelation of the leisure industry and China's culture and economy during the 21st century, and also calls for the strengthening of relevant academic research[12]. It was widely cited in 2005. Li Wenbo's ' Leisure Sports Consumption Research: An Interpretation of Culture and Sociology', which was published in Jiangxi Social Sciences in 2004, proposes that sports culture has a unique cultural symbolic significance[13]. It was widely cited during the period 2006-2008.

Wu Yigang's 'Current Situation and Problems of Leisure Sports Research at Home and Abroa'" was published in Journal of Shanghai Institute of Physical Education in 2003. It studies changes within leisure sports and their general development of leisure sports at both the domestic and international level, and also acknowledges limiting factors in each of these respects[14]. It was quoted extensively in 2009. Tian Hui's 'Leisure, Leisure Sports and Its Development Trend in China' was published in Sports Science in 2006. It traces the origin and evolution of leisure sports in China and also explains their significance and content[15]. It was widely cited in 2013. Chen Yuzhong's ' Future Trend of Leisure Sports Development in China' was published in Journal of Shanghai Institute of Physical Education in 2007. It explores the conditions and historical stages of the rise of leisure sports in China, and provides insight into their future prospects[16]. It was widely cited in 2010.

These frontier contributions have made a vital contribution by establishing a basis for the future development of leisure sports research.

5. Conclusions and Future Research

5.1 Conclusions

The scientific knowledge map of core journal papers on leisure sports from the past 20 years is subject to analysis by Citespace 5.0 software, and this confirms that research on leisure sports spiraled upwards from the beginning of the 21st century, before peaking in 2010. Although a decline then followed, a relatively constant annual output was maintained. The research process of leisure sports was hierarchical and closely related to social development and policy dynamics.

Leisure sports research institutions mainly focused on professional sports colleges and comprehensive universities, while financial and economic colleges and normal universities paid less attention to leisure sports. There was less cooperation among scientific research institutions, as effective cooperation mechanisms were still absent. Researchers of leisure sports, including Chen Rong, Yu Kehong, Shi Zhenguo, Lu Feng, Tian Hui, Chen Yuzhong and Zhou Aiguang, contributed to the study of leisure sports by offering different perspectives and methods. But cooperation between different authors was still not sufficient. The hotspots and frontiers of leisure sports research were mainly based on policy, history and culture, and focused on industry, history and comparative sports. This research generally tended to be systematic, pluralistic and innovative.

Although substantial achievements have been made in the research of leisure sports in China, there are still some problems and shortcomings: The connotation of leisure sports research is still not sufficiently deep, and most contributions struggle to extend beyond a relatively superficial concept analysis or assessment of prospects. The results of interdisciplinary research have been insufficient to form horizontal and vertical cooperation mechanisms. In addition, there is also a gap between concrete developments and research. For example, as national fitness has improved, leisure sports items, including yachts, hot balloons, racing cars and RV camping, have also increased; this development has not, however, been reflected in domestic research. Although there have been individual breakthroughs in research paradigms, methods and perspectives, the overall impact has fallen short of what is required.

5.2 Future Research

In the future, the research on leisure sports can further improve the insufficiency of this research. The depth of connotation of leisure sports research is insufficient. It can be traced from the meta-research level, correctly define the meaning of leisure sports, and carry out theoretical innovation. Interdisciplinary research on leisure sports Too few results should be led by the government level, so that different disciplines can strengthen cooperation and form a sustainable cooperation mechanism; research on leisure sports should be paid attention to, and many emerging leisure sports should receive more attention, such as: Outdoor sports, yachts, motor homes, etc.; enrich the research paradigm of leisure sports, emphasizing the theoretical depth and practical value of leisure sports research through the perspective of qualitative research.

When compared against previous research, this research offers a significant methodological innovation. Its application of visualization software has shown how it is possible to provide an objective representation of the frontiers and hotspots of leisure sports development in China over the last 20 years. Visualization analysis saves a considerable amount of time in accessing the literature, and also makes it possible to extract research hotspots and deficiencies in a certain research field from a large number of papers. In future, it can be both applied to other disciplines and also used to open up new areas of research into leisure sports.

References

[1] Wang Liwei J., Zhang Yongjun, Fan Suxiang. (2007) Leisure Sports: A Historical Category of Interactive Economic Development. Journal of Chengdu Institute of Physical Education, 33 (5), pp. 5-8.

[2] Yu Guangyuan M. (2005) On the Generally Leisure Society. Beijing: China Economic Publishing House, pp. 12-16.

[3] Li Xiangru M., Ling Ping, Lu Feng. (2011) Introduction to Leisure Sports. Beijing: Higher Education Press, pp. 11.

[4] China's annual average GDP Z. https://www.kuaiyilicai.com/stats/global/yearly_per_country/g_gdp_per_capita/chn.html.

[5] Lu Feng J., Liu Xishan, Wen Xiaoyuan. (2006) Classification of Leisure Sports Activities. Journal of Wuhan Institute of Physical Education, 40 (12), pp. 59.

[6] Li Xiangru M., Ling Ping, Lu Feng. (2011) Introduction to Leisure Sports. Beijing: Higher Education Press, pp. 160.

[7] Yu Kehong J., Liang Ruowen. (2003) On Leisure Sports from the Definition of Leisure. China Sports Science and Technology, 39(1), pp. 21-23.

[8] Tian Hui J., Zhou Hong. (2006) Leisure and leisure sports and their development trend in China. Sports Science, 26 (4), pp. 67-70.

[9] Chen Rong J. (2002) Cultural Significance of the Rise of Leisure Sports.Sports Culture Guide, (2), pp. 12-13.

[10] Shi Zhenguo J., Tian Yupu. (2004) Leisure, Leisure Sports. Sports Culture Guide, (8), pp. 45-46.

[11] Lu Feng J. (2004) Discrimination of the Concept of Leisure Sports. Journal of Chengdu Institute of Physical Education, 30 (5), pp. 32-34.

[12] Ma Huidi J. (2001) The 21st Century and Leisure Economy, Leisure Industry and Leisure Culture. Research on Dialectics of Nature, 17 (1), pp. 48-52.

[13] Li Wenbo J. (2004). Research on Leisure Sports Consumption: An Interpretation of Culture and Sociology. Jiangxi Social Science, (9), pp. 154-157.

[14] Wu Yigang J. (2003). Current situation and problems of leisure sports research at home and abroad. Journal of Shanghai Institute of Physical Education, 27 (3), pp. 39-43.

[15] Tian Hui J., Zhou Hong. (2006). Leisure and leisure sports and their development trend in China. Sports Science, 26 (4), pp.67.

[16] Chen Yuzhong J. (2007). Future Trend of Leisure Sports Development in China. Journal of Shanghai Institute of Physical Education, 31 (1), pp. 9.

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

An Effective and Efficient Machine-Learning-Based Framework for Supporting Event Detection and Analysis in Complex Environments

Alfredo Cuzzocrea[a,1], Enzo Mumolo[b,2]

[a] iDEA Lab, University of Calabria, Rende, Italy
[b] University of Trieste, Trieste, Italy

ARTICLE INFO

Article History:
Submitted 4.8.2019
Revised 4.30.2019
Second Revision 5.20.2020
Accepted 7.21.2020

Keywords:
fall detection
fall classification
Neural Network
MEMS accelerometer
Monte Carlo algorithm

ABSTRACT

In this paper we describe a falls detection and classification algorithm for discriminating falls from daily life activities using a MEMS accelerometer. The algorithm is based on a shallow Neural Network with three hidden layers, used as fall/non fally classifier, trained with daily life activities features and fall features. The novelty of this algorithm is that synthetic falls are generated as multivariate random Gaussian features, so only real daily life features must be collected during some day of normal living. Moreover, the features related to synthetic fall events are generated as complement of normal features. First of all, the features acquired during daily life are clustered by Principal Component Analysis and no Fall activities shall be recorded. The complement set of the normal features is found and used as a mask for Monte Carlo generation of synthetic fall. The two feature sets, namely the features recorded from daily life activities and those artificially generated are used to train the Neural Network. This approach is suitable for a practical utilization of a Neural Network based fall detection characterized by high Recall-Precision rate.

1. Introduction

The detection of falls of the elderly and people with deseases like epilepsy or Parkinson or simple people with motor difficulties, is today a problem of great public interest. This generated a wide range of research and led to the development of various falls detection and tele-monitoring systems to allow prompt intervention when a fall occurs. Studies in the past years have shown that 1/3 of Senior citizens over age 65 ([28]) are often victims of undetected falls and most of their injuries are due to a lack of intervention. Generally speaking the classification between daily life activities and falls is a difficult task because many daily life activities look like fall (for example running, sitting in a car or lying in bed) and many falls may look like daily life activities. All the types of errors of these systems are of great importance. In case of False Positive error, users are not motivated to use the falls detection system because in many normal life activities are wrongly detected as falls. In those cases, the operator soon gets tired of the false alarms. On the other hand, in case of False Negatives errors it happens that the system leads to lack of interventions in case of fall. Of course this situation is followed by problems of serious injury and also of mortality.

Generally speaking, while it is quite simple to gather accelerometer data during Normal Daily Living (Daily Life Activities or DLA), it is very difficult if not impossible to collect data during Falls, so the question is: how can the false positive rate be reduced if enough fall data is not available? The answer to this question is that we produce synthetic fall data starting from data collected during normal daily living.

Our approach reduces the amount of False Positives compared to threshold based systems by performing extensive training of a neural network if large quantities of features are gathered during normal life activities.

In the last decade, there has been a great deal of research that has examined the use of inertial sensors such as accelerometers and/or gyroscopes to realize systems for automatic detection of falls. The goal of these systems is to detect the falls of patients who have any difficulty in walking to quickly alert the operators who provide a suitable assis-

alfredo.cuzzocrea@dia.units.it (A. Cuzzocrea); mumolo@units.it (E. Mumolo)

DOI reference number: 10-18293/JVLC2020-N1-023

tance. The characteristics these systems must have are on one side low cost, consumption and size, and on the other side high performance. While the first characteristics are always better met, the performance is still unsatisfactory. Typically, performance is measured in false positives and false negatives. False positives, if excessive, could demotivate the use of these systems because the recipient of a fall warning could get tired of it. Clearly, false negatives are the most dangerous for the patient's health because they correspond to missed falls. Normally these errors are measured by Sensitivity (ability to detect actual falls) and Specificity (ability to avoid false positives) or equivalently Recall and Precision. From the literature, the existing technology for falls detection systems can be roughly classified into three categories.

- With wearable sensors: These systems use a triaxial accelerometer or gyroscope or a combination of both to estimate the posture of the subject's body. The sensors are placed in different places, such as the waist, the thigh, the wrist, the shoes. Many systems use the smart-phones inertial sensors.
- With environmental sensors: The environmental sensors are nothing more than sensors positioned around the subject. Floor sensors such as pressure sensors on the mat, microphones, infrared sensors, microwave motion detectors are used to detect the fall. The classifiers of neural networks are used to classify daily activities.
- With image processing: In such methods, a camera is used to monitor body postures. Falls are detected using various image processing techniques such as pattern matching, posture recognition, skeleton extraction, background subtraction, optical flow processing, etc.

A fall can be described as the rapid change from standing or sitting towards an elongated position in earth or almost elongated [27]. This definition has been used in many studies. The paper is organized as follows. In Section 2 we describe the state of the art. In Section 3 we provide the main features of the accelerometer. In Section 4 we introduce preliminary definitions useful to describe our work. In Section 5 we describe our main approach. In Section 6 we provide the proposed metrics for performance evaluation. In Section 7 we provide the experimental evaluation and assessment of our proposed framework. Finally, in Section 8 we discuss concluding remarks and future work of our research. A preliminary version of this paper appears in the short paper [7].

2. Previous Work

The algorithms used in systems with wearable sensors can be divided into approaches that use thresholds based heuristics and approaches that use machine learning tools [26]. The latter may be k-Nearest Neighbor (kNN), neural networks, hidden Markov models, or the two classes classification schemes based on the Support Vector Machines (SVM) classifiers. In all cases, however, it is important to have both falls and daily life activity features. In the case that the approaches are based on thresholds, the availability of both types of data is important to find optimal thresholds, while in the other case the data are important for correctly training the machine learning algorithms.

The thresholds-based heuristics approaches are methods that use thresholds and appropriate functions derived from inertial parameters. The simplest approach to detecting a fall could be to detect the ground position of the person, by means of a horizontal inclination detection sensor. This method is suitable for monitoring "isolated subjects" but less suitable for the detection of falls of an elderly person in his home environment as the hours of sleep are not regular. Therefore this method provides many "false positives". A complementary solution is to detect the person lying on the floor, using floor tiles equipped with sensors. But when the falls do not end on earth, or if the floor does not have these sensors, obviously they are not detectable. When it falls, the person often hits the ground or an obstacle. The "shock impact" causes an intense inversion of the polarity of the acceleration vector in the direction of the trajectory, which can be detected with an accelerometer or a shock detector, which is actually a threshold accelerometer. Although most of the falls occur in the "front" plane (forward or backward), the direction of the trajectory of fall and is obviously variable from one fall to another. Also the position of the sensor on the body relative to the point of impact modifies the signal recorded at the moment of shock. The lack of movement can be used to detect the fall as, after the "serious" fall, in which the person can be seriously injured, they often remain immobilized in one position. A motion / vibration sensor, positioned on the body (e.g., wrist or ankle), can be used or, again more simply, infrared sensors of presence disseminated in the home. The choice of latency time (the delay before the decision) that should be long enough to reduce "false positives", which translates into a longer delay before intervention, represents a critical problem for these approaches. As discussed earlier, during a fall there is a temporary "fall free" period, during which the vertical speed increases linearly over time due to gravitational acceleration. If you measure the vertical speed of the normal movements of the person (getting up, lowering, sitting down), you can discriminate these speeds from what you do during the fall, which would exceed an appropriate threshold. The intrinsic of analytical methods lies in the choice of this threshold, which if too low causes "false positives" and if too high causes "false negatives". Also this threshold differs from subject to subject. Image processing of video signals can also be used to detect one fall identifying the lying posture using analysis of the visual scene or detecting brusque movements using the revelation of movements with respect to the background. The latter method usually consists in subtracting successive images to keep only the variations, which are then sorted according to their direction and or their width. While these techniques are well established in controlled environments (e.g., laboratory), they must be modified in environments uncontrolled where parameters such as lighting or framing can be arbitrary. Furthermore, if the subject moves in a 3-dimensional space, it may need more complex techniques, namely the use of 2 cameras ("VisionStereo"). These image techniques are absolutely feasible at present, both technically and economically, thanks to the presence on the market of low cost cameras (web cam), with the possibility to transmit images in wireless mode over short distances and availability of appropriate processing algorithms. However the acceptance of these technologies for images poses big problems of privacy, as it requires the positioning of video cameras in the living space of the person, and in particular in the bedroom and the bathroom.

An alternative to heuristics approaches, machine learning methods can be used to detect falls. These methods are based on the data acquired on the real working system, used in a preliminary training phase. "Supervised" or "unsupervised" classification algorithms can be used. In case of classification algorithms with "supervised" learning, the person wearing the device performs a series of voluntary actions in order to identify the parameters in the normal cases. In the case of "unsupervised" learning, it is possible to record the movements of the person, within a few hours or several days, and then perform a statistical analysis of the measured speed. These approaches to developing fall detection algorithms are based on observing of the data (the training period) and then on the classification. The choice of classification algorithms is very broad. If you use a supervised method, a simpler choice is to train a neural network, which will then be used to automatically classify the signal. Only the situations encountered during training can be recognized, all others can be shuffled into a class called "others" if the algorithm " is " unsupervised ", falls can be isolated if the training period is much longer than the fall event. Furthermore, it is likely that the first event fall is not detected since its class is still unknown before its premiere appearance.

Regarding the thresholds based heuristics approaches, Bourke *et al.* [3] uses signals from triaxial accelerometers mounted on the trunk and the thigh to distinguish falls from the *Activities of Daily Living* (ADLs). They propose a higher fall threshold (UFT) and a lower fall threshold (LFT) in an attempt to optimize the balance of false positives and false negatives. Likewise, Kangas *et al.* [16] attached a triaxial accelerometer to the waist, wrist and head of volunteers who performed simulated drops and ADLs in laboratory. Their algorithms considered the phases of pre-impact, impact and post-impact of the fall, separately and in combination, and achieved up to 100% of specificity and sensitivity of 95%, using a single sensor mounted at the waist. However, this algorithm has not been tested in real environments. The only study that examined its accuracy in the real world was conducted by Bagala *et al.* [9], which evaluated fall detection methods (including the Bourke and Kangas algorithms described above) using data from real falls, achieving much better results. In laboratory settings, the development of improved algorithms for automatic fall detection in the elderly requires an understanding of real-life fallout scenarios in older adults and the integration of such information into the design of laboratory experiments. The common fall scenarios are often absent in the majority, if not in all, of the previous laboratory experiments of fall, and the consequent discrepancy in the sensor data is, perhaps, the main cause of the lack of accuracy of the fall detection algorithms, when tested on real scenarios. In [12] a system is described that uses a triaxial accelerometer and gyroscope. The detection algorithm uses three thresholds: one to recognize pre-fall situations, one to detect the maximum of the acceleration vector module and one to detect the maximum angular velocity. The algorithm described in [13] also uses three thresholds, one for the local minima of the acceleration module and two for the local maxima of the acceleration module and the angular velocity module. Systems based on machine learning algorithms use classifiers that are trained with both ADL and falls data. However, in a realistic context, due to the lack of sufficient availability of falls data and the lack of knowledge and understanding of what could be the falls, approaches based on the detection of anomalies and classification of a single class can be used. These techniques can not identify falls directly because fall data is not available for classifier training. However, they can identify falls indirectly by classifying them as abnormal activities. In these approaches, therefore, abnormal activities are classified as deviations from normal behavior. Naturally, the concept of normal activities must be clearly defined to identify abnormal activities. Moreover, even if the data of normal activities are not sufficient, then these techniques can produce excessive false positives. Recent research projects [5], [25], [19] show that falls can be identified without actually acquiring them. As evidenced by Klenk et al. [14], simulated falls differ significantly from real-world falls. Thus, having simulated falls in the training data set could lead to achieve classifiers that show different behaviors with real-world falls However, other authors such as Zhou and others [24] have presented a method to detect falls using transitions between activities to model falls. Zhou and others trained supervised classifiers using the normal activities collected by a mobile device, then used transitions between these activities to train a One-class Support Vector Machine (OSVM) and showed that it performs better than an OSVM trained only with activities normal. Micucci et al. [25] evaluates methods of detecting falls that do not require dropping data during training on different data sets collected using the smart-phone accelerometer. Their results show that in most cases, the One Class k-Nearest approach Neighbor classifier OCNN behaves better or equivalently to supervised SVM and KNN classifiers that require both types of data, i.e. data for normal and abnormal activities. In other words, Micucci et al. use the one-class k-Nearest Neighbor (kNN) classifier and the one-class SVM classifier. These classifiers have been trained only with ADL and FALL instances. If the anomaly score is higher than a given threshold, the new instance is classified as an anomaly / fall, otherwise is classified as an ADL. Micucci et al compares the anomaly detectors with a two classes kNN and a two-classes radial basis SVM. These classifiers have been trained and tested with both instances of ADL and FALL. This is the case that we are looking for in a real scenario. The main contribution of [25] is the discovery that to design an effective method of detection of falls, it is not necessary to acquire data of falls but it is sufficient to classify the test data as anomalous. As for the HMM models, the traditional way to detect unseen abnormal activities appears as a model of normal activity using an HMM, appears the likelihood of a test sequence with the trained models and if it is below a pre-defined threshold then identify it as an anomalous activity [23]. Another common method to detect anomalous activities is to model the normal activities by a common HMM instead of modeling them separately. In [20] two HMM algorithms are presented that are normal HMM, in which the system noise covariance of the normal dynamics is used to determine the region with highest likelihood which are far from normality based on which events can be classified as 'not normal'. Their results show high detection rates for falls on two activity recognition data sets, albeit with an increase in the number of false alarms. In [19], Khan *et al.* experimentally show that this approach can give better results than supervised classification with limited fall data. When the number of fall data increases, the performance of supervised classifiers improves, but falling data collection can take a long time.

3. Accelerometer Features

The output of the MEMs accelerometer are the three component of the acceleration vector according to the three axis x, y, z, namely a_x, a_y, a_z, each of them related to the current time instant. From this signal, many features are extracted, see for example [17, 30]. We first compute the modulus of the acceleration vector, namely $Acc = \sqrt{a_x^2 + a_y^2 + a_z^2}$. Let us give a look to Figure 1 which is the time evolution of Acc for a typical fall. In this case, it is a Fall backward while trying to sit on a chair, taken from Mobifall v.2. It is worth noting that the overall time frame is the typical fall time. Here we choose the following measures: the maximum value of the modulus, labeled as ***Peak*** in Figure 1, the length between the two arrows, labeled as ***Base***, and the modulus of the slopes of the three components within the signal frame. The slope is computed as follows. Calling t^i_1, t^i_{1+N} the first and last time instant of the N samples i-th frame, let us consider the values:

$$
\begin{aligned}
max_{a_x} &= max\{a_x(t^i_1), \ldots, a_x(t^i_{1+N}\}, \\
min_{a_x} &= min\{a_x(t^i_1), \ldots, a_x(t^i_{1+N})\}, \\
max_{a_y} &= max\{a_y(t^i_1), \ldots, a_y(t^i_{1+N}\}, \\
min_{a_y} &= min\{a_y(t^i_1), \ldots, a_y(t^i_{1+N})\}, \\
max_{a_z} &= max\{a_z(t^i_1), \ldots, a_z(t^i_{1+N})\}, \\
min_{a_z} &= min\{a_z(t^i_1), \ldots, a_z(t^i_{1+N})\}.
\end{aligned}
$$

Then, in the interval t^i_1, t^i_{1+N}, the slopes of the three components are: $slope_x = max_{ax} - min_{ax}$, $slope_y = max_{ay} - min_{ay}$, $slope_z = max_{az} - min_{az}$. The modulus of the slope component is

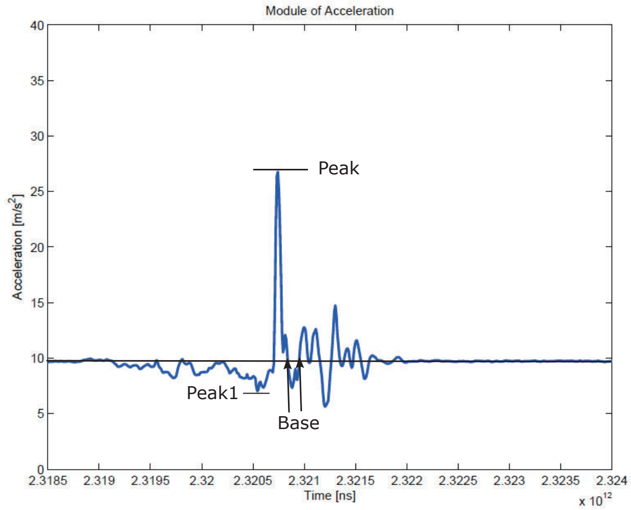


Figure 1: Features Extracted from the Graph of the Module

$$Slope = \sqrt{slope_x^2 + slope_y^2 + slope_z^2} \tag{1}$$

Another feature we use is the Ratio of the Peak over Base, as described in (2)

$$Ratio = \frac{Peak}{Base} \tag{2}$$

These features have been chosen because they require very low computation, and so they can be used also on embedded processors with very little computational power.

4. Preliminary Definitions

We now make some preliminary definitions useful in Section 5. Assume we use N features describing Falls and ADL.

The Features Space (FsS) is an hyper-cuboid with 2^N vertices and $2 \cdot N$ sides where all the original features points lie. Calling $max(feature_i)$, and $min(feature_i)$ respectively the maximum and minimum values the $i-th$ feature can reach in the current case, the length of the first side of the hypercuboid is $max(feature_1) - min(feature_1)$, of the second side is $max(feature_2) - min(feature_2)$ and so forth. The two vertices V_1 and V_2 with respectively the minimum and maximum Euclidean distance from the origin have coordinates $V_1 = (min(feature_1), min(feature_2), \ldots, min(feature_N))$, $V_2 = (max(feature_1), max(feature_2), \ldots, max(feature_N))$ respectively. Of course, if we have only two features, FsS is a rectangle and if we have three features, FsS is a 3D cuboid.

$DLAS \in FsF$ is a set whose elements are vectors of features collected in one or more days of Daily Living Activities (DLA).

$FAS \in FsF$ is a set whose elements are vectors of features collected during Fall Events.

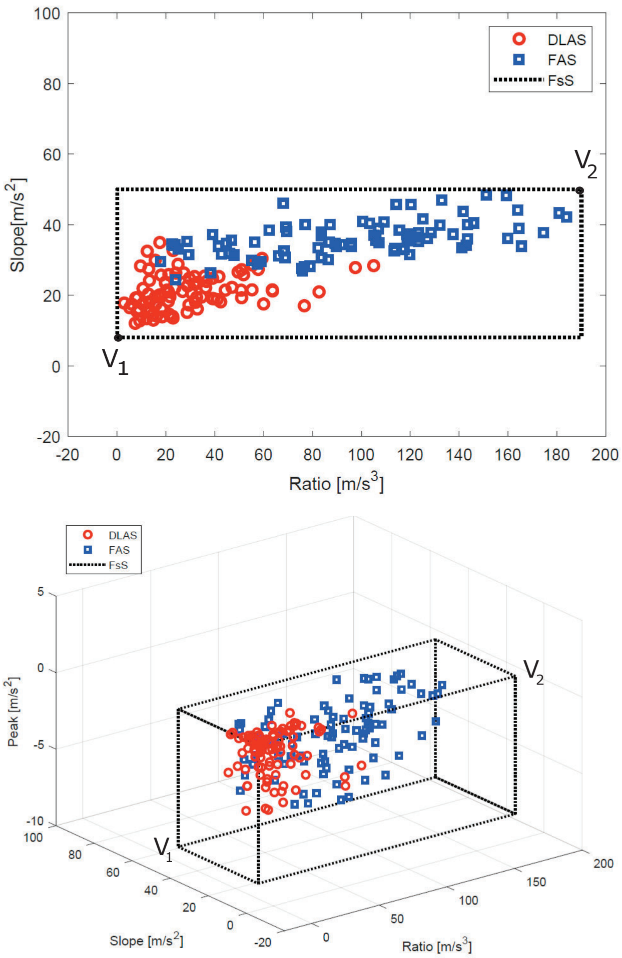


Figure 2: Example of 2D and 3D Features Obtained in Daily Living Activities and Fall Events

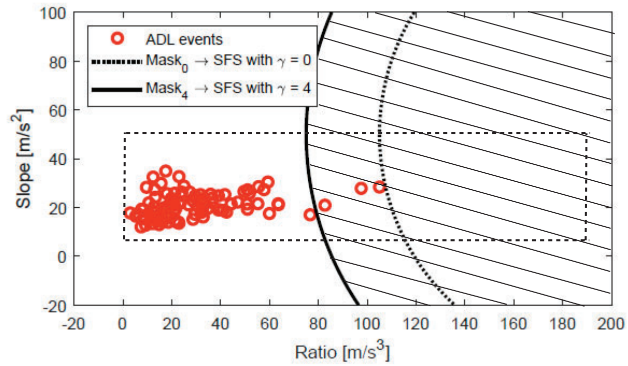


Figure 3: Example of Two Synthetic Fall Sets with Different Cardinality

In Figure (2) we report an example of FsF, $DLAS$, FAS and vertices V_1, V_2 for two and three features respectively. $SynFS \in FsF$ is a set whose elements are features representing Synthetic Falls, and C_{SFS} is its Cardinality, i.e. the number of its elements. The contours of such sets will be used as $Masks$ in Monte Carlo synthetic generation of falls. As stated before, our assumption is that $SynFS$ can be viewed approximately as a complement to $ADLS$, provided that certain conditions are verified. $SynFS$ is represented by an N-dimensional sphere with center in V_2, which is the vertices of FsS most distant from the origin. The radius of SFS, initially equal to zero, is found with an iterative approach which increase its value until the desired number of elements of $DALS$ are included in it. In other words, $SynFS$ is defined as follows:

$$SynFS = \{z | z \in D\bar{L}AS \cup SEDLAS \wedge C_{SFS} \leq \gamma\}$$

where $SEDLAS$ set contains some elements of $DLAS$, chosen as explained shortly.

Let us look at the Figure 2 and Figure 3. Here we assume that only two features are used, so the sets can be drawn as a 2D plane. The points represented with squares are ADL elements. The circles' boundaries are related to some SFS with different Cardinalities and define two $Masks$ which will be noted as $Mask_{Cardinality}$. We recall now the definition of Delation, which is a useful operator from Binary Morphology [18, 22]. The Delation of a given matrix A by a structuring element B, represented with the $\oplus$ symbol is defined as follows: $A \oplus B = \{x | B \cap A \neq \emptyset\}$.

5. Description of the Approach

Our goal is to generate synthetic falls by using a Monte Carlo algorithm. In other words N-dimensional random vectors from Gaussian Distribution are generated and filtered by the Masks described above so that only the vectors falling within the Mask are retained. In order to reduce the computational complexity of the Monte Carlo algorithm, binary operations are performed.

First of all, Principal Component Analysis [1, 15] of the data contained in $DLAS$ set is performed in order to cluster DLA data. To clarify the algorithm description let us assume that only two features are used. Using the data of Figure 2 by PCA we approximate the shape of the data with the ellipse depicted in Figure 4. It is worth observing that the major and minor axis of the ellipse are the first and second components of the data.

The binary operation starts by a binary version of the PCA ellipse, which is simply a projection of the PCA ellipse on a 100×100 binary matrix. It is worth noting that here we make use of binary data for complexity reduction. The ellipse is filled with ones. This is shown in Figure 5, where the binary matrix corresponding to Figure 4 is depicted. The set whose elements are the pixels inside the binary ellipse is called BE for binary ellipse. Then, a mask is generated with Morphological Delation using n-dimensional polytope structuring elements. It is worth noting that we use delation instead cicles because the shapes are easier to binarize. In the 2D example we use an octagon. Starting from an octagon in V_2, a rough approximation of the spherical boundaries shown in Figure 2 are obtained using the following it-

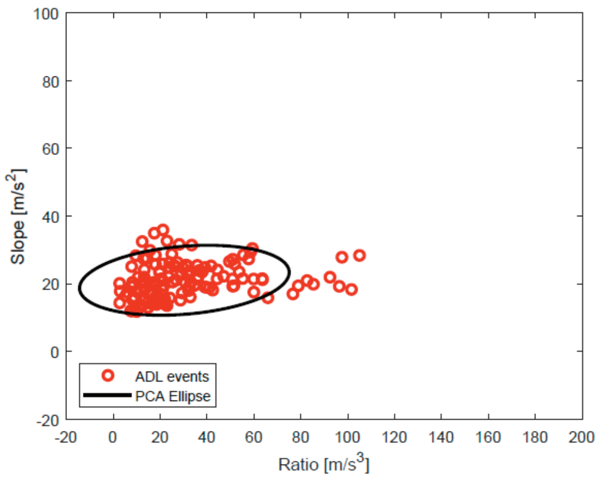


Figure 4: Shape of the Data Captured by PCA

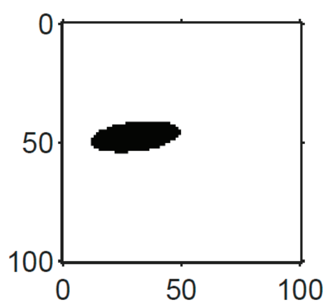


Figure 5: PCA Shape as a Binary Matrix

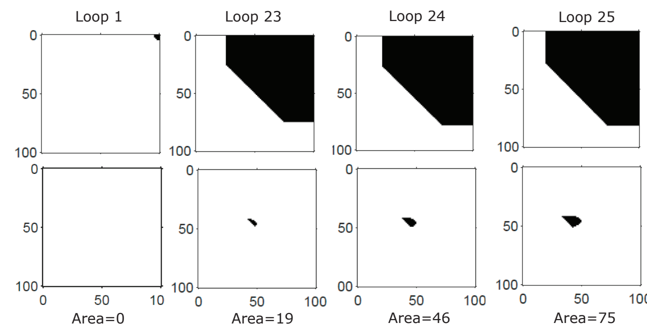


Figure 6: Portion of the Sequence of Masks for Increasing Walues of γ

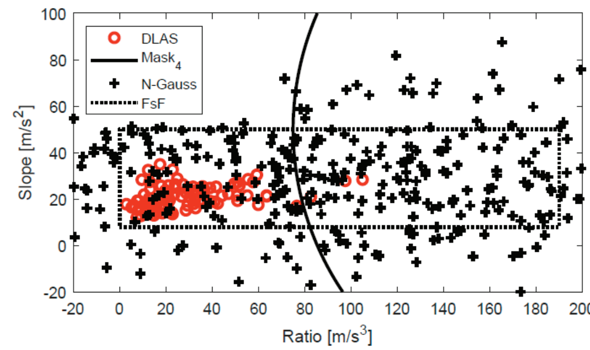


Figure 7: Example of Random Bi-Variate Gaussian Generation

eration: $mask = mask \oplus octagon$. Let M be the set of pixels set to one in the mask. Then we estimate the number of pixels of the intersection between the mask and the binary ellipse by performing the set operation $Len(M \cap BE)$. Now we introduce the threshold γ such that the number of pixels in the intersection between mask and binary ellipse be $\leq \gamma$. A loop is performed such as:

$$\textbf{while} \quad Len(M \cap BE) \leq \gamma \quad \textbf{then } M = M \oplus B \tag{3}$$

where B is an octagon structuring element. In the upper panel of Figure 6, a sequence of Masks (the black surface) for different values if γ is reported. The panel at the bottom of Figure 6 shows the corresponding result of the set operation $M \cap E$. The number of pixels of $M \cap E$ increases until the area is greater than the threshold γ.

Synthetic falls, finally, are generated as random vectors $X = [X_1 \dots X_n]^T$, where $X_1 = \text{feature}_1, \dots, X_N = \text{feature}_N$, according to the Gaussian multidimensional probability distribution $N - Gauss$ reported in (4).

$$p(x; \mu, \Sigma^2) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\right)} \tag{4}$$

In (4) the mean is estimated as

$$\mu = [\frac{max(\text{feature}_1) + min(\text{feature}_1)}{2}, \dots$$

$$\dots, \frac{max(\text{feature}_N) + min(\text{feature}_N)}{2}] \tag{5}$$

so $\mu \in \mathbf{R}^n$. Morover, in (4) the covariance matrix is an $N \times N$ symmetric matrix. The diagonal elements are the variances of each feature. It is estimated as follows:

$$\Sigma(i,i) = \frac{(max(\text{feature}_i) - min(\text{feature}_i)^2)}{4}. \tag{6}$$

All the other elements are equal to zero. In Figure 7 we report an example of random generation of features according to the bi-variate Gaussian distribution.

Finally we evaluate the synthetic fall features as intersection of the random bi-variate features with the feature Space and the binary Mask. The synthetic falls features are reported in the example shown in Figure 8 where also the real fall features are reported for a first comparison.

In Figure 9 we recall that the neural network has input data derived from the sensors. Then we have three hidden layers and one exit layer. The dimension of this network allows to perform the network training using usual back-propagation.

The key point of our approach is the following. The neural network is trained with three features, namely Ratio, Base

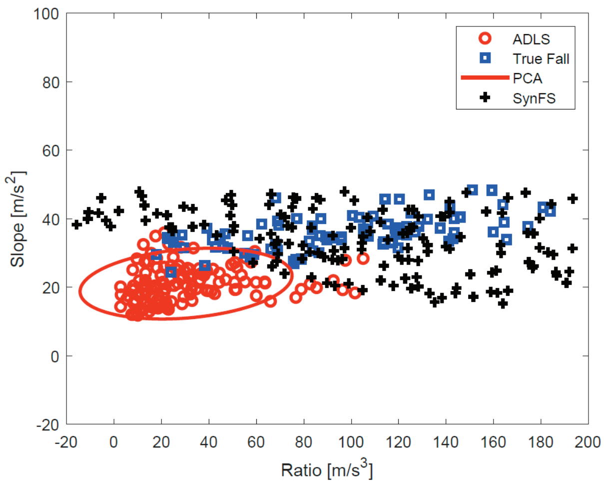


Figure 8: Example of 2D Synthetic Fall Features

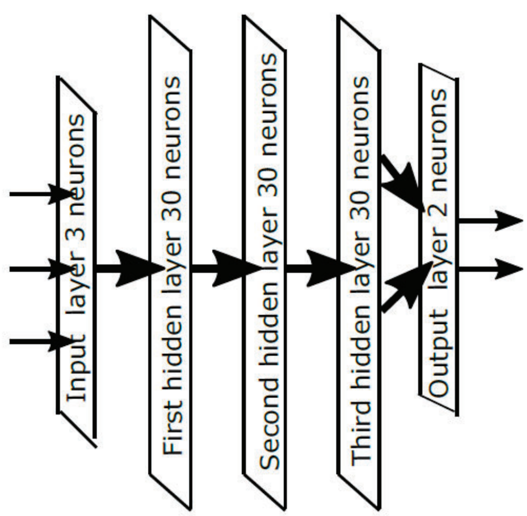


Figure 9: Neural Network Structure Trained with Real DLA and Synthetic Falls

and Peak1 extracted from ADL and falls. The DAL output of the network is set to zero during daily living features and the FALL output is set to one for synthetic falls features. When a real fall happens, the network should be able to detect it by looking which output is greater tahn the other.

It is worth recalling that the problem with training such neural network classifier is, in practice, we have only real ADL features and it is not easy to have real falls. To solve this difficulty, we generate features which are classified as Falls. In other words, we generate features which are points in the set that is complement to the set of the DAL set.

6. Metrics for Performance Evaluation

The starting point for measuring the quality of a classifier is to obtain the rate of false positive (fp), false negative (fn), true positive (tp) and true negative (tn) from the classifier. In our case of a falls detector, let us suppose that there is a fall. If the detector detects it, a **tp** is measured. If it does not detect it, we have a **fn**. For example, on 100 actual fall events, the detector could have 80 **tp** and 20 **fn**. If there has not been a fall, the detector could say that there was a fall (**fp**) or that there was no fall (**tn**). So on 100 non-fall events, we could have 80 **tn** and 20 **fp**.

In other words:

- TP (true positive): This is a situation in which a fall occurs and the system correctly detects it.
- FN (false negative): In this situation we have the fall happens, but the device does not announce it.
- TN (true negative): This is the situation in which a fall does not occur and the system correctly detects that there has not been a fall.
- FP (false positive): In this situation the fall does not happen but the device incorrectly announces that it has detected a fall.

These measures are also called:

tp → hit fn → correct rejection
fp → false alarm
tn → miss

It is sometimes convenient to measure errors in a more concise way. The most used measures are:

$$Precision = \frac{tp}{tp + fp}$$

This parameter measures the following quantity: the proportion of positive responses that have really fallen.

$$Recall = \frac{tp}{tp + fn}$$

This parameter represents the system's ability to detect a fall every time it occurs. The algorithm is good if the recall approaches 1 because in this case there are no false negatives. In other words, this parameter measures the proportion of falls that have been correctly identified.

7. Experimental Results

First we give a look to the data sets used for experiments. The first is the MobiAct v2.0 data set [10]. MobiAct contains data of four different types of falls and nine different daily living activities from a total of 57 subjects with more than 2500 trials. As well as being used to obtain experimental results, in this paper MobiAct is used as representative data set in all the Figures.

The second is UMAFall, which contains data from 17 subjects performing 8 different types of ADL and 3 different

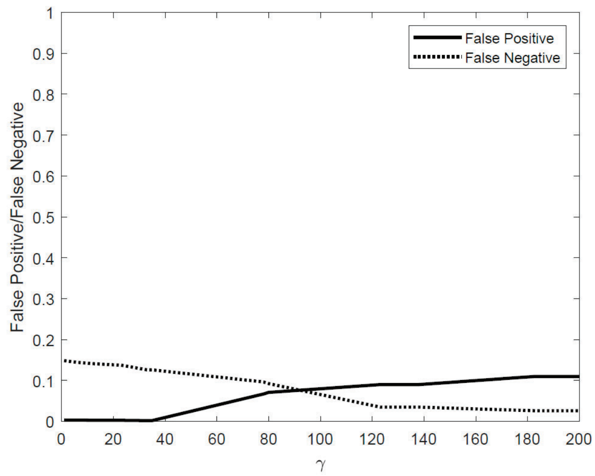


Figure 10: False Positive and False Negative Obtained with Real ADL and Falls Data

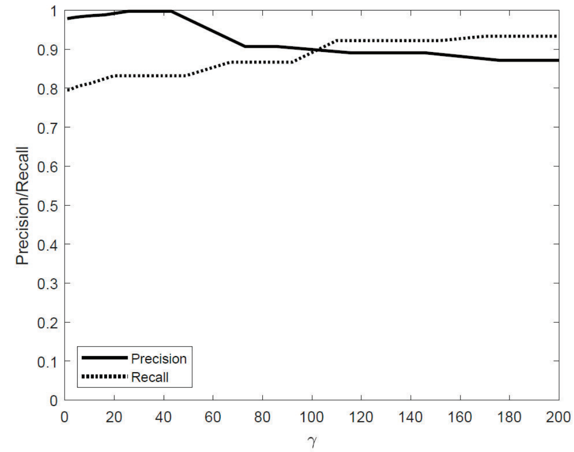


Figure 12: Recall and Precision Obtained with Real ADL and falls Data

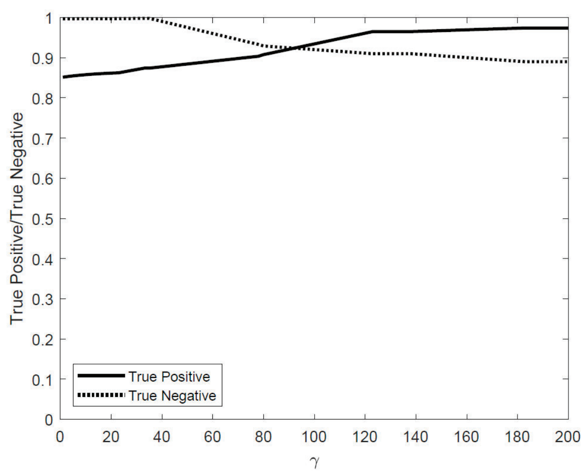


Figure 11: True Positive and True Negative Obtained with Real ADL and Falls Data

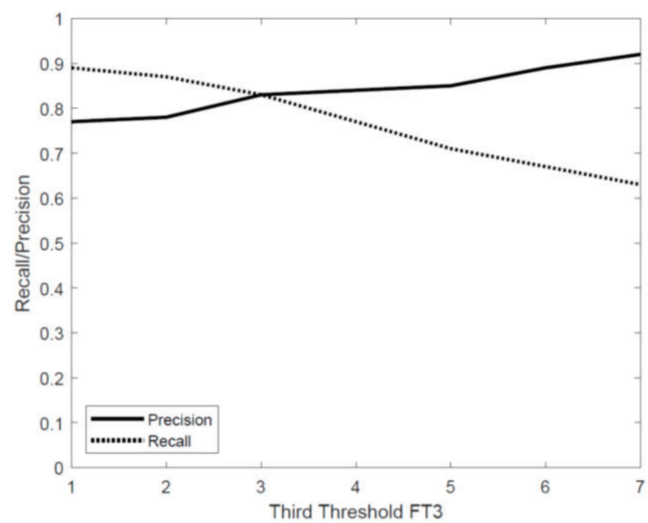


Figure 13: Recall and Precision Performances of the Bourke Algorithm for Various Values of the Third Threshold

types of falls. Sensing point of MobiAct and UMAFall is the right trouser pocket.

The third data set used for experimental result is the SisFall [11, 29], selecting data related to the waist sensor point. It was generated with 38 participants performing repetitions of 19 ADL and 15 fall types.

We first obtain False Positive, and Negative as well as True positive and True Negative. All the results reported in the following are averaged over these data sets. Of course as usual we try an input signal from the test section of the data sets and we look if the output is correct or not. These results are reported in Figure 10 and in Figure 11.

Then we obtain the values of Recall and Precision, reported in Figure 12.

These performances have been compared with the Bourke algorithm [3], which is three threshold based. To find the optimum values of the thresholds, we noticed that the most important threshold is the third one, so we found its performances at various values of the third threshold. The results are reported in Figure 13 which shows that the values of Recall/Precision are quite lower than our algorithm.

8. Final Remarks and Future work

Many fall detection systems are based on thresholds applied on features derived from inertial sensors. In this paper we report a novel algorithm which is able to achieve high performance, namely an equal Recall and Precision value more of 90% and a false error rate in this point less than 8%. The main feature of this algorithm are that it only requires collecting features in periods of normal daily living and then, from these features, it estimates features of artificial falls. The availability of many features which describe normal and fall events allow to use Machine Learning approaches which are very powerful classifiers provided that sufficient amount of data is given. An important aspects of the described approach is that we use only one type of inertial sensor, namely the accelerometer. Future work will be focused on the fusion of the described results obtained with only an accelerometer with other types of inertial sensors, for example a gyroscope or a magnetometer. In this way many other data could be given to the neural network. The computation complexity is very low because it needs only to compute a trained neural network, so it can be performed in real time. Moreover, we used only three features. It would be interesting to see how much the performance increase if other features are used, hence leading to a feature hyperspace as indicated in Section 3. Another possible direction of research consists in making our algorithm compliant with emerging features of novel big data systems (e.g., [21, 31, 4, 32, 2, 6, 8]).

References

[1] Abdi, H., Williams, L.J., 2010. Principal component analysis. WIREs Comput. Stat. 2, 433–459.

[2] Bonifati, A., Cuzzocrea, A., 2006. Storing and retrieving xpath fragments in structured P2P networks. Data Knowl. Eng. 59, 247–269.

[3] Bourke, A., Lyons, G., 2008. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Med Eng Phys. 30, 84–90.

[4] Braun, P., Cameron, J.J., Cuzzocrea, A., Jiang, F., Leung, C.K., 2014. Effectively and efficiently mining frequent patterns from dense graph streams on disk, in: 18th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, KES 2014, Gdynia, Poland, 15-17 September 2014, pp. 338–347.

[5] C, M., R, I., I, P., M, C., 2014. Detecting falls as novelties in acceleration patterns acquired with smartphones. PLoS ONE .

[6] Cuzzocrea, A., Moussa, R., Xu, G., 2013. Olap*: Effectively and efficiently supporting parallel OLAP over big data, in: Model and Data Engineering - Third International Conference, MEDI 2013, Amantea, Italy, September 25-27, 2013. Proceedings, pp. 38–49.

[7] Cuzzocrea, A., Mumolo, E., Tessarotto, M., 2019. Towards an effective and efficient machine-learning-based framework for supporting event detection in complex environments, in: Proceedings of the 43rd IEEE Computer Society Signature Conference on Computers, Software and Applications, COMPSAC 2019, Milwaukee, WI, USA, July 15-19, 2019.

[8] Cuzzocrea, A., Russo, V., 2009. Privacy preserving OLAP and OLAP security, in: Wang, J. (Ed.), Encyclopedia of Data Warehousing and Mining, Second Edition (4 Volumes). IGI Global, pp. 1575–1581.

[9] F, B., C, B., A, C., L, C., K, A., JM, H., W, Z., J., K., 2012. Evaluation of accelerometer-based fall detection algorithms on real-world falls. PLoS One 7, 1–9.

[10] G., V., C., C., T., M., M., P., Tsiknakis, M., 2016. The mobiact dataset: Recognition of activities of daily living using smartphones, in: International Conference on Information and Communication Technologies for Ageing Well and e-Health, pp. 143–151.

[11] Gasparrini, S., Cippitelli, E., Spinsante, S., Gambi, E., 2014. A depth-based fall detection system using a kinect® sensor. Sensors 14, 2756–2775.

[12] Guo, H.W., Hsieh, Y.T., Huang, Y.S., Chien, J.C., Haraikawa, K., Shieh, J.S., 2015. A threshold-based algorithm of fall detection using a wearable device with tri-axial accelerometer and gyroscope, in: Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2015 International Conference on, IEEE. pp. 54–57.

[13] Huynh, Q.T., Nguyen, U.D., Irazabal, L.B., Ghassemian, N., Tran, B.Q., 2015. Optimization of an accelerometer and gyroscope-based fall detection algorithm. J. Sensors .

[14] J, K., C, B., F, L., S, N., W, M., W, A., W, Z., JM, H., van Lummel RC, L, C., U, L., 2011. Comparison of acceleration signals of simulated and real-world backward falls. Medical Engineering & Physics , 368–373.

[15] Jolliffe, T.T., Cadima, J., 2016. Principal component analysis: A review and recent developments. Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences , 1–16.

[16] Kangas, M., Konttila, A., Lindgren, P., Winblad, I., Jämsä, T., 2008. Comparison of low-complexity fall detection algorithms for body attached accelerometers. Gait Posture 28, 285–291.

[17] KANSIZ, A.O., GUVENSAN, M.A., TURKMEN, H.I., 2013. Selection of time-domain features for fall detection based on supervised learning, in: Proceedings of the World Congress on Engineering and Computer Science 2013 Vol II, pp. 1–6.

[18] Kennell, L.R., Ives, R.W., Gaunt, R.M., 2006. Binary morphology and local statistics applied to iris segmentation for recognition, in: ICIP, IEEE. pp. 293–296.

[19] Khan, S.S., Karg, M.E., Kulic, D., Hoey, J., 2012. Towards the detection of unusual temporal events during activities using hmms, in: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 1075–1084.

[20] Khan, S.S., Karg, M.E., Kulic, D., Hoey, J., 2014. X-factor hmms for detecting falls in the absence of fall-specific training data, in: Ambient Assisted Living and Daily Activities - 6th International Work-Conference, IWAAL 2014, Belfast, UK, December 2-5, 2014. Proceedings, pp. 1–9.

[21] Li, K., Jiang, H., Yang, L.T., Cuzzocrea, A. (Eds.), 2015. Big Data - Algorithms, Analytics, and Applications. Chapman and Hall/CRC.

[22] Li, S.Z., Jain, A.K. (Eds.), 2009. Encyclopedia of Biometrics. Springer US.

[23] Lühr, S., Venkatesh, S., West, G.A.W., Bui, H.H., 2004. Explicit state duration HMM for abnormality detection in sequences of human activity, in: PRICAI 2004: Trends in Artificial Intelligence, 8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand, August 9-13, 2004, Proceedings, pp. 983–984.

[24] M., Z., S., W., Y., C., Z., C., Z., Z., 2012. An activity transition based fall detection model on mobile devices. In: Park J., Jin Q., Sang-soo Yeo M., Hu B. (eds) Human Centric Technology and Service in Smart Space. Lecture Notes in Electrical Engineering, vol 182. Springer, Dordrecht .

[25] Micucci, D., Mobilio, M., Napoletano, P., Tisato, F., 2017. Falls as anomalies? an experimental evaluation using smartphone accelerometer data. Journal of Ambient Intelligence and Humanized Computing , 87–99.

[26] Mirchevska, V., Lustrek, M., Gams, M., 2014. Combining domain knowledge and machine learning for robust fall detection. Expert Systems 31, 163–175.

[27] Noury, N., Rumeau, P., Bourke, A., ÓLaighin, G., Lundy, J., 2008. A proposal for the classification and evaluation of fall detectors. IRBM 29, 340 – 349.

[28] Shi, T., Sun, X., Xia, Z., Chen, L., Liu, J., 2016. Fall detection algorithm based on triaxial accelerometer and magnetometer, vol. 24, no.2, pp157-163, 201. Engineering Letters 24, 157–163.

[29] Sucerquia, A., López, J.D., 2016. Sisfall: A fall and movement dataset. Sensors , 1–14.

[30] Wu, S., Wang, Z., 2016. Applying online feature selection for fall detection, in: IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1–5.

[31] Wu, Z., Yin, W., Cao, J., Xu, G., Cuzzocrea, A., 2013. Community detection in multi-relational social networks, in: Web Information Systems Engineering - WISE 2013 - 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part II, pp. 43–56.

[32] Yang, C., Liu, J., Hsu, C., Chou, W., 2014. On improvement of cloud virtual machine availability with virtualization fault tolerance mechanism. The Journal of Supercomputing 69, 1103–1122.

Journal of Visual Language and Computing

Volume 2020, Number 1