Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

Enhanced Emotion Detection and Analysis in Human-Robot Interactions: An Innovative Model and Its Experimental Validation

Alfredo Cuzzocrea[a,b,*], Alessia Fantini[c,d] and Giovanni Pilato[d]

[a] iDEA Lab, University of Calabria, Rende, Italy
[b] Dept. of Computer Science, University of Paris City, Paris, France
[c] University of Pisa, Pisa, Italy
[d] CNR, Istituto di Calcolo e Reti ad Alte Prestazioni, Consiglio Nazionale delle Ricerche, Palermo, Italy

ARTICLE INFO

Article History:
Submitted: 7.1.2023
Revised: 7.26.2023
Accepted : 7.28.2023

Keywords:
Emotion Detection
Human-Robot Interaction
ChatGPT
Dialogues Generation

ABSTRACT

This study presents an ongoing project on developing an Italian textual dataset for emotion recognition in human-robot interaction (HRI). The main goal is constructing a dataset using a well-defined methodology that generates custom interactions. To accomplish this, we employed ChatGPT to assist us in developing the dialogues, which human psychology experts then reviewed. Our analysis primarily focused on assessing various aspects of the dataset, including the distribution of context types, gender representation, consistency between context and emotion, and the quality of interaction. By "quality," we refer to how the generated text accurately reflects the intended manifestation of emotions. Based on the analysis, we modified the dialogues to emphasize specific emotions in particular contexts. The findings from this preliminary study were significant, providing valuable insights to guide the generation of subsequent conversations and facilitate the creation of a more comprehensive dataset. A case study is also outlined with the aim of enabling increasingly realistic interactions in HRI scenarios.

1. Introduction

Emotions play a crucial role in Human-Robot Interaction (HRI), but one of the most challenging tasks for robots in such interactions is recognizing emotions [47], [19]. Emotions are multidimensional, and their comprehension relies on the context in which they are expressed. Understanding emotions necessitates considering context, which challenges Natural Language Processing (NLP) research. Context enables us to predict emotions to some extent. For instance, attending a party, securing a new job, or embarking on a trip are highly likely to evoke the feeling of "joy." Conversely, experiencing a loss or arguing with a loved one is often associated with "sadness."

While emotions can overlap and vary among individuals, certain objective situations tend to be consistently linked to specific emotions. Therefore, providing examples of context-related emotions can aid in identifying emotions accurately. In a study on conversational context modeling [48], the authors highlight the significant enhancement that considering context can bring to NLP systems. Consequently, constructing a specific and rich contextual information dataset is vital within data-driven models.

A relevant number of contributions in the literature focus on developing datasets for emotion recognition. However, the majority of these datasets typically encompass only a limited number of emotions, often centered around Ekman's basic emotions. Some examples are EmotionX [54] , Affect-Intensity Lexicon and Emotion Dataset (AILA) [42], CrowdFlower's Emotion Dataset [1], Friends [30] , EmoBank [11]. Furthermore, many approaches build dataset using news paper, books or dialogues found on the Internet, including those found from social media, e.g. SemEval-2018 Task 1: Affect in Tweets (AIT-2018) [43], Sentiment140 [25], Emotion Intensity Dataset (EmoInt) [41], The International Survey on Emotion Antecedents and Reactions (ISEAR) [52]. Others use movies, e.g. The Stanford Sentiment and Emo-

*This research has been made in the context of the Excellence Chair in Big Data Management and Analytics at University of Paris City, Paris, France

alfredo.cuzzocrea@unical.it (A. Cuzzocrea); alessia.fantini@phd.unipi.it (A. Fantini); giovanni.pilato@icar.cnr.it (G. Pilato)

ORCID(s): 0000-0002-7104-6415 (A. Cuzzocrea); 0009-0007-0337-2423 (A. Fantini); 0000-0002-6254-2249 (G. Pilato)

DOI reference number: 10.18293/JVLC2023-N1-031

tion Classification (SSEC) [53, 44] or physiological signals, e.g. The DEAP (Database for Emotion Analysis using Physiological Signals) [33]. In addition to this, the general problem has been stirred-up by several research efforts in the community (e.g., [58, 12, 40, 57, 32]).

Regarding Italian datasets, there are fewer contributions and often from tweets, some of the most widely used include SEMEVAL-ITA-2018 [13], ITA-EVALITA-2020 [7], EmoLexIta [16], The STS-ITA (Sentences in the Wild - Italian) [9], or news articles e.g. News-ITA [50]. A lexicon based approach has been also used for sentiment classification of books reviews in the Italian language [15].

Regarding the primary contributions in the existing literature, we intentionally chose to exclude data from social media or newspaper articles. Such content often exhibits language that may not align well with natural interactions. Instead, for our specific usage scenarios, such as Human-Robot Interactions, we utilized examples of interactions between individuals that emphasize the emotions we wish to focus on. This particular aspect holds significant importance in our study because the dialogue structure enables us to furnish the robot with examples of interactions closely resembling those that occur in real-world settings. Through the generation of custom dialogues, we were able to provide precise contextual scenarios in which particular emotions might arise. Also, the labeling was not done directly by us: this is another of the challenges highlighted by [48] in conversational context. We asked the ChatGPT to generate dialogues in which a specific emotion, such as *joy*, emerges; subsequently, we monitored and possibly adjusted or validated the associated labeling. An additional crucial aspect of our research is that we go beyond merely incorporating basic emotions; we have identified and labeled a comprehensive set of fourteen emotions, considering those likely to arise during Human-Robot Interaction (HRI) across different contexts, including home, medical settings, school, and everyday life. These emotions are *joy, sadness, anger, fear, surprise, disgust, frustration, embarrassment, boredom, nervousness, melancholy, guilt, hope, and stress*. In conclusion, from our standpoint, an ideal emotional dataset should demonstrate a well-balanced representation of data from various perspectives. To accomplish this objective, we conducted additional analysis on the dialogues generated using ChatGPT. This involved calculating several metrics, including the distribution of gender, the types of contexts, the consistency between emotions and contexts, and, overall, an evaluation of the interaction quality. By following this main trend, it is worth to notice that, recently, there has been a trendy interest on the issue of coupling HRI with *big data*, as it emerges from recent studies (e.g., [29, 10, 56]).

The remainder of the paper is organized as follows. The next Section explores some works on topics related to the theme of our study; in Section 3 we illustrate the methodology that we used to build the dataset; in Section 4 we discuss about results, and a sample of the collected and modified dialogues as well as the subsequent analysis is reported; then in Section 5 we present a case of study; subsequently in Section 6 a brief discussion is given about the dataset characteristics; in the end conclusions and future work are illustrated.

2. Related Work

There are several contributions in the literature in the area of developing models for emotion recognition in conversational agents. An interesting review was conducted by [46]. The study provides a systematic review of approaches in building an emotionally-aware chatbot (EAC). In [55] the authors use multi-task learning to predict the emotion label and generate a valid response for a given utterance. Their model consists of a self-attention based encoder and a decoder with dot product attention mechanism to generate response with a specified emotion and produce more emotionally relevant responses. In order to improve user interaction in [23] a model was developed with the task of generating empathetic and personalized dialogues, giving the machine the ability to respond emotionally and in accordance with the user. Other studies have tried to create empathetic generative chatbots that could simulate the responses provided by mental health professionals, mapping the emotions of the interlocutor and generating appropriate responses [27], [14], [20]. In general, a lot of efforts have been made to try to make conversational models more fluid by providing an ontology for the description of all concepts that may be relevant conversational topics, as well as their relationships to each other [26] or providing context and increasing the domain of knowledge [51], [59], [45], [39], [28]. Other interesting contributions train agents using datasets that include everyday conversations in order to dynamically answer people's questions. In [4] the chatbot model is based on the Simple Dialog and Daily Dialog datasets, which are then merged into a single dataset. Also, in [61] the authors train a model on 147 million conversational exchanges extracted from Reddit comment chains over a period from 2005 to 2017. In [49] the authors propose a new benchmark for generating empathetic dialogues and a new dataset based on emotional situations (EmpatheticDialogues) that includes 25k conversations.

Balanced dataset construction is critical for emotion recognition in general, but in [60] the authors point out that this is even more important when dealing with textual datasets. The reason is twofold: on the one hand because of the manual annotation of the data, a costly and time-consuming process, on the other hand because of the assignment of labels, which most of the time is related to subjectivity. In [62] they propose a Knowledge Enriched Transformer (KET), where contextual utterances are interpreted using hierarchical self-attention and external commonsense knowledge is dynamically leveraged using a context-aware affective graph attention mechanism At the same time, another crucial element is the techniques used for emotion recognition. In [31] the authors show the classification of emotion detection methods proposed in the literature: Keyword-based detection, Learning-based detection and Hybrid detection. In Keyword-based detection, emotions are detected based on the related set of keywords found in the input text. In Learning-based detection method, emotions are detected based on previous training result. In Hybrid detection method, emotions are detected on the basis of a combination of detected key words, learned patterns, and other additional information. In [21] authors reviewed some deep learning-based technologies commonly utilized in Textual Emotion Recognition (TER). The most commonly used are: Word Embedding, Basal Neural Networks and Derived Variations, Knowledge Enhanced Representation and Transfer Learning for Emotion Recognition. Textual emotion in dialogue is dynamic and strongly correlated with context information, thus the TER of dialogue is more challenging. There are several methods that try to address this problem, as they reported in their survey. The first is context modeling, as each utterance is highly dependent on contextual clues, and good results will not be obtained if the sentence-level deep TER algorithm is directly applied to the dialogue. In particular, context modeling can be summarized as either a flatten context modeling or hierarchical context modeling. By flatten context modeling, context utterance and current utterance are concatenated, and all tokens are flattened into a word sequence. Then, a neural network receives this sequence of words for contextual learning and final prediction. However, the flattened context processing turns the word sequence too long and ignores the temporal step, disrupting the hierarchy. In hierarchical context modeling, each utterance in dialogue is embedded into a vector representation by utterance-level encoder: context information is further extracted by hierarchy context-level encoder. Another method is the Dynamic Emotion Modeling that mainly focus on tracking the contextual information and exploring the overall tone in dialogue.

Lastly, to the best of our knowledge, there are no other studies that have used ChatGPT to construct text datasets. ChatGPT is used in several areas, as highlighted in this interesting review [24]. The main industries that are exploiting the potential of ChatGPT are e-commerce companies, healthcare organizations, financial institutions, customer service, call centers, and potentially could be greatly exploited by education industry. Developers also benefit a lot from ChatGPT, e.g., for bug fixing, identifying potential errors in the system, and making new pieces of source code based on samples, but it seems that it has not been used to create text datasets.

3. The Proposed Methodology

Our work aims to construct a dataset in Italian for emotion recognition based on dialogues. Initially, we established methodological criteria to guide the generation process, and then we utilized ChatGPT to expedite the data generation phase. After generating the dialogues, professional psychologists reviewed each conversation to assess the dataset's suitability from various perspectives. Our analysis focused on examining the consistency between the intended emotion and the context, the distribution of genders, the types of contexts produced, and the quality of interaction. In this context, interaction quality refers to the appropriateness of language concerning specific emotions. The methodology employed in this study consists of three stages: the procedure for generating dialogues, data analysis, and subsequent improvements.

3.1. Main Procedure

We created twenty-five dialogues for each of the fourteen emotions in our dataset. The instruction given to ChatGPT was to generate a brief conversation, approximately five lines long, between two individuals, wherein a specific emotion is evident. Additionally, we generated five more dialogues for each emotion, requesting ChatGPT to avoid using the word associated with that emotion. These particular dialogues were labeled as "Without Word (W.W.)".

This was done to test whether ChatGPT could generate discussions in which, e.g., sadness emerged without having the word “sadness" in the text. The goal is to create data that increasingly reflect real situations to train robots that can recognize emotions based on context and not just by recognizing specific words. The small number is because this is a pilot study to build a more extensive dataset later. Finally, the original dialogues generated were retained, but we created a copy to edit them after performing the analysis. Both the Web interface and the API provided by OpenAI were used. This has made it possible to obtain different styles of narrations of the events. Gpt 3.5-turbo model was used, with the following role: “You are a writer assistant who produces dialogue that accurately reflects emotion".

3.2. Analysis

Dialogues were analyzed considering four factors: consistency between context and emotion, gender distribution, type of contexts, and quality of interaction. By **consistency (C)** between context and emotion, we mean whether the context generated is consistent with the feeling expressed. For example, the context of an argument with the boss is a context compatible with the emotion of anger. So for each dialogue, we assessed whether or not there was consistency. We counted the percent relative frequency.

$$C = \frac{N yes}{N dialogues} \cdot 100$$

Similarly, for **gender distribution(GD)**, we counted how many times the gender “Neutral (N), Masculine (M) and Feminine (F)" occurred in the dialogues and we calculated the percent relative frequency.

$$GD = \frac{N gender(N or M or F)}{N totgender} \cdot 100$$

Regarding the **type of context(TC)**, we created classes and counted how many belonged to each class; then, we calculated the percent relative frequency.

$$TC = \frac{N contextX}{N totcontexts} \cdot 100$$

The classes identified are *Work*, *Leisure*, *Luck*, *Interpersonal sphere*, *Generic*. In some cases, we identified a specific category, e.g., in the “Disgust" dialogues, we identified the category “Animals and Objects," as several scenarios expressed disgust for objects or animals.

Finally, for the **quality of interaction(QoI)**, we analyzed the appropriateness of language in expressing a specific emotion. This was evaluated with three values: “*Sufficient*”, “*Not much*”, “*No*”. By “Sufficient (S)” we mean that the language appears natural enough and reflects in the terms used the emotion. By "Not much (NM)" we mean that the language is not very natural and it does not entirely reflect the emotion, e.g., using words that also represent other emotions, but all in all, it is acceptable. By "No (N)," we mean confusion, unusual terms, and/or language that does not reflect the specific emotion. Also, for this parameter, we calculated the percent relative frequency.

$$QoI = \frac{NValue(SorNMorN)}{Ntotinteractions} \cdot 100$$

3.3. Further Improvements

Following the analysis, we implemented several modifications addressing grammatical and content-related aspects. An important focus was placed on examining the distribution of different contexts and selecting those most relevant to interpersonal and social scenarios for inclusion in the dataset, which will be built after this pilot study.

To ensure diverse scenarios, we initially requested the generation of five potential social scenarios where a specific emotion could manifest. This approach enabled us to identify scenarios aligned more closely with Human-Robot Interaction (HRI). Once an interesting scenario was generated, we requested modifications to emphasize social interaction. A dialogue was created and expanded for each of these scenarios as needed. However, the model often struggled to broaden the conversation without repetitively using emotion-specific terms. To address this, we requested the model to replace such terms with expressions that could serve as metaphors or equivalent phrases.

When changing scenarios, there were instances where certain emotions were confused. For example, when requesting the feeling of anger, the generated dialogues often expressed frustration, frequently using the term "frustrating" in the text, and vice versa. Similar confusion occurred with stress and nervousness. To overcome this, we initially asked for a clear definition to differentiate between the two emotions. Once provided with the description, we instructed the model to generate scenarios where each emotion could emerge distinctly. As a result, the developed scenarios became more specific, effectively distinguishing between the two emotions. The same approach was applied to stress and nervousness. This process underscores the significance of human experts intervening in all phases to guide ChatGPT in generating more focused dialogues.

4. Empirical Results

The results will be shown first according to a global view and then in detail for each emotion.

4.1. Global Analysis

With respect to **consistency**, 86% of the generated contexts are consistent with emotion. An example of consistency is this:

- Person 1: Ciao, come stai oggi? (*Hello, how are you today?*)
- Person 2: Non molto bene, sinceramente. (*Not very well, honestly.*)
- Person 1: Oh, mi dispiace sentirlo. C’è qualcosa che ti preoc*cupa? (Oh, I’m sorry to hear that. Is something bothering you?*)
- Person 2: Sì, ho avuto un brutto litigio con un caro amico ieri sera. Siamo finiti per dire cose terribili l’un l’altro e ora mi sento come se avessi perso una parte di me stesso. (*Yes, I had a bad fight with a close friend last night. We ended up saying terrible things to each other and now I I feel like I’ve lost a part of myself.*)
- Person 1: Capisco. Litigi del genere sono sempre molto difficili da gestire. Ma forse potresti cercare di parlare con il tuo amico e chiarire la situazione? (*I understand. Fights like that are always very difficult to deal with. But maybe you could try to talk to your friend and clear the air?*)
- Person 2: Ho provato a chiamarlo stamattina, ma non ha risposto. Non so se vuole ancora parlarmi. (*I tried to call him this morning, but he didn’t answer. I don’t know if he still wants to talk to me.*)
- Person 1: Mi dispiace tanto. Spero che alla fine riuscirete a risolvere le cose. Se hai bisogno di parlare, sai che puoi contare su di me. (*I’m so sorry. I hope you will eventually be able to work things out. If you need to talk, you know you can count on me.*)
- Person 2: Grazie, apprezzo molto il tuo sostegno. Ma per ora mi sento solo triste e confuso. (*Thank you, I really appreciate your support. But for now I just feeling sad and confused.*)

In this dialogue, the context is an argument with a friend, and effectively the sadness that emerges is consistent with the type of context. In some cases, the generated contexts did not reflect the corresponding emotion. An example:

- Person 1: Ciao, come stai? (*Hello, how are you?*)
- Person 2: Non molto bene, devo dire. (*Not very well, I must say.*)
- Person 1: Che cosa è successo? (*What happened?*)
- Person 2: Stamattina mi sono svegliato con un forte mal di stomaco. Ho passato tutto il mattino in bagno. (*I woke up this morning with a bad stomachache. I spent the whole morning in the bathroom.)*

- Person 1: Ah, mi dispiace sentirlo. Cosa pensi possa essere? (*Ah, I'm sorry to hear that. What do you think it might be?)*
- Person 2: Non ne sono sicuro, ma credo sia stato qualcosa che ho mangiato ieri sera. Ho provato un nuovo ristorante messicano, ma credo che il cibo non fosse fresco. (*I'm not sure, but I think it was something I ate last night. I tried a new Mexican restaurant, but I think the food wasn't fresh.*)
- Person 1: Oh no, che schifo. Ti senti meglio adesso? (*Oh no, that's gross. Do you feel better now?*)
- Person 2: Un po' meglio, ma non ancora del tutto. Devo ancora prendere qualche medicina. (*A little better, but not quite yet. I still have to take some medicine.*)
- Person 1: Sì, è meglio prendersi cura di te stesso. Spero tu ti riprenda presto. (*Yes, you better take care of yourself. I hope you recover soon.*)

In this case, except for "that's gross", the entire dialogue is centered on a context showing a reaction to physical discomfort. The person feels vomiting because of physical discomfort, as he has probably eaten bad food, and therefore not because of an emotional reaction of disgust.

Regarding **gender**, there is a strong imbalance in the dataset: N=30% M=56% F=14%

In addition, in a couple of cases, the gender count was canceled because the same person was first male and then female. Here is an example of a dialogue about frustration:

- Person 1: Ho lavorato duramente su questo progetto ma non ha (grammar error) ottenuto il successo sperato. (*I worked hard on this project but it did not (in the Italian version-grammar error) achieve the success I had hoped for.*)
- Person 2: Mi dispiace sentirti così deluso (indicates that person 1 is male). Cosa pensi sia andato storto? (*I'm sorry to feel so disappointed (in the Italian version indicates that person 1 is male). What do you think went wrong?*)
- Person 1: Non ne sono sicuro, ho messo tutta me stessa (female gender) ma sembra che non sia abbastanza. (*I'm not sure, I put all of myself (in the Italian version-female gender) but it seems like it's not enough.*)
- Person 2: Non scoraggiarti, ogni esperienza è una lezione imparata. Magari hai bisogno di un po' di tempo per riflettere e riprovarci con un approccio diverso. (*Don't be discouraged, every experience is a lesson learned. Maybe you need some time to reflect and try again with a different approach.*)

The **context** overall appears heterogeneous but it is unbalanced when observed in relation to specific emotions. For example, for the emotion "Joy," only three types of context were generated. Specifically, ten contexts are about *success* (e.g., passing a university exam, promotion at work), ten are about *leisure* (e.g., traveling, starting a yoga class), four are about *luck* (e.g., winning the lottery), and only one is about *Personl life situations* (receiving a gift). The type of context will be discussed in depth in the description of each emotion.

Regarding the **quality of interaction**, the adherence of text to the desired manifestation of emotions was evaluated. In 65% of the dialogues, we can define the quality of interaction as "sufficient". However, some changes were added later either in terms of grammatical corrections or to make the dialogue more fluid and natural. In 25% of cases, there is a poor fit between text and emotion. Finally, in 10% of the dialogues, the text was completely garbled or did not reflect the desired emotion. Here are some examples of the three categories:

Sufficient: Boredom

- Friend 1: "Cosa c'è che non va, sembri distratta?" (*"What's wrong, you seem distracted?"*)
- Friend 2: "Sì, sto solo pensando ad altro. Questa lezione mi fa venire la noia." (*"Yes, I'm just thinking about something else. This class is making me bored."*)
- Friend 1: "Capisco come ti senti, anche io sto trovando difficoltà a restare concentrata." (*"I understand how you feel, I am also finding it hard to stay focused."*)
- Friend 2: "Sì, vorrei solo essere altrove ora. Anche voi pensate la stessa cosa, giusto?" (*"Yes, I just wish I was somewhere else right now. You also feel the same way, right?"*)
- Friend 1: "Sì, credo che tutti vorremmo fare altro." (*"Yes, I think we all would like to do something else."*)

Not much: Anger

- Person 1: "Non ci posso credere, ho perso l'autobus per la terza volta questa settimana!" (*"I can't believe I missed the bus for the third time this week!"*)
- Person 2: "Ma come hai fatto?" (*"But how did you do it?"*)
- Person 1: "Non mi hai visto? Mi hai tenuto a parlare e l'autobus è passato sotto il mio naso!" (*"Didn't you see me? You kept me talking and the bus passed right under my nose!"*)
- Person 2: "Non è colpa mia se sei sempre in ritardo!" (*"It's not my fault you're always late!"*)
- Person 1: "Ma certo che è colpa tua! Non riesci mai a smettere di parlare e poi ti lamenti se arrivo sempre tardi!" (*"Of course it's your fault! You can never stop talking and then you complain that I'm always late!"*)

- Person 2: "Ok, ok, calmati! Non c'è bisogno di arrabbiarsi!" (*"Okay, okay, calm down! No need to get angry!"*)
- Person 1: "Ma come faccio a non arrabbiarmi? Questo mi fa perdere tempo e soldi!" (*"But how can I not get angry? This wastes my time and money!"*)
- Person 2: "Hai ragione, mi dispiace. Cercherò di essere più attento la prossima volta." (*"You're right, I'm sorry. I'll try to be more careful next time."*)

No: Hope

- Character A: "Spero solo di non sembrare troppo stressato/a stasera." (*"I just hope I don't look too stressed out tonight."*)
- Character B: "Non preoccuparti, sei bellissimo/a e la serata sarà fantastica." (*"Don't worry, you look beautiful and the evening will be great."*)
- Character A: "Speriamo che ci siano delle sorprese piacevoli stasera, vorrei che fosse tutto diverso dal solito." (*"Hopefully there will be some pleasant surprises tonight, I'd like everything to be different than usual."*)
- Character B: "Stasera sarà diversa dal solito, perché sarà proprio come ci piace. Semplice e piena di speranze!" (*"Tonight will be different than usual, because it will be just the way we like it. Simple and hopeful!"*)

4.2. Single Emotion Analysis

Below we show the analysis of each of the 14 emotions according to the 4 parameters outlined in the methodology section.

- **JOY**
  - Consistency = 100%
  - Gender = N 12% M 80% F 8%
  - Contexts = 10 Success, 10 Leisure, 4 Luck, and only 1 is about personal life situations
  - Quality of interaction = Sufficient 64% Not much 36%
- **SADNESS**
  - Consistency = 92%
  - Gender = N 46% M 54% F 0
  - Contexts = Heterogeneous mainly generic and interpersonal
  - Quality of interaction = Sufficient 88% Not much 12%
- **ANGER**
  - Consistency = 100%
  - Gender = N 12% M 55% F 33%
  - Contexts = Heterogeneous, sometimes reactions out of proportion to the context
  - Quality of interaction = Sufficient 88% Not much 12%
- **FEAR**
  - Consistency = 100%
  - Gender = N 24% M 72% F 4%
  - Contexts = Mostly related to horror contexts (shadows, animals, running away from someone) - Absence of contexts related to more interpersonal or social fear, such as fear of the future.
  - Quality of interaction = Satisfactory 72% Not much 28%
- **SURPRISE**
  - Consistency = 100%
  - Gender = N 24% M 76% F 0%
  - Contexts = Heterogeneous
  - Quality of interaction = Satisfactory 88% Not much 12%
- **DISGUST**
  - Consistency = 96%
  - Gender = N 72% M 28% F0
  - Contexts = Highly related to foods, insects, objects. No examples related to people's behaviors or abstract concepts. Only in two cases is there a reference to disgust as a result of a person's behavior.
  - Quality of interaction = Sufficient 84% Not much 16%
- **FRUSTRATION**
  - Consistency = 28%: in three cases there is confusion with *anger*
  - Gender = N 46% M 50% F 4%
  - Contexts = Heterogeneous, sometimes reactions out of proportion to the context
  - Quality of interaction = Sufficient 80% Not much 20%
- **EMBARRASSMENT**
  - Consistency = 68% sometimes there is confusion with *guilt*.
  - Gender = N 44% M 48% F 8%
  - Contexts = Heterogeneous
  - Quality of interaction = Sufficient 76% Not much 24%

- **BOREDOM**
  - Consistency = 92%
  - Gender = N 16% M 56% F 28%
  - Contexts = Heterogeneous, mainly leisure time
  - Quality of interaction = Sufficient 56% Not much 28% No 16%
- **NERVOUSNESS**
  - Consistency = 88%
  - Gender = N0 M 53% F 47%
  - Contexts = Heterogeneous
  - Quality of interaction = Sufficient 68% Not much 20% No 12%
- **MELANCHOLY**
  - Consistency = 88%
  - Gender = N 40% M 60% F0
  - Contexts = Heterogeneous
  - Quality of interaction = Sufficient 68% Not much 16% No 16%
- **GUILT**
  - Consistency = 92%
  - Gender = N 40% M 40% F 20%
  - Contexts = 24% relate to work contexts, while most are related to interpersonal or social situations (e.g., arguing with a friend, neglecting family, telling a lie, etc...)
  - Quality of interaction = Satisfactory 52% Not much 44% No 4% . In many dialogues the language appears out of proportion to the emotion
- **HOPE**
  - Consistency = 100%
  - Gender = N 52% M 36% F 16%
  - Contexts = 28% relate to work contexts, 44% relate to medical contexts, 28% relate to interpersonal or social situations
  - Quality of interaction = Satisfactory 28% Not much 64% No 8% . Often the language seems to belong more to fear or nervousness and not to hope. Here is an example:

- Student 1: "Sto preparando questo esame da giorni, spero di ottenere un buon voto." (*"I've been preparing for this exam for days, I hope to get a good grade."*)
- Student 2: "Sono sicuro che andrà tutto bene, hai studiato tanto e sai quello che fai." (*"I'm sure you'll do well, you've studied hard and you know what you're doing."*)
- Student 1: "Sì, ma ho paura di non ricordare tutte le informazioni durante l'esame." (*"Yes, but I'm afraid I won't remember all the information during the exam."*)
- Student 2: "Non preoccuparti, vai tranquillo e non lasciare che l'ansia ti prenda il sopravvento. Spero che otterrai la valutazione che meriti." (*"Don't worry, go easy and don't let anxiety get the best of you. I hope you will get the grade you deserve."*)
- Student 1: "Grazie per il supporto! Ho davvero bisogno di sentirlo. Speriamo che andrà tutto bene." (*"Thank you for the support! I really need to hear it. Hopefully everything will be okay."*)
- Student 2: "Sarà così, cerca di rilassarti e di essere positivo. La speranza è la chiave del successo." (*"It will, just try to relax and be positive. Hope is the key to success."*)

- **STRESS**
  - Consistency = 84% . Sometimes confusion with nervousness
  - Gender = N 0 M 83% F 17%
  - Contexts = 56% relate to work contexts, the remaining are heterogeneous
  - Quality of interaction = Satisfactory 60% Not much 12% No 28%

4.3. Analysis of Dialogues “Without Words”

As for dialogues generated without the word expressing the emotion emerging, most generated contexts **consistent** with the emotion. Not very consistent were the dialogues of Boredom, Fear, Surprise and Hope. In contrast, only 1 out of 5 of the Disgust dialogues was found to be consistent. The other 4 were so confused that they could not be classified. Here is an example:

- Person 1: "Che schifo...quella lì sembra venuta direttamente dal nido delle rane!" (*"Gross...that one looks like it came straight from the frogs' nest!"*)
- Person 2: "Davvero, mi ha fatto venire il voltastomaco. Però non possiamo farci niente al riguardo... forse conviene che andiamo a prendere un po' d'aria fresca." (*"Really, it made me sick to my stomach. We can't do anything about it though...maybe we'd better go get some fresh air."*)
- Person 1: "Sì, direi che hai ragione. Quando si vedono situazioni simili, l'unica cosa da fare è portare il nonno in braccio in giro per la città, almeno la compagnia è più gradevole!" (*"Yes, I would say you are right. When you see situations like that, the only thing to do is to carry Grandpa around town, at least the company is more pleasant!"*)

Furthermore, right among the dialogues of Disgust-W.W. a stereotype about Neapolitans emerged. Here is the dialogue:

- Person 1: "Oh, guardate quella lì...è più volgare di un bidone di spazzatura a Napoli!" (*"Oh, look at that one...she's grosser than a garbage can in Naples!"*)
- Person 2: "Sì, mi ha messo i brividi appena l'ho vista. Comunque, perché non ci beviamo un po' d'acqua insieme? Così ci togliamo lo schifo di bocca!" (*"Yeah, she gave me the creeps as soon as I saw her. Anyway, why don't we have some water together? That way we can get the filth out of our mouths!"*)
- Person 1: "Mi pare un'ottima idea, non vedo l'ora di liberarmi di questa sensazione." (*"That sounds like a great idea, I can't wait to get rid of this feeling."*)

It is not only not at all sufficient from the point of view of language, but a stereotype clearly emerges. Regarding **gender** and **contexts**, the number of dialogues is small to draw specific inferences, however, we can say that they seem to reflect the general trend. As for the **quality of interaction**, it appears worse than the basic dialogues, that is, the non-W.W. dialogues. In fact, in 43% of the cases the quality of interaction was rated as "sufficient", in 32% of the cases "not very much", and in 25% "no". The sum of "not very much" and "no" is also 57% thus exceeding the percentage of those considered sufficient.

5. Case Study

The dialogues created with ChatGPT were used as dialogue prototypes to improve human-robot interaction. Specifically, we used the RiveScript [3] dialogue engine to create a simple system that could give the ability to create a natural language interaction system associated with specific situations. RiveScript employs a concise set of rules that, when combined, can provide effective chatbot personalities. By writing triggers in a simplified regular expression format, complex sets of word patterns can be efficiently matched in a single step, enhancing the chatbot's capabilities. The core library is compact, self-contained, and it is aimed at receiving human input and delivering "intelligent", even if pattern-matching based, responses. This adaptability allows RiveScript to be utilized according to individual needs, empowering users with flexibility. One of the main advantages of the framework is that itadopts a straightforward plain text scripting language that is easy to learn and allows quick writing. Its line-based structure is readily understandable for maintenance purposes without using cumbersome XML code or complex symbols, which may hide the code structures.

Starting with the stimulus-response pair, which is the basic element of the RiveScript knowledge base, ChatGPT was asked to generate a set of N sentences similar to the input (trigger) sentences and N sentences semantically similar to the output answers of the bot. The phrases generated as triggers are provided into a subsymbolic layer that is used to identify what the user is saying. To build such a sub-symbolic layer, we have exploited the Spacy library [2], which easily allows to compute a sub-symbolic semantic similarity between two sentences. At the same time, the response phrases are used to enhance the variety of chatbot replies to make the interaction more engaging. To be more specific, we formulate the same dialogue (context, type of emotion, type of interaction) while increasing the variety of possible words (as it might happen in real life). We thus exploit "fast data generation" to improve the robot's listening skills and avoid monotonicity in responses.

An advantage is that we can exploit the dataset for emotion recognition as an additional element for emotion detection according to a given context. Specifically, the system is coded as a finite state machine that identifies the sequence of sentences uttered by the user based on the sentences given by the chatbot and contextualizes the situation, providing an extra element to associate the contextual interaction with a possible emotion, which has somehow been implicitly associated with the dialogue. In this manner, the agent will have a "basic" dialogue skeleton and a set of possible variations, all of them considering a specific context and emotion related to the interaction. This also improves the quality of the interaction, the associative process regarding emotions and context, and the possibility of understanding the situation.

As mentioned above, for each sentence in each chatgpt-generated dialogue, chatGpt was then asked to provide a number N of similar sentences. These sentences play a similar role to the emotional beacons introduced in [47]. In particular, given $s_{i,d}$ the *i-th* sentence of the *d-th* dialogue, a set of N sentences $\{b^j_{i,d}\}$ is generated with $j = 1, \cdots, N$ where N is a positive integer that has been set to $N = 10$ for our case study. Each sentence $\{b^j_{i,d}\}$ together with each "seed" sentence $s_{i,d}$ is then encoded subsymbolically. For efficiency reasons, we used the Spacy library [2] and the Italian language model "it_core_news_sm". Each sentence u_k provided to the system by the user is then compared sub symbolically with one of the possible sentences in the chatgpt-generated dialogues $s_{i,d}$ and the related "beacons" $\{b^j_{i,d}\}$ which are expected to be said by the user: the semantic similarity between the sentence u_k and all sentences $s_{i,d}$ and $\{b^j_{i,d}\}$ is then computed. If the calculated similarity exceeds a certain experimentally established threshold T_h, the corresponding dialogue d with which the sentence can be associated is identified. If several sentences belonging to different dialogues provide semantic similarity greater than the threshold T_h, the dialogue with the highest similarity value is selected.

Once a specific dialogue has been identified, the system tries to maintain the same structure provided by chatGPT to identify a possible context that is already known. The comparison then between what the user says and the sentences, along with their beacons, that the system expects continues until the presumed conclusion of the dialogue and recognition of the situation. If, during the conversation, the calculation of semantic similarity within the selected dia-

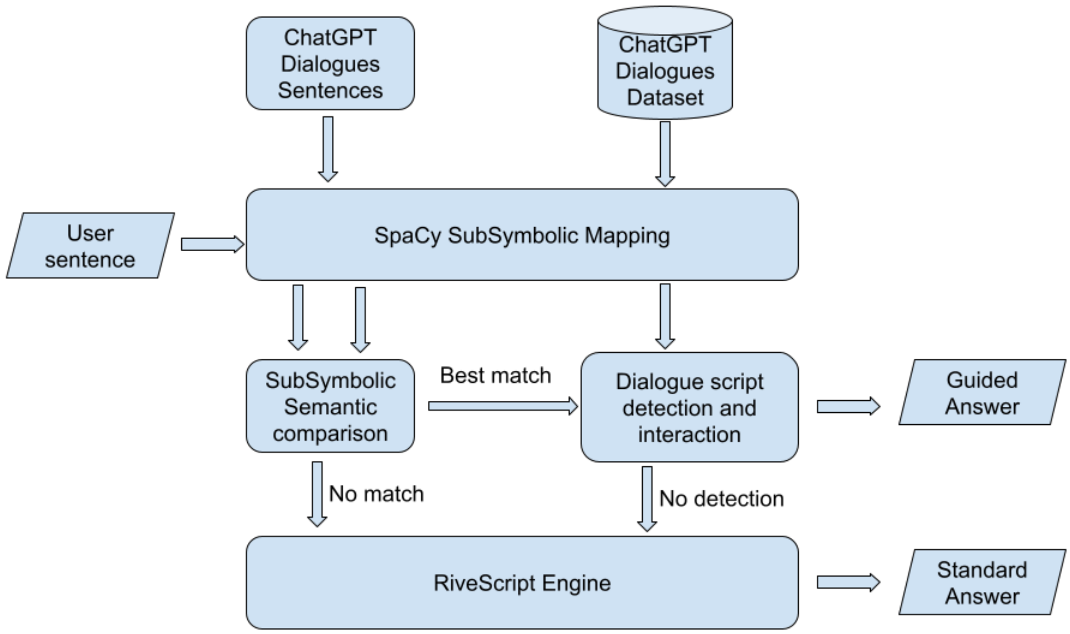

Figure 1: A schema of the interaction process for the case of study.

logue gives a value below the minimum value required by the threshold T_h, the system restarts from scratch, trying to identify another possible discussion. If, in the end, the sub-symbolic comparison of semantic similarity does not yield any value above the predetermined T_h threshold value, the control of the interaction passes directly to the RiveScript engine, which continues with the chat according to the usual standard rule-based chatbot mode. In any case, each sentence spoken by the user is still sub symbolically encoded and compared with the extended set of sentences in the dataset of known dialogues to identify possible occurrences of proper contexts to be recognized for determining the emotions expressed by the user during the interaction with the robot. A schema of the adopted architecture for this case of study is reported in Fig. 1.

6. Critical Discussion

The analysis of the dataset revealed both strengths and weaknesses, providing valuable insights for constructing the larger dataset. The main strength lies in the consistency between context and emotion, which ensures high reliability in automatically generating dialogues and facilitates fast data generation. However, as the results indicate, the generated conversations need to undergo validation by a human operator, as they were not consistently consistent. Additionally, special attention must be given to the types of contexts included in the dataset, aiming for maximum heterogeneity. This can be achieved by incorporating increasingly relevant contexts to interpersonal and social spheres, thereby reflecting realistic Human-Robot Interaction (HRI) scenarios.

The analysis of gender distribution highlighted a significant imbalance favoring male gender representation. This finding allows for addressing the bias and encourages broader reflections on training AI systems that strive for greater diversity. Furthermore, the presence of stereotypes concerning the city of Naples in some dialogues should also be carefully considered.

Lastly, the quality of interaction frequently necessitates modifications by the human operator. This may involve rectifying occasional grammatical errors, aligning the language with the intended emotion, and enhancing the naturalness of the dialogues.

Regarding the case study, we believe that a system designed in this way can be of help in improving HRI by having some specific strengths. One of these is the presence of context in dialogues: this can help the robot place certain emotions in certain contexts making the interaction more consistent. Another element is the variety of words expressing the same concept: this allows the robot to respond in a relevant way even when there are word variations as there are in natural interactions. Clearly, the effectiveness of the model will be the subject of subsequent studies, which will aim to identify measurement tools to establish the actual improvement of HRI.

7. Conclusions and Future Work

We conducted an initial study to guide the development of an Italian dataset designed for emotion recognition. Once the methodology and procedure were established, we leveraged ChatGPT to generate dialogues rapidly. Working alongside psychology experts, we meticulously examined 420 conversations covering 14 emotions to assess the dataset's balance across various dimensions, including consistency between context and emotion, gender distribution, types of generated context, and interaction quality. The findings revealed the advantages and limitations of utilizing automated dialogue generation systems. It became evident that the construction of the dataset cannot disregard human control.

The most significant advantage observed was the speed of data generation, with most cases demonstrating consistency between the generated contexts and intended emotions. However, exercising control over the dialogues remains crucial to ensure heterogeneity and a stronger focus on interpersonal and social aspects. The study also highlighted a notable gender distribution imbalance, predominantly favoring masculine representations. Addressing this disparity will generate dialogues explicitly requesting feminine and neutral genders, thereby achieving a more balanced dataset.

Furthermore, numerous modifications were made to the dialogues concerning language usage and interaction quality, encompassing grammatical, structural, and content corrections. Despite these adjustments, an additional advantage emerged—the ability to create dialogues from scratch, directing ChatGPT to generate dialogues that align with predetermined criteria established by the researchers. Moving forward, this study's insights will inform the development of a larger, well-balanced dataset tailored specifically for (HRI) scenarios.

Finally, the case study allowed us to observe how fast data generation is, making it possible to create more examples, thus improving the robot's listening skills and avoiding the monotonicity of responses. Certainly, however, a deeper evaluation of effectiveness will be the subject of subsequent studies.On the other hand, we aim at integrating our framework with emerging challenges due to novel *big data trends* (e.g., [8, 18, 5, 38, 17, 35, 36, 37, 6, 22, 34]).

Acknowledgements

The authors would thank Mr. Antonino Asta for his partial contribution to this work.

References

[1] , . Crowdflower. 2016. the emotion in text. https://www.figure-eight.com/data/ sentiment-analysis-emotion-text/.

[2] , . Industrial-strength natural language processing. https://spacy.io/.

[3] , . Rivescript. https://www.rivescript.com/.

[4] Bachtiar, F.A., Fauzulhaq, A.D., Manullang, M.T.R., Pontoh, F.R., Nugroho, K.S., Yudistira, N., 2022. A generative-based chatbot for daily conversation: A preliminary study, in: 7th International Conference on Sustainable Information Engineering and Technology 2022, pp. 8–12.

[5] Balbin, P.P.F., Barker, J.C.R., Leung, C.K., Tran, M., Wall, R.P., Cuzzocrea, A., 2020. Predictive analytics on open big data for supporting smart transportation services. Procedia Computer Science 176, 3009–3018.

[6] Barkwell, K.E., Cuzzocrea, A., Leung, C.K., Ocran, A.A., Sanderson, J.M., Stewart, J.A., Wodi, B.H., 2018. Big data visualisation and visual analytics for music data mining, in: 22nd International Conference Information Visualisation, IV 2018, Fisciano, Italy, July 10-13, 2018, IEEE Computer Society. pp. 235–240.

[7] Basile, V., Maria, D.M., Danilo, C., Passaro, L.C., et al., 2020. Evalita 2020: Overview of the 7th evaluation campaign of natural language processing and sspeech tools for italian, in: Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), CEUR-ws. pp. 1–7.

[8] Bellatreche, L., Cuzzocrea, A., Benkrid, S., 2010. *F&A*: A methodology for effectively and efficiently designing parallel relational data warehouses on heterogenous database clusters, in: Data Warehousing and Knowledge Discovery, 12th International Conference, DAWAK 2010, Bilbao, Spain, August/September 2010. Proceedings, Springer. pp. 89–104.

[9] Braunhofer, M., Elahi, M., Ricci, F., 2015. User Personality and the New User Problem in a Context-Aware Point of Interest Recommender System. pp. 537–549. doi:10.13140/RG.2.1.2493.8001.

[10] Budiharto, W., Andreas, V., Gunawan, A.A.S., 2020. Deep learning-based question answering system for intelligent humanoid robot. J. Big Data 7, 1–10.

[11] Buechel, S., Hahn, U., 2017. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Association for Computational Linguistics, Valencia, Spain. pp. 578–585. URL: https://aclanthology.org/E17-2092.

[12] Caruccio, L., Cirillo, S., Deufemia, V., Polese, G., Stanzione, R., 2022. Data analytics on twitter for evaluating women inclusion and safety in modern society, in: Proceedings of the 1st Italian Conference on Big Data and Data Science (ITADATA 2022), Milan, Italy, September 20-21, 2022, CEUR Workshop Proceedings.

[13] Caselli, T., Novielli, N., Patti, V., Rosso, P., 2018. Sixth evaluation campaign of natural language processing and speech tools for italian: Final workshop (evalita 2018), in: EVALITA 2018, CEUR Workshop Proceedings (CEUR-WS. org).

[14] Catania, F., Spitale, M., Fisicaro, D., Garzotto, F., 2019. Cork: A conversational agent framework exploiting both rational and emotional intelligence., in: IUI Workshops.

[15] Chiavetta, F., Bosco, G.L., Pilato, G., et al., 2016. A lexicon-based approach for sentiment classification of amazon books reviews in italian language. WEBIST (2) 2016, 159–170.

[16] Corazza, M., Menini, S., Cabrio, E., Tonelli, S., Villata, S., 2020. A multilingual evaluation for online hate speech detection. ACM Transactions on Internet Technology (TOIT) 20, 1–22.

[17] Coronato, A., Cuzzocrea, A., 2022. An innovative risk assessment methodology for medical information systems. IEEE Trans. Knowl. Data Eng. 34, 3095–3110.

[18] Cuzzocrea, A., Martinelli, F., Mercaldo, F., Vercelli, G.V., 2017. Tor traffic analysis and detection via machine learning techniques, in: IEEE BigData, 2017, IEEE Computer Society. pp. 4474–4480.

[19] Cuzzocrea, A., Pilato, G., 2021. A composite framework for supporting user emotion detection based on intelligent taxonomy handling. Logic Journal of the IGPL 29, 207–219.

[20] Das, A., Selek, S., Warner, A.R., Zuo, X., Hu, Y., Keloth, V.K., Li, J., Zheng, W.J., Xu, H., 2022. Conversational bots for psychotherapy: A study of generative transformer models using domain-specific dialogues, in: Proceedings of the 21st Workshop on Biomedical Language Processing, pp. 285–297.

[21] Deng, J., Ren, F., 2021. A survey of textual emotion recognition and its challenges. IEEE Transactions on Affective Computing .

[22] Filali, H., Riffi, J., Boulealam, C., Mahraz, M.A., Tairi, H., 2022. Multimodal emotional classification based on meaningful learning. Big Data Cogn. Comput. 6, 95.

[23] Firdaus, M., Thangavelu, N., Ekbal, A., Bhattacharyya, P., 2022. I enjoy writing and playing, do you: A personalized and emotion grounded dialogue agent using generative adversarial network. IEEE Transactions on Affective Computing .

[24] George, A.S., George, A.H., 2023. A review of chatgpt ai's impact on several business sectors. Partners Universal International Innovation Journal 1, 9–23.

[25] Goel, A., Gautam, J., Kumar, S., 2016. Real time sentiment analysis of tweets using naive bayes, in: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), IEEE. pp. 257–

261.
[26] Grassi, L., Recchiuto, C.T., Sgorbissa, A., 2022. Knowledge-grounded dialogue flow management for social robots and conversational agents. International Journal of Social Robotics 14, 1273–1293.
[27] Huang, J.Y., Lee, W.P., Chen, C.C., Dong, B.W., 2020a. Developing emotion-aware human–robot dialogues for domain-specific and goal-oriented tasks. Robotics 9, 31.
[28] Huang, M., Zhu, X., Gao, J., 2020b. Challenges in building intelligent open-domain dialog systems. ACM Transactions on Information Systems (TOIS) 38, 1–32.
[29] Jiang, M., Zhang, L., 2015. Big data analytics as a service for affective humanoid service robots, in: INNS Conference on Big Data 2015, San Francisco, CA, USA, 8-10 August 2015, Elsevier. pp. 141–148.
[30] Joshi, A., Tripathi, V., Bhattacharyya, P., Carman, M.J., 2016. Harnessing sequence labeling for sarcasm detection in dialogue from TV series 'Friends', in: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, Association for Computational Linguistics, Berlin, Germany. pp. 146–155. URL: https://aclanthology.org/K16-1015, doi:10.18653/v1/K16-1015.
[31] Kao, E.C.C., Liu, C.C., Yang, T.H., Hsieh, C.T., Soo, V.W., 2009. Towards text-based emotion detection a survey and possible improvements, in: 2009 International conference on information management and engineering, IEEE. pp. 70–74.
[32] Kaur, R., Majumdar, A., Sharma, P., Tiple, B., 2023. Analysis of tweets with emoticons for sentiment detection using classification techniques, in: Distributed Computing and Intelligent Technology - 19th International Conference, ICDCIT 2023, Bhubaneswar, India, January 18-22, 2023, Proceedings, Springer. pp. 208–223.
[33] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2011. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing 3, 18–31.
[34] Lee, S.J., Lim, J., Paas, L., Ahn, H.S., 2023. Transformer transfer learning emotion detection model: synchronizing socially agreed and self-reported emotions in big data. Neural Comput. Appl. 35, 10945–10956.
[35] Leung, C.K., Braun, P., Hoi, C.S.H., Souza, J., Cuzzocrea, A., 2019a. Urban analytics of big transportation data for supporting smart cities, in: Big Data Analytics and Knowledge Discovery - 21st International Conference, DaWaK 2019, Linz, Austria, August 26-29, 2019, Proceedings, Springer. pp. 24–33.
[36] Leung, C.K., Chen, Y., Hoi, C.S.H., Shang, S., Cuzzocrea, A., 2020a. Machine learning and OLAP on big COVID-19 data, in: 2020 IEEE International Conference on Big Data (IEEE BigData 2020), Atlanta, GA, USA, December 10-13, 2020, IEEE. pp. 5118–5127.
[37] Leung, C.K., Chen, Y., Hoi, C.S.H., Shang, S., Wen, Y., Cuzzocrea, A., 2020b. Big data visualization and visual analytics of COVID-19 data, in: 24th International Conference on Information Visualisation, IV 2020, Melbourne, Australia, September 7-11, 2020, IEEE. pp. 415–420.
[38] Leung, C.K., Cuzzocrea, A., Mai, J.J., Deng, D., Jiang, F., 2019b. Personalized deepinf: Enhanced social influence prediction with deep learning and transfer learning, in: IEEE BigData, 2019, IEEE. pp. 2871–2880.
[39] Li, Q., Li, P., Ren, Z., Ren, P., Chen, Z., 2022. Knowledge bridging for empathetic dialogue generation, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 10993–11001.
[40] Lim, M.J., Yi, M.H., Shin, J.H., 2023. Intrinsic emotion recognition considering the emotional association in dialogues. Electronics 12.
[41] Mohammad, S., Bravo-Marquez, F., 2017. WASSA-2017 shared task on emotion intensity, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Copenhagen, Denmark. pp. 34–49. URL: https://aclanthology.org/W17-5205, doi:10.18653/v1/W17-5205.
[42] Mohammad, S.M., 2017. Word affect intensities. arXiv preprint arXiv:1704.08798 .
[43] Mohammad, S.M., Bravo-Marquez, F., Salameh, M., Kiritchenko, S., 2018. Semeval-2018 Task 1: Affect in tweets, in: Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA.
[44] Mohammad, S.M., Sobhani, P., Kiritchenko, S., 2017. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT) 17, 1–23.
[45] Nuruzzaman, M., Hussain, O.K., 2020. Intellibot: A dialogue-based chatbot for the insurance industry. Knowledge-Based Systems 196, 105810.
[46] Pamungkas, E.W., 2019. Emotionally-aware chatbots: A survey. arXiv preprint arXiv:1906.09774 .
[47] Pilato, G., D'Avanzo, E., 2018. Data-driven social mood analysis through the conceptualization of emotional fingerprints. Procedia computer science 123, 360–365.
[48] Poria, S., Majumder, N., Mihalcea, R., Hovy, E., 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access 7, 100943–100953.
[49] Rashkin, H., Smith, E.M., Li, M., Boureau, Y.L., 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. arXiv preprint arXiv:1811.00207 .
[50] Rollo, F., Bonisoli, G., Po, L., 2022. Supervised and unsupervised categorization of an imbalanced italian crime news dataset, in: Information Technology for Management: Business and Social Issues: 16th Conference, ISM 2021, and FedCSIS-AIST 2021 Track, Held as Part of FedCSIS 2021, Virtual Event, September 2–5, 2021, Extended and Revised Selected Papers, Springer. pp. 117–139.
[51] Santhanam, S., Shaikh, S., 2019. A survey of natural language generation techniques with a focus on dialogue systems-past, present and future directions. arXiv preprint arXiv:1906.00500 .
[52] Scherer, K.R., Wallbott, H.G., 1994. " evidence for universality and cultural variation of differential emotion response patterning": Correction. .
[53] Schuff, H., Barnes, J., Mohme, J., Padó, S., Klinger, R., 2017. Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 13–23.
[54] Shmueli, B., Ku, L.W., 2019. Socialnlp emotionx 2019 challenge overview: Predicting emotions in spoken dialogues and chats. arXiv preprint arXiv:1909.07734 .
[55] Varshney, D., Ekbal, A., Bhattacharyya, P., 2021. Modelling context emotions using multi-task learning for emotion controlled dialog generation, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pp. 2919–2931.
[56] Vogt, J., 2021. Where is the human got to go? artificial intelligence, machine learning, big data, digitalisation, and human-robot interaction in industry 4.0 and 5.0. AI Soc. 36, 1083–1087.
[57] Wang, Y., 2020. Iteration-based naive bayes sentiment classification of microblog multimedia posts considering emoticon attributes. Multim. Tools Appl. 79, 19151–19166.
[58] Wani, A.H., Hashmy, R., 2023. A supervised multinomial classification framework for emotion recognition in textual social data. Int. J. Adv. Intell. Paradigms 24, 173–189.
[59] Young, T., Cambria, E., Chaturvedi, I., Zhou, H., Biswas, S., Huang, M., 2018. Augmenting end-to-end dialogue systems with commonsense knowledge, in: Proceedings of the AAAI conference on artificial intelligence.
[60] Zad, S., Heidari, M., James Jr, H., Uzuner, O., 2021. Emotion detection of textual data: An interdisciplinary survey, in: 2021 IEEE World AI IoT Congress (AIIoT), IEEE. pp. 0255–0261.
[61] Zhang, Y., Sun, S., Galley, M., Chen, Y.C., Brockett, C., Gao, X., Gao, J., Liu, J., Dolan, B., 2019. Dialogpt: Large-scale generative pre-training for conversational response generation. arXiv preprint arXiv:1911.00536 .
[62] Zhong, P., Wang, D., Miao, C., 2019. Knowledge-enriched transformer for emotion detection in textual conversations. arXiv preprint arXiv:1909.10681 .