Deep Data Anaylizing Method Based on Scale Space Theory

Ye Hu
College of Mechanical Engineering
Ningbo Dahongying University
Ningbo, China
Tongji University，Shanghai, China
pxhuye@126.com

Guangqun Chen
Ningbo Dahongying University
Ningbo, China
College of Mechanical Engineering
Guangqunchen2016@126.com

Kun Gao
Zhejiang Business Technology Institute, China
Zhejiang Wanli University
Ningbo, China
kungao@live.com

***Abstract*—Scale space theory has been introduced into the field of big data, but its research is still not deep enough and perfect because of the lack of universal theory and method. With deepening of big data processing applications, the research becomes more and more urgent. In view of the above question, this paper studies pervasive multi-scale data analysis theory and method. On one hand, we give the definition and partition of data scale as well as the relationship of multi-scale data set between the upper scale and lower scale based on concept hierarchy theory. On the other hand, we clarify the definition of multi-scale data analysis, study essence and classification method. Previous studies show that the proposed method has high coverage rate, high accuracy rate, lower error rate of support estimation degree and greater improvement the efficiency than the traditional algorithm.**

Keywords: Multi-Scale Space; Big Data; Frequent Item-set; Association Rules;

I. INTRODUCTION

Broadly speaking, the scale is research object of units or measurement tools [1-3]. According to the data as the research object of the data analysis, scale is also a measurement unit of data [4-6]. And statistical measurement scales to variables as benchmark data for similar and data analysis of scale measurement based quasi should be the intrinsic properties of the data [7-9]. When researchers have a survey on the category data often correspond to data in the category of a property set, the attribute set can usually form a partial order structure explicit in the concept hierarchy, according to the concept hierarchy related concept is used to divide the data can be formed with the multi-scale characteristic of the data set [10-12].

From the present study, analysis and research of multi scale data is not fully expanded, the data types involved are more concentrated in spatial data, which limits the development space of multi scale analysis data [13-15]. In order to expand the theory of multi scale analysis research scope and depth in the field of data, this article from the following 3 aspects:

1) will the basic theory of multi-scale science into the general data set, put forward a more universal theory of multi-scale data;

2) describes the definition of multi-scale analysis of the data, essence and classification;

3) proposed the multi-scale data analysis algorithm frame and gives its theoretical basis, then analysis the algorithm framework is applied to association rules further, we propose a multi-scale analysis of association rules algorithm, multi-scale data derivation of implicit association rules, the single scale data sets through knowledge push and push down, switch to other scales Level of knowledge, an analysis can be obtained multi-level knowledge. And the accuracy of the algorithm theoretically evaluated; finally, the real data set for experiment to prove the feasibility and efficiency of the algorithm.

II. MULTI-SCALES FOR DATA

A. *Concept Hierarchy*

Definition 1: concept hierarchy

Concept hierarchy C is a partial order set (C, ≼), where C is a finite set of concepts, ≼ denote C contains a partial order relation between concepts.

A category attribute data sets that have a distinct concept hierarchy partial ordering: every property c_i(i=1,…n) in attribute set C = {c_1,…,c_i,…c_n}can be used as a concept in limited concept set. Corresponding to the partial order relationship in limited concept set, attributes can form a partial order relation depend on domain knowledge. A attribute $c_i \in C$ is corresponding to a number of specific attribute values after the instances, denoted as $V_{ci} = \{v_1, v_2, \ldots, V_m\}$ (v_j present a specific discrete values or continuous intervals). Highly abstract of those attribute values in semantic formed attributes concept h_i, called as attribute values $v_j \in V_{hi}$ (j = 1, ..., m), belong to semantically h_i, denoted as: $V_j \overset{s}{\in}_{c_i}$. In practical applications, the geographical scope of attributes concept set may be formed concept hierarchy: ($C_{location}$, ≼) = {county≼ township ≼village ≼ province≼ city}; Time Category attribute set can also form hierarchical concept: (C_{time}, ≼) = {day≼ month ≼ years}.

B. *The Basic Concept of Data Scale*

In the concept hierarchy (C, ≼), the partial order ≼ indicates concept involved range, the relative size of particle size, or the relative length of time amplitude, then call this concept hierarchy (C, ≼) with a multi-scale feature. For example, the concept hierarchy ($C_{location}$, ≼) = {county ≼ township ≼ village ≼ province≼ city} represents geographical area range from small to large; The concept (C_{time}, ≼) = {day≼ month ≼ years} represents time amplitude from short to long. All of these concepts have the ability to represent the relative size of the data granularity, so they are all equipped with multi-scale characteristics. Using the concept of hierarchical multi-scale features as the standard for dividing, the data set may be formed multi-scale data sets.

Definition 2: Data scale division

The attribute set of a certain category of data set DS forms a concept hierarchy with multi-scale characteristics: (C, ≼) ={c_1≼...c_i≼ ... ≼c_n}, the set of attribute values of a certain concept c_i(i=1,…,n) in the finite concept set C is $V_{c_i} = \{v_1^i, \dots, v_{m_i}^i\}$; According to the different attribute value $v_j^i (j = 1, \dots, m_i)$ of the concept c_i divide DS with the same attribute values to form an independent sub-data set, denoted as $ds_{c_i - v_j^i}$;Data set DS is divided into m_i sub data sets, which form a set of data with the concept of c_i as the partition granularity. This process is called the data scale division based on the concept c_i.

In the process of data scaling, the essential of scale is a unit of measurement with a certain semantic meaning, and its semantics is related to the concept of the concept hierarchy with multi-scale characteristics.

Definition 3: Data Scale

The concept c_i that is based on the classification of data set DS is the data scale of the data set in this division, denoted as $Scale_{DateSet}$=c_i.

Definition 4: Benchmark Scale Data Set

In the results of division for data set DS based on the concept c_i of concept hierarchy (C, ≼), all sub data set $ds_{c_i - v_j^i}$ is the meta scale data set under the data scale $Scale_{DateSet}$=c_i for data set DS. If another scale data sets can be obtained through merging or decomposing from the meta scale data set, then the metascale data set is called the benchmark scale data set, the concept c_i corresponding to the benchmark scale data set is benchmark scale, denoted as Benchmark-Scale.

In general, all meta scale data set of a data scale $Scale_{DataSet}$=c_i can be denoted as the $dataset_{c_i}$to represent the data partition of the scale by the granularity is c_i. Using different concepts in the concept hierarchy as data scale to divide the same data set in multi-scale and multi granularity can obtain multi resolution and multi-scale data set. To investigate data in the manner of multi-scale and multi granularity can obtain implicit knowledge under different scales and different levels, so as to achieve the purpose of analyzing data from multiple scales. The proposed related concepts of data scale and multi-scale data set are similar to the data cube [16] in the representation. However, the multi-scale data and data cube have different focus in nature. The former concerns multi-scale characteristics of the data set itself, while the latter concerns the aggregation values.

C. The relationship between multi scale data sets

Definition 5: upper scale data sets and lower scale data sets

The data set DateSet is divided by data scale $Scale_{DataSet}$=c_x and $Scale'_{DataSet}$=c_y, the obtained data sets are $dataset_{c_x}$ and $dataset_{c_y}$. If c_x≼c_y, then all meta scale data set $dataset_{c_y}$under data scaleScale'$_{DataSet}$=c_y are upper scale data set for all meta scale data set $dataset_{c_x}$ under data scale $Scale_{DataSet}$=c_x, all meta scale data set $dataset_{c_x}$ under data scaleScale'$_{DataSet}$=c_x are lower scale data set for all meta scale data set $dataset_{c_y}$ under data scale Scale'$_{DataSet}$=c_y.

The upper and lower scale data sets are relative concepts. The concept level of data scale in upper scale data set is higher, partition granularity is larger, data set containing data meaning is more generalization; The concept level of data scale in lower scale data set is lower, partition granularity is smaller, so it contains the meaning of data more clear and specific. The upper scale data sets clearly contain more data than the lower scale data set, so the upper scale data set is large scale data sets than the lower scale data set; correspondingly, lower scale data set is small scale data set than the upper scale data set. Large scale data sets and small scale data sets corresponding to the data scales are called large scale and small scale, but also for the relative concept. If the data scale $scale_{dataset}$=c_x and scale'$_{dataset}$=c_y in concept hierarchy c_x≼c_y, then $c_x$is a small scale, c_y is large scale, denoted as $scale_{datase}$≼scale'$_{dataset}$.

III. MULTI-SCALE DATA ANALYSIS

A. Definition of Multi-scale Data Analysis

The main task of multi-scale data analysis has two aspects, that is, the multi scale implementation of data and knowledge of multi scale mining. The former belongs to the data preprocessing, the data partition of the scale can be realized; the latter needs improvement technique in detail, in the form of multiple scale data knowledge discovery, analysis and knowledge derivation are connected with each other. The definition of multi-scale data analysis is given in this paper.

Definition 6: Multi scale data analysis

Multi-scale data analysis refers to the data were multi-scale processing, form a set of multi-scale data, use, or improved data analysis techniques explore multi-scale data of implicit knowledge and analysis of, derived data of various scales form behind the knowledge between relations.

B. The essence of multi-scale data analysis

Scaling is the core content of multi-scale science research, literature [17] has needle of spatial data analysis put forward three kinds of scale transformation of the way, summed up includes two aspects, namely multi-scale multi scale conversion of data and knowledge conversion. In fact, the research method of multi-scale spatial data analysis is also applicable to the analysis of multi scale data. Is not difficult to see that the conversion of data at multiple scales of the principle is simple, but requires multi-scale forms of data were analyzed, workload is big; knowledge of scaling the workload is small, only a single scale of data analysis, but need to solve the problem of scale effect, complexity theory. Scale effect is the conclusion that a certain scale cannot be applied to another scale [18], that is, when the study object to change its scope of performance, size or amplitude, the analysis results will be changed with [19,20]. Similarly, in a certain scale data focuses on knowledge cannot be no difference in applicable in other scales in the data, we must use domain knowledge or multi-scale data set between relationship of analysis results are deduced, naturalization, is it possible to achieve true knowledge of multi-scale conversion. It is obvious that the multi-scale transformation of knowledge is the essence of multi-scale data analysis. The definition of knowledge scale transformation is given below.

Definition 7: knowledge scale conversion

Let SCALE, SCALE' are data scale of data set DS. The process of knowledge transfers from the SCALE to SCALE' is called knowledge scale transformation, denoted as $SCALE \xrightarrow{knowledge} SCALE'$ .If SCALE is large than SCALE' SCALE≼SCALE', then the knowledge scale conversion from SCALE to SCALE' is called as scale down, denoted as SCALE → ↓ SCALE'; on the contrary, the knowledge scale conversion from SCALE' to SCALE is called as scale up, denoted as SCALE'→ ↑ SCALE.

Due to data collected in the actual application, usually only appears to single scale, and there is no need to analysis data according to multiscale expression form. So in this paper, we will take the multi scale transform of knowledge as the research essence of multi-scale data analysis: research methods or algorithms. Knowledge of a single scale data set is derived and calculated, which is suitable for the knowledge of other scale data sets within a certain error range [21,22].

C. *Classification of Multi Scale Data Analysis Algorithms*

Based on the knowledge of scale transformation of the definition and study field scale conversion classification [23,24], from knowledge of the multi-scale conversion direction angle, multi-scale data analysis algorithm scale on Pushover analysis algorithm and scale pushdown analysis algorithm:

1) scale up analysis algorithm: using scale data from the lower concentration of knowledge, and domain knowledge, lower scale data set between the relationship, derived upper scale data hidden knowledge, and not the upper scale data set for direct analysis. Scale push analysis algorithm is designed to use the data of the micro one-sided knowledge of the macro comprehensive knowledge of the data.

2) scales down analysis algorithm: using data from the upper scale concentration of knowledge, domain knowledge and, lower scale data set between the relationship, is the lower scale data hidden knowledge, and not lower scale data set for direct analysis. The scale down push analysis algorithm is designed to utilize the data from the macroscopic and comprehensive knowledge of the details of the data to derive the local knowledge.

IV. Multi Scale Data Analysis Algorithm

A. *Algorithm Framework*

Based on the benchmark scale and scale conversion mechanism, this paper proposes the basic framework of Multi-Scale Data Analyzing Algorithm. The basic idea of the framework is: first select benchmark scale, on benchmark data sets used in data analysis algorithm to get the results of the analysis; then for any other target scale so, through the use of scale conversion mechanism and the conversion methods Standard BS analysis results or knowledge inversion to target scale so, direct access to the target scale data set behind the implicit knowledge of the approximate results, rather wrong target scale data set for direct analysis, the ultimate realization of multi-scale data analysis. The specific steps of the algorithm framework are as follows:

Algorithm 1: the framework of multi-scale data analysis algorithm

INPUT: data set with the characteristic of multi-scale

OUTPUT: results of data analysis on the target scale

1. Select the benchmark scale of data set;
2. Use data analysis algorithm on data set of benchmark scale;
3. While the target scale is not benchmark scale
4. If target scale ≼benchmark scale
5. Run scale down analysis algorithm: benchmark scale → ↓ target scale
6. Else if benchmark scale ≼target scale
7. Run scale up analysis algorithm: benchmark scale → ↑ target scale
8. Exit while all data have been executed

In the algorithm, if the target scale is smaller than the benchmark scale, call scale down algorithm analysis, conduct knowledge scale down conversion, through benchmark scale down analysis, and reverse conduct the results on the target scale; if the target scale is bigger than the benchmark scale, of the analysis show; if so scale is larger than the benchmark scale, call scale up algorithm analysis, conduct knowledge scale up conversion, through the benchmark scale analysis, and reverse conduct the results on the target scale; If the interested scale has been analyzed, the algorithm ends and returns to the analysis of the results of each scale.

V. Conclusion

In this paper, the basic idea of multi-scale science is introduced into the field of data analysis. Data classification and data scale concept hierarchy scale definition is proposed based on the relationship between the upper and lower scale data set between multi-scale data sets is given, laid the foundation for the theory of multi-scale data; the core of knowledge conversion based on multi-scale, given the definition and classification of multi-scale data analysis; proposed the multi-scale data analysis of algorithm framework, and gives the theoretical basis of knowledge transformation in the framework of multi-scale algorithm, and this algorithm is applied to multi-scale analysis of association rules, proposed the multi-scale analysis of association rules algorithm, numerical analysis results with benchmark data sets. And the benchmark data set for target scale data set weight that is the target scale data sets of association rules behind the realization of the cross scale derived knowledge, provides the possibility for multi-scale decision.

Acknowledgment

This research is supported in part by Zhejiang Provincial Natural Science Foundation of China (No. LY16F020012), Ningbo Key Laboratory of intelligent home appliances (No.

2016A22008), Key Research Center of Philosophy and Social Science of Zhejiang Province-Modern Port Service Industry and Creative Culture Research Center, Zhejiang Province Public Technology Research and Industrial Project (No. 2015C32124), Projects in Science and Technique of Ningbo Municipal (No. 2014C10047 and No. 2012B82003), National Natural Science Foundation of China (No.61300202) and Natural Science Foundation of Shanghai (No. 13ZR1452900). Kun Gao is the corresponding author.

REFERENCES

[1] Felzenszwalb, P., McAllester, D., and Ramanan, D.: 'A discriminatively trained, multi-scale, deformable part model', in Editor (Ed.)^(Eds.): 'Book A discriminatively trained, multi-scale, deformable part model' (IEEE, 2008, edn.), pp. 1-8

[2] Kolda, T.G., and Sun, J.: 'Scalable tensor decompositions for multi-aspect data mining', in Editor (Ed.)^(Eds.): 'Book Scalable tensor decompositions for multi-aspect data mining' (IEEE, 2008, edn.), pp. 363-372

[3] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., and Euler, T.: 'Yale: Rapid prototyping for complex data mining tasks', in Editor (Ed.)^(Eds.): 'Book Yale: Rapid prototyping for complex data mining tasks' (ACM, 2006, edn.), pp. 935-940

[4] Wang, Z., Simoncelli, E.P., and Bovik, A.C.: 'Multi-scale structural similarity for image quality assessment', in Editor (Ed.)^(Eds.): 'Book Multi-scale structural similarity for image quality assessment' (Ieee, 2003, edn.), pp. 1398-1402

[5] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H.: 'The WEKA data mining software: an update', ACM SIGKDD explorations newsletter, 2009, 11, (1), pp. 10-18

[6] Riedel, E., Gibson, G., and Faloutsos, C.: 'Active storage for large-scale data mining and multimedia applications', in Editor (Ed.)^(Eds.): 'Book Active storage for large-scale data mining and multimedia applications' (Citeseer, 1998, edn.), pp. 62-73

[7] Huan, T., Sivachenko, A.Y., Harrison, S.H., and Chen, J.Y.: 'ProteoLens: a visual analytic tool for multi-scale database-driven biological network data mining', BMC bioinformatics, 2008, 9, (Suppl 9), pp. S5

[8] Eldawlatly, S., Jin, R., and Oweiss, K.G.: 'Identifying functional connectivity in large-scale neural ensemble recordings: a multi-scale data mining approach', Neural computation, 2009, 21, (2), pp. 450-477

[9] Li, T., Li, Q., Zhu, S., and Ogihara, M.: 'A survey on wavelet applications in data mining', ACM SIGKDD Explorations Newsletter, 2002, 4, (2), pp. 49-68

[10] Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A.: 'Comparing community structure identification', Journal of Statistical Mechanics: Theory and Experiment, 2005, 2005, (09), pp. P09008

[11] Fu, T.-c.: 'A review on time series data mining', Engineering Applications of Artificial Intelligence, 2011, 24, (1), pp. 164-181

[12] Pal, S.K., and Mitra, P.: 'Pattern recognition algorithms for data mining' (CRC press, 2004. 2004)

[13] Tsoumakas, G., Katakis, I., and Vlahavas, I.: 'Mining multi-label data': 'Data mining and knowledge discovery handbook' (Springer, 2009), pp. 667-685

[14] Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., and Onnela, J.-P.: 'Community structure in time-dependent, multi-scale, and multiplex networks', science, 2010, 328, (5980), pp. 876-878

[15] Han, J., Kamber, M., and Pei, J.: 'Data mining: concepts and techniques' (Elsevier, 2011. 2011)

[16] X. Luo, Zheng Xu, J. Yu, and X. Chen. Building Association Link Network for Semantic Link on Web Resources. IEEE transactions on automation science and engineering, 2011, 8(3), 482-494.

[17] C. Hu, Zheng Xu, et al. Semantic Link Network based Model for Organizing Multimedia Big Data. IEEE Transactions on Emerging Topics in Computing, 2014, 2(3), 376-387.

[18] Zheng Xu et al. Semantic based representing and organizing surveillance big data using video structural description technology. The Journal of Systems and Software, 2015, 102, 217-225.

[19] Zheng Xu et al. Knowle: a Semantic Link Network based System for Organizing Large Scale Online News Events. Future Generation Computer Systems, 2015, 43-44, 40-50.

[20] Zheng Xu et al. Semantic Enhanced Cloud Environment for Surveillance Data Management using Video Structural Description. Computing, 98(1-2):35-54, 2016.

[21] Zheng Xu et al. Crowdsourcing based Social Media Data Analysis of Urban Emergency Events. Multimedia Tools and Applications, 10.1007/s11042-015-2731-1.

[22] C. Hu, Zheng Xu, et al. Video Structured Description Technology for the New Generation Video Surveillance System. Frontiers of Computer Science, 2015, 9(6): 980-989.

[23] Zheng Xu et al. Crowdsourcing based Description of Urban Emergency Events using Social Media Big Data. IEEE Transactions on Cloud Computing,10.1109/TCC.2016.2517638.

[24] Zheng Xu et al. Generating Temporal Semantic Context of Concepts Using Web Search Engines. Journal of Network and Computer Applications, 2014, 43, 42-55.