PICKT: A Solution for Big Data Analysis

Tianrui Li(✉), Chuan Luo, Hongmei Chen, and Junbo Zhang

School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
{trli,hmchen}@swjtu.edu.cn, luochuan@my.swjtu.edu.cn, JunboZhang86@163.com

Abstract. Emerging information technologies and application patterns in modern information society, e.g., Internet, Internet of Things, Cloud Computing and Tri-network Convergence, are growing in an amazing speed which causes the advent of the era of Big Data. Big Data is often described by using five V's: Volume, Velocity, Variety, Value and Veracity. Exploring efficient and effective data mining and knowledge discovery methods to handle Big Data with rich information has become an important research topic in the area of information science. This paper focuses on the introduction of our solution, PICKT, on big data analysis based on the theories of granular computing and rough sets, where P refers to parallel/cloud computing for the Volume, I refers to incremental learning for the Velocity, C refers to composite rough set model for the Variety, K refers to knowledge discovery for the Value and T refers to three-way decisions for the Veracity of Big Data.

Keywords: Big data · Rough set · Granular computing · Incremental learning

1 Introduction

Enormous amounts of data are generated every day with the amazing spread of computers and sensors in a range of domains, *e.g.*, search engines, social media, health care organizations, insurance companies, financial industry, retail, and many others [1,2]. Now we are in the era of Big Data, which is characterized by 5Vs, *i.e.*, Volume, Velocity, Variety, Value and Variety. Volume means the amount of data that needs to be handled is very huge. Velocity means that the speed of data processing is very high. Variety means that the data is varied in nature and there are many different types of data that need to be properly combined to make the most of the analysis. Value means high yield will be achieved if the big data is processed correctly and accurately. Veracity means the inherent trustworthiness of data, namely, the uncertainty about the consistency or completeness of data and other ambiguities [3]. Big data is currently a fast growing field both from an application and from a research point of view.

Granular computing is a newly computing paradigm in the realm of computational intelligence, which offers a multi-view model to describe and process

D. Ciucci et al. (Eds.): RSKT 2015, LNAI 9436, pp. 15–25, 2015.
DOI: 10.1007/978-3-319-25754-9_2

uncertain, imprecise, and incomplete information. According to the selection of a suitable granule and the granularity conversion in the complex problem space, granular computing provides a granular-based approximate solution for mining big data [4]. There are two key issues in granular computing. One is how to select the right level of granularity in problem solving. The story of "The Blind Men and the Elephant" tells us the importance of how to select the right level of granularity. Like an adage says, "One person standing in the right place at the right time possesses the power to change the world." Here is an example to illustrate the importance of the selection of the right level of granularity. The length granularity could be Kilometer, Meter, Centimeter, Micron, ... If you are asked to use the ruler (30 cm) to measure the length of one book, it surely does the job. If you are asked to measure the distance across a desk, you can simply use the ruler. You may have to pick it up and move it several times, but it still does the job. But what if you are asked to measure the distance from one side of campus to another side or the diameter of a cell? The ruler is NOT the right tool for that job. The other one is how to change granularity efficiently in problem solving. Granularity variation often appears when the data evolves with time in real situations, *e.g.*, the variation of attributes, objects or attributes' values in information systems. If you stay in the wrong granularity, like using the ruler to measure the diameter of a cell, you may not easily obtain the solutions of problems. Then how you can change the granularity efficiently by the previous knowledge to get the solutions of problems becomes an important issue in knowledge discovery.

This paper aims to introduce our solution, PICKT (see Fig. 1), on big data analysis based on the theories of granular computing and rough sets, where P refers to parallel/cloud computing for the Volume, I refers to incremental learning for the Velocity, C refers to composite rough set model for the Variety, K refers to knowledge discovery for the Value and T refers to three-way decisions for the Veracity of Big Data. The PICKT (German) means the peck (English). We compare the process of big data analysis to the activity that a woodpecker searches food in the forest. Thus, we use PICKT as a short name to represent our solution for big data analysis.

2 Parallel Computing for the Volume

Volumes of data have increased exponentially due to ubiquitous information-sensing mobile devices, remote sensing, software logs, cameras and wireless sensor networks as well as the increasing capacity to store information [5]. It is not uncommon for people to deal with petabytes of data, and the analysis is typically performed over the entire data set, not just a sample, which has become a huge challenge not only in science but also in business use [6]. As a powerful mathematical tool that can be used to process inconsistent information, Rough Set Theory (RST) has been used successfully for discovering patterns in a given data set through a pair of concepts, namely, the upper and lower approximations [7]. The efficient computation of these approximations is vital to improve

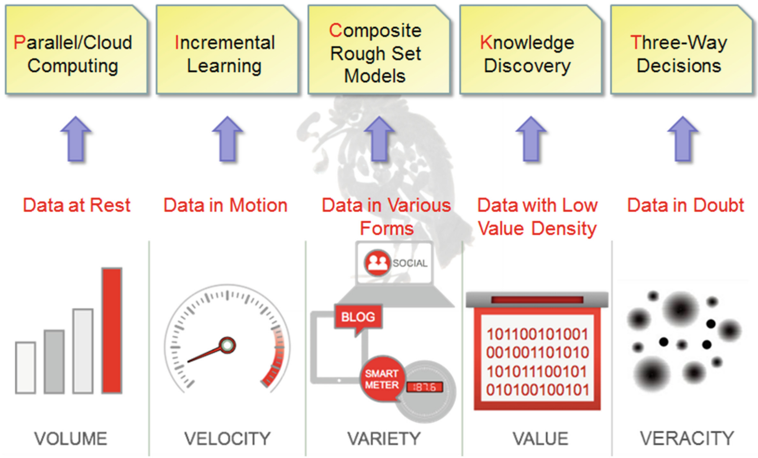

Fig. 1. PICKT—A solution for Big Data Analysis

the performance and obtain the benefits that rough sets can provide for data mining and related tasks.

In the big data environment, we firstly presented a parallel method for computing rough set approximations by the parallel programming model, MapReduce, which is a framework for processing huge data sets using a large number of computers (nodes) [8]. The algorithms corresponding to the parallel method based on MapReduce were designed. Speedup, scaleup and sizeup were used to evaluate the performances of the proposed parallel algorithms. Experiments on the real and synthetic data sets validated that they could deal with large data sets effectively in data mining. Following that, we compared the parallel algorithms of computing approximations of rough sets and knowledge acquisition on three representative MapReduce runtime systems, e.g., Hadoop, Phoenix and Twister [9]. Also, we presented a parallel and incremental algorithm for updating knowledge based on RST in cloud platform [10].

Secondly, we developed three parallel strategies based on MapReduce to compute approximations by introducing the matrix representations of lower and upper approximations to process large-scale incomplete data with RST [11]. The first strategy is that a Sub-Merge operation is used to reduce the space requirement and accelerate the process of merging the relation matrices. The second strategy is that an incremental method is applied to the process of merging the relation matrices and the computational process is efficiently accelerated since the relation matrix is updated in parallel and incrementally. The third strategy is that a sparse matrix method is employed to optimize the proposed matrix-based method and the performance of the algorithm is further improved.

As an extension of classical RST, Dominance-based Rough Set Approach (DRSA) can process information with preference-ordered attribute domain and has been successfully applied in multi-criteria decision analysis and other related tasks [12]. We further studied the parallel approach for computing approximations in DRSA based on the classification of information systems [13]. Its idea is to decompose the computation of approximations based on the decomposition of computing basic knowledge granules and concept granules (upward and downward unions of decision classes). By this strategy, it guarantees each process element can do its part of work concurrently. Following that, approximations of DRSA are obtained by composing those interim results computed in parallel. Experimental evaluation on a multi-core environment showed that the parallel method can dramatically decrease the time of computing approximations in DRSA.

Fourthly, we presented a unified parallel large-scale framework for feature selection based on rough sets in big data analysis [14]. Its corresponding three parallel methods were proposed, *e.g.*, model parallelism (MP), data parallelism (DP), and model-data parallelism (MDP). In this case, we only considered heuristic feature selection whose core is to calculate the significance measures of features based on rough sets. Then we presented a unified representation of feature evaluation functions. Furthermore, we showed the divide-and-conquer methods for four representative evaluation functions, and designed MapReduce-based and Spark-based Parallel Large-scale Attribute Reduction (PLAR) algorithms [15]. Subsequently, we accelerated the process of feature selection by introducing granular computing (GrC) theory and presented PLAR-MDP algorithm by combining PLAR with MDP. Finally, we verified the effectiveness of the proposed algorithms by experimental evaluation on the big data computing platforms, *e.g.*, Hadoop and Spark, using UCI datasets and astronomical datasets. It was also shown that PLAR-MDP can maximize the performance of data processing by combining with MP, DP and GrC methods.

3 Incremental Learning for the Velocity

In real-life applications, data in information systems will evolve over time in the sense that (1) records (objects) are continuously being added or deleted, (2) features (attributes) are frequently being added or removed, and (3) criteria values (attribute values) are continually being updated over time. Tracking these dynamic changes may contribute to improve the ability of decision making. Incremental learning is regarded as the process of using previously acquired learning results to facilitate knowledge maintenance by dealing with the new added-in data set without re-implementing the original data mining algorithm, which leads to considerable reductions in execution time when maintaining knowledge from the dynamic database [16]. Based on GrC and RST, there have been many attempts at introducing incremental learning techniques for updating knowledge in a dynamic data environment.

For the variation of objects, we presented several incremental approaches for updating approximations and knowledge under the extended models of rough sets. Firstly, we proposed a novel incremental model and approach to induce interesting knowledge under both accuracy and coverage from dynamic information systems when the object set varies with time [17]. Its feature is to calculate both accuracy and coverage by matrixes to obtain the interesting knowledge. Furthermore, we presented an optimization incremental approach for inducing interesting knowledge in business intelligent information systems [18].

Variable Precision Rough Set model (VPRS) has been used to solve noise data problems with great success in many applications [19]. Then, we proposed incremental methods for updating approximations under VPRS when the information system is updated by inserting or deleting an object through investigating the property of information granulation and approximations under the dynamic environment. The variation of an attribute's domain was also considered to perform incremental updating for approximations under VPRS [20].

To handle information with preference-ordered attribute domain, we further proposed an incremental approach for updating the approximations of DRSA under the variation of the object set [21]. The idea is that we only focus on the objects whose P-generalized decision is updated after the addition or deletion of the object. Then, approximations of DRSA can be updated incrementally instead of re-computing from scratch according to the alteration of P-dominated sets and P-dominating sets of those objects. We also extended the proposed method to update approximations based on set-valued ordered information systems in terms of inserting or deleting an object [22].

However, the previous methods can only be used to deal with the categorical data. We presented a new dynamic method for incremental updating knowledge based on the neighborhood rough sets to deal with dynamic numerical data when multiple objects enter into and get out of the information system simultaneously [23]. Following that, we introduced an incremental algorithm for updating approximations of rough fuzzy sets under the variation of the object set in fuzzy decision systems [24]. Moreover, we developed an algorithm to maintain approximations incrementally when adding or removing some objects in composite information systems [25].

For the variation of attributes, based on Chan's work to update approximations by the lower and upper boundary sets [26], we presented a method for updating approximations in incomplete information systems under the characteristic relation-based rough sets when an attribute set varies over time [27]. Then, we proposed an incremental approach for updating approximations in set-valued information systems under the probabilistic neighborhood rough set model [28]. Furthermore, we developed an incremental algorithm for updating approximations under the DRSA by introducing a kind of dominance matrix to update P-dominating sets and P-dominated sets [29]. We also presented incremental approaches for updating approximations in set-valued ordered decision systems by introducing the dominant and dominated matrices with respect to the dominance relation for computing approximations of upward and downward unions of decision classes [30]. We discussed the updating mechanisms for feature selection with the variation of attributes and presented fuzzy rough set approaches for incremental feature selection in hybrid information systems, which consist of different types of data (*e.g.*, boolean, categorical, real-valued and set-valued) [31].

In fact, the ever-changing attribute values may lead some or all of obtained knowledge to failure. For the variation of attribute values, we defined the attribute values' coarsening and refining in information systems as the semantic level changes, and then proposed an incremental algorithm for updating the approximations in case of attribute values' coarsening and refining [32]. Then, we presented a method to dynamically maintain approximations of upward and downward unions when attribute values alter in incomplete ordered decision systems [33]. We also developed incremental algorithms for computing approximations in the set-valued decision systems with the addition and removal of criteria values [34]. Furthermore, we presented matrix-based incremental algorithms for updating decision rules when coarsening or refining attribute values [35]. We have developed "iRoughSet", a toolkit for incremental learning based on RST [36] based on the above research works.

4 Composite Rough Set Model for the Variety

In many real-world applications, attributes in information systems generally have different types, *e.g.*, categorical ones, numerical ones, set-valued ones, interval-valued ones and missing ones. However, most of the previous rough set models fail to deal with information systems with more than two different types of attributes. We firstly defined composite information systems. Then, we introduced an extension of rough set model, composite rough set (CRS), which may deal with multiple different types of attributes and provide a novel approach for complex data fusion. Furthermore, we introduced a matrix-based method to compute the CRS approximations, positive, boundary and negative regions intuitively from composite information systems, which may help to combine different types of data to make the most of the analysis [37]. Finally, we developed parallel algorithms based on Single-GPU and Multi-GPU for computing CRS approximations and carried out experiments on UCI and user-defined datasets to verify the effectiveness of the proposed algorithms [38].

However, CRS cannot be employed directly to mine the useful knowledge hidden in the Incomplete Composite Information System (ICIS). We then introduced a new composite rough set model based on a composite binary relation by the intersection of tolerance and dominance relations. This model can be applied to the ICIS directly when all criteria and attributes are considered at the same time [25]. In addition, considering the preference relations between objects may be different under different criteria in the ordered information system, we proposed a composite dominance relation based on the integration of multiple dominance relations and then presented a composite DRSA model for the composite ordered information system [39].

5 Knowledge Discovery for the Value

Knowledge Discovery in Database (KDD) is regarded as the process of discovering previously unknown and potentially interesting patterns in large databases, which includes the storage and accessing of such data, scaling of algorithms to massive data sets, interpretation and visualization of results, and the modeling and support of the overall human machine interaction [40]. However, the previous KDD process models do not pay attention to the data collection, which is vital to apply the KDD techniques in some real applications, *e.g.*, information security and medical treatment. Therefore, we proposed an extension of the KDD process model (see Fig. 2) of Fayyad et al. which incorporates the step of data collection that employs techniques to collect data for real applications according to the current discovery task and previous mining results [41]. The cost of data collection is generally huge, *e.g.*, making different kinds of investigations for nuclear safeguards information management. By applying the proposed model, we can not only reduce the cost including storage, preprocessing, *etc.*, but also directly enhance the ability and efficiency of knowledge discovery, *e.g.*, reducing the time required to execute a discovery. Then, we illustrated a case study by the reduct method in RST to show in what situation it can be used in practice and how it can support KDD applications better. We have developed "RSHadoop", a rough set toolkit for massive data analysis on Hadoop for large-scale knowledge discovery based on RST [42].

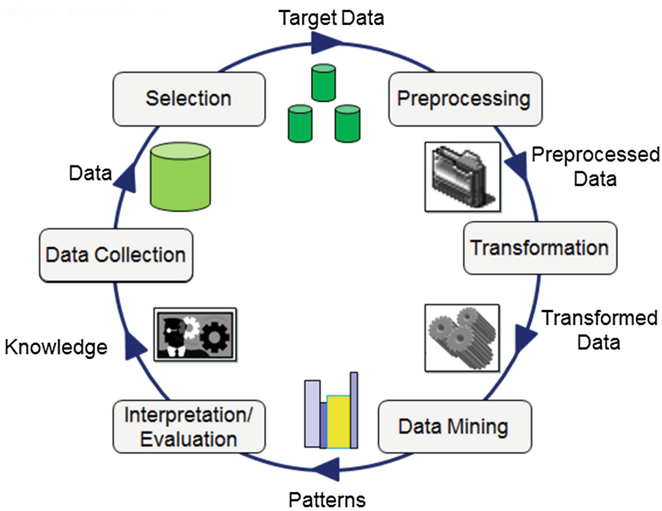

Fig. 2. An updated process model of KDD

6 Three-Way Decisions for the Veracity

In real application situations, there are many biases, noise or abnormality in data. Some people believe that veracity in data analysis is the biggest challenge when compared to things like volume and velocity. The theory of three-way decisions proposed by Yao is an effective mathematical tool to deal with the veracity [43]. Decision-theoretic rough set (DTRS) provides a three-way decision framework for approximating a target concept, with an error-tolerance capability to handle uncertainty problems by using a pair of thresholds with probability estimation. The three-way decision rules of acceptance, rejection and deferment decisions can be derived directly from the three regions implied by rough approximations [44].

In a dynamic data environment, we considered the situation of the variation of objects. With the addition of different objects, different updating patterns of conditional probability will appear, and then the decision regions implied by rough approximations will change dynamically in a non-monotonic way. To address this problem, we presented the incremental estimating methods for the conditional probability in the DTRS [45]. Then, we designed the incremental mechanisms for updating rough approximations with an increase or a decrease of conditional probability. On the other hand, since the rough approximations will be varied inevitably with the data changed continuously, the induced three-way decision rules will be altered accordingly. We developed four different maintenance strategies of the three-way decision rules which can update decision rules by modifying partial original rule sets without forgetting prior knowledge, thereby avoiding re-computing decision rules from the very beginning [46,47]. Furthermore, we proposed an incremental approach for updating probabilistic rough approximations, *i.e.*, positive, boundary and negative regions, in incomplete information systems [48]. Finally, we presented a dynamic DTRS approach for updating rough approximations when objects and attributes are altered simultaneously by defining the equivalence feature vector and matrix, which may avoid some time-consuming operations of sets [49].

7 Conclusion

This paper introduced the PICKT solution for big data analysis based on the theories of GrC and RST. However, it is still far away from the target of fully using big data. In our future research work, we will continue to study the efficient algorithms for mining big data in different situations and develop a platform for big data analysis based on cloud computing and the research output from GrC and RST.

Acknowledgments. This work is supported by the National Science Foundation of China (Nos. 61175047 and 71201133) and NSAF (No. U1230117).

References

1. Li, T.R., Lu, J., Luis, M.L.: Preface: Intelligent techniques for data science. Int. J. Intel. Syst. **30**(8), 851–853 (2015)
2. Dealing with data: Science **331**(6018), 639–806 (2011)
3. Jin, X.L., Wah, B.W., Cheng, X.Q., Wang, Y.Z.: Significance and challenges of big data research. Big Data Res. **2**(2), 59–64 (2015)
4. Yao, Y.Y.: Granular computing: past, present and future. In: Proceedings of GrC 2008, pp. 80–85 (2008)
5. Big data. Nature 455(7209), 1–136 (2008)

6. Wu, X.D., Zhu, X.Q., Wu, G.Q., Ding, W.: Data mining with big data. IEEE Trans. Knowl. Data Eng. **26**(1), 97–107 (2014)
7. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Inf. Sci. **177**, 3–27 (2007)
8. Zhang, J.B., Li, T.R., Ruan, D., Gao, Z.Z., Zhao, C.B.: A parallel method for computing rough set approximations. Inf. Sci. **194**, 209–223 (2012)
9. Zhang, J.B., Wong, J.S., Li, T.R., Pan, Y.: A comparison of parallel large-scale knowledge acquisition using rough set theory on different MapReduce runtime systems. Int. J. Approx. Reasoning **55**(3), 896–907 (2014)
10. Zhang, J.B., Li, T.R., Pan, Y., Luo, C., Teng, F.: Parallel and incremental algorithm for knowledge update based on rough sets in cloud platform. J. Softw. **26**(5), 1064–1078 (2015). (in Chinese)
11. Zhang, J.B., Wong, J.S., Pan, Y., Li, T.R.: A parallel matrix-based method for computing approximations in incomplete information systems. IEEE Trans. Knowl. Data Eng. **27**(2), 326–339 (2015)
12. Greco, S., Matarazzo, B., Slowinski, R.: Rough set theory for multicriteria decision analysis. Eur. J. Oper. Res. **129**, 1–47 (2001)
13. Li, S.Y., Li, T.R., Zhang, Z.X., Chen, H.M., Zhang, J.B.: Parallel computing of approximations in dominance-based rough sets approach. Knowledge-Based Systems (in Press, 2015)
14. Zhang, J.: Research on efficient feature selection and learning algorithms for big data. Ph.D. dissertation, Southwest Jiaotong University (2015)
15. Zhang, J.B., Li, T.R., Pan, Y.: PLAR: Parallel large-scale attribute reduction on cloud systems. In: Proceedings of PDCAT 2013, pp. 184–191 (2013)
16. Raghavan, V., Hafez, A.: Dynamic data mining. In: Proceedings of the IEA of AI & ES, pp. 220–229 (2000)
17. Liu, D., Li, T.R., Ruan, D.: An incremental approach for inducing knowledge from dynamic information systems. Fundamenta Informaticae **94**(2), 245–260 (2009)
18. Liu, D., Li, T.R., Ruan, D., Zhang, J.B.: Incremental learning optimization on knowledge discovery in dynamic business intelligent systems. J. Glob. Optim. **51**(2), 325–344 (2011)
19. Ziarko, W.: Variable precision rough set model. J. Comput. Syst. Sci. **46**(1), 39–59 (1993)
20. Chen, H.M., Li, T.R., Ruan, D., Lin, J.H., Hu, C.X.: A rough-set based incremental approach for updating approximations under dynamic maintenance environments. IEEE Trans. Knowl. Data Eng. **25**(2), 274–284 (2013)
21. Li, S.Y., Li, T.R., Liu, D.: Dynamic maintenance of approximations in dominance-based rough set approach under the variation of the object set. Int. J. Intel. Syst. **28**, 729–751 (2013)
22. Luo, C., Li, T.R., Chen, H.M., Liu, D.: Incremental approaches for updating approximations in set-valued ordered information systems. Knowl.-Based Syst. **50**, 218–233 (2013)
23. Zhang, J.B., Li, T.R., Ruan, D., Liu, D.: Neighborhood rough sets for dynamic data mining. Int. J. Intel. Syst. **27**, 317–342 (2012)
24. Zeng, A.P., Li, T.R., Zhang, J.B., Chen, H.M.: Incremental maintenance of rough fuzzy set approximations under the variation of object set. Fundamenta Informaticae **132**(3), 401–422 (2014)
25. Li, S.Y., Li, T.R., Hu, J.: Update of approximations in composite information systems. Knowl.-Based Syst. **83**, 138–148 (2015)
26. Chan, C.: A rough set approach to attribute generalization in data mining. Inf. Sci. **107**(1–4), 169–176 (1998)

27. Li, T.R., Ruan, D., Geert, W.: A rough sets based characteristic relation approach for dynamic attribute generalization in data mining. Knowl.-Based Syst. **20**(5), 485–494 (2007)
28. Zhang, J.B., Li, T.R., Ruan, D.: Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems. Int. J. Approx. Reasoning **53**, 620–635 (2012)
29. Li, S.Y., Li, T.R., Liu, D.: Incremental updating approximations in dominance-based roughsets approach under the variation of the attirbute set. Knowl.-Based Syst. **40**, 17–26 (2013)
30. Luo, C., Li, T.R., Chen, H.M.: Dynamic maintenance of approximations in set-valued ordered decision systems under the attribute generalization. Inf. Sci. **257**, 210–228 (2014)
31. Zeng, A.P., Li, T.R., Liu, D., Zhang, J.B., Chen, H.M.: A fuzzy rough set approach for incremental feature selection on hybrid information systems. Fuzzy Sets Syst. **258**, 39–60 (2015)
32. Chen, H.M., Li, T.R., Qiao, S.J., Ruan, D.: A rough set based dynamic maintenance approach for approximations in coarsening and refining attribute values. Int. J. Intel. Syst. **25**(10), 1005–1026 (2010)
33. Chen, H.M., Li, T.R., Ruan, D.: Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining. Knowl.-Based Syst. **31**, 140–161 (2012)
34. Luo, C., Li, T.R., Chen, H.M., Lu, L.X.: Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values. Inf. Sci. **299**, 221–242 (2015)
35. Chen, H.M., Li, T.R., Luo, C., Horng, S.J., Wang, G.Y.: A rough set-based method for updating decision rules on attribute values' coarsening and refining. IEEE Trans. Knowl. Data Eng. **26**(12), 2886–2899 (2014)
36. http://sist.swjtu.edu.cn:8080/ccit/project/iroughset.html
37. Zhang, J.B., Li, T.R., Chen, H.M.: Composite rough sets for dynamic data mining. Inf. Sci. **257**, 81–100 (2014)
38. Zhang, J., Zhu, Y., Pan, Y., Li, T.: A parallel implementation of computing composite rough set approximations on GPUs. In: Lingras, P., Wolski, M., Cornelis, C., Mitra, S., Wasilewski, P. (eds.) RSKT 2013. LNCS, vol. 8171, pp. 240–250. Springer, Heidelberg (2013)
39. Luo, C., Li, T.R., Chen, H.M., Zhang, J.B.: Composite ordered information systems. J. Chin. Comput. Syst. **35**(11), 2523–2527 (2014). (In Chinese)
40. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: The KDD process for extracting useful knowledge from volumes of data. Commun. ACM **39**(11), 27–34 (1996)
41. Li, T.R., Ruan, D.: An extended process model of knowledge discovery in database. J. Enterp. Inf. Manage. **20**(2), 169–177 (2007)
42. http://sist.swjtu.edu.cn:8080/ccit/project/rshadoop.html
43. Yao, Y.: An outline of a theory of three-way decisions. In: Yao, J.T., Yang, Y., Słowiński, R., Greco, S., Li, H., Mitra, S., Polkowski, L. (eds.) RSCTC 2012. LNCS, vol. 7413, pp. 1–17. Springer, Heidelberg (2012)
44. Yao, Y.Y.: Three-way decisions with probabilistic rough sets. Inf. Sci. **180**, 341–353 (2010)
45. Luo, C., Li, T.R., Chen, H.M.: Dynamic three-way decisions method based on probabilistic rough sets. In: Yu, H., Wang, G.Y., Li, T.R., Liang, J.Y., Miao, D.Q., Yao, Y.Y. (eds.) Three-Way Decisions: Methods and Practices for Complex Problem Solving. Science Press, Beijing (2015). (in Chinese)

46. Luo, C., Li, T., Chen, H.: Dynamic maintenance of three-way decision rules. In: Miao, D., Pedrycz, W., Slezak, D., Peters, G., Hu, Q., Wang, R. (eds.) RSKT 2014. LNCS, vol. 8818, pp. 801–811. Springer, Heidelberg (2014)
47. Luo C.: Research on approaches for dynamic knowledge acquisition from incomplete data. PhD dissertation, Southwest Jiaotong University (2015) (In Chinese)
48. Luo, C., Li, T.: Incremental three-way decisions with incomplete information. In: Cornelis, C., Kryszkiewicz, M., Ślęzak, D., Ruiz, E.M., Bello, R., Shang, L. (eds.) RSCTC 2014. LNCS, vol. 8536, pp. 128–135. Springer, Heidelberg (2014)
49. Chen, H.M., Li, T.R., Luo, C., Horng, S.J.: Wang G.Y: A decision-theoretic rough setapproach for dynamic data mining. IEEE Trans. Fuzzy Syst. (2015). doi:10.1109/TFUZZ.2014.2387877