

A Two- Step CBR Method Based on Sequential Data

Quan Xiao

*School of Information Technology, Jiangxi University of Finance and Economics,
330032 Nanchang, China
xiaoquan@foxmail.com*

Abstract

Case-Based Reasoning (CBR) is widely used for problem solving in many fields, but there still exist limitations for the problems with dynamic characteristics. This work tries to introduce approaches in Sequence Pattern Mining (SPM) to extend the capability of CBR for solving problems described with sequential data. We propose a SPM algorithm named wGSP, which takes user's different concerns on events into consideration by weight setting, to discover typical patterns in case base. Then the support information of cases to typical patterns is used to depict cases and facilitate efficient case retrieval. The contribution of this work lies in two aspects: firstly it is an improvement to traditional CBR method for coping with sequential data based cases with better interpretability and higher efficiency; secondly, it provides flexibility for parameters setting in SPM to satisfy the personalized preferences of users. Finally through a calculating instance, the advantages and effectiveness of the two-step CBR method based on sequential data is illustrated.

Keywords: *case-based reasoning, case retrieval, sequential data, sequential pattern mining*

1. Introduction

CBR is an important branch of AI research as a problem solving schema. The idea of CBR is analogical reasoning, which comes from the cognitive mode when human solving problems: the experiences of solving problems in the past can be used as references to deal with new similar problem [1]. On account of good explainability and operability of CBR, so far it has been successfully applied to many fields, such as emergency preparedness [2], bankruptcy prediction [3], fault diagnosis [4], software design [5], image retrieval [6], stock selection [7] *etc.*

However, most of current CBR researches are based on feature attributes, in which way the cases are described and stored in case base statically to certain extent [7]. As a result, the problem solving capability of CBR for sequential and dynamic cases is relatively weak. Actually, a lot of problems in the real world are time sequence related, and the traditional CBR methods based on feature attributes are just available for solving static or slightly dynamic problems. Therefore, it is necessary for CBR to intensify its capability for solving complex problems with sequential characteristics in a reasonable and efficient way.

Sequential data are very common in DM researches. Problems such as customer buying behavior prediction, Web visiting mode prediction, disease diagnosis, network intrusion detection are all involved with sequential data and their mining [8]. It may be a good idea to take advantage of approaches in SPM to improve the capability of CBR for solving complex time sequence related problems [9].

Specific to cases described by sequential data, in this paper we introduce SPM in DM field and propose a two-step CBR method as the problem solving framework. For the first step we propose a weighted SPM algorithm named wGSP and get typical case patterns in

case base. For the second step, we measure the similarity between cases for case retrieval by means of the information of sequential data in cases supporting to typical case patterns. Our contribution is that we extend the problem solving capability of CBR to time sequential cases, and the new CBR cycle based on typical case patterns can increase the efficiency of case retrieval and satisfy the customized preferences of users.

The remainder of this paper is organized as follows. In the next section, we introduce related researches of CBR and SPM in order to state the necessity of this work. Section 3 describes the main methodology of this paper, which includes two steps of weighted SPM and similarity measuring of sequential cases. Before drawing conclusions, in Section 4 a calculating instance is illustrated and discussed in detail.

2. Related Researches

2.1. CBR

Roger Schank's research on cognitive science and the proposition of concept "script" are widely considered as the origin of CBR thoughts [10]. The basic idea of CBR derives from the cognitive mode of human when facing with problems. When a new problem is coming, we will usually recall the similar problems solved in the past, and then reuse and revise the useful methods and experiences to solve the new problem.

In CBR systems, case base is the foundation of operating, where abundant domain related instances are stored as cases and described with specific forms such as feature attributes set and text. In terms of the reasoning module, the stored cases in accord with the current problem context can be retrieved and then be revised as the final solution. As the 4R reasoning mode proposed by Aamodand Plaza [11], a typical CBR working process includes Retrieve, Reuse, Revise and Retain of case as shown in Figure 1. If the Representing of the problem case in considered, the CBR working process can be described as a 5R model [12]. CBR has good adaptability to combine with other AI approaches. Such as Robert and Melody investigated the combination of CBR and RBR (Rule-Based Reasoning) [13]. Cardie [14] added MBR (Model-Based Reasoning) to CBR. Baumeister *et al.* [15] studied the CBR based on inductive machine learning. Ahn and Kim [16] managed to enhance the predicting capability of CBR by GA (Genetic Algorithm).

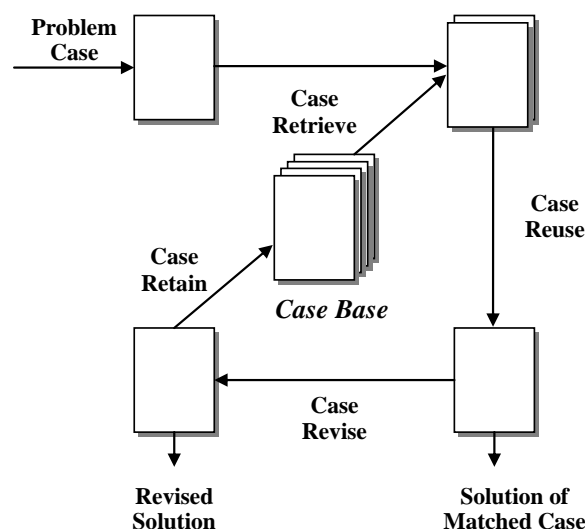


Figure 1. 4R Model of CBR

A case in CBR is usually composed of two parts: one is the problem description part which illustrates the context when the case happens, and the other is the solution description part which denotes the outcome of the corresponding problem [17]. Cases in the existing researches mostly describe the two parts statically and rarely considered the time characteristic in problem solving process. As CBR has its limitation when handling problems with timeline and process features, it is seldom applied to dynamic problem fields so far. "CBR systems exploiting the temporal dimension of cases are not so numerous, case descriptors are not compulsorily described with time stamps" [18]. But dynamic problems are very common in reality, especially complex problems will almost refer to the timeline features, so the extension of CBR to support dynamic and time sequential is an inevitable requirement [19].

2.2. SPM

A CBR system based on sequential data can be manifested as sequences of actions or events on logical layer and a transaction dataset in time order on data layer, so methods in SPM can be applicable for sequential data based CBR systems. SPM is first studied by Agrawal and Srikant [20]. The difference between SPM and association rule mining is that target data in SPM are sequential ones with time attributes, through which frequently emerged actions or event sequences can be mined. Sequential data are very common in real world, for example, route planning [21], customer's buying behavior [22], aeroengine monitoring [23] and medications prediction [24] are all significant target problems of SPM.

Applications of SPM are based on implements of the mining algorithm, and a large number of literatures about SPM are focused on SPM algorithms. The existing SPM algorithms include three categories of Apriori based approaches, vertical format approaches and projection based approaches [8], and specific algorithms include: AprioriAll, AprioriSome, DynamicSome, GSP, MFS, PSP, SPADE, cSPADE, SPAM, LAPIN-SPAM, FreeSpan and PrefixSpan *etc.*

As different algorithms have their own application conditions, it is an important issue to choose appropriate algorithm for SPM in accordance with the contexts, characteristics and constraints of the problem to be solved. In many SPM algorithms, the items in sequential database are treated equally, *i.e.* items in item set are deemed equally important, and thus it is easily ignored whether user is interested and the extent in the item. In CBR systems based on sequential data, it is necessary to weight the importance of different items on the basis of the concern extent, so that the SPM can focus on the events or activities significant to applications, avoiding wasting resources for mining on the vast nonsensical data.

3. Methodology

3.1. Weighted SPM

(1) Weights setting

In the sequential case base, the sequences record the happenings of events in cases. The events set $\{Event\}$ in all cases is the item set of sequential database, denoted as E . According to the extent of user's concerning on different events, the events in $E = \{e_1, e_2, \dots, e_n\}$ can be weighted, thus E is transformed to weighted events set E_W :

$$E_W = (E, W) = \{U_{i=1}^n (e_i, w_i)\} = \{(e_1, w_1), (e_2, w_2), \dots, (e_n, w_n)\} \quad (4.1)$$

Then, the events sequence $ES = \langle es_1, es_2, \dots, es_m \rangle$ can be transformed as weighted events sequence:

$$ES_w = \langle \bigcup_{i=1}^m (es_i, w_{es_i}) \rangle = \langle (es_1, w_{es_1}), (es_2, w_{es_2}), \dots, (es_m, w_{es_m}) \rangle \quad (4.2)$$

where the weight of events set es_i in ES is a function of the contained events' weights, having $w_{es_i} = f(W_{es_i})$. If the weight of es_i calculated as the arithmetic mean of the contained events' weights in es_i , then we get:

$$w_{es_i} = \frac{\sum_{j=1}^{|es_i|} weight(es_{ij})}{|es_i|} \quad (4.3)$$

where $weight(es_{ij})$ is the weight of event j in events set es_i .

The weight of events sequence ES is the function of the weight of contained events set, having $w_{ES} = g(W_{ES})$. If the weight of ES calculated as the arithmetic mean of the weights of all contained events sets in ES , then we get:

$$w_{ES} = \frac{\sum_{i=1}^{|ES|} w_{es_i}}{|ES|} \quad (4.4).$$

(2) Calculation of weighted support degree

After weight setting for events, event sets and event sequence, the definition of support degree should also be revised. The support degree now is no longer the simple sum of supported times of an event sequence in sequential database, because the occurrence frequencies of more important event sequence and less important event sequence have different meanings to SPM. Consequently, the traditional definition of support degree should be changed to weighted one, so that the searching of important event sequences in sequential database can be optimized in the process of event SPM. The weighted support degree is defined as follows.

Given an event sequence database ESD , for event sequence $esd \in ESD$ and event sequence $ES = \langle es_1, es_2, \dots, es_m \rangle$, if esd support ES , then denote $Sup(esd, ES) = 1$, else denote $sup(esd, ES) = 0$. The weighted support degree of ES in ESD is:

$$wSup(ES) = \frac{\sum_{i=1}^{|esd|} (w_{esd_i} \cdot Sup(esd, ES))}{\sum_{i=1}^{|esd|} w_{esd_i}} \quad (4.5)$$

According to this definition, the higher the weight of an event sequence is, the greater influence or contribution to the result of SPM will be made. The weighted support degree defined here is dimension independent, and the calculation of un-weighted support degree and support count is the special case of setting weights of all events as equal.

(3) wGSP algorithm

In the process of CBR, there commonly exist some constraints, such as industry field, temporal interval or geographic location, so there may be some restricted conditions when implementing event SPM. Among the related SPM algorithms, GSP has good adaptability for restricted conditions, and the implementation process is clear compared to other

algorithms, so lots of researches are based on GSP. However, in the current GSP based algorithms, the sequence support degree is calculated by un-weighted approach, which cannot satisfy the differentiated demands of user's concerns. For this reason, based on the weight setting of events and calculating of weighted support degree presented above, we propose wGSP algorithm for weighted SPM and described as follows.

Step1: Initialization. Given the event sequence database ESD , set the minimum support degree threshold value $minsup$, and the weight W of events.

Step2: Solving 1-event-sequence pattern set ESL_1 . For each event e in event set E , calculates the weighted support degree $wSup(e)$ of e in ESD . If $wSup(e) \geq minsup$, then add event e into ESL_1 .

Step3: Generating candidate k-event-sequence pattern set ESC_k . For the solved (k-1)-event-sequence pattern set ESL_{k-1} , to perform join operator for each two event patterns in ESL_{k-1} , and add the joint result to initial k-event-sequence pattern set ESC'_k . Then delete the k-event-sequence patterns in ESC'_k who has at least one (k-1)-sub-sequence not contained in ESC'_k , thus the k-event-sequence pattern set ESL_k is obtained. The rule of join operator is: for the two event sequence patterns $es_1, es_2 \in ESL_{k-1}$ participated in join operating, if the sequence after es_1 deleting the first item is equal to es_2 deleting the last item, the add the last item of es_2 to the end of es_1 as the result of join operating, otherwise do not perform join operating.

Step4: Solving k-event-sequence pattern set ESL_1 . For each k-event-sequence pattern in ESC_k , to calculate the weighted support degree, if above $minsup$, then add it into ESL_k .

Step5: Loop iteration. Loop to Step3 until k-event-sequence pattern cannot be found, then continue to Step6.

Step6: Result output. Output event sequence patterns in all event sequence pattern set $\bigcup_{i=1}^n ESL_i$ as the final result of event SPM.

3.2. Sequence Similarity Measure

(1) Analysis of sequence-patternsupporting

In the analogical reasoning process of CBR, an important work is to measure the similarity between cases. The event sequence records a series of events happened chronologically with uncertain length, and it cannot be described by explicit feature attributes as common cases, so the similarity measuring methods in traditional CBR

methods cannot directed implemented to sequential data. To consider the similarity matching of event sequence from another point of view, the typical event patterns in case base are derived from the event SPM, if each mined event pattern is deemed as one typical feature of event, then through investigating the values of events under these typical features, the similarity between event sequences can be measured.

Setting all event sequence patterns mined through event SPM as $ESL = \langle esl_1, esl_2, \dots, esl_m \rangle$, all event sequences ES in case base represented as rows and ESL represented as columns, we construct sequence-pattern support matrix SPS , the element sps_{ij} of which represents the support status of the event sequence of its row to the event pattern of its column, having:

$$sps_{ij} = \begin{cases} 1, & es_i \xrightarrow{sup} esl_j, \\ 0, & otherwise. \end{cases} \quad (4.6)$$

(2) Similarity measure of event sequence

The construction of sequence-pattern support matrix SPS actually transforms the event sequences with uncertain length which cannot be described as feature attributes to feature vectors with the same length. Each dimension of the feature vector expresses the support status of the corresponding event sequence to event sequence pattern, and the length of feature vector is the total count of mined event sequence patterns. If an event sequence es_i is supported by event sequence pattern esl_j , then es_i is called to have esl_j feature, and all event sequence patterns that support es_i is called feature set F_{es_i} of es_i .

According to the similarity measure idea of Tversky, we define the similarity function of event sequence as:

$$sim(es_i, es_j) = \frac{|ESL_i \cap ESL_j|}{|ESL_i \cap ESL_j| + 1/2 \cdot (|ESL_i - ESL_j| + |ESL_j - ESL_i|)} \quad (4.7)$$

Based on sequence-pattern support matrix, the similarity of es_i and es_j is calculated as:

$$sim(es_i, es_j) = \frac{A}{A + 1/2 \cdot B}, \quad (4.8)$$

where $\begin{cases} A = |esl_k | sps_{ik} = 1 \wedge sps_{jk} = 1|, \\ B = |esl_k | (sps_{ik} = 1 \wedge sps_{jk} = 0) \vee (sps_{ik} = 0 \wedge sps_{jk} = 1) | \end{cases}$

4. Instance Analysis

4.1. Calculating Instance

Below is an instance in entrepreneurship illustrated the process of CBR based on sequential data. Problem case $Case_p$ is a problem of how to implement entrepreneurship in the front of a start-up company. The event sequence happened for the company is $\langle \text{market research, trial operation} \rangle$. In the case base, there stored some entrepreneurship cases happened in the past in different companies. The sequential data of case base is shown in Table 1.

Table 1. Sequential Data of Case Base

Case ID	Sequential Data
<i>Case₁</i>	<MR, (MR, CR)>
<i>Case₂</i>	<MR, (BP, SR, TO), (PA, CR)>
<i>Case₃</i>	<(MR, RD, TO)>
<i>Case₄</i>	<MR, (BP, TO), CR>
<i>Case₅</i>	<CR>
<i>Case₆</i>	<TO, MR>

Note: PA- patent application, MR- market research, BP- Business planning, RD- R&D, SR- staff recruitment, TO- trial operation, CR- Capital raising

In terms of the two-step CBR method proposed above, we first implement the weighted SPM for the case base. The weights of events in case base set by user are shown in Table 2.

Table 2. Weight Setting of Events

Event	PA	MR	BP	RD	SR	TO	CR
Weight	0.5	0.8	0.9	0.9	0.4	0.7	0.9

Set the minimum support degree threshold value $minsup=25\%$, the result of event SPM is shown as Table 3.

Table 3. Result of Event SPM

ID	Event Sequence Pattern	wSup
<i>ESL₁</i>	<MR>	83.8%
<i>ESL₂</i>	<BP>	27.4%
<i>ESL₃</i>	<TO>	68.8%
<i>ESL₄</i>	<CR>	58.6%
<i>ESL₅</i>	<MR, BP>	27.4%
<i>ESL₆</i>	<MR, TO>	55.3%
<i>ESL₇</i>	<MR, CR>	42.4%
<i>ESL₈</i>	<BP, TO>	27.4%
<i>ESL₉</i>	<BP, CR>	27.4%
<i>ESL₁₀</i>	<TO, CR>	27.4%
<i>ESL₁₁</i>	<MR, (BP, TO)>	27.4%
<i>ESL₁₂</i>	<MR, BP, CR>	27.4%
<i>ESL₁₃</i>	<MR, TO, CR>	27.4%
<i>ESL₁₄</i>	<(BP, TO), CR>	27.4%
<i>ESL₁₅</i>	<MR, (BP, TO), CR>	27.4%

Table 4 is the sequence-pattern support matrix SPS of cases and the result of similarity measure.

Table 4. Sequence-Pattern Support Matrix And Similarity Measure Result

	ESL ₁	ESL ₂	ESL ₃	ESL ₄	ESL ₅	ESL ₆	ESL ₇	ESL ₈	ESL ₉	ESL ₁₀	ESL ₁₁	ESL ₁₂	ESL ₁₃	ESL ₁₄	ESL ₁₅	A Value	B Value	Similarity
Case ₁	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	4	0.3333
Case ₂	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	3	11	0.3529
Case ₃	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	3	1	0.8571
Case ₄	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	3	11	0.3529
Case ₅	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0.0000
Case ₆	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0.8000
Case _p	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	Similar case		Case ₃

4.2. Explanations and Discussions

Among all cases in the case base, *Case₃* is the most similar one to problem case *Case_p*. The similarity is 0.8571, there are three same event sequence pattern they support collectively, and only one supported pattern is different. Other cases and their similarity to *Case_p* are *Case₆* (0.8000), *Case₂* (0.3529), *Case₄* (0.3529), *Case₁* (0.3333), *Case₅* (0.0000). Thus, in the CBR system, the case description and solution of *Case₃* will be retrieved for further revision and provide solution for the solving of problem case *Case_p*. If valuable and distinguished, the solved *Case_p* will be added into the case base, thereby the problem solving capability of the CBR system is enhanced by the incremental learning of case base.

The two-step CBR method in this work is an improvement both to traditional CBR and to SPM methods in coping with dynamic problems.

(1) Improvements to CBR

Compared to traditional CBR methods, our method has advantages in the supporting of sequential data. On one hand, it can extend the solve area of CBR to sequential data described cases. Previous CBR methods usually consider the feature description of cases, such as which features are included and the strength value under each feature. Besides, the case retrieval methods are on the basis of feature attributes, where the orders of events and items are ignored in most cases. As in the calculating instance above, the most similar case is *Case₃*, because of the high consistence on event order descriptions between *Case₃* and *Case_p*. If we measure the similarity between cases with traditional feature methods regardless of event sequences, the most similarity case will be *Case₆*, as their events contained are the same, even though the happening orders of events are completely inconsistent. It is obviously unreasonable to the solving of problems where the order of events matters.

On the other hand, the proposed case retrieval method based on typical event patterns can enhance the efficiency of case retrieval in a flexible way. Figure 2 shows our improvement to the traditional CBR closed cycle.

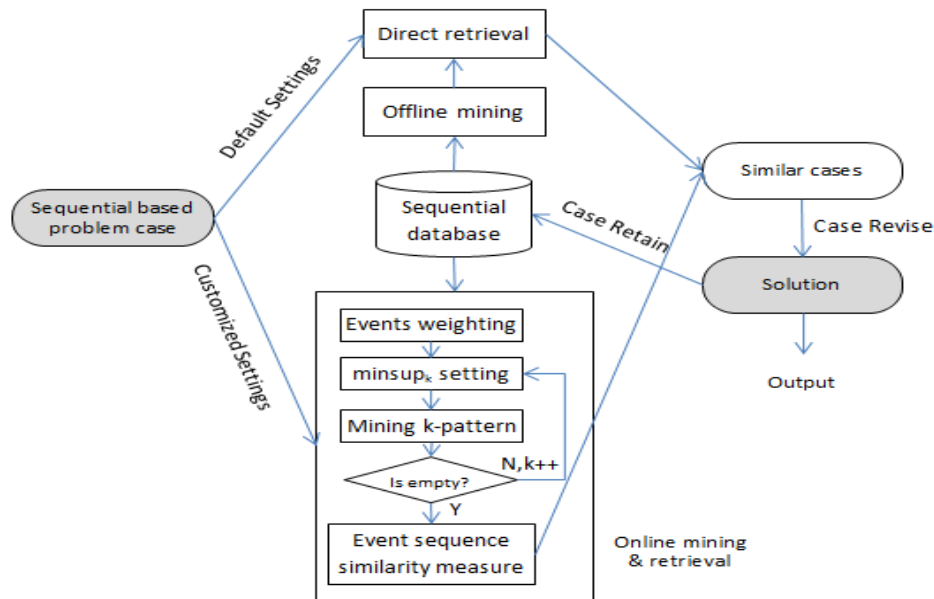


Figure 2. The Improved CBR Cycle

In the two-step CBR method, if user has not special requirements for the weights of events, the first step can be done offline with default settings. In this way, an intermediate database for sequence-pattern support information can be built, where the result of event SPM is stored. If it is needed to measure the similarity of event sequences, the support information can be selected from the intermediate database for calculation. Through the construction of intermediate database, the scanning of whole case base for each retrieval operation can be avoided. It plays a role of temporary buffer pool, thereby offline sequence matching can be implemented to increase the processing efficiency.

The mining of typical patterns is substantially a refining process to the case base, which is an important issue of CBR cycle. Typical patterns represent the generality and experience summary of massive similar cases happened in the past, and is beneficial to the future problem solving. Additionally, the re-mining and refining of case base are of great significance to knowledge discovery.

(2) Improvements to SPM

The method in this work enables flexible settings of weight and support degree in process of SPM. At the same time of user submitting the event sequence of problem $Case_p$ to CBR system, the weights of events concerned which dynamically reflect the concerning extent on different events can also be submitted. In many SPM approaches, the items in item set are deemed as equally important, so whether user is interested in the items and the extent are ignored. This may be feasible in some situations such as buying sequence pattern analysis, because shops only need to analyze the buying order of consumers without concerning on the importance of goods to consumers. Nonetheless more situations are not the case, as different items often have different importance, which is the internal feature of real world database.

Actually, for many problems the concern extent of deciders on different items is discrepant. As far as the entrepreneurship problem, technology-based firms may pay more attention to events in patent application and R&D, while sale-based enterprises will lay emphasis on events in marketing management. As a consequence, to set weights for different events regarding interests of enterprises and focus the SPM on the more significant events by weighing, can avoid the mining and searching for massive meaningless events.

From the perspective of information retrieval, to match sequential data directly is a hard problem. Our method transforms sequential data into feature attribute ones, which

can greatly increase the efficiency of case retrieval for CBR. The minimum support degree threshold minsup of weighted SPM can also be adjusted dynamically. If the user's problem is an especial one and he wants to match non-typical target cases in case retrieval, then he can lower the value of minsup , in which way more typical patterns can be mined and more cases that satisfy user's especial requirement can be retrieved. Conversely, if user expects to discover the regularity of problem solving in masses of cases, than he can increase the value of minsup . This strategy is suitable for the field that user is unfamiliar with.

5. Conclusions

CBR imitates human's mode of thinking to cope with new problems, and is an effective AI methodology. In traditional CBR systems, the case bases are mostly described as feature attributes, and the case retrieval process is the match of attributes between problem case and target cases. But in reality there exist a large number of problems described with sequential data, the reasoning and learning capability of CBR can be very helpful to problem solving if it can support sequential data, however, there is lack of researches on the issue. In this paper, we try to integrate CBR and sequential data to investigate case retrieval, the essential issue of CBR, on the basis of sequentially described case base. As the matching of sequential data in the whole database is a hard problem, we design a two-step CBR process. The first step is SPM, where the mining of typical patterns can realize case learning of CBR and the mined sequence pattern can service as feature description for the second step of case retrieval. The SPM method in this work is weighted and based on wGSP algorithm, where user can set weights to events according the difference of concerns. About the matters of case retrieval, we transform the cases described with sequential data to feature data in terms of sequence-pattern support matrix, as thus case can be retrieved effectively and quickly. Through the instance analysis, we demonstrate the availability of our method.

The work done in this paper is exploration for CBR to support sequential data. On one hand we improve the solving capability of CBR for dynamic problems, while on the other hand a corresponding efficient case retrieval strategy is proposed. In the future, there are still plenty of issues worth being studied further about the topic of our work, such as the optimizing of weighted SPM algorithm, heuristic approaches of sequential data matching, etc.

Acknowledgements

This work has been supported by National Natural Science Foundation of China (NSFC, Grant No. 71361011), Science and Technology Project of Education Department of Jiangxi Province (Grant No.GJJ11734) and JXUFE Research Project. The author would be like to thank the reviewers for their helpful comments and suggestions for improving the manuscript.

References

- [1] P. Perner, "Mining Sparse and Big Data by Case-based Reasoning", *Procedia Computer Science*, vol. 35, (2014), pp. 19-33.
- [2] Z. Liao, X. Mao, P. M. Hannam and T. Zhao, "Adaptation methodology of CBR for environmental emergency preparedness system based on an improved genetic algorithm", *Expert Systems with Applications*, vol. 39, no. 8, (2012), pp. 7029-7040.
- [3] C. L. Chuang, "Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction", *Information Sciences*, vol. 236, (2013), pp. 174-185.

- [4] A. Yan, W. Wang, C. Zhang and H. Zhao, "A fault prediction method that uses improved case-based reasoning to continuously predict the status of a shaft furnace", *Information Sciences*, vol. 259, no. 2, (2014), pp. 269-281.
- [5] R. Rocha, R. Azevedo and Y. Sousa, "A Case-based Reasoning System to Support the Global Software Development", *Procedia Computer Science*, vol. 35, (2014), pp. 194-202.
- [6] E. Rashedi, H. Nezamabadi and S. Saryazdi, "Long term learning in image retrieval systems using case based reasoning", *Engineering Applications of Artificial Intelligence*, vol. 35, (2014), pp. 26-37.
- [7] H. Ince, "Short term stock selection with case-based reasoning technique", *Applied Soft Computing*, vol. 22, (2014), pp. 205-212.
- [8] H. Shyur, C. Jou and K. Chang, "A data mining approach to discovering reliable sequential patterns", *The Journal of Systems and Software*, vol. 86, (2013), pp. 2196-2203.
- [9] Y. Guo, J. Hu and Y. Peng, "Research on CBR system based on data mining", *Applied Soft Computing*, vol. 11, (2011), pp. 5006-5014.
- [10] R. C. Schank and R. Abelson, "Scripts, plans, goals and understanding", Hillsdale, NJ: Erlbaum, (1977).
- [11] A. Aamod and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and systems approaches", *AI Commun*, vol. 7, (1994), pp. 39-59.
- [12] E. Reyes, S. Negny and G. Robles, "Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning Application to process engineering design", *Engineering Applications of Artificial Intelligence*, vol. 41, (2015), pp. 1-16.
- [13] H. C. Robert and Y. K. Melody, "An integrated approach of rule-based and case-based reasoning for decision support", In *Proceedings of the 19th Annual Conference on Computer Science*, (1999), pp. 255-267.
- [14] C. Cardie, "Using decision trees to improve case-based reasoning learning", In *Proceedings of the Tenth International Conference on Machine Learning*, (1993), pp. 5-16.
- [15] J. Baumeister, M. Atzmulle and F. Puppe, "Inductive learning for case-based diagnosis with multiple faults", *ECCBR*, (2002), pp. 28-42.
- [16] H. Ahn and K. J. Kim, "Global optimization of case-based reasoning for breast cytology diagnosis", *Expert Systems with Applications*, vol. 36, (2009), pp. 724-734.
- [17] G. Zhu, J. Hu and J. Qi, "An integrated feature selection and cluster analysis techniques for case-based reasoning", *Engineering Applications of Artificial Intelligence*, vol. 39, (2015), pp. 14-22.
- [18] A. Mille, "From case-based reasoning to traces-based reasoning", *Annual Reviews in Control*, no. 30, (2006), pp. 223-232.
- [19] A. Cordier, M. Lefevre and P. Champin, "Trace-Based Reasoning - Modeling interaction traces for reasoning on experiences", *Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference*, (2013), pp. 363-368.
- [20] R. Agrawal and R. Srikant, "Mining sequential patterns", *Proceeding of the 11th International Conference on Data Engineering*. Washington DC: IEEE Computer Society, (1995), pp. 3-14.
- [21] C. Tsai and B. Lai, "A Location-Item-Time sequential pattern mining algorithm for route recommendation", *Knowledge-Based Systems*, vol. 73, (2015), pp. 97-110.
- [22] N. Desai and A. Ganatra, "Buying Scenario and Recommendation of Purchase by Constraint Based Sequential Pattern Mining from Time Stamp Based Sequential Dataset", *Procedia Computer Science*, vol. 45, (2015), pp. 166-175.
- [23] A. Palacios, A. Martinez and L. Sanchez, "Sequential pattern mining applied to aeroengine condition monitoring with uncertain health data", *Engineering Applications of Artificial Intelligence*, vol. 44, (2015), pp. 10-24.
- [24] A. P. Wright, A. T. Wright and A. B. McCoy, "The use of sequential pattern mining to predict next prescribed medications", *Journal of Biomedical Informatics*, vol. 53, (2015), pp. 73-80.

Author



Quan Xiao, received his Ph.D. degree in management science and engineering from Huazhong University of Science and Technology, China in 2010. He is currently a lecturer in the School of Information Technology, Jiangxi University of Finance and Economics, China. His research interests include decision support systems, risk identification and online customer behaviors.

