Reexamination of CBR Hypothesis

Xi-feng Zhou[1,2,3,4], Ze-lin Shi[1,2,3], and Huai-ci Zhao[1,2,3]

[1] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2] Key Laboratory of Optical-Electronics Information Processing, Chinese Academy of Sciences, Shenyang 110016, China
[3] Key Laboratory of Image Understanding and Computer Vision, Liaoning Province, Shenyang 110016, China
[4] Graduate School of the Chinese Academy of Sciences, Beijing 100049, China
xifeng_2002@163.com, {zlshi,hczhao}@sia.cn

Abstract. Most of the recent literature on complexity measures in textual case-based reasoning examined alignment between problem space and solution space, which used to be an issue of formulating CBR hypothesis. However, none of existing complexity measures could dispel the specter of predefined class label that does not appear in public textual datasets available, or clarify the correctness of the proposed solutions in the retrieved cases most similar to a target problem. This paper presented a novel alignment measure to circumvent these difficulties by calculating rank correlation between most similar case rankings in problem space and most similar case rankings in solution space. We also examined how to utilize existing alignment measures for textual case retrieval and textual case base maintenance. Empirical evaluation on Aviation Investigation Reports from Transportation Safety Board of Canada showed that rank correlation alignment measure might become a promising method for case-based non-classification systems.

1 Introduction

In the classic paper of D. B. Leake in 1996, he proposed two assumptions on that the case-based reasoning (CBR) approach is based, one of which is similar problems have similar solutions [1] that is usually called similarity assumption or CBR hypothesis. Afterwards in 1999, D. B. Leake and D. C. Wilson presented problem-solution regularity [2], which describes the relationship between problem descriptions and solutions, as a formulation of the aforementioned CBR hypothesis. Although many of the subsequent contributions have focused on the alignment between problem space and solution space, there have been few attempts to calculate alignment between most similar case rankings in problem side and most similar case rankings in solution side, instead of alignment between problem side similarities and solution side similarities.

Recently, the concept of complexity, which is equivalent to alignment, has been introduced in order to measure the extent to which CBR hypothesis hold true in TCBR [3, 4, 5, 6, 7, 8, 9]. Unfortunately, there are some limitations that are needed to be overcome in existing complexity measures for TCBR. In addition, few of them can clarify the correctness of proposed solutions in the retrieved cases most similar to a

I. Bichindaritz and S. Montani (Eds.): ICCBR 2010, LNAI 6176, pp. 332–345, 2010.

target problem (query) from different system designs. It is necessary to reexamine the problem of formalizing the CBR hypothesis in case-based non-classification systems in order to make these issues clear and to be solved.

In particular, three limitations of the existing alignment measures for case-based non-classification systems that are in the majority of current CBR systems, such as textual CBR (TCBR), may be noted:

- The class label that to a certain extent represent/determine the correctness of a textual solution, on which evaluation of effectiveness of existing complexity measures also rely, does actually not appear in public textual datasets available, like Aviation Investigation Reports (AIR) from Transportation Safety Board of Canada.
- Performance indicators don't have a standard definition because the goal of developing TCBR systems is different from diverse application domains, and at present there is not consensus about TCBR system framework.
- Although some of these alignment measures can provide case base profile for case authoring, they have not shown any further improvement in non-classification case retrieval and non-classification case base maintenance.

The computation of alignment measure or complexity measures for TCBR systems can be categorized into three classes in terms of different variables used, such as similarity scores, retrieval sets, case features. In essence, alignment measure or complexity measure is an informal definition of CBR hypothesis that describe the extent to which similar problems have similar solutions hold true in case-based reasoning systems. However, to our knowledge, there are few contributions that calculate alignment with case rankings.

In this paper, our primary motivation is to examine how to utilize alignment score to provide support for textual case retrieval and textual case base maintenance. A novel alignment measure is presented in order to formalize CBR hypothesis with partial rankings correlation between most similar case rankings in problem side and most similar case rankings in solution side. Inspired by the concept of correlation, we redefined a few performance indicators for evaluating the output of a non-classification CBR system, like TCBR.

This paper is organized as follows. Section 2 reviews related work about formulation of CBR hypothesis and positions our work in context. Section 3 argues that it is difficult to define the correctness of a predicted textual solution, and we circumvents this issue by evaluating partial rankings correlation between most similar case rankings in problem side and most similar case rankings in solution side. We provide empirical results to substantiate this idea in the following experimental section. Section 4 proposes a novel technique for textual case retrieval and textual case base maintenance based on correlation. Section 5 discusses our evaluation method and presents empirical findings on how well the alignment measure proposed by us can clarify the importance of each nearest neighbor. We highlight our main contributions and conclude in Section 6.

2 Related Work

Although many researchers considered CBR hypothesis as the main assumption that underpin case-based reasoning (CBR) as a suitable problem-solving methodology for a particular domain, there are still not standard explanation or formulation about it. To date there appeared many explanations about CBR hypothesis suggested by different researchers that can be classified two categories, qualitative description and quantitative formulation. To our knowledge, the earliest explanation given by J. L. Kolodner in 1993 was that the intuition of case-based reasoning was that situations recur with regularity [10]. The regularity was further enlarged by D. B. Leake in 1996 who claimed that the CBR approach was based on two tenets about the nature of the world, which are similar problems have similar solutions and the types of problems one encounters tend to recur [1]. In 1997, B. Faltings made a possible explanation that presumed a problem with similar features as an earlier one was likely to have the same solution [11]. These studies have emphasized qualitative description and theoretical analysis opposed to quantitative formulation of CBR hypothesis. Although B. Faltings used probability theory to assure CBR hypothesis to be correct on the average, his explanation was still a qualitative description.

In 1999 D. B. Leake revisited CBR hypothesis and proposed problem-solution regularity and problem-distribution regularity in explanation of his earlier two tenets, which was the first time to formulate CBR hypothesis in a quantitative way as far as I know. He made a clear explanation on CBR hypothesis with mathematical equation for problem-solution regularity. Problem-solution regularity has become an oracle of subsequent definition of alignment measures, which captures how well problem similarities approximate solution similarities in practice [2]. For example, Case Alignment [4], Similarity Profile [6], MST and Weighted Correlation [9], etc. belongs to this category. The research has tended to focus on alignment between problem side similarities and solution side similarities (Fig. 1).

$$
\begin{array}{ccc}
< C_1, SIM_p(C_t, C_1) > & < C_2, SIM_p(C_t, C_2) > & \cdots < C_n, SIM_p(C_t, C_n) > \\
\Updownarrow & \Updownarrow & \Updownarrow \\
< C_1, SIM_s(C_t, C_1) > & < C_2, SIM_s(C_t, C_1) > & \cdots < C_n, SIM_s(C_t, C_n) >
\end{array}
$$

Fig. 1. Alignment between similar problems sequence and similar solutions sequence

The other two categories of method for calculating alignment appeared in recent years. GAME [5, 8] adopted case base image metaphor as a case clustering method with textual case features, which could be used to compare clusters alignment in both problem and solution space. Although considerable research on similarity value of retrieval result <CaseID, SimilarityValue> has been devoted to formalize CBR hypothesis for case-based classification or non-classification systems, rather less attention has been paid on the third category of methods for a long time, which utilized case ID of retrieval result to measure alignment (Fig. 1). This situation held in line until L. Lamontagne proposed Case Cohesion (alignCohesion) alignment measure [3] in 2006, which measures level of overlap in retrieval sets, in order to evaluate the competence of different system designs. However, threshold must be set using trial and error to help to select the number of nearest neighbors in problem side and solution side in this method. Therefore, it would be a matter of interest to learn how to formulate CBR hypothesis with alignment measure based on case ID rankings between the retrieval sets in this paper.

$$alignCohesion(t) = \frac{|RS_p(t) \cap RS_s(t)|}{|RS_p(t) \cup RS_s(t)|} \tag{1}$$

After analyzing its equations for calculating alignment scores, we found that Case Alignment [8] (alignMassie) could not get to the maximum alignment value when problem space is completely in alignment with solution space, which is shown in our following experiments. Although ref. [9] presented two parameters-free complexity measures, the effectiveness of these measures need to correlate with human judgments like precision or classification accuracy, just like ref. [3] and ref. [8] did, that does not actually exist in public textual datasets available.

$$Align(t, c_i) = 1 - \frac{D_s(t, c_i) - D_{s\min}}{D_{s\max} - D_{s\min}} \tag{2}$$

$$alignMassie(t) = \frac{\sum_{i=1}^{k}(1 - D_p(t, c_i)) * Align(t, c_i)}{\sum_{i=1}^{k}(1 - D_p(t, c_i))} \tag{3}$$

Although there exist many explanations about CBR hypothesis, few of them is designed to formalizing it for case retrieval and case base maintenance in non-classification case-based systems. Our aim is to formulate CBR hypothesis for non-classification case-based systems by emphasizing the importance of case rankings, and utilize it to guide non-classification case authoring, case retrieval and case base maintenance.

3 Definition of the Correctness for Non-classification Solutions

Unlike classification domains where a solution is a class label associated with a group of cases, textual problem descriptions map into unique solutions. Especially during textual case retrieval, we do not know whether a retrieved textual solution to a query is right or not because it is usually a posterior knowledge available only after the retrieved textual solution is applied to solve the query. Before that, we can not say the textual solution is correct to any other textual problems. This is the property of uniqueness of textual solutions.

However, each case in the neighborhood of the query has a right solution. For example, in TCBR, one can say a textual solution is only correct to its corresponding textual problem description because it is the known case already existed in case base. Therefore, we can get to know whether the nearest neighbors (cases) of the query in problem side can be solved or how difficult it is to solve those nearest neighbors (cases) using CBR methods with other cases in case base. It is not easy to acquire knowledge of whether a textual case (problem) is successfully solved or how difficult it is to solve the textual case (problem).

Because CBR is the approach we adopted to solve problems, the difficulty to solve a problem should be measured according to whether the basic tenet of this approach, similar problems have similar solution, hold true. In ICCBR 2008, Raghunandan M. A. etc. has pointed out that complexity reflects the degree to which we can expect a CBR system to be competent in answering a new problem by retrieving solutions to similar problem in the repository and that the complexity of a domain in CBR is a measure of the difficulty of the problem being faced. In ICCBR 2009, they presented a weighted correlation method (alignCorr) for calculating alignment measures that were designed to formulate similar problems have dissimilar solutions. The following are the equations to calculate alignCorr [9].

$$alignCorr(t) = wtCorr(PS(t), SS(t), PS(t)) \tag{4}$$

$$wtCorr(x, y, w) = \frac{wtCov(x, y, w)}{\sqrt{wtCov(x, x, w) * wtCov(y, y, w)}} \text{[1]} \tag{5}$$

$$wtCov(x, y, w) = \frac{\sum_i (w_i (x_i - m(x, w))(y_i - m(y, w)))}{\sum_i w_i} \tag{6}$$

$$m(x, w) = \frac{\sum_i w_i x_i}{\sum_i w_i} \tag{7}$$

Inspired by aforementioned idea, we suggested employing complexity to measure the difficulty to solve a case in case base as a nearest neighbor of a query. Now, the problem is how we take advantage of this information to define the correctness of the prediction for the query.

In Leave-One-Out evaluation for a textual case base, each case in case base can be a target case that is composed of target problem and target solution. All the nearest neighbors of the query (target problem) in descending problem similarity scores in problem side are called Similar Problems Sequence (SPS). Similarly, all the nearest neighbors of the target solution in descending solution similarity scores in solution side are called Similar Solutions Sequence (SSS). If the *ith* most similar solution in SSS can solve the *ith* similar problem in SPS, we can say the *ith* nearest neighbor of the query are aligned. Assuming C1 is one most similar case in top k nearest neighbors to the query, its position in SPS is denoted as Position(C1, SPS), and its

[1] After communicating with Raghunandan M.A., we rectified the weighted correlation function in his original paper.

position in SSS is denoted as Position(C1, SSS). Check if the two positions are equal. If so, C1 is aligned for the query. If not, C1 is not aligned for the query. If each of the top k nearest neighbors is aligned for the query, we can say the target case completely respects CBR hypothesis. That means when the solution of the most similar case to the query in problem side is adopted to solve the query, it would be guaranteed that the best solution with max utility is selected. However, this is not the case. In addition, this kind of description is too informal to be an alignment measure. There must be some measures defined that can address this issue.

Our primary aim is to make sure every case in case base respects alignment between problem space and solution space. In other words, the position of a nearest neighbor of a target case in SPS approximates the position of the nearest neighbor of the target case in SSS. If not, we would penalize them with a penalty parameter. The measure that implemented this function is rank correlation.

4 Comparing Partial Rankings

In this paper, we introduced Kendall profile metric [12] as a new alignment measure, which is a generalization of the Kendall's τ rank correlation coefficient for comparing partial rankings. In statistics, a ranking can be seen as a permutation of a set of objects. Rank correlation is the study of relationships between different rankings on the same set of objects. Rank correlation coefficient measures the correspondence between two rankings and assesses its significance. Kendall's tau rank correlation coefficient is a measure of the similarity of the orderings of the data when ranked by each of two quantities. However, Kendall's τ rank correlation coefficient measures only similarity between full lists, not that between top k lists. In 2006, R. Fagin etc. generalized it for calculating distance between top k lists, namely Kendall profile metric [12].

In this paper, the top k nearest neighbors of a case (the target problem) in problem space and the top k nearest neighbors of the case (the target solution) in solution space are our concerns. We consider SPS and SSS of the case as two partial rankings with an emphasis on their top k elements and employ Kendall profile metric to calculate alignment scores of the case, as shown in equation (8).

$$K_{prof}(\delta_1,\delta_2)=\sum_{\{i,j\}\in P}\bar{K}_{i,j}^{(1/2)}(\delta_1,\delta_2) \tag{8}$$

In the equation (8), δ_1,δ_2 are two partial rankings with domain D. In addition, $P=\{\{i,j\}\mid i\neq j \text{ } and \text{ } i,j\in D\}$. And $\overline{K}_{i,j}^{(1/2)}(\delta_1,\delta_2)$ is a penalty for δ_1,δ_2 for $(i,j)\in P$, whose calculation method is detailed in ref. [12] on page 633.

Going through the equation (8), it would be noticed the value obtained is a number greater than 0. When SPS and SSS are reverse orders to each other, equation (8) reaches its maximum value. Therefore, the equation calculating alignment scores with rank correlation, namely alignRank, is normalized to [-1,1] as an correlation measure in the following equation.

$$alignRank(t) = 1 - \frac{K_{prof}(\delta_1, \delta_2)}{\max(K_{prof}(\delta_1, \delta_2))} \tag{9}$$

The value of alignRank greater than 0 would indicate that the top k elements in SPS and the top k elements in SPS are well aligned around the target case, whereas a value equal or less than 0 would indicate poor alignment. Specially, we consider well aligned cases respect CBR hypothesis, and poor aligned cases don't respect CBR hypothesis. The bigger the value of alignRank is, the higher the degree of CBR hypothesis hold true. Due to the needs of computing similarity with other cases in case base, rank correlation becomes a global alignment measure.

Given k value, alignment score between the top k elements in SPS and the top k elements in SSS of a case could be regarded as a property of each case in case base. But it is a derived one, instead of a component property like problem description, solution, justification of solution, or result. As a result, we had to acquire this property before retrieval or maintenance. We had better complete it after the case base is stable or the case base is not changing within a relatively long time, at least not during retrieval or maintenance. However, it is impossible to directly compute rank correlation of a target problem that does not appear in case base, not to mention using it for textual case retrieval or textual case base maintenance.

Before retrieval or maintenance, it is necessary to compute nearest neighbors of each case in case base in problem and solution side, respectively. After determining k, the number of nearest neighbors that would be used to vote during retrieval, we just regard k as a parameter and pass it with two rankings of each case to the procedure for calculating alignRank with normalized Kendall profile metric.

The rank correlation of each of the k nearest neighbors of the target problem is the only information available when retrieving a relevant solution to the target problem. This information could be used to predict the possible rank correlation of case rankings of the target problem. Now, the problem is how we can utilize the rank correlation of k nearest neighbors to vote in retrieval or maintenance?

We can check whether the output of rank correlation procedure for each case is positive. If the majority of the votes are positive, then the predicted correlation of the target case (query) is positive. Otherwise, the correctness is low. Just liked we defined for alignRank, if the prediction value for rank correlation of the target problem is positive, we could consider the nearest neighbors has the most similar and effective solution to the target problem with higher guarantee or confidence because relative majority of its top k nearest neighbors also respect CBR hypothesis. Similarly, this method can be applicable to maintain a textual case base.

Aforementioned process is just how we take advantage of the value of alignRank for assuring the correctness of a predicted textual solution through the rank correlation of the nearest neighbors of a query. At this very moment, we circumvent the problem of evaluating a predicted textual solution.

In a word, although it is difficult to define the correctness of a predicted textual solution for a query in non-classification domains, we really know whether the top k nearest neighbors of the query can be solved or not. As long as the case base respects CBR hypothesis, we can predict the query with most similar case rankings in a simple way and guarantee the prediction is the best choice for the query.

5 Empirical Evaluations

As we can see, alignment score can be regarded as a measure of the correctness of a predicted solution if the case base respects CBR hypothesis. Similar problems having similar rank correlation could guarantee the whole case base has well alignment. The bigger the rank correlation of the case base is, the more accurate the prediction is. Then next issue is how to testify the usefulness of rank correlation alignment measure proposed by us.

5.1 Datasets Preparation

TCBR 2007 workshop proposed a few challenges to create a TCBR system that might support investigators in tasks such as the following:

a. Authoring a new investigation report.
b. Proposing safety actions in response to a new incident.
c. Discovering recurring unsolved problems.

However, it is difficult to achieve any one of them. There are many sections in investigation reports that we can not determine which should be authored in a new one. Safety actions do not appear in each of all the investigation reports, which can be seen in part of 576 cases in Challenger 1.0 [14]. It is not easy for knowledge engineers to decide which section could be a problem and which section could be a solution, not mention to discovering recurring unsolved one.

Therefore, in this paper we try to take advantage of the sections appeared in the majority of cases in case base, and regard one of them as a problem description, another one as a solution to constitute a <problem, solution> case pair. These sections are Summary, Other Factual Information, and Analysis appeared in all 568 cases from the case base used in Challenger 1.0. There are 118 cases that own all six subtitles in the 568 cases if you also want to take other sections into account, such as Other Findings about Causes and Contributing Factors, Other Findings about Risks, and Safety Action Taken.

There are so many similarity measures available for text up to now, such as Jaccard Coefficient, Dice Coefficient, Cosine Coefficient, and Overlap Coefficient, etc. Due to variability in vocabulary usage and uniqueness of solutions, we choose to jump over feature-based case representation into featureless similarity measures, such as compression-based similarity for text, which has proved to be effective in spam filtering [15]. Two compression-based similarity measures for text that have been implemented in JCOLIBRI CBR framework was used to estimate problem similarities and solution similarities and to obtain the most similar case rankings in problem side and solution side. Hence, in our experiments, there is no feature extraction and selection.

In the following experimental evaluation, our main task is to investigate the alignment relationship between two of these sections according to aforementioned four alignment measures and check which one of them can demonstrate the case base profile in a reasonable and clear way. The outcome can be used as an input for the following textual case authoring, textual case retrieval, and textual case base maintenance.

5.2 Performance Indicators for TCBR Based on Correlation

Before evaluation, some performance indicators for TCBR should be first defined according to whether predicted correlation is positive. It is generally accepted that evaluation is a challenge for TCBR systems. However, we can predict the correlation between SPS and SSS of a target case (query) according to the correlation between SPS and SSS of each of its k nearest neighbors. Now we can redefine precision, recall, and accuracy according to whether the prediction about correlation of a case is right or not, just like we did in information retrieval (IR) before. The goal of IR is to retrieve relevant documents. The performance indicators of precision and recall are defined according to whether the retrieved document is relevant that depends on human judgments. However, in public TCBR datasets at present, these kinds of human judgments are unknown. We redefined the relevance according to whether the predicted solution (case) respects CBR hypothesis or not that is determined by the voting of correlation of a few nearest neighbors for a query.

We can define three performance indicators without the need of human judgments about the correctness of a prediction that is not a class label. In the definition of the various indicators below, we denote a as the total number of the retrieved cases whose correlation between SPS and SSS is positive which is identical with the positive correlation of the target case (query). We denote b as the number of the retrieved cases whose correlation between SPS and SSS is negative which is not identical with the positive correlation of the target case (query). We denote c as the number of the retrieved cases whose correlation is positive which is not identical with the negative correlation of the target case (query), and denote d as the number of cases whose correlation is negative which is identical with the negative correlation of the target case (query), then we can have following equations of performance indicators.

$$precision: p = \frac{a}{a+b} \tag{10}$$

$$recall: r = \frac{a}{a+c} \tag{11}$$

$$accuracy: acc = \frac{a+d}{a+b+c+d} \tag{12}$$

Once defined these three performance indicators, we can use the traditional CBR methods for non-classification case retrieval and non-classification case-based case base maintenance.

5.3 An Illustration Using Artificial Data

In order to testify the feasibility of our rank correlation alignment method, for a first evaluation we have used a very simple artificial test domain, due to the huge complexity of the AIR application domain.

Assuming there are 7 cases in a case base, C1, C2, C3, C4, C5, C6, C7, the SPS and SSS of each case are computed according to 3-NN rank correlation alignment. The correlation of these cases are +, +, -, +, -, -, +, respectively. The SSS of C7 is C1, C2, C3, C4, C5, and C6.

Considering C7 as a target case, we execute 3-NN retrieval. If the SPS is C1, C2, C5, C6, C4, C3, then the predicted correlation of C7 is + according to the majority votes from top 3 nearest neighbors C1(+), C2(+), and C5(-). The prediction is correct. Comparing the position of the most similar case C1 in SPS and SSS, we could find they are the opposite ones.

Similarly, if the SPS of C7 is C6, C5,C1, C2, C3, C4, then the predicted correlation of C7 in 3-NN retrieval is -, according to the majority votes from C6(-), C5(-), and C1(+). The prediction is wrong. At the moment, if we take the most similar case C6 as the candidate solution for C7, we will make a wrong decision. The negative correlation of C6 just makes us avoid this kind of mistake.

Of course, there are still many conditions of SPS needed to be considered that we would not further explain due to the limitation of the length of this paper. But the contribution to retrieval made by correlation has already been clearly clarified.

5.4 Textual Case Authoring Based on Correlation

Case selection is one of the tasks of case authoring [3]. Decisions about what textual descriptions should be included in the case base must be made, especially when there are no explicit instructions about which section should be considered as the problem or the solution among many sections in an investigation report. This task is not like other tasks, such as vocabulary selection, case structuring, similarity metrics and retrieval strategy selection. If there were a domain expert, this would not be a problem. However, knowledge engineers have no idea about the relationship among different sections. Alignment measures could help to find the hidden relationship between two of these sections. This is just what we want alignment measures to do.

In order to prove our method to be effective for alignment computation, we select the same textual content (Analysis) as problem and solution. Now, if average alignment score of the case base is 1, then our method is right. Otherwise, if the alignment score can't get to 1, then our method is not so perfect that it would need to be improved. This test method can also be applied to other alignment measures so that we can check if the alignment can take into the situation account when the problems and the solutions are completely paralleled.

Considering Summary, Other Factual Information, Analysis as problem description respectively, and Analysis as solution, we could organize every Air Investigation Report from Canadian Transportation Safety Board into different <problem, solution> case pairs. Of course, we could use all kinds of alignment measures. If the alignment score between two sections are greater than that between the other two sections for the whole case base through most of alignment measures, we would consider the first two sections as case pair for using in later processing, rather than the latter two ones.

We will compute and compare average alignment score of the case base with three different alignment measures available, Case Alignment (alignMassie), Case Cohesion (alignCohesion), and our Rank Correlation (alignRank). We choose 67 cases in 2000 as our experimental data. The results are as follows:

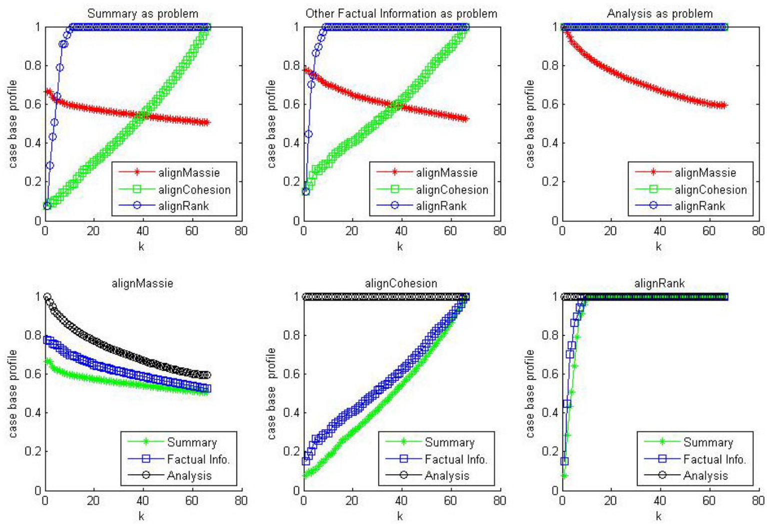

Fig. 2. Case base profile with three different alignment measures for three different subtitles as problem and Analysis as solution

As we can see in Fig. 2, the average alignment score of case base obtained from alignMassie is descending with the number of nearest neighbors that contribute to vote. On the contrary, the average case base alignment score of alignCohesion and alignRank is ascending. Specially, alignRank converge quickly to the max value 1. That is a outstanding outcome. The explanation we can present is that the formula of alignMassie descends with k. But the other two measures ascend with k. In addition, that our alignRank can converge to max alignment score means alignment score don't change with k when it has enough nearest neighbors to vote for his prediction of correlation. We think it is a virtue for selecting which section to constitute a case pair. As the case stands, when the problems and the solutions are completely paralleled alignRank and alignCohesion can demonstrate this situation, but alignMassie can not.

5.5 Textual Case Retrieval Based on Correlation

Rank correlation alignment could be used for textual case retrieval, which assumes the retrieved case have the same correlation with a query. We obtained our experimental results in Fig. 3 using 55 cases in 2000 and 2001 taken from 118 cases of Challenger 1.0 developed by J. A. Recio-Garcia etc. in TCBR'07 workshop [14].

We can see from Fig. 3 our rank correlation alignment measure, alignRank, acquire higher precision and recall than other three alignment measures after 5-NN and 3-NN respectively. alignCorr have the same precision with alignRank in 3-NN. However, alignRank outweighs alignCorr in the precision of 1-NN, 5-NN, 7-NN, and 9-NN. In addition, alignRank overruns alignCorr in recall and accuracy for all k-NN. alignMassie is an exception because the trend in direction of precision and accuracy is first up then down but the range of change is pretty small so that we can think alignMassie

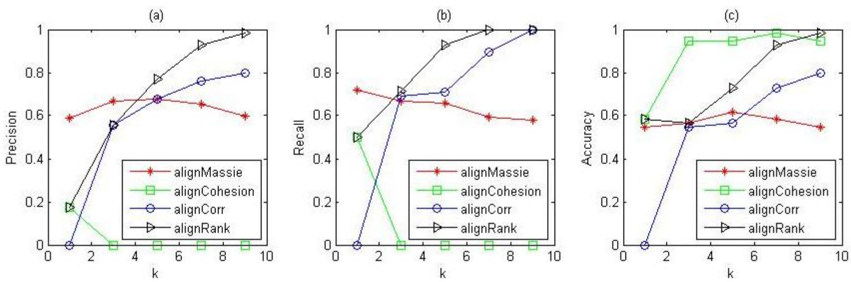

Fig. 3. Comparison of retrieval performance among different alignment measures

is not sensitive to the number of nearest neighbors, k. It is not coincident with human judgment that the more retrieved case is, the more accurate the prediction is. Therefore we are not sure about its contribution to prediction. By the way, most of time the performance alignMassie obtained can't match that alignRank did or alignCorr did.

It should be noted that we adopted the function f(x)=2x-1 to transform the value of both alignMassie and alignCohesion into [-1,1] that represented correlation relationship. This transformation for alignCohesion led to more than half of the retrieved set is covered with negative correlation that means the majority of cases in case base don't respect CBR hypothesis. Although the alignCohesion achieved higher accuracy than other three alignment measures, it is still unacceptable because its precision and recall is too low which means the cases in case base are negative correlation or the majority of cases in case base don't respect CBR hypothesis. It is necessary to reconsider the formula design from transformation to correlation. This would be our further work.

5.6 Textual Case Base Maintenance Based on Correlation

Rank correlation alignment could also be used to maintain case base. It is necessary to adapt the ICF (Iterative Case Filtering) and CBE (Case Base Editing) method implementation because maintenance algorithms in JCOLIBRI CBR framework are just designed for classification applications.

The earlier work in alignment measures or complexity measures suggested these measures could be applied to case base maintenance. For example, ref. [4] pointed out Case Alignment allows us to make informed maintenance decisions about individual cases. Ref. [8] mentioned these measures could identify neighborhoods with poorly aligned problem and solution concepts. However, none of them give an illustration about this issue. In this section, we will show an example application of alignment measures in textual case base maintenance.

A case in case base whose correlation is not identical to its k nearest neighbors is considered as a redundancy. Considering it as a criterion for maintaining case base, we can obtain following results shown in Table. 1.

Table 1. Comparison of accuracy of a predicated correlation before and after maintenance with extended ICF method

Alignment measure	Percent reduced	Accuracy in 3-NN case base maintenance (118 textual cases)	
		Before maintenance	After maintenance
alignMassie	53.39%	71.19%	43.64%
alignCohesion	96.61%	99.15%	75%
alignCorr	48.31%	54.24%	24.59%
alignRank	77.12%	93.22%	59.26%

The experimental results showed that it looks like correlation alignment don't support textual case base maintenance with our definition of redundancy for k nearest neighbors of a query. Surprisingly, alignCohesion outperformed the other alignment measures in the accuracy after maintenance. However, the percent reduced (96.61%) is too large to accept because there are little cases available for prediction. Of course, if there are too many cases that are not positive correlation, the prediction becomes meaningless because the majority of cases in case base don't respect CBR hypothesis. Further experiments will be needed to investigate the effect applying alignment to textual case base maintenance on more textual case bases in future work.

6 Conclusions

The main contributions of this paper are threefold. Firstly, we highlight the importance of case rankings both in problem side and in solution side and prove that it can contribute to the correctness of a predicted non-classification solution. In terms of this observation, we define three performance indicators for evaluating non-classification case-based systems. Secondly, we present a novel alignment/complexity measure based on Kendall profile metric for comparing partial rankings. Thirdly, we present a new algorithm for non-classification case retrieval and case base maintenance that can guarantee that the k nearest neighbors in problem space can be correctly solved by the k nearest neighbors in solution space without human feedback or human judgments. However, we are not sure about whether this method is applicable to textual case base maintenance because our preliminary results don't support our initial presupposition. It is necessary to redefine noise case and redundant case for textual case base according to alignment scores in future work.

Acknowledgments

We thank the anonymous ICCBR reviewers for their very valuable comments and Raghunandan M.A. for his proposal about the rectification in equation (5).

References

1. Leake, D.B.: CBR in context: The present and future. In: Leake, D.B. (ed.) Case based reasoning: Experiences, lessons, and future directions, pp. 3–30. MIT, Cambridge (1996)
2. Leake, D.B., Wilson, D.C.: When Experience Is Wrong: Examining CBR for Changing Tasks and Environments. In: Althoff, K.-D., Bergmann, R., Branting, L.K. (eds.) ICCBR 1999. LNCS (LNAI), vol. 1650, pp. 218–232. Springer, Heidelberg (1999)
3. Lamontagne, L.: Textual CBR Authoring using Case Cohesion. In: TCBR 2006 - Reasoning with Text, Proceedings of the ECCBR 2006 Workshops, pp. 33–43 (2006)
4. Massie, S., Wiratunga, N., Craw, S., Donati, A., Vicari, E.: From Anomaly Reports to Cases. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 359–373. Springer, Heidelberg (2007)
5. Chakraborti, S., Beresi, U., Wiratunga, N., Massie, S., Lothian, R., Watt, S.: A Simple Approach towards Visualizing and Evaluating Complexity of Textual Case Bases. In: TCBR 2007 – Beyond Retrieval, Proc. of the ICCBR 2007 Workshops (2007)
6. Hullermeier, E.: Credible case-based inference using similarity profiles. IEEE Transactions on Knowledge and Data Engineering 19(6), 847–858 (2007)
7. Massie, S., Craw, S., Wiratunga, N.: When Similar Problems Don't Have Similar Solutions. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 92–106. Springer, Heidelberg (2007)
8. Raghunandan, M.A., Wiratunga, N., Chakraborti, S., Massie, S., Khemani, D.: Evaluation Measures for TCBR Systems. In: Althoff, K.-D., Bergmann, R., Minor, M., Hanft, A. (eds.) ECCBR 2008. LNCS (LNAI), vol. 5239, pp. 444–458. Springer, Heidelberg (2008)
9. Raghunandan, M.A., Chakraborti, S., Khemani, D.: Robust Measures of Complexity in TCBR. In: McGinty, L., Wilson, D.C. (eds.) ICCBR 2009. LNCS (LNAI), vol. 5650, pp. 270–284. Springer, Heidelberg (2009)
10. Kolodner, J.L.: Case-Based Reasoning. Morgan Kaufmann, San Francisco (1993)
11. Faltings, B.: Probabilistic Indexing for Case-Based Prediction. In: Leake, D.B., Plaza, E. (eds.) ICCBR 1997. LNCS, vol. 1266, pp. 611–622. Springer, Heidelberg (1997)
12. Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., Vee, E.: Comparing Partial Rankings. SIAM J. Discrete Math. 17(3), 628–648 (2006)
13. JCOLIBRI CBR Framework, Group for Artificial Intelligence Applications, Complutense University of Madrid, http://gaia.fdi.ucm.es/projects/jcolibri/jcolibri2/index.html
14. Recio-Gacia, J.A., Diaz-Agudo, B., Gonzalez-Calero, P.A.: Textual CBR in jCOLIBRI: From Retrieval to Reuse. In: TCBR 2007 – Beyond Retrieval, Proc. of the ICCBR 2007 TCBR Workshops (2007)
15. Delany, S.J., Bridge, D.G.: Catching the Drift: Using Feature-Free Case-Based Reasoning for Spam Filtering. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 314–328. Springer, Heidelberg (2007)