Case Retrieval Optimization of Case-based reasoning through Knowledge-Intensive Similarity Measures

Surjeet Dalal
Research Scholar
Suresh Gyan Vihar University
Jaipur Rajasthan (India)

Dr. Vijay Athavale
Director & Professor
Devraj Groups Technical Campus
Ferozepur Punjab (India)

Keshav Jindal
Assistant Professor
BPR College of Engineering
Gohana Haryana(India)

ABSTRACT
Case based reasoning has become the emerging field of Artificial Intelligence area. It is mostly used in designing the real time application having the decision support capability. It reassembles with human reasoning approach. This reasoning approach contains four phases. It stores the solution of past problems faced in form the case in its case base. In this paper we have discussed about the case retrieval phase of case based reasoning approach. All efficiency of the CBR system depends on the case retrieval process. There are various strategies are used in this phase of case based reasoning. Nearest neighbour & Induction retrieval algorithms are discussed. These algorithms are very simple but inefficient in larger case base & incomplete case. In this paper we will discuss Knowledge-Intensive Similarity measure retrieval strategies for the case base reasoning system & model the knowlededge-intensive similarity measure by using myCBR tool. The basic purpose of our work is to over the bottlenecks of other retrieval strategies.

Keywords- Case-based Reasoning, Case retrieval, similarity measures, Knowledge-intensive similarity measures, myCBR.

1. INTRODUCTION
Case-based Reasoning is one of emerging field of Artificial intelligence research area. It is mostly used in problem solving in the artificial intelligence applications. Case-based reasoning may be defined the approach which utilize the experience gained from solving past problems [1]. This approach maintains all information of past problem solving experience that is called the case. The collection of all these past experiences is stored in form of case base. There are various factors which define the efficiency of this approach [2]. The major factor is the numbers of past experiences stored in case base. The new problem should be identified in term of the experience of past problems faced. The new upcoming problem is considered as new case. The strategies of finding the similar case for the new case regarding the past case stored in case base is another major factor of defining the efficiency of the case-based reasoning approach. The evaluation of selected case & indexing of suggested case for future use are another factor of defining the performance of case-based reasoning system.

The case-based reasoning finds out the solution of new problem in 4 REs phases. In first phase, regarding the new coming problem which is considered as new case, particular case is selected from the cases stored in the case base of the case-based reasoning system. Then selected case is modified with respect to the new case to produce suggested case. This case is tentative solution of the problem. Next the suggested case is revised to validate the solution of the problem. This phase check the solution if it is the optimized solution for solving the problem or not. If it is not found the adequate to fulfil the constraints of the problem, then it is repaired. In this way, the solution is found out for the particular problem. In last phase the solution is stored in case base of this system for future. All these phases are shown in the figure as shown below:

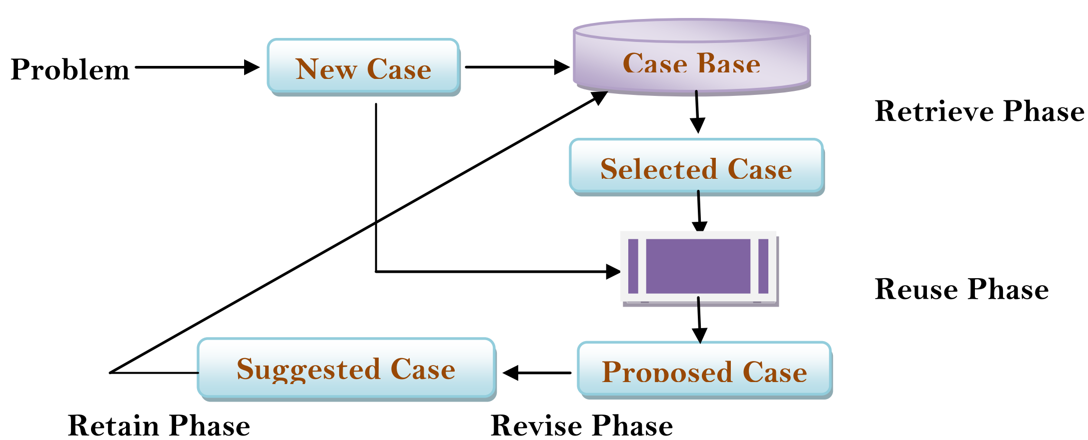

Figure 1: Case-based reasoning

The case-based reasoning has many advantages over other reasoning approaches such like rule based reasoning [4]. This reasoning approach bears a resemblance to human reasoning. It provides the facility of taking the decision such as human beings take decision in real time. The case-based reasoning enhances the learning as the solutions of precedent problems faced are stored in the case base. This approach learns from both success & failures of solutions of the previous problems. These past experiences are being reused for solving new coming problem. The process of knowledge acquisition is easily handled in this approach. But in case of other reasoning approach, the knowledge acquisition process is not so easily & also costly. Other major advantages of this approaches over the other reasoning approaches is the worth of the solution. In the revise phase of case-based reasoning approach, the proposed solution is revised according constraints of the problem. Then the proposed solution is repaired according to constraints. It is also modified for fulfilling the constraints of the problem. This phase of case-based reasoning boost the excellence of the solutions & extends the effectiveness of this approach. The errors of previous solutions do not propagate in the future problem's solutions [5]. It can be also applied in those domains where the information about problems is incomplete & insufficient for finding the adequate rules or algorithms to solve them.

There are a number of major concerns of case-based reasoning approach. These major concerns of this approach are listed as below: -

- What is structure of the cases?
- What are selection strategies for finding similar case?
- How is the case being retrieved?
- How is the selected case being revised?
- How is the suggested case being stored in case base?
- How is suggested case being indexed for faster access.

In this paper, the all concerns regarding the case retrieval phase of this approach is discussed in next sections. There are a number of the algorithms i.e. nearest neighbour algorithm used for selecting the similar case from the case base. In next sections of this paper, the concept of knowledge-intensive similarity measures will be applied for efficient case retrieval in the case-based reasoning system.

2. RELATED WORK

The efficient case retrieval is major factor of determining the performance of case-based reasoning system. This phase of this approach involves the process of finding similar case similar case from stored case in the case base. The concept of similarity measure is used for finding similar case among stored cases [8].

The similarity measurement can be performed in various ways. Both local & global similarity measurement is calculated on basis of attribute-value pairs of the case. The mostly case-based reasoning application use nearest neighbour algorithm is applied to search the similar case among the various cases in case base of this system. This task is finished in two stepladders. In first step, the relevant cases are chosen through facilitating of indexing of the case base. After selection of the relevant cases, the similarity measurement concept is applied to pick the most similar case to new problem [10].

Mingyang Gu et al. provided the comparison of various similarity measurement methods used in case-based reasoning system. The author classified the methods into three categories on root of features found in stored case & query- Case-biased (oriented on features in stored cases), Query-biased (oriented on features in query) & Equally-biased (oriented on features in both stored cases & query). In this paper, the performances of all these methods were analyzed on various sizes of samples [13].

Zhi-Ying Zhang et al. developed the case retrieval model oriented on Artificial Neural Network (ANN) & Nearest Neighbour (NN) algorithm. The first phase of case retrieval (indexing) was handled through the artificial neural network. Finally the similar cases were picked by applying the nearest neighbour algorithm. But the nearest neighbour algorithm is not efficient in large case base [11].

Hugh R. Osborne et al. had proposed the case base similarity framework which was capable of performing the similarity measurement among the various cases. This framework emphasised on sensibility of cases rather than problem description. It did not support the adaptability of the solutions. Hence this framework was not found efficient in similarity measurement [9].

Chuanmin Mi et al. had built the Grey Incidence Theory based framework for finding the nearest neighbour case for new problem. The degree of grey incidence was applied to trace the similar case. In this framework, the concept of analytic hierarchical process (AHM) was introduced to calculate the weight of the case. The operation of framework was oriented on the calculation of degree of the grey incidence. It was found efficient in case of banking domain applications [15].

Mohadam F.M.M et al. worked on hashing indexing technique in the case retrieval process. The author had replaced the sequential indexing technique with hashing indexing techniques for fast access of cases. It generated the hash key for stored case. The model was developed on basis of this hashing indexing technique [16].

Titilola O. Fanoiki et al. had proposed the case-based reasoning approach in which clustering and similarity relations had been applied in measuring the similarity between the new case & stored case. The stored cases were organized in form of clusters. Every cluster contains the same features cases on bases of case base relations [17].

Du Hui et al. proposed improving method based on self-organizing maps (SOM) for case retrieval. Through the thought of SOM networks, clusters were designed, according the visual clustering output similar case group were recognize the most similar case; Self-organizing maps (SOM) afford enhanced solutions to cluster of high dimension data [18].

All these research works are major steps in improvement in case retrieval phase of the case-based reasoning but they were oriented on nearest neighbor algorithms. This algorithm is found inefficient in case large case base. Hence in this paper, the concept of knowledge-intensive is being implemented for fast case retrieval through myCBR tool.

3. CASE RETRIEVAL PHASE

The case retrieval is one of key phase of case based reasoning system. It may be defined as the process of probing the case which is contiguous to the present case contained by a case base. The proficient approach is mandatory to discover the relevant case. The retrieval strategy determines how a case is judged to be apposite for retrieval and a mechanism to control how the case base is searched. The selection strategy is required to find out which is the best case to retrieve, by determining how close the current case is to the existing cases stored in the case base of case based reasoning system.

The case retrieval phase is subdivided into further subtasks. This phase is divided into 4 subtasks as given below:

- Recognize features: At this step, the problem is being acknowledged & all explanation concerning problem is being composed.
- Investigate: Depending on index based information of stored cases in case base, objective solution is being searched.
- Initially match: If direct search is not possible, then on basis of similarity we calculate the similarity between the new case & stored cases in case base
- Select: After calculating similarity between the cases, final selection of the proposed solution is performed regarding the selection strategy [1].

In this way, the more fairly accurate case is being elected for new problem. The case selection tactic depends on the novel problem faced. Regarding the problem as new-fangled case, the appropriate case is being searched out from the case base. There are a number of factors which are considered for selecting the case retrieval method. The main factor is the number of cases to be searched for the new problem. The second factor is availability of domain specific knowledge. Next factor is the simplicity of determining weightings for individual features of the particular cases. Last main factor depends on the indexing all cases which should be indexed by the same features or whether each case may have features that vary in importance. The stored cases are indexed by a particular labels, the new situation is recognized as a key into that index and traverses apposite indexing paths to locate relevant cases. In fact the particular indexing scheme selection is most concern for the case retrieval in case-based reasoning. These indexing scheme search memory using those index levels, and choose the best of the retrieved cases [2].

The output of indexing process is the collection of relevant cases with respect to new case. After this, concept of the similarity measurement between the cases is way of filtering similar case from output of indexing process. Similarity measurement is quite difficult task to perform. Similarity may be defined as amount that reflects the strength of relationship between two substances. The value of this measurement is frequently having range of either -1 to +1 or normalized into 0 to 1. The cases can be distinguished from each other on basis of similarity measure values & they may be grouped using k-means clustering. The basic advantage of grouping of cases is that the characteristics of each group can be recognized. It describes the behaviour of the groups or clusters. Grouping also may give more efficient organization and retrieval of information. It also helps in predicting the behaviour of the new case & simplifying the data that we have into more reasonable relationship. These factors show the significance of similarity measurement for case retrieval process. There are a lot of techniques that use the concept of similarity measurement in various case retrieval algorithms.

4. VARIOUS CASE RETRIEVAL ALGORITHMS

There exists a numeral of case retrieval algorithms applicable in case based reasoning system. These algorithms are based on the similarity metric that allows resemblance between cases stored in case base. The nearest neighbour retrieval algorithm & induction retrieval algorithms are two chief algorithms are used in this process. Nearest-neighbour retrieval is a straightforward approach that computes the similarity relevant cases found through indexing. The case is elected on worth of weighted computation of its feature. When the value of weighted calculation of its features is greater than other cases, then meticulous case is elected from the case base.

In other words, there are multiple case elected though indexing from case base then case4 will be measured as the nearest neighbour among these cases from the case base due to similarity(NewCase, case4) > similarity(NewCase, case1), similarity(NewCase, case4)> similarity(NewCase, case2) and similarity (NewCase, case4)> similarity (NewCase, case3).

The Nearest-Neighbor algorithm is basically oriented of similarity value. For every case, initialize value of total similarity to 0. For each case retrieved from database calculate the value of sim($f_{NewCase}$, f_{casek}) first by following formula:

$$sim(f_{NewCase}, f_{caseK}) = \frac{\sum_{k=1}^{n} (f_{NewCase} * f_{CaseK})}{\sum_{k=1}^{n} \sqrt{(f_{NewCase2}} * \sum_{k=1}^{n} \sqrt{f_{caseK2}}}$$

Using sim($f_{Newcase}$, f_{Casek}) then calculate the similarity value over all signifance weight as given below:

$$Similarity(N, K) = \frac{\sum_{i=1}^{n} w_i * sim(f_N, f_K)}{\sum_{i=1}^{n} w_i}$$

Where f_N & f_K are value of features related for new case & particular case that is stored in case base. The w_i is the significance weight of a feature & sim is the similarity function of features. Next compare the total similarity values of all the cases and find the nearest case for new case.

The main advantage of the nearest neighbour retrieval algorithm is that it is much uncomplicated in the completion. One supplementary gain of nearest neighbour retrieval algorithm is that preindexing is not necessitated. The preindexing process is also time-consuming. But main limitation of this algorithm is that it is slow when number of cases stored in case base is large. In the nearest neighbour algorithm, the speed is major problem for case based reasoning. As the size of case base grow, then performance of this approach goes down. Second induction retrieval approach, failure takes place due to absent of information require in case.

Another approach is the inductive retrieval algorithm that is widely applied in CBR applications. This method determines specific features complete the superlative work in sensitive cases and generates a decision tree type structure to organize the cases in memory. This approach seems very effective when a single case feature is required as a solution, and when that case feature is dependent upon others. Its fast retrieval speed is major advantages of this algorithm. It fails in incomplete & data missing cases & it is also dependent on time consuming preindexing process. In mostly real time applications, the information fields are unpredictable so these fields on totally independent on each other. These algorithms are not oriented on domain specific knowledge. Due to all these facts, these algorithms are not so much efficient in the case retrieval in case based reasoning system.

5. APPLYING KNOWLEDGE-INTENSIVE SIMILARITY MEASURES IN CASE RETRIEVAL PHASE

The main intend of a similarity measure is to evaluate two cases and to calculate a numeric value which represents the level of similarity. Similarity functions can be used to calculate the similarity between numbers based on domain explicit criteria instead of only relying on information about the mathematical distance of the values. The similarity table represents a very dominant demonstration of similarity measures because it represents all the possibility to define separate similarity values for all possible value combinations. The basic purpose of the similarity table is to implement only an exact-match comparison and similarity function represents a simple distance metric. Mostly, the similarity tables can be used for all distinct value types where the value range is defined by an explicit list of a fixed set of values, i.e. the values need not necessarily be symbols. Mostly Case-based reasoning applications are oriented on uncomplicated, broad applicable distance metrics; many application domains necessitate knowledge-intensive similarity measures where domain-specific knowledge is used to estimate the cases' utility more precisely

Knowledge-Intensive Similarity Measures uncover the similarity by encoding more specific domain knowledge about the utility of cases into the similarity measure. Domain specific knowledge is evaluated to determine the features of a case that are important for retrieving that case in the future. In mostly times, various features of a case will have different levels of significance or involvement to the success levels associated with that case. It can be more effective for searching. It improves efficiency and the competence of a CBR system significantly. It retrieves more constructive case in decreased adaptation effort and faster way.

The local similarity measure concerns only solitary feature of the entire domain & it could be characterized as low-level knowledge about the primary similarity function. So for acquiring such common domain knowledge, at slightest a partial understanding of the domain is compulsory. Basically, they are used to state the power of each single attribute on the utility estimation. The last element of the described similarity representation is calculated by comparing new problem as a new case with stored solution as existing cases to find out the final similarity value which is known as the so-called global similarity measure. This measure is represented by an aggregation function computing the final similarity based on the local similarity values computed previously and the attribute weights. It perform the global similarity measure for case into a set of independent local similarity measures sim_i for each attribute a_i and an accurate incorporation of the resulting similarity values has proven its value. By using a weighted sum as incorporation function the similarity between two instances N and M of case may be computed as follows:

$$Sim_i = \sum_{i=0}^{n} w_i * sim_i(N_i, M_i) with \sum_{i=0}^{n} w_i = 1$$

There are a lot of advantages for global similarity measure for case into a set of independent local similarity measure. If the local similarity is performed then it is very time consuming. The adaptation knowledge is being analyzed manually and the relevant knowledge has to be encoded into the similarity measure.

6. IMPEMENING KNOWLEDGE INTENSIVE SIMILARITY MEASURES WITH myCBR

The DFKI had launched open-source case-based reasoning tool *myCBR*. The *myCBR* has a number of the features. The myCBR provides very easy way to develop the case based reasoning applications. It supports fast prototyping & combining state-of-the-art CBR functionality. It can work as standalone & also as plugin of Protégé. The Protégé OWL editor has a specific tab as Similarity Measure Editor shown below in figure 2.

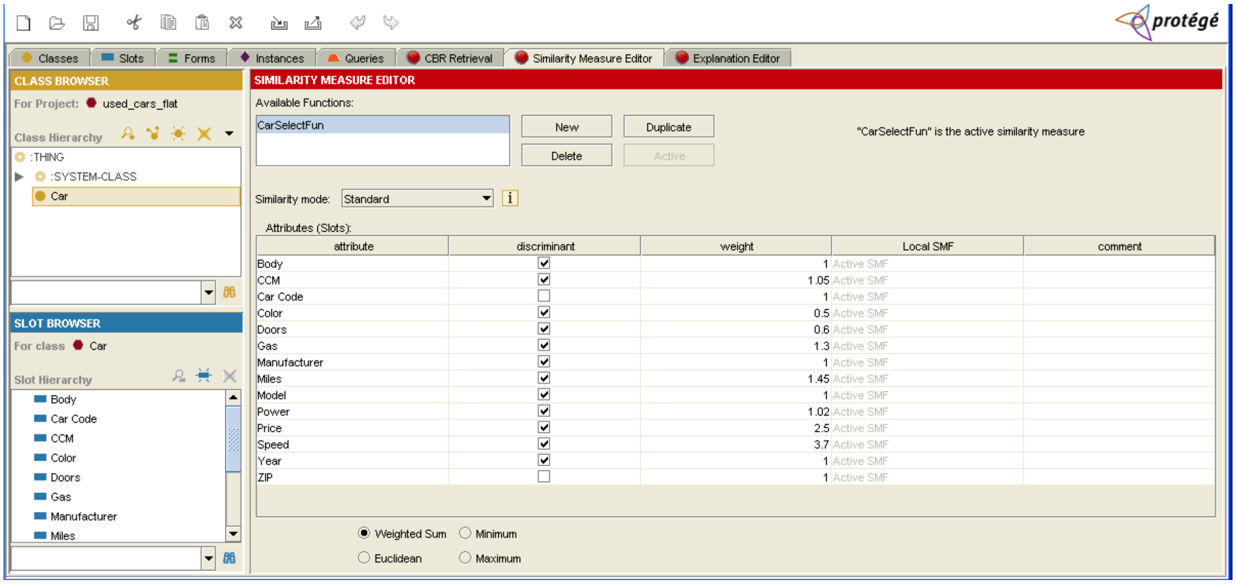

Figure 2: Similarity Measurement Editor

In Similarity Measure Editor, the developer can inscribe the active similarity measure function. The developer can customize the attribute & classify the weight age of various features. When a new problem faces then depending on the parameters it show the CBR Retrieval show the selected case with similarity value even some fields of the case is pending. It means that it has resolved the problem which occurred during Induction Retrieval algorithm. The induction retrieval becomes failure when information about case is missing. But it is not mandatory for knowledge intensive similarity measure case representation in case-based reasoning system. It also implements the global similarity measure with every local attribute as shown below in figure 3.

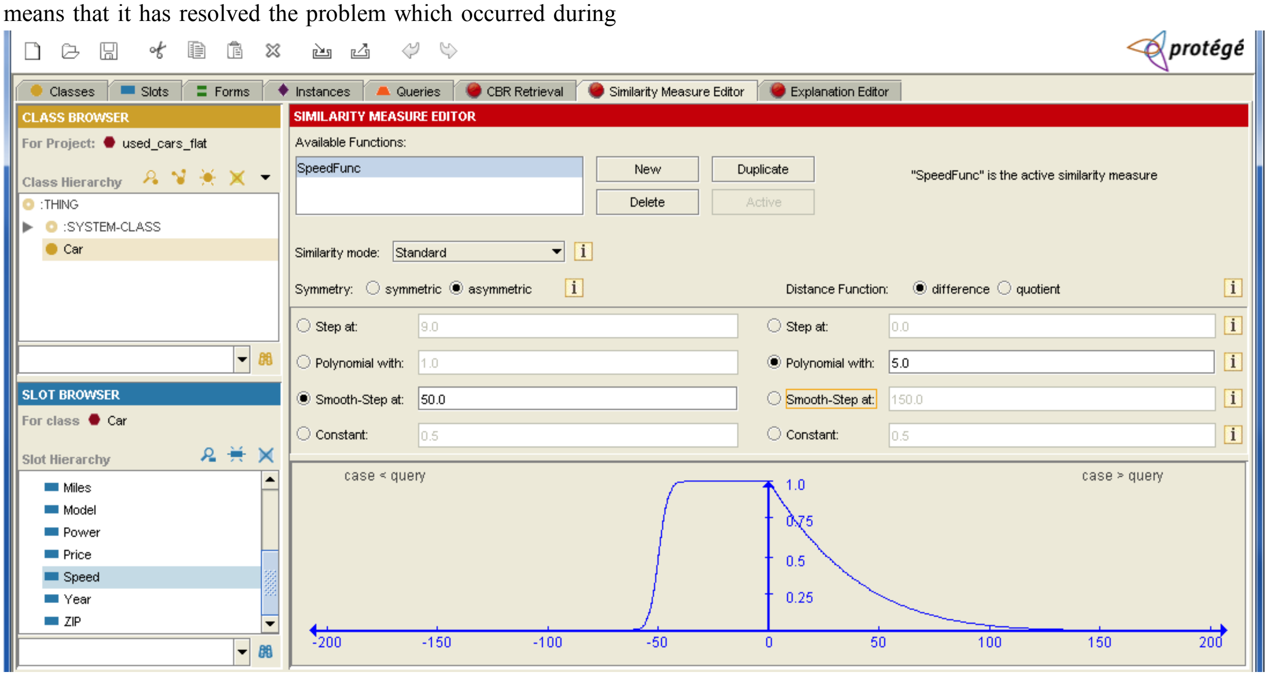

Figure 3: Similarity Measurement

In given diagram we can see the similarity values of one attribute from range of min to max value. It shows the similarity values at the specific point (X, Y). Depending on the Query parameters, the myCBR show the query results with the similarity values.

In CBR Retrieval, the parameters of new case are being defined. The more similar case is retrieved on basic of domain-specific knowledge concept. The result contains list of similar cases arranged in ascending order as shown in the figure 4.

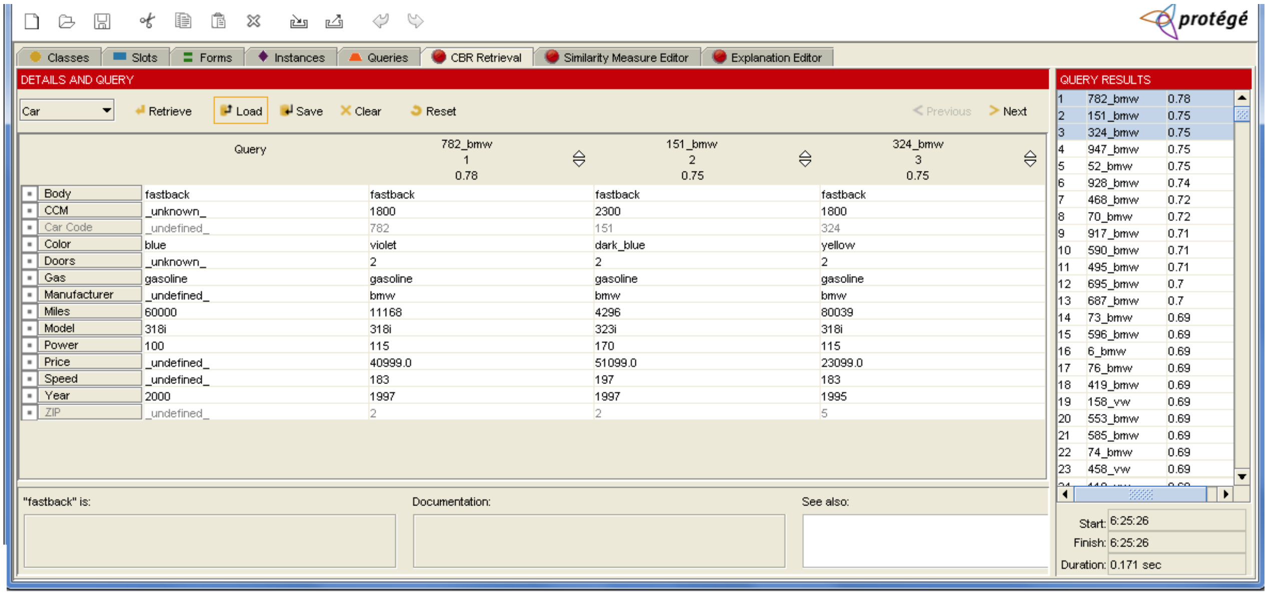

Figure 4: CBR Retrieval

7. CONCLUSION

The knowledge-intensive similarity measures contribute a significant function during human problem-solving processes. This approach is basically based on feature weights and feature-specific local similarity measures. It retrieves the case which is relevant to the problem. In comparison of nearest neighbor retrieval, comparison of the new problem with the case is not a mathematical calculation attribute by attribute. So it does not provide so relevant case that is required to solve the problem in efficient way. But this approach composes global similarity measure with set of local similarity measure. With help of myCBR the developer can implement the knowledge-intensive similarity measure for the case retrieval in case based reasoning system.

8. REFERENCES

[1] A. Aamodt and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and system approach" *AI Communications* 7(1), 39–59, 1994.

[2] Janet L. Kolodner, "An Introduction to Case-Based Reasoning" *Artificial Intelligence Review* 6, 3-34, 1992.

[3] Zhi-We Ni, Shan-Lin "Integrated Case-based Reasoning" *Proceedings of the Second International Conference on Machine Learning and Cybernetics, XI*; 2-5 November 2003.

[4] Leake, David, "CBR in Context: The Present and Future (http:/ / www. cs.indiana.edu/~leake/papers/p-96-01_dir.html/paper. html)", In Leake, D., editor, Case-Based Reasoning: Experiences, Lessons, and Future Directions. AAAI Press/MIT Press, 1-30, 1996.

[5] Simon C.K. Shiu, "Case-Based Reasoning: Concepts, Features and Soft Computing" Applied Intelligence 21, 233–238, 2004.

[6] Frode Sqrmo, "Explanation in Case-Based Reasoning–Perspectives and Goals" Artificial Intelligence Review (2005) 24: 109–143, 2005.

[7] David W. Aha, "Advances in conversational case-based reasoning", *The Knowledge Engineering Review*, Vol. 20:3, 247–254, 2006.

[8] Ramon Lopez De Mantaras, "Retrieval, reuse, revision and retention in case-based reasoning", *The Knowledge Engineering Review*, Vol. 20:3, 215–240, 2006.

[9] Hugh R. Osborne, "A Case Base Similarity Framework" Proceedings of the 3rd Eurepean Workshop on Case-based Reasoning, EWCBR'96, Advance in Case-based Reasoning, Lecture Notes in Artificial Intelligence 1168, Springer Verlag, 1996.

[10] I.Y. Lodhi, "Optimizing Retrieval Process and using Neural Networks for Adaption Process in Case Based Reasoning Systems" Proceeding of IEEE INMIC 2003, 354-360, 2003.

[11] Zhi-Ying Zhang, "A Model for Retrieval base on ANN and Nearest Neighbor Algorithm" Proceeding of Seventh International Conference on Machine Learning & Cybernetics, 142-147, 2008.

[12] Armin Stahl, "Optimizing Retrieval in CBR by Introducing Solution Similarity" Proceeding of the International Conference on Artificial Intelligence (IC-AI'02), 13–23, 2004.

[13] Mingyang Gu, "Comparing Similarity Calculation Methods in Conversational CBR" Proceedings of the International Conference on Information Reuse and Integration IRI -2005 IEEE, 427-4323, 2005.

[14] Sae-Hyun Ji, "Similarity measurement method of case-based reasoning for conceptual cost estimation" Proceedings of the 13th International Conference on Computing in Civil and Building Engineering (icccbe 2010) University of Nottingham 2010.

[15] Chanumin Mi, "Study on Case Retrieving in Case-based Reasoning Based on Grey Incidence Theory and Its Application in Bank Regulation". Proceedings of Fuzzy Systems, 2008. FUZZ-IEEE 2008, 1530-1533, 2008.

[16] Mohamad Farhan Mohamad Mohsin "The Development of Hashing Indexing Technique in Case Retrieval" Proceedings of International Symposium in Information Technology (ITSim), 1045-1050, 2010.

[17] Titilola O. Fanoiki, "Case-Based Reasoning Retrieval and Reuse Using Case Resemblance Hypergraphs" Proceedings of IEEE International Conference on Fuzzy System FUZZ 2010, 1-7, 2010.

[18] Du Hui, "An improving method of CBR retrieval based on self-organizing map" Proceeding of IEEE International Conference on Intelligent Computing and Intelligent Systems ICIS-2009, 616-620, 2009.

[19] Stahl, A, "Learning of Knowledge-Intensive Similarity Measures in Case-Based Reasoning, Dissertation", Verlag dissertation.de, Band 986, (http://www.dfki.uni-kl.de/~stahl/papers/diss.pdf), 2004.

[20] Stahl, A, "Rapid prototyping of CBR applications with the open source tool myCBR" In Bergmann, R., Altho_, K.D., eds.: Advances in Case-Based Reasoning, Springer Verlag 2008.

[21] D Bahls, "Explanation Support for the Case-Based Reasoning Tool *myCBR*" Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence 1844-1846, 2007.