ISSN 2250-3153

iAssist:An Intelligent Online Assistance System

Khanapure V.M[1], Prof. Chirchi V.R[2]

[1]P.G Department(CNE), MBES,COE Ambajogai, M.S.India,
[2] IEEE Member P.G Department(CNE), MBES,COE Ambajogai, M.S.India,
(khanapure.vaishu@gmail.com,vr.chirchi@gmail.com)

Abstract- World Wide Web is the most useful source of information. A search engine can support only the initial stages of the search process, i.e., it can locate Web sites where relevant information is available. But, most of the search engines are keyword-based and are not much useful within a Web site to help the user to identify his preferred service. For this purpose, many companies use intelligent assistance systems (e.g., case-based systems) to improve customer service quality. We develop iAssist, an intelligent online assistance system, to automatically find problem solution at terns from the past customer–representative interactions.When a new customer request arrives, iAssist searches and ranks the past cases based on their semantic relevance to the request, groups the relevant cases into different clusters using symmetric matrix factorization, and summarizes each case cluster to generate recommended solutions.

Index Terms- Case clustering, case summarization, intelligent assistance, semantic similarity.

I. INTRODUCTION

The 70% of the customers hit the road not because of the price or product quality issues but because they do not like the customer service. Current customer service (also called helpdesk, call center, etc.) involves a lot of manual operations, which require customer service representatives to master a large variety of malfunction issues. Moreover, it is difficult to transfer knowledge and experience between representatives. Thus, many companies attempt to build intelligent helpdesk systems to improve the quality of customer service.

It is also many online intelligent systems but mainly suffer from keyword matching technologies and error-level information at the solution time. So we have proposed a new algorithm called semantic role parser, similarity score calculation. The main objectives of this automatically find the problem solution.

Given a new customer request, one common scenario of an intelligent helpdesk system is to find whether similar requests have been processed before. Helpdesk systems usually use databases to store past interactions (e.g., descriptions of a problem and recommended solutions) between customers and companies.

In a case-based system case collects all the information provided by the user during a recommendation session such as the user's queries to the product catalogues, the selected products, and, in case the user is registered, some stable user-related preferences and demographic data. However, these case-based systems face the following two challenges.

A. Case retrieval measures

Given a new request from a customer, most case-based systems search and rank the documents of past cases based on their relevance to the request. Many methods have been proposed to determine the relevance of past cases to requests in database, and to perform similarity search. However, these methods usually use traditional keyword-matching-based ranking schemes, which have difficulty in capturing the semantic meanings of the requests and the past cases.

For example, given a request ―can you switch the computers? most case-based systems would return past cases related to network switches. In addition, when the description of the cases or items becomes complicated, these case-based systems also suffer from the curse of dimensionality, and the similarity/distance between cases or items becomes difficult to measure. New similarity measurements that are able to understand the semantic meanings in the requests and the past cases are thus needed.

B. Result representation

Most case-based systems return a list of past cases ranked by their relevance to a new request. Customers have to go through the list and examine the cases one by one to identify their desired cases. This is a time-consuming task if the list is long. A possible solution is to organize the past cases into different groups, each of which corresponds to a specific context or scenario. This would enable the customers to identify their desired contexts at a glance. It is also necessary to generate a short and concise summary for each context to improve the usability.

II. LITERATURE SURVEY

A. Case-based systems

Some case-based systems is based on keyword matching, which lacks the semantic analysis of customer requests and existing cases[4][6].Thus new similarity measurement are needed that are able to understand the semantic meaning in the request & past cases[1][5].

B. Database search and ranking

Similarity is measure based on Keyword matching, which have difficulty to understand text deeply[7]. For finding answers quickly once a new request arrives, cases are rank based on semantic importance[2].

In database search, many methods have been proposed to perform similarity search and rank results of a query. However, similar to the case based systems, the similarity is measured based on keyword matching, which have difficulty to understand the text deeply.

C. Clustering search results

Search results are long list, so it is time consuming process[8]. To find the better solution for problem, Online Helpdesk System first cluster the top ranking cases[3][5].

Since existing search engines often return a long list of search results, clustering technologies are often used in search result organization. However, the existing document-clustering algorithms do not consider the impact of the general and common information contained in the documents. In our work, by filtering out this common information, the clustering quality can be improved, and better context organizations can then be obtained.

D. Document summarization

Information contain in different document often overlap with each other[9]. Therefore, it is necessary to find an effective way to merge the document while recognizing and removing redundancy[3].

Multidocument summarization is the process of generating a summary by reducing documents in size while retaining the main characteristics of the original documents . We utilize the idea of a request-focused multidocument summarization and propose a new summarization method to summarize each cluster of the past cases and generate reference solutions, which can better assist customers to find their desired solutions.

III. PROPOSED IMPLEMENTATION

A. Existing System

In *existing system*, a help desk is a place that a user of information technology can call to get help with a problem. In many companies, a help desk is simply one person with a phone number and a more or less organized idea of how to handle the problems that come in. In larger companies, a help desk may consist of a group of experts using software to help track the status of problems and other special software to help analyze problems (for example, the status of a company's telecommunications network).

Typically, the term is used for centralized help to users within an enterprise. A related term is call center, a place that customers call to place orders, track shipments, get help with products, and so forth. The World Wide Web offers the possibility of a new, relatively inexpensive, and effectively standard user interface to help desks (as well as to call centers) and appears to be encouraging more automation in help desk service.

Some common names for a help desk include: Computer Support Center, IT Response Center, Customer Support Center, IT Solutions Center, Resource Center, Information Center, and Technical Support Center.

The above *current customer service* (also called helpdesk, call center, etc.) involves a lot of manual operations, which require customer service representatives to master a large variety of malfunction issues. Moreover, it is difficult to transfer knowledge and experience between representatives. Thus, many companies attempt to build intelligent helpdesk systems to improve the quality of customer service.

Existing customer service or helpdesk systems were dealing with keyword-matching based ranking scheme for case retrieval and results will be in a list format i.e. these case based systems faces two main challenges :

1) Case retrieval measures: most case-based systems use traditional keyword-matching-basedranking schemes for case retrieval and have difficulty to capturethe semantic meanings of cases and

2) Result representation: mostcase-based systems return a list of past cases ranked by theirrelevance to a new request, and customers have to go through thelist and examine the cases one by one to identify their desired cases**.**

Example: Apache Lucene keyword based text ranking engine.

B. Proposed System

Pre-processing of customer request and pas cases
Sentence-level semantic analysis to compute semantic similarity between customer request and past cases
Case Relevance calculation
Top-ranking case clustering using SNMF algorithm
Multidocument summarization for each case cluster
Reference solution in a summarized format

Figure 1: Proposed Intelligent Assistance system

Example: Can you switch those two computers?: Table II shows the top-ranking case samples retrieved by Lucene and iAssist, respectively. While looking at the ranking results, we find that Lucene takes the word "switch" as the keyword; thus, some high-ranking past cases actually discuss about the equipment switch. However, these cases are not semantically related to the user's request at all. For iAssist, the word "switch" is a verb, and the corresponding semantic role is "rel." Therefore, the cases related to the equipment switch are unlikely to be ranked high. In addition, iAssist ranking is able to find some informative content that Lucene cannot. For example, some cases related to "move machines" ("machines" and "computers" are related words) and "exchange computers" are highly ranked in iAssist, although the word of "switch" or "computer" does not appear in the text itself.

After clustering the top-ranking cases, we find that the request of switching computers can be either changing the location of two computers or changing the users and accounts for the computers. Sample summaries by iAssist are shown in Table III.

ISSN 2250-3153

TABLE II: TOP-RANKING CASE SAMPLES BY LUCENE AND iASSIST IN SCENARIO 1

Request	Can you switch those two computers?
Lucene	Top-ranking case sample 1: Subject: KVM switches/extra cables order request We would like to purchase following switches and cables: -eight port KVM w/corresponding 8 VGA/(ps/2) switch cable sets -6 VGA/(ps/2) additional sets for our current CYBEX switches Could we get quotes please?
	Top-ranking case sample 2: Subject: research 2-port usb switch Steve: When we deploy Tim's new machine in his office,we will need another 2-port KVM switch, like the one we purchased for the Lab.
iAssist	Sample Top-ranking case 1: Subject: Moving people and computers from ECS 257 The following moves have been apoproved.Move 94 on desk D-322 to 238 also remove D-322. Deploy 1 new Dell to 238. If you need help identifying these machines/desk,Andriy or I can help you.
	Top-ranking case sample 2: Subject: my computer I have an approval from Becky to exchange computers, and I need some time to make a backup of important information.Then I will bring both computers on Tuesday.

C. Advantages of iAssist System

Many companies use intelligent assistance systems (e.g., case-based systems) to improve customer service quality. We develop iAssist, an intelligent online assistance system, to automatically find problem solution at terns from the past customer–representative interactions. When a new customer request arrives, iAssist searches and ranks the past cases based on their semantic relevance to the request, groups the relevant cases into different clusters using symmetric matrix factorization, and summarizes each case cluster to generate recommended solutions.

Many case based system suffer from keyword matching technologies and error-level information at the solution time. So we have proposed a new algorithm called semantic role parser, similarity score calculation. The main objectives of this automatically find the problem solution. The high performance of iAssist benefits from the proposed approaches of semantic case ranking, case clustering using the symmetric matrix factorization, and the request-focused multidocument summarization.

D. Disadvantages of iAssist System

Helpdesk is critical to every enterprise's IT service delivery. iAssist, an intelligent online assistance system, to automatically find problem–solution pat-terns given a new request from a customer by ranking, clustering, and summarizing the past interactions between customers and representatives, Which is time consuming process. When Sentence-Level Semantic Similarity Calculation takes more time.

E. Application of iAssist System

The main focus is the central support for

1. Office and back office products,
2. Hardware and network infrastructure
3. Customer specific applications support.

IV. FRAMEWORK

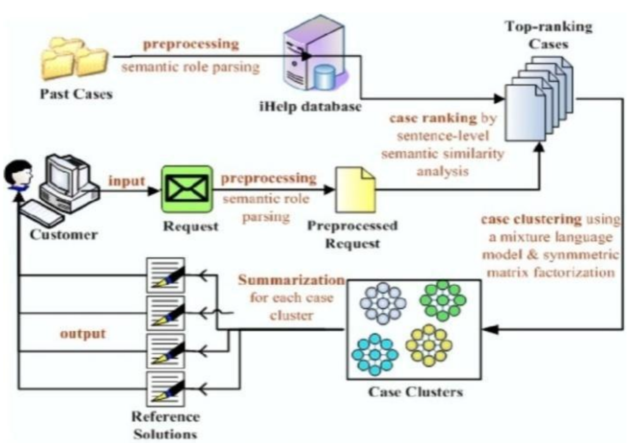

Fig.1. Framework of the iHelp system

Fig. 1 shows the framework of Online Helpdesk System. The input of the system is a request by a customer and a number of past cases. First of all, the past cases are cleaned by removing formatting char-acters and stopping words; then, each of the cases is trunked into sentences and passed through a semantic role parser in the preprocessing step. Then, in the case-ranking module, the past cases are ranked based on their semantic importance to the preprocessed input request. Other than searching and ranking the relevant cases, Online Helpdesk

System also groups the top-ranking cases into clusters using SNMF. Finally, a brief summary for each case cluster is generated as a reference solution to the customer.

V. REQUEST-BASED SEMANTIC CASE RANKING

To assist users in finding answers quickly once a new request arrives we propose a method to calculate the semantic similarity between the sentences in the past cases and the request based on the semantic role analysis.

A. Sentence-Level Semantic Similarity Calculation

Given sentences *Si* and *Sj* , we now calculate the similarity between them. Suppose that *Si* and *Sj* are parsed into frames by the semantic role labeler, respectively. For each pair of frames *fm* ∈ *Si* and *fn* ∈ *Sj* , we discover the semantic relations of terms in the same semantic role using WordNet [12]. If two words in the same semantic role are identical or of the semantic relations such as synonym, hypernym, hyponym, meronym, and holonym, the words are considered as ―related.

Let *{r*1*, r*2*, . . . , rk}* be the set of *K* common semantic roles between *fm* and *fn*, *Tm*(*ri*) be the term set of *fm* in role *ri*, and *Tn*(*ri*) be the term set of *fn* in role *ri*. Letting |*Tm* (*ri*)| ≤ |*Tn*(*ri*)|, we compute the similarity between *Tm*(*ri*) and *Tn*(*ri*) as

$$rsim(Tm(ri),Tn(ri)) = \frac{\sum_j tsim(t_{ij}^m, ri)}{|Tn(ri)|} \quad (1)$$

Then, the similarity between *fm* and *fn* is

$$fsim(f_{m,} f_n) = \frac{\sum_{i=1}^{k} rsim(T_m(r_i), T_n(ri))}{K} \quad (2)$$

Therefore, the semantic similarity between Si and Sj can be calculated As follows:

$$Sim(S_{i,} S_j) = \max_{f_m \in S_{i,} f_n \in S} fsim(f_{m,} f_n) \quad (3)$$

where each similarity score is between zero and one.

VI. TOP -RANKING CASE CLUSTERING

To better facilitate users to find the solutions of their problems, iHelp first clusters the top-ranking cases and then generates a short summary for each case cluster. Although the top-ranking cases are all relevant to the request input by the customer, these relevant cases may actually belong to different categories. For example, if the request is “my computer does not work.” the relevant cases involve various computer problems, such as system crash, hard disk failure, etc. Therefore, it is necessary to further group these cases into different contexts.

A. Semantic role parsing

Sentence-level semantic analysis can better capture the relationships between sentences, and we use it to construct the sentence similarity matrix by computing the pair wise sentence similarity.

A semantic role is a description of the relationship that a constituent plays with respect to the verb in the sentence"[10]. Semantic role analysis plays a very important role in semantic understanding. In Online Helpdesk System, we use NEC SENNA [10] as the semantic role labeler, which is based on Prop Bank semantic annotation [11]. The basic idea is that each verb in a sentence is labeled with its propositional arguments, and the labeling for each particular verb is called a “frame”. Therefore, for each sentence, the number of frames generated by the parser equals the number of verbs in the sentence. There is a set of abstract arguments indicating the semantic role of each term in a frame

B. Symmetric nonnegative matrix factorization

Once we obtain the similarity matrix of the relevant cases, clustering algorithms need to be performed to group these cases into clusters. Most document-clustering algorithms deal with a rectangular data matrix (e.g., document-term matrix and sentence-term matrix), and they are not suitable for clustering a pairwise similarity matrix. In our work, we propose the SNMF algorithm to conduct the clustering.

Problem Formulation and Algorithm Procedure:

Given a matrix of pairwise similarity *W*, we want to find *H* such that

$$\min_{H \geq 0} J = \| W - HH^T \|^2$$

where the matrix norm $\| X \|^2 = \sum_{ij} X_{ij}^2$ is Frobenius norms.

To derive the updating rule for above Eq with non-negative constraints, $h_{ij} \geq 0$ we introduce the Lagrangian multipliers λ_{ij} and let $L = J + \sum_{ij} \lambda_{ij} H_{ij}$.

The first order KKT condition for local minima is

$$\frac{dL}{dH_{ij}} = \frac{dJ}{dH_{ij}} + \lambda_{ij} = 0$$, and $\lambda_{ij} H_{ij} = 0, \forall i, j$.

Note that

$$\frac{dJ}{dH} = -4WH + 4HH^T H$$.

Hence the KKT condition leads to the fixed point relation:

$$(-4WH + 4HH^T H)_{ij} H_{ij} = 0$$

Using gradient descent method, we have

$$H_{ij} \leftarrow H_{ij} - \in_{ij} \frac{dJ}{dH_{ij}}$$

Setting

$$\in_{ij} = \frac{H_{ij}}{(8HH^T H)_{ij}}$$

we obtain the NMF style multi-plicative updating rule for SNMF:

$$H_{ij} \leftarrow \frac{1}{2}[H_{ij}(1 + \frac{(WH)_{ij}}{(HH^T H)_{ij}}]$$

Hence, the algorithm procedure for solving SNMF is: given an initial guess of H, iteratively update H using above Eq.until convergence. This gradient descent method will converge to a local minima of the problem.

Within-Cluster Sentence Selection

After grouping the sentences into clusters by the SNMF algorithm, in each cluster, we rank the sentences based on the sentence score calculation as shown in below Eqs. The score of a sentence measures how important a sentence is to be included in the summary.

$$Score(S_i) = \lambda F_1(S_i) + (1-\lambda) F_2(S_i)$$

$$F_1(S_i) = \frac{1}{N-1} \sum Sim(S_{i,} S_j)$$

$$F_2(S_i) = Sim(S_i, T)$$

where F1(Si) measures the average similarity score between sentence Si and all the other sentences in the cluster Ck, and N is the number of sentences in Ck. F2(Si) represents the similarity between sentence Si and the given topic T.

VII. CONCLUSION

iAssist, an intelligent online assistance system, to find problem–solution patterns given a new request from a customer by ranking, clustering, and summarizing the past interactions between customers and representatives. Many case based system suffer from keyword matching technologies and error-level information at the solution time. So we have proposed a new algorithm called semantic role parser, similarity score calculation. The main objectives of this automatically find the problem solution..

REFERENCES

[1] B. K. Giamanco, Customer service: The importance of quality customer service. [Online]. Available: http://www.ustomerservicetrainingcenter.com

[2] S. Agrawal, S. Chaudhuri, G. Das, and A. Gionis, “Automate ranking of database query results,” in *Proc. CIDR*, 2003, pp. 888–899.

[3] D.Wang, S. Zhu, T. Li, and C. Ding, “Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization,” in *Proc. SIGIR*, 2008, pp. 307–314.

[4] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, “When is nearest neighbor meaningful?” in *Proc. ICDT*, 1999, pp. 217–235.

[5] D. Radev, E. Hovy, and K. McKeown, “Introduction to the special issue on summarization,” *Comput. Linguist.*, vol. 28, no. 4, pp. 399–408, Dec. 2002.

[6] D. W. Aha, D. Mcsherry, and Q. Yang, “Advances in conversational case-based reasoning,” *Knowl. Eng. Rev.*, vol. 20, no. 3, pp. 247–254, Sep. 2005.

[7] R. Agrawal, R. Rantzau, and E. Terzi, ―Context-sensitive ranking,‖ in *Proc. SIGMOD*, 2006, pp. 383–394.

[8] A. Leuski and J. Allan, “Improving interactive retrieval by combining ranked list and clustering,” in *Proc. RIAO*, 2000, pp. 665–681.

[9] X. Liu, Y. Gong, W. Xu, and S. Zhu, “Document clustering with cluster refinement and model selection capabilities,” in *Proc. SIGIR*, 2002, pp. 191–198.

[10] R. Collobert and J.Weston, “Fast semantic extraction using a novel neural network architecture,” in *Proc. ACL*, 2007, pp. 560–567.

[11] M. Palmer, P. Kingsbury, and D. Gildea, “The proposition bank: An annotated corpus of semantic roles,” *Comput. Linguist.*, vol. 31, no. 1, pp. 71–106, Mar. 2005.

[12] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.

[13] X. Liu, Y. Gong, W. Xu, and S. Zhu, “Document clustering with cluster refinement and model selection capabilities,” in Proc. SIGIR, 2002, pp. 191–198.

[14] D. Radev, E. Hovy, and K. McKeown, “Introduction to the special issue on summarization,” Comput. Linguist., vol. 28, no. 4, pp. 399–408, Dec. 2002.

[15] Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, Aug. 2000.

[16] D.W.Aha, D. Mcsherry, and Q. Yang, “Advances in conversational case-based reasoning,” Knowl. Eng. Sep. 2005.

[17] D. Bridge, M. H. Goker, L. Mcginty, and B. Smyth, “Case-based recommender systems,” *Knowl. Eng. Rev.*, vol. 20, no. 3, pp. 315–320,Sep. 2005

[18] Dingding Wang, Tao Li, Shenghuo Zhu, and Yihong Gong,”iHelp: An Intelligent Online Helpdesk System” in IEEE TRANSACTIONS 2011