Divided Pretreatment to Targets and Intentions for Query Recommendation

Yangyang Kang, Yu Hong[*], Li Yu, Jianmin Yao, and Qiaoming Zhu

School of Computer Science & Technology, Soochow University
No.1 Shizi Street, Canglang District, Suzhou City, Jiangsu Province, 215006
`tianxianer@gmail.com`

Abstract. We propose a query recommendation method called "Divided Pretreatment to Targets and Intentions for Query Recommendation", which concentrates on the structure, elements and composition of a query. Based on the recognition of query targets and query intentions by a classifying method, the clusters of query intentions are built following the clue of consistent and similar query targets. After that, query recommendations are generated by simple substitution of peer intentions. This method aims to explore a simple and efficient mechanism, which only analyzes and processes query itself and its internal attributes. The experiment demonstrates that, accuracy of "query targets" and "query intentions" recognition is 73.11%, while that of intention clustering reaches 55.67%. The p@1 value of query recommendation gets 57.83%.

Keywords: query recommendation, query target, query intention.

1   Introduction

Query Recommendation is an information retrieval technology to recommend identical or related queries for a particular query. Its core mission is to understand user's query intention through machine learning, mine the expressional forms of query with similar intentions, and finally realize mutual recommendation of the synonymous or associated queries. The ultimate purpose is to use the synonymous or associated queries as medium to share and mutually recommend better search results.

Currently, the main research form of query recommendation is to express user's potential query intention with the correlation information within the framework of information retrieval and outside the query such as relevant or pseudo-relevant text content or click and browsing retrieval behaviors, then measure query semantic consistency and intention association, and recommend approximate or related queries finally. However, we ignore to analyze or understand the intention from query itself because of sparse information, fuzzy structural relations and optional component forms.It is often used as a medium to auxiliarily mine external reference resources which contain richer and potential intention informations. Therefore, query itself was long considered "be unrefined" for intention analysis and query recommendation.

---

[*] Corresponding author.

However, query is a direct intention expression of user, which describes real and appropriate meaning of intention. At the same time, query itself contains less noisy informations when semantic ambiguity and misspelling is ignored. In contrast, using external resources associated with the query for intention analysis will introduce a lot of noise which will mislead intention analysis and calculation. So, if there is an effective measure of expansion and disambiguation, the direct use of query itself can achieve the mining of correlation intention and query recommendation from the perspective most closed to user as well as intention expressed in behavior. Therefore, we propose a query recommendation method based on divided pretreatment to query targets and query intentions. This method includes three basic steps.

— Classifying and recognizing the words of query which describe the target and the intention.
— Clustering the words of query intention to mine consistent or similar intention.
— Using intention cluster to realize a simple substitution of peer intentions to form query recommendation.

Query target refers to the target entity, behavior or status which user retrieves. Query intention refers to the operations of intention or the motivations of retrieval which impose on the query target. For Example 1, there is a query "*Where is the No. 18 bus station?*"..Its query target is "*No. 18 bus station*" and query intention is "*where*".

We can effectively overcome the sparsity and ambiguity of the query informations, through classifying and recognizing the targets and intentions of query with large-scale query samples. As the sparsity of the information which describes the intention, we can expand by different intention words with common target. There are queries such as "How to get to the 18 bus station?", "Where is the location of the 18 bus station" and "Where is the position of the 18 bus station" with common target "18 bus station" in large-scale query samples. We expand the intention "*where*" in example 1 to {"*where*", "*how to get to*", "*location*", "*position*"} to enrich the description of original query intention. As description of target can cause ambiguity possibility, we can take different targets pointed to by same intention to disambiguate. For example 2, there is a query "*Apple Quote*" which query target is "*Apple*" and query intention is "*Quote*". The target "*Apple*" is ambiguous because it may refers to the electronic product or fruit. So, we gather the queries with consistent intention words such as "*Iphone Quote*", "*Ipod Quote*", "*Computer Quote*", "*Mainboard Quote*". By extracting the description of targets, the lexical meanings {"*Iphone*", "*Ipod*", "*Computer*", "*Mainboard*"} of target "*Apple*" in original query are formed to disambiguate effectively. While, "*Fruit*" is often gone with the word "Price", such as "*Banana Price*" rather than "*Banana Quote*".

In the rest of the paper, we will first introduce the related work in Section 2. Then, we give the framework of a new query recommendation method based on divided pretreatment to targets and intentions in Section 3. Section 4 will show the modeling process of classifying and recognizing the targets and intentions of query. The method of intention cluster will be given in Section 5. We present and discuss the experimental results in Section 6. Finally, we conclude the paper in Section 7.

2   Related Work

According to the source of corpus, the methods of query recommendation are divided into two categories: *document-based* approaches and *log-based* approaches.

   Document-based approaches exploit the words or phrases related to the query to constitute candidate recommended queries. Yanan Li et al. [1] proposed a division of *document-based* approaches into three categories. The first approach based on global documents is to find the words closed to the query through analyzing the relations between words in all documents, then construct the recommended queries. But it is difficult to be realized in real large-scale data. The second approach based on local document is to find the words or phrases related to the query in relevant documents. But it is difficult to get the relevant documents. The solution is to select *top-n* pseudo-relevant documents from search engine but it will introduce the noise of irrelevant documents. Nick et al. [2] proposed a hidden markov random walk model based on large-scale click logs to obtain the relevant documents that have not been clicked.The last approach based on manual editing corpus will have accurate results but deal with the new word in the network difficultly.

   Currently, *log-based* approach is the main direction of research on query recommendation. Cao et al. [3] divided log-based approaches into two categories as *session-based* approach and *click-based* approach after summary of previous work. For *session-based* approach, Huang et al. [4] recommended mutually between queries through mining the queries co-occurred in the same session. For *click-based* approach, the similar queries were recommended through bipattite graph model which was built up with the user's click and history information. Jimin Wang et al. [5] proposed a method based on the numbers of common URLs clicked between queries. The basic assumption is that the more same clicked URLs two queries share, the more similar they are. Yanan Li et al. [6] proposed a method based on weighted SimRank to mine the indirect correction and semantic relation between queries. Mei et al. [7] apply Hitting Time to large-scale bipattite graph to rank the queries and obtain the consistent semantic queries. As the sparse of user clicks in query logs, the solution is to compute iteratively by transferring similarity with improved SimRank algorithm or reduce dimensions. But the calculation is complex which is difficult to be applied to the general search engine with high real-time demand.

   The purpose of *document-based* approach and *log-based* approach is to better understand user query intention and recommend more high-quality queries. But the understanding of query intention is difficult because the sparsity and ambiguity of queries will cause the fuzziness and diversity of intentions and the deviation of retrieval results. Therefore, research on recognition of query intention has become popular in recent years.

3   Framework of Query Recommendation

The main framework is shown in figure 1 which contains two main parts: *offline-sys* and *online-sys*. For *offline-sys*, we first extract queries from query logs and utilize "*Targets*

& Intention Recognition Model" to extract target or intention words in each query. Then, we utilize "*Intention Cluster Model*" to build intention clusters. For *online-sys*, if there is a query under test, we first extract the target and intention of the query. Then, we determine the similarity of query intention and priori intention clusters with intention similarity matching. The most similar intention cluster is chosen as a candidate. Finally, we measure similarity on each word in candidate intention cluster with query intention test and rank intentions according to similarity. At last, the most similar intention word and query target will be combined to form recommended queries.
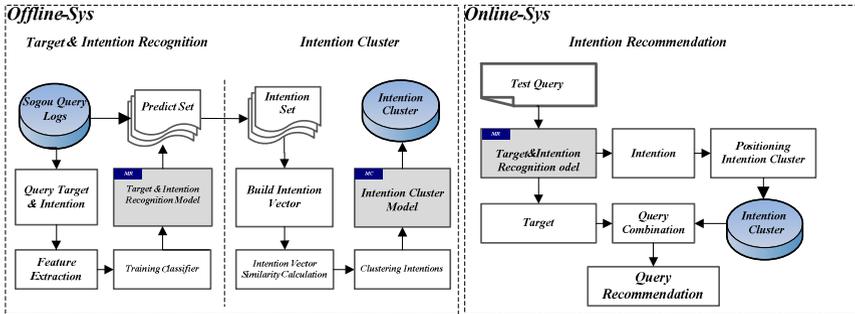

Fig. 1. Framework of Query Recommendation

4   Target and Intention Recognition Model

4.1   Query Preprocessing

Based on the rules of classification proposed by Broder et al. [8], *Beaze-Yates* divided query into three categories: informational, not informational and ambiguous. In this paper, we only focus on the informational queries because not informational queries doesn't contain target word (Such as query "*Quote*") and   ambiguous queries only give target or intention word(such as query "*Apple*") with amiguous.

Thus, we propose three filtering rules to recognize and filter the not informational and ambiguous queries in preprocessing. 1)Filtering the queries which contain letter, numeral, punctuation or URL. Because considering the internal structure and the componential role information of queries from a semantic perspective, the noise brought about by above symbols should be avoided. 2)Filtering "short queries". In experiment, we extract the queries with more than two words to form query sets. 3)Filtering queries like title, news and notice. Such queries belong to navigational queries in nature. Search engine can identify and recommend a consistent navigational link with maximum string matching.

By taking the refined filter rules and remove the reduplicate queries, we collect *1.9* million queries from the query logs of *Sogou2008* to auxiliary query analysis and recommendation research. As the query resources coming from Chinese retrieval system *Sogou*(*Sogou 2008*) and the advantages of *ICTCLAS* dealing with query segmentation, we use *ICTCLAS* to realize the basic division of query features words.

4.2   Feature Selection

We should consider from multiple perspectives such as *lexical-based* perspective and *context-based* perspective to identify a word. Following, feature extraction method will be introduced separately.

- **Lexical-Based Perspective**

We choose four empirical classification features (shown in Table 1) from the *lexical-based* perspective. First, the word itself is an effective feature. The words such as "*免费*", "*在线*" and "*下载*" have obvious intention performance, while these words like "*大学*", "*计算机*", and "*单位*" express lower probability of intention than probability of target. Secondly, the head or tail character of word can be used as classification feature. For example, the character "*最*" among the words as "*最新*", "*最好*", "*最贵*" with a strong emotional color, can better transfer user's query intention. In contrast, another group of words "*在线*", "*热线*" and "*路线*" etc. reflect intention by the tail character "*线*". Finally, part of speech is an effective classification feature. The main grammatical structure of Chinese is *Subject-Verb-Object*, which Subject and Object consist of noun phrase and Verb consist of verbal phrase. Nouns and noun phrases often reflect the targets, while Verbs and verbal phrases often reflect the specific intention. Similarly, adjectives and adverbs appeared in the short text chapter (such as query) often express an emphasis on target state of query which can also reflect user's intention.

Table 1. Features of Classification

Feature	Description
Lexical-based	Word itself
	The head character of word
	The tail character of word
	Part of speech
Context-based	Word Frequency
	Word Position

- **Context-Based Perspective**

From the *context-based* perspective, we proposed two features as frequency and position. Frequency was considered as feature on the basis of an assumption that intention word cannot be used as independent query. It can combine with a large number of different types of target words to form a complete query description. For example, user does not often use "*下载*" as query independently because its description of itention is fuzzy. By using a combination of large different targets such as "*电影*", "*视频*", and "*软件*" etc, we can form a complete query semantics. So, the intention words are widely distributed, high frequency, and able to combine with various categories of words. In contrast, the target words have a single combination with other words, and the frequency of occurrence is much lower than intention words. For

instance, target word "*电影《黑衣人》*" can only collocate a limited number of intention words such as "*在线*", "*观看*", "*下载*", and "*光盘*" or used as query directly and independently. Its probability of occurrence in the large-scale samples is much lower than intention words like "*在线*" and so on.

The position of word in the query used as classification feature comes from the phenomenon when user build query, they are used to put some words with purpose in the head or tail of the query, for example "*在线观看电影黑衣人III*" or "*Facebook的上市时间是几号*". So, the position of word in query is an effect feature. In this paper, the position feature is relative position after normalizing with calculating process as follows. Given a specific query w, $Ord(w)$ represents the order of w in the query and $Seq(Q)$ represents the number of words in the query. So the position feature of w is calculated as formula (1).

$$P_w = Ord(w)/ Seq(Q), \qquad\qquad 公式(1)$$

5   Intention Cluster and Recommend

The purpose of intention cluster is to cluster the query words with consistent or related query motivation, so as to realize the reorganization and recommendation of query description by taking advantage of intention relation between internal clusters.

Intention cluster can be divided into two categories: 1) *Consistency-based* query intention cluster. It aims at cluster the words that are same with user's query intention; 2) *Preference-based* query intention cluster. It aims at cluster the words that meet user preference. The former is similar to the phenomenon which identify the expression of intention in different words but synonymous. The latter focus on mining associated intention words which point to similar goals.

5.1   Intention Vector

Intention set is a collection of intention words. If intention words matched and compared alone without considering whole query, it is difficult to identify consistent or associated intentions from semantic . We can only match the consistency of word type that cannot help reach the core purpose of intention cluster. Therefore, we propose the concept of intention vector to build the associated network of intentions.

The construction of intention vector comes from the rules that query can be conceptualized as the form of "<Target + Intention>". Intention should be combined with the specific target that makes sense. For example, the intention word "*治疗*" is unable to transfer any description of intention in the absence of target. Only together with explicit "*治疗对象*" such as "*胃病*", "*失眠*" and "*产后综合症*", the whole query has a complete semantic. On the basis on the rule, we propose a "skillfully deflected" description of intention vector. The description itself is not the lexical meaning of intention words or such as WordNet hyponymy lexical semantic relations features, but through large-scale query samples to describe the targets indirectly which

high-frequencly co-occurred with intention words. For example, given a query intention word "治疗" or "症状", the form of vector constructed is shown in Table 2.

By building the intention vectors, we convert the correction intention measure into the similarity measure between vectors. The contact with intention is more closely if the intersection of intention vector is larger between different intention words.

Table 2. Intention Vector

Intention Word	Intention Vector
治疗	[癌症　骨刺　神经　胃病　胆固醇　……　手足口病]
症状	[妊娠　胃病　忧郁　癌症　手足口病　……　胃溃疡]

5.2   Intention Similarity Calculation

From the description of intention above, we attempt to use the correlation matching of intention based on VSM and Language Model. And such correlation is used as reference standard of cluster. The VSM[9] calculate with cosine similarity, and the Language Model[10] calculate with Kullback Leibler Divergence[11].

Through calculating similarity of intention with vector space model and measuring similarity of probability distribution between intentions with *KL* divergence, we realize the numerical measure of the correlation between intentions. With this measure, we use Apcluster clustering algorithm to cluster the intention words so as to form intention clusters and provide reference data for query recommendation.

- **Vector Space Model + Cosine Similarity**

Using VSM to describe intention, we need to estimate the numerical weight which gives to each dimensional target in the model. In this paper, we use *TF-IDF* to calculate the weight values. Assuming that intention word I_j co-occurred with n target words and each target presented by O_i, then the intention vector of I_j presents $V(I)=\{I_1, I_2, \ldots I_j, \ldots I_m\}$. All different queries in the query sets can be represented as a vector $V(I)=\{I_1, I_2, \ldots I_j \ldots I_m\}$. The weight values of vector use the idea of *TF-IDF* for reference, which considers the target words to make a contribution to the intention words. As shown in formula (2), $w(O_i, I_j)$ represents the weight values that O_i co-occurs with I_j and $C(O_i, I_j)$ represents the times of O_i co-occurs with I_j in the large-scale query samples. $N(I)$ shows the number of all intention vectors. $N(O_i, I_j)$ represents the number of intention vectors that O_i and I_j co-occurs.

$$w(O_i, I_j) = \frac{C(O_i, I_j)}{\sum_{O_i \in V(I_j)} C(O_i, I_j)} \log \frac{N(I)}{\sum_{I_j \in V(I)} N(O_i, I_j)} \qquad 公式(2)$$

TF value reflects the contribution that target word O_i makes to the intention word I_j. The larger the value is, the stronger the features of target are. But, *IDF* value reflects the contribution that target word O_i makes to the global intention. The larger the value is, the weaker the features of target are. Through *TF-IDF* calculating, we obtain each feature weight of intention vector. For two different query intention words, we can

calculate its cosine similarity. The larger the value is, the more similar the intentions are.

- **Language Model +** *KL* **Divergence**

Besides, we introduce the language model to calculate the probability that intention words generate the targets words. *KL* divergence is used to measure whether the probability of intention vectors distributed is approximate.If two intention vector $V(I_i)=\{O_{i1}, O_{i2}, \ldots , O_{in}\}$ and $V(I_j)=\{O_{j1}, O_{j2}, \ldots ,O_{jm}\}$ is given, their *KL* divergence is calculated with the formula(3). In the equation, $p(o|I_i)$ represents the probability of target word *o* co-occurs with intention word I_i.

$$D(I_i \| I_j) = \sum_{o \in (V(I_i))} p(o \mid I_i) \log \frac{p(o \mid I_i)}{p(o \mid I_j)} \qquad 公式(3)$$

As the zero problem caused by sparse date for language model, we carry on further smoothing after normalizing probability in experiment. When the molecule is zero, we set $0*\log(0/p) = 0$. And when $p(w|I_j)$ is zero, we set $p(o|I_j) = 1/C_j$ which is the reciprocal of C_j. The C_j is the total number of all targets occur in the $V(I_j)$.In order to balance the deviation brought about by each sub-directions, we add two-way values $D(I_i\|I_j) + D(I_j\|I_i)$ as *KL* balanced index. The smaller the value is, the closer the probability distribution of intention vector is and the more similar they are.

5.3   Intention Recommendation

Based on the intention cluster above, we adopt a simple transductive approach for query recombination and recommendation. Given a query, we identify the intention words and targets, mine the targets co-occurred in the global query samples firstly. By building the description of intention vectors, we use VSM and language model to describe intention separately with cosine similarity and *KL* divergence to measure the similarity between intention words and prior intention clusters. The most similar cluster is chosen as candidate recommended set according to the similarity ranking. Finally, we combine the candidate intention words with the query target unorderedly to form new queries and recommend as related queries.

However, not all the intention words in the cluster are suitable for recombination and recommendation. The reason is the similarity measure cannot avoid focusing on the local property of cluster. In other words, the correction can only represent part of the intentions with the query intention under test. This phenomenon is a common problem that most clustering application technologies are difficult to avoid. For details, if clustering cohesion is too strong, the similarity reaches consistently in fact, which contributes to detect the queries with different words but synonymous for recommending synonymous queries. But at the same time, the detection of relevant intention will be ignored. That is, the kind of intentions have relevant preference such as "*Apple Mobile*" and "*Apple Computer*" but not consistent intention. In contrast, if clustering cohesion is weak, then the intention words in cluster must have a more fine-grained division. At the moment, the relevance always comes from one point or

locality of cluster internal, and the rest parts are weak associated or unrelated intention words. As a result, we take the following steps to select the intention words of cluster and use for query recombination and recommendation. 1) Adjust cluster parameters until the number of intention words in cluster reaches 10, and then the cohesion is not strong. Apcluster should iterate over 200 times. 2) Match the most prior cluster and the similarity between query intention under test and each intention word of the cluster, and then to rank according to the similarity. 3) Select the top n intention words of sort for query recombination and recommendation. The value of n will be introduced in experiment part.

6   Experiment Setting and Result Analysis

In this section, we will introduce the experimental corpus and evaluating methods firstly; Then, give the main systems include Target and Intention Recognition System, Intention Cluster System and Query Recommendation System which participate in the test; Finally, we give the test results of each system and the corresponding analysis.

6.1   Corpus

The experimental data is *Sogou2008* query logs which contains user query information within a month. After filtering queries, we obtain *1,902,402* no-repeat informational queries and group the queries with same clicked URLs.

For the classification experiment, we take *968* groups queries for manually label with an average of 5. The queries in the same group must have at least one common click. On this basis, we invite three volunteers to label targets and intentions. These volunteers do research on query recommendation more than one year.

For the intention cluster experiment, *1,981* intention words are selected randomly and clustered. The form of labeling is to compare the intention words in cluster which provided by the automatic intention clustering system to intention words under test by human-labeling (three volunteers + cross-validation).

For query recommendation experiment, we extract *2,000* queries randomly and divide equally to six sample space for query recommendation test. The recommended results of each sample space are tested by a volunteer and taking *macro-average* representation of overall performance.

6.2   Evaluation Method

In the experiment of classification, we use the *precision*, *recall*, *F-value* and *global accuracy* for evaluation. Through statistical, the percentage of target words to intention words in query samples is close to 2:1. The distribution is reasonable because targets words are often more than intention words in the queries. Therefore, *baseline1* is the accuracy of target reaching 66.7%, which assumes that all query words are target words with the recall of targets is 100%. Correspondingly, *baseline2* is the

accuracy of intention reaching 33.3% , which assumes that all query words are intentions for the prior probability with the recall of intention is also 100%.

For the evaluation of intention cluster, we take the proportion of correct intention words in clusters as *precision* and obtain the global performance by taking *macro-average* of all clustering performance. For query recommendation, we use *global precision(G-P), consistent precision(C-P), relevant precision(R-P)* and $P@n(1{\leq}n{\leq}10)$ to evaluate. Consistent precision refers to the recommended queries have the consistent semantics with original query(such as "位置" and "地点"), that is different words but synonymous. Relevant refers to the recommended queries is relevant with original query(such as "诊断" and "治疗"). The consistency and relevance is judged by human. *P@n* refers to the precision of query recombination and recommendation which use the most similar *top-n* intention words in cluster.

6.3   Experiment Results and Analysis

In this section, we give the experimental results of target and intention classification system, intention cluster system and query recommendation system as well as the analysis.

- **The Performance of   Target and Intention Recognition**

Table 3 shows the experimental results of target and intention classification system. *Sys-1* is *lexical-based* approach used only and *Sys-2* uses *Context-base* approach only. *Sys-3* combines two approaches above. We use *NaiveBayes* classifier and 10 *cross-validations* of the data set. Results display the similarity of global accuracy between *sys-1*and *sys-2*. But combined two kinds of features, the results improve about 3%~4%. The reason benefits from the advantage of indentifying intention words based on lexical meaning and target words based on context property.

Table 3. Targets and Intention Identify Results

System	Targets			Intention			Gobal Accuracy
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)
Sys-1	77.8	74.7	76.2	55.7	59.8	57.7	69.55
Sys-2	69.0	95.2	80.0	68.0	19.2	30.0	68.86
Sys-3	77.2	83.4	80.2	63.2	53.7	58.1	73.11
Baseline1	66.7	100	80.0	0	0	0	66.7
Baseline2	0	0	0	33.3	100	49.9	33.3

Besides, the recall of target words in *sys-2* is highest reaching 95.2%. Correspondingly, its recall of intention words is lowest only 19.2%. Because of the *context-based* approach have poor effect on identifying the impact intention words. For example, for query "北京电信", the target word is "电信" and intention word is "北京". Its overall meaning is to find the telecom located in Beijing but not the telecom in other place. Here, "北京" is an implicit intention word different from the explicit

intention words such as "下载", "浏览" and "报价". The frequency of implicit words is low and the position in the query is limited that makes no obvious distinction with target words. So using the feature alone will cause the misjudgment of implicit intention words. Especially, the numbers and frequency of implicit intention words are more than the explicit intention words. As a result, *sys-2* loses a lot of intention words because of the misjudgment. In contrast, *sys-1* can improve the recall of intention words but not lose the recall of target. So the lexical-based approach is effect for intention identification and part of speech and morphology is the most effective feature.

Considering the distribution of target and intention in real, the measure of comprehensive system performance should take F-value and global accuracy into consideration. As shown in table 3, *sys-3* has a more performance than above two. Moreover, the classifier mixed with features performs better than baseline1 and baseline2. Especially, the ability of identifying the intentions has improved 9.8 percent while that of target improve litter. As the target words distribute high is queries priori, the performance of baseline itself is already high. So the ability of identifying target words of *sys-3* has reached the condition for further treatment.

- **The Performance of Intention Cluster**

We construct *1,981* intention vectors to realize two kinds of intention cluster systems. The Apcluster algorithm was chosen because of the advantage without setting the numbers of clustering categories in advance. *Sys-VSM* refers to use Vector Space Model and *Sys-KL* uses Language Model to calculate KL Divergence. The experimental results of query intention cluster are verified by three volunteers. As shown in table 4, #1#2#3 represents three experimental samples. Obviously, the performance of *Sys-VSM* is better than *Sys-KL*.

Table 4. Intention Cluster Results

Volunteer	Sys-VSM			Sys-KL		
	P(%)	P(%)	P(%)	P(%)	P(%)	P(%)
	#1	#2	#3	#1	#2	#3
VolunteerA	48.63	57.98	51.11	24.05	37.08	32.52
VolunteerB	45.28	53.83	54.52	27.67	35.42	37.34
VolunteerC	44.18	55.19	49.95	28.32	34.19	35.55
Average	**46.03**	**55.67**	**51.86**	**26.68**	**35.56**	**35.14**

From the data of table 4, we can find the results of manual judgment swing greatly, and the most optimal performance is 55.67%. It demonstrates that the diversity of intentions have brought uncertainties to volunteers. At the same time, the *VSM-based* approach has smaller granularity of clustering so that the ability of distinguish the clustering intention clusters is stronger. But the global precision of *KL* Divergence based on language model is lower. And at the same experimental samples, the number of *KL* cluster is less than the VSM approach. It indicates the loose of semantic relation between queries and a more diverse form of combination between queries.

Besides, the results illustrate that not all intention words of cluster can be recommended equally, but ranking according to the similarity of intentions and recommend the prior.

- **The Performance of Query Recommendation**

The performance of query recombination and recommendation according to similarity is illustrated in Fig.2, whose horizontal axis P@n represents top n(1≤n≤10) intention words and vertical axis represents mean precision(average of all P@n). The label "Eva$i$" （1≤i≤6） indicates different volunteers' accessments. As shown in the graph, recommendation precision goes down with n increases, and it reaches overall highest at P@1,which is 57.8% on average. We attribute it to the recommendation mechanism, which always chose the center-closest intention word first. And these words are always most representative for the cluster, which covers a wide range of meanings and is easily accepted by users. For example, an intention cluster contains words as follows: Example3: "看病(see a doctor)", "挂号(register)", "诊断(diagnose)", "开药(prescribe)" and "住院(hospitalization)". (cluster center is "看病(see a doctor)").
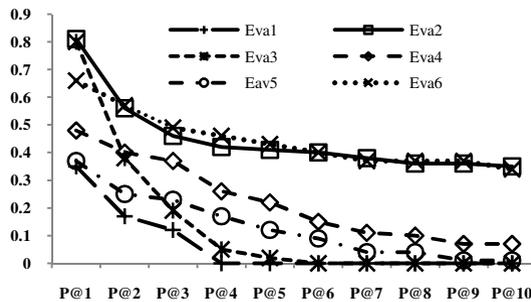

Fig. 2. Recommand System Performance

The cluster center "*see a doctor*" is most representative, thus has certain relations to all other words constitute queries, like "*how to register*", "*who to diagnose*" or "*where for hospitalization*". And it is reasonable to be a decent recommendation of highest probability.

However, Precisions drop quickly after *P@4*, which is caused by the relations between words outside the center, which were not absolutely consistent or related. As words outside center are related to the center in different aspects, and the candidate is only related to the center and a certain aspect, the results turn out to be other aspects are of little consistency. For example, a candidate intention is related to "*hospitalization*" in the above cluster, then "*where to register*", which is constituted by "*register*", is not a good recommendation for "*where for hospitalization*". So, with the increasing n, words far from center and related aspects have higher probability of recommendation, which leads to lower precision.

Apart from that, table 5 lists the average precision of *P@n*(1≤n≤5) in six sample spaces. The overall distribution is not gaussian distribution, which shows unbalance

on two ends instead. The reason is, the different effects caused by different kinds of recommendations. Recommendations with high performance are words with obvious intentions, like "download", while those with low performance are words modifying the targets, like place names. Simple substitution of the latter ones may cause distortions of user intentions. Thus, a classification of queries according to user emphasis is needed to improve the system's performance.

<div style="display: flex; gap: 2em;">

Table 5. P@N Values

	P@1	P@2	P@3	P@4	P@5
Eva1	0.35	0.17	0.12	0	0
Eva2	0.81	0.56	0.46	0.42	0.41
Eva3	0.80	0.38	0.19	0.05	0.02
Eva4	0.48	0.4	0.37	0.26	0.22
Eva5	0.37	0.25	0.23	0.17	0.12
Eva6	0.66	0.57	0.49	0.46	0.43

Table 6. Average Recommend Accuracy

	G-P	C-P	R-P
Eva1	0.028	0.011	0.018
Eva2	0.336	0.013	0.324
Eva3	0.065	0.015	0.050
Eva4	0.123	0.113	0.010
Eva5	0.065	0.016	0.048
Eva6	0.331	0.004	0.327

</div>

Average precision of the system is shown in table 6. It shows a low overall precision, which is caused by not only the influences of different relations with cluster center (already discussed above), but also the irrational combination of target and intention words. Simple substitution of intention words in our method suffers from the problem of unsmooth semantics and unreasonable logics, which affects about 40% of all the intention words. For example, intention words "precious" and "expensive" are related, but when combined with target word "air", the situation is different. It is easy to understand "precious air", but not for "expensive air". Thus, the proper combination of intention words and target words are of vital importance to effective query generation, and is an essential part of our future work.

7   Conclusions

Our paper proposed a new query recommendation which concentrates on the analysis and application of query itself. It recognizes the target words and intention words by means of classification to describe their semantics based on massive query samples. After that, synonymous or related intention words are obtained by clustering for recommendation. Experiments show that our method can divided target and intention word effectively, and it gets *55.67%* performance. *P@1* reached *57.83%* respectively.

However, there is still a large room for improvement. The experimental results show that the intention clustering performance influences queries recommendation a lot, and the clustering performance depends on the feature selection and matching algorithms. Therefore, future work will focus on feature selection method to enhance the intention description method and try to use semantic intention sample matching algorithm. For example, dividing the entity roles in the intention description, and employed into the similarity matching process. In addition, future work will also

analyze the combination of target words and intention words, in order to form a fluent and logical query.

References

1. Li, Y., Wang, B., Li, J.: A survery of Query Suggestion in Search Engine. Journal of Chinese Information Processing 24(6), 75–84 (2010)
2. Craswell, N., Szummer, M.: Random Walks on the Click Graph. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), Amsterdam, The Netherlands, pp. 239–246 (2007)
3. Cao, H., Jiang, D., Pei, J., et al.: Context-Aware Query Suggestion by Mining Click-Through and Session Data. In: Proceedings of the 14th ACM SIGKDD, pp. 875–883. ACM, New York (2008)
4. Huang, C.K., Chien, L.F., Oyang, Y.J.: Relevant term suggestion in interactive web search based on contextual information in query session logs. Journal of the American Society for Information Science and Technology 54(7), 638–649 (2003)
5. Wang, J., Peng, B.: User Behavior Analysis for a large-scale Search Engine. Journal of The China Society For Scientific and Technical Information 25(2), 154–162 (2006)
6. Li, Y., Xu, S., Wang, B.: Chinese Query Recommendation by Weighted SimRank. Journal of Chinese Information Processing 24(3), 3–10 (2010)
7. Mei, Q., Zhou, D., Church, K.: Query Suggestion Using Hitting Time. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008), Napa Valley, California, pp. 469–478 (October 2008)
8. Broder, A.: A taxonomy of web search. SIGIR Forum, 3–10 (2002)
9. Wong, S.K.M., Ziarko, W., Wong Patrick, C.N.: Generalized Vector Spaces Model in Information Retrieval. In: Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA, pp. 18–25 (1985)
10. Song, F., Croft, W.B.: A General Language Model for Information Retrieval. In: Proceedings of the Eighth International Conference on Information and Knowledge Management, New York, USA, pp. 316–321 (1999)
11. Lavrenko, V., Allan, J., Deguzman, E., et al.: Proceedings of the Second International Conference on Human Language Technology Research, San Francisco, USA, pp. 115–121 (2002)