Optimization of a Search Engine for an Organized and Effective Browsing

[1]K.Kalaivani, [2]K.Lavanya, [3]Dr.S.Uma

Abstract- In web search applications, queries are submitted to search engines to represent the information needs of users. Discovering the number of diverse user search goals for a query and depicting each goal with some keywords automatically. In the existing work propose a novel approach to infer user search goals by analyzing search engine query logs. First propose a novel approach to infer user search goals for a query by clustering our proposed feedback sessions. Second we propose a novel optimization method to map feedback sessions to pseudo-documents which can efficiently reflect user information needs. In the end, we cluster these pseudo documents to infer user search goals and depict them with some keywords. In proposed system k means clustering algorithm is computationally difficult, in order to overcome the k means clustering problem, enhancement a Fuzzy c-means clustering (FCM) algorithm to group the pseudo documents and it also measure the similarity between the pseudo terms in the documents, it improves the feedback sessions results than the normal pseudo documents. The FCM algorithm divides pseudo documents data for dissimilar size cluster by using fuzzy systems. FCM choosing cluster size and central point depend on fuzzy model. The FCM clustering algorithm it congregate quickly to a local optimum or grouping of the pseudo documents in well-organized wayAnnotation of the search results of the database we proposed a annotation based results. We used Automatic annotation approach that first aligns the data units on a result page into different groups such that the data in the same group have the same semantic results. Finally measures the clustering results classified average precision (CAP) to evaluate the performance of the restructured web search results.

Keywords-pseudo-documents, k means clustering, Fuzzy c-means clustering (FCM), annotation,classified average precision (CAP)

1. Introduction

In web search based applications user enters the query in the website to search the efficient information. The needs of the information may differ from each user and goal to achieve the user need are still becomes difficult. Because the user given queries may not understandable by system or it becomes less sometimes queries may not exactly represented by users. To achieve the user specific information needs many uncertain queries may cover a broad topic and dissimilar users may want to get information on different point of view when they submit the same query. User information need is to desire and obtain the information to satisfy the needs of each user. To satisfy the user information needs by considering the search goals with user given query. We cluster the user information needs with different search goal .Because the interference and examination of user search goals with query might have a numeral of advantages by improving the search engine significance and user knowledge. So it is necessary to collect the different user goal and retrieve the efficient information on different aspects of a query. Capture different user search goals in information retrieval outcome becomes changes than the normal query based information retrieval.

Major advantages from this search engine based query results are the first restructure web pages result according to user search goals. X. Wang and C.-X Zhai et.al [1] presented a user boundary that organizes Web search consequences into hierarchical category. Text classification algorithms are used to repeatedly classify subjective search consequences into an obtainable group organization. A user study compared our new category interface with the typical ranked list interface of search results. Here the author group user search goal with similar results

Hua-Jun Zeng et.al [2] suggested a query based search results for user goal and the rank list of documents return by a certain Web search engine, it first extracts and ranks most important phrases as candidate cluster names, base on a regression model learned beginning human labeled training data. Assigned the documents to applicable most important phrases to form candidate clusters, and the final cluster are generated by assimilation these candidate clusters. But this method only produces the result with higher level of the documents only and it doesn't make the results for all search based user goals.

Clustering search results is an efficient method to systematize investigates results, which allows a user to find the way into applicable documents quickly. Wang and Zhai [1] learning "interesting aspects" of a topic beginning Web search logs and organize search results therefore and generate further significant cluster labels using history query words entered by users. However, this method has limits because the numeral of dissimilar clicked URLs of a query may be small. To discover the user information automatically at different point of view with user given query and collects the similar search goal result with URL first we collect similar feedbacks sessions.

Reorganize the web structure environment based on the link in the URL clicked and unclicked by the URL. First the infer user search goals for a query by clustering the similar documents from the web search. Then, map feedback sessions from the pseudo documents to collect the similar pages of links to satisfy user goals and retrieve the user information. The K means clustering algorithm can be used to cluster the similar pseudo documents and group them according to the search goal. In the proposed system we used the Fuzzy c-means (FCM) is a method of clustering which allows one piece of documents data to belong to two or more clusters.

Represent the URL using Feedback session that includes the URLs, it consists of the clicked URL and Unclicked URL links. Usually language because users will scan the URLs single by single from top to down, we can believe that in addition the three clicked URLs, four unclicked ones in the rectangular box contain also be browsed and evaluate by the user and they be supposed to rationally be a division of the user feedback. Within the feedback session, the clicked URLs tell what user necessitates and the unclicked URLs reflect what user does not care about. It is supposed to be well-known that the unclicked URLs after the last clicked URL should not be included into the feedback sessions since it is not certain whether they were scanned or not. Each feedback session can tell what a user requires and what he/she does not care about. Furthermore, there is abundance of miscellaneous feedback sessions in user click-through logs. Consequently, for inferring user search goals, it is well-organized to analyze the feedback sessions than to observe the search results or clicked URLs straight.

Main contributions of this work are,

1. First collect the different user query based results in the web based search engine and then generated the different feedback session are collected infer user search goals for a query by clustering the similar documents from the web search.
2. After the collection of feedback session introduce a method to collect the similar pages of links to satisfy user goals and retrieve the user information.
3. Then measuring the semantic information makes result better than the normal keywords information. Measuring the similarity between the pseudo terms here used a cosine similarity measure.
4. After that similarity measure again group the pseudo documents using FCM clustering algorithm. After clustering the search result, we annotating the search result.
5. Finally compare the results with parameters classified average precision (CAP), Average Precision (AP), Voted AP(VAP) and risk to evaluate the performance of the restructured web search results.

2. Related works

The problem of clustering investigate results has been investigate in a numeral of previous works. All of the previous work apply clustering algorithms which first group documents into similar groups according to content similarity, and produce expressive summary for clusters. Though, these summaries are often illegible which construct it difficult for Web users to recognize relevant clusters.

Zamir and Etzioni [3] introduced a Suffix Tree Clustering (STC) which first identifies sets of documents that split general phrases, and following that create clusters according to these phrases. The extraction procedure of applicant phrase is similar to STC but we supplementary calculate a number of significant properties to identify salient phrases, and make use of learning methods to rank these salient phrases. Some topic finding or text trend analysis mechanism is also related to our method. The dissimilarity is that we are specified titles and short snippets somewhat than whole documents. For the meantime, we train regression model for the ranking of cluster which is closely related to the efficiency of users' browsing.

Web search engines challenge to satisfy users' information needs by standing web pages with reverence to queries. But the realism of web search is that it is frequently a procedure of querying, learning, and reformulating. A sequence of interactions among user and search engine can be essential to satisfy a solitary information need [4].

Though users query search engines in order to achieve tasks at a diversity of granularities, issue numerous query as they effort to accomplish tasks. R. Jones and K.L. Klinkner [5] learning real sessions manually labelled into hierarchical tasks, and demonstrate that timeouts, anything their length, are of incomplete utility in identifying task boundaries, achieving a greatest precision. Though, their method only identifies whether a pair of queries belongs to the same goal or mission and does not mind what the goal is in aspect.

U. Lee, Z. Liu, and J. Cho [6] study the "goal" at the back based on a user's Web query, in order that this goal can be used to get better the excellence of a investigate engine's results. Preceding studies encompass mainly focused on manual query-log investigation to recognize Web query goals. Identify the user goal automatically with no any explicit feedback from the user. User search goals represented by a number of keywords can be utilized in query suggestion [7], [8], [9]; thus, the suggested queries can assist user to form their query more accurately.

A previous exploitation of user click-through logs is to get user implicit feedback to expand training data when knowledge ranking functions in information retrieval. Adapt a recovery system to challenging groups of users and exacting collections of documents promise further improvement in retrieval quality for at least two reasons. Since physically adapting retrieval function is instance consuming or even not practical, investigate on automatic adaptation by means of machine learning is in receipt of a great deal notice.

T. Joachims [10] explore and evaluate strategies for how to mechanically produce training example for learning retrieval functions from experiential user behavior. Yet, implicit feedback is more hard to interpret and potentially noisy. First examine which types of implicit feedback can be dependably extracted from experiential user behavior, in particular clickthrough data in WWW search. To assess the reliability of implicit feedback signals, we conduct a user study. The learn is intended to examine how users interrelate with the list of ranked consequences from the Google search engine and how their behavior can be interpret as significance judgments.

Thorsten Joachims did lots of works on how to use implicit feedback to get better the retrieval quality [11], [12]. In our effort we believe feedback sessions as user implicit feedback and suggest a novel optimization method to merge both clicked and unclicked URLs in feedback sessions to discover what users really necessitate and what they do not mind. One submission of user search goals is restructuring web examine results. There are also some related works focuses on organize the search results [13], [1], [2]. In this work we infer user search goals beginning user click-through logs and reorganize the search results according to the inferred user search goals and then finally measure the results.

3. Existing system

In the overview, the main goal of this system is discovering the number of diverse user search goals for a query and depicting each goal with some keywords automatically. Here first propose a novel approach to infer user search goals for a query by clustering the proposed feedback sessions. It consists of both clicked and unclicked URLs and ends with the last URL that be clicked in a single session. Then, propose a novel optimization method to map feedback sessions to pseudo-documents which can efficiently reflect user information needs. At last, here cluster these pseudo documents to infer user search goals and depict them with some keywords. Given that the evaluation of clustering is also an important problem, also propose a novel evaluation criterion classified average precision (CAP) to evaluate the performance of the restructured web search results.

Feedback session

Feedback session is a session for web search is a sequence of consecutive queries to satisfy a single information require and some clicked investigate results focal point on inferring user search goals for a exacting query. Consequently the single session contain simply one query is introduce, which distinguish from the conservative session. For the moment, the feedback session is based on a solitary session, though it can be comprehensive to the entire session. It consists of both clicked and unclicked URLs and ends with the last URL that be clicked in a single session. It is forced that previous to the last click, all the URLs have been scanned and evaluate by users. Each feedback session can tell what a user requires and what he/she does not care about. Additionally, there are plenty of diverse feedback sessions in user click-through logs. Consequently, for inferring user search goals, it is additional efficient to examine the feedback sessions than to examine the investigate consequences or clicked URLs in a straight line. To represent the feedback session efficiently some demonstration methods needed, because each and every user based search goal feedback sessions are differs and their corresponding log files also changed.

Represent a feedback session to Pseudo-Documents with Binary vector technique to characterize a feedback session search consequences are the URLs return by the search engine when the question "the sun" is submits, and "unclicked" is defined as "0" in the click sequence. The binary vector [0110001] can be second-hand to symbolize the feedback session, where "1" represent "clicked" and "0" represents "unclicked.

Building a pseudo documents

In the primary step, we primary augment the URLs with extra textual inside by extracting the titles and snippets of the returned URLs appear in the feedback session. Each URL in a feedback session is representing by a little text subsection that include of its title and snippet. Then, a number of textual processes are implementing to persons text paragraphs, for instance transforming all the letters to lowercases, stemming and remove stop words. At the end, every URL's title and snippet are generated by a Term Frequency-Inverse Document Frequency (TF-IDF) vector, consequently

$$T_{u_i} = \{T_{W_1}, T_{W_2}, \dots \dots, T_{W_n}\}^T \longrightarrow (1)$$

$$S_{u_i} = \{S_{W_1}, S_{W_2}, \dots \dots, S_{W_n}\}^T \longrightarrow (2)$$

Where

T_{u_i}- TF-IDF vectors of the URL's title

$S_{u_i}$are the TF-IDF vectors of the URL's snippet .

u_i-i^{th}URL in the feedback session.

W_j={1; 2; . . . ; n} -j^{th} term appear in the enriched URLs. Each term in the URL is defined as a word or a numeral in the vocabulary of document collections. t_{wj} and s_{wj}characterize the TF-IDF significance of the j^{th} term in the URL's title and snippet, correspondingly. Taking into consideration that URLs' titles and snippets have dissimilar significances, we symbolize the enriched URL by the weighted sum of $T_{ui}$and S_{ui} , namely,

$$F_{u_i} = T_{u_i}\omega_t + S_{u_i}\omega_s = \{f_{W_1}, f_{W_2}, \dots\dots, f_{W_n}\}^T \rightarrow (3)$$

Where F_{ui} means the feature representation of the i^{th}URL in the feedback session, $\omega_t$is weight of the titles and ω_s is the snippets, respectively. In order to obtain the feature demonstration of a feedback session, suggest an optimization method to merge both clicked and unclicked URLs in the feedback session. Attain such a $F_{f_s}$with the purpose of the calculation of the distance between$F_{f_s}$and each $F_{uc_m}$is minimize and the sum of the distance between F_{f_s} and each F_{uc_l} is maximize. On the basis of supposition that the terms in the vectors are self-governing, perform optimization on each dimension separately,

$$F_{f_s} = \left[f_{f_s}(\omega_1), \dots\dots.. f_{f_s}(\omega_n)\right)]^T \longrightarrow (4)$$

Infer user search goals and represent them with a number of significant keywords. Then the similarity between the pseudo documents is evaluated as the cosine similarity score

$$Sim_{\ i,j} = \cos\left(f_{f_{s_i}}, f_{f_{s_j}}\right) = \frac{f_{f_{s_i}} f_{f_{s_j}}}{|\, f_{f_{s_i}}||f_{f_{s_j}}|} \longrightarrow (5)$$

$$Dis_{\ i,j} = 1 - Sim_{\ i,j} \longrightarrow (6)$$

Cluster pseudo-documents with K means clustering technique

In this investigate we cluster pseudo-documents by K-means clustering which is straightforward and efficient. Because we not recognizable with the precise figure of user search goal for every query, we position K to be five different values.

$$F_{center_i} = \frac{\sum_{k=1}^{C_i} F_{f_{S_k}}}{C_i}, \left(F_{f_{S_k}} \subset Cluster\ i\right) \longrightarrow (7)$$

where F_{center_i}- i^{th}cluster's center and C_i is the numeral of the pseudo-documents in the i^{th}cluster. $F_{center_i}$is utilize to finish the investigate goal of the i^{th} cluster. Finally, the conditions with the highest values in the $F_{center_i}$are second-hand as the keywords to represent user search goals, it is a keyword based explanation is that the extracted keywords be able to in addition be utilized to form a more significant query in query suggestion and thus can represent user information needs most effectively.

4. Proposed system

To improve the clustering result and to improve the performance of the system, in the proposed system we used the effective clustering technique with annotation search result.

Cluster pseudo-documents with FCM clustering technique

Fuzzy c-means (FCM) is a technique of clustering which allows one piece of pseudo documents data to belong to two or more clusters. This clustering process is used to solve how the data with similar pseudo documents are clustered according to cosine similarity of the pseudo terms in the documents. In this algorithm the same given data or pseudo documents does not go completely to a well definite cluster, based on the fuzzy membership function only the pseudo documents the cluster groups are formed in efficient manner with possible number of the groups at user feedback sessions. In the FCM approach, instead, the same given datum does not belong exclusively to a well defined cluster, but it can be placed in a focal point way. In this case, the membership function follows a flatter line to designate that each datum may go to frequent clusters with different standards of the membership constant. In fuzzy clustering, each position has a degree of belong to clusters, as in fuzzy logic, rather than belong totally to just one cluster. Thus, points on the edge of a cluster might be *in the cluster* to a smaller degree than points in the midpoint of cluster. It is based on selection of the degree membership function,

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\,||x_i\text{-}x_j||^2 \qquad 1 \leq m \leq \infty$$

where m is any real numeral greater than 1, u_{ij} is the degree of membership of x_i documents based data in the cluster j, x_i isthe $i^{th}$of measured data, C_j is the numeral of the pseudo-documents in the j^{th}cluster and ||*|| is any norm expressing the similarity between any measured data of the pseudo documents and the centre. Fuzzy separating is carried out concluded iterative optimizations of the objective function with the modernize of membership u_{ij} and the cluster centresc_j by:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left\{\frac{||x_i-c_i||}{||x_i-c_k||}\right\}^{\frac{2}{m-1}}}$$

This iteration will stop,whenmax$_{ij}\{|\mu_{ij}^{(k+1)} - \mu_{ij}^{(k)}|\} < s$, where s is a termination criterion between 0 and 1, whereas k is the repetition steps. This technique converges to a local smallest or a saddle point of J_m.

The algorithm steps are described as follows:

1) Initialize U=[U_{ij}]matrix, $U^{(0)}$ pseudo documents

2) At each and every position K calculate the midpoint vectors of each pseudo documents

$C^{[K]}$ =[C_j] with $U^{(k)}$

$$C_j = \frac{\sum_{i=1}^{N}\mu_{ij}^{m} - x_i}{\sum_{i=1}^{N}\mu_{ij}^{m}}$$

3) Update $U^{(k)}$, $U^{(k+1)}$ the pseudo documents datapoints with membership function

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left\{\frac{||x_i-c_i||}{||x_i-c_k||}\right\}^{\frac{2}{m-1}}}$$

4) if || $U^{(k+1)}$ - $U^{(k)}$|| <∈ then stop .It satisfies the condition then group the pseudo documents. Otherwise else to step 2 and again find the best pseudo documents data with user search goal.

Annotation of results

Data alignment is to put the data units of the same concept into one group so that they can be annotated entirely. Whether two data units belong to the same concept is determined by how similar they are based on the features

(i) Presentation style similarity (SimP): It is the average of the style feature scores (FS) over all six presentation style features (F) between d_1 and d_2

$$SimP(d_1, d_2) = \sum_{i=1}^{6} FS_i \,/6$$

(ii) Data type similarity (SimD): It is determined by the common sequence of the component data types between two data units. The longest common sequence (LCS) cannot be longer than the number of component data types in these two data units. Accordingly, let t_1 and t_2 be the sequences of the data types of d_1 and d_2, correspondingly, and TLen(t) represent the number of component types of data type t, the $SimD$ i.e., data type similarity between data units d_1 and d_2 is

$$SimD(d_1, d_2) = \frac{LCS(t_1, t_2)}{Max(Tlen(t_1), Tlen(t_2))}$$

(iii) Adjacency similarity (SimA): The adjacency similarity between two data units d_1 and d_2 is the average of the similarity between d_p^1 and d_p^2 and the similarity between d_s^1 and d_s^2 that is

$$SimA(d_1, d_2) = (Sim'(d_1^p, d_2^p) + Sim'(d_1^S, d_2^S))/2$$

In a returned result page containing multiple search result records (SRRs), in the result the data units corresponding to the same concept (attribute) often share special common features. These common features are usually associated with the data units on the result page in certain patterns. In this work, we present an automatic annotation approach that first aligns the data units on a result page into different groups such that the data in the same group have the same semantic. After that, for each group we annotate it from different aspects and aggregate the different annotations to predict a final annotation label for it.

5. Experimentation results

Before conclusion of the results and remarks of the paper the major part is the evaluation of the results from the experiments with classification results from each user search goal inference us a major problem , since user search goals are not predetermined and there is no ground truth. It is necessary to develop a metric to evaluate the performance of user search goal inference objectively. In this section finally measure the performance of the FCM with annotation and accessible pseudo documents based clustering Measure the performance of the system with parameters like Classified Average Precision (CAP), Voted AP (VAP) which is the AP of the class including more clicks namely, risk to avoid classifying search results and average precision (AP).The corresponding AP,VAP,CAP and Risk values are measured Between user search Goal using K-Means and User search Goal using proposed FCM with annotation are shown in Figure 1,2,3 and 4. It shows that the User search goal usingFCM with annotation results are better than User search goal using K-means clustering approach.

Average precision (AP)

In order to be appropriate the assessment method to large-scale data, the solitary sessions in user click-through logs are second-hand to reduce physical work. Since beginning user click-through logs, we can get disguised significance feedbacks, specifically “clicked” means applicable and “unclicked” means inappropriate. A probable evaluation principle is the average precision (AP) which evaluate according to user implicit feedbacks. The average of precisions compute at the position of each applicable document in the ranked sequence is called as AP.

$$AP = \frac{1}{N^+}\sum_{r=1}^{N} rel(r)\frac{R_r}{r}$$

whereN^+ is the numeral of applicable (or clicked) documents in the retrieved ones, r is the rank, N is the total numeral of retrieved documents, rel() is a binary function on the relevance of a given rank, and R_r is the number of relevant retrieved documents of rank r or less.

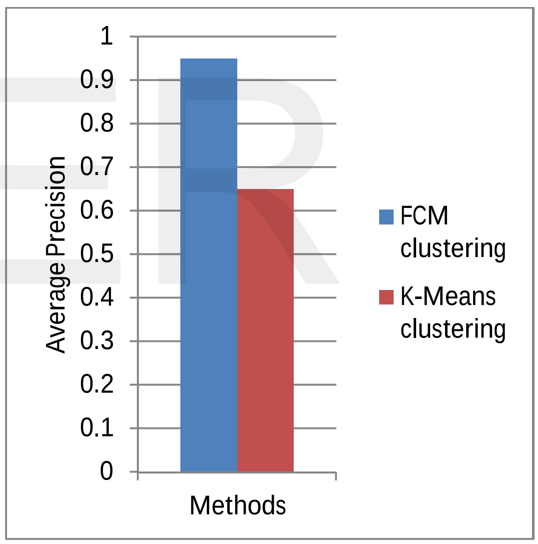

Figure 1: Average Precision (AP) comparison

Voted AP (VAP)

VAP of the modernized search result the AP of class 1, It is defined as ,

$$VAP = \frac{1}{NC}\sum_{r=1}^{NC} rel(r)\frac{R_r}{r}$$

Where, N means total numeral of retrieved documents with class label one ,rel() is a binary function on the relevance of a given rank, and R_r is the number of relevant retrieved documents of rank r or less.

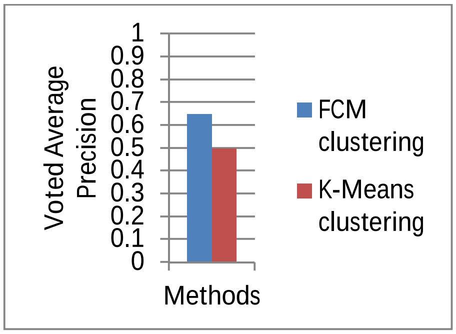

Figure 2: Voted Average Precision (AP) comparison

Classified Average Precision (CAP)

By introducing the above Risk Extend VAP and propose a new criterion Classified AP(CAP)

$$CAP = VAP * (1 - risk)^{\gamma}$$

Where γ is used to adjust the influence of Risk on CAP. CAP select the AP of the class with the aim of user is interested with the most clicks/votes and takes the risk of wrong classification into account.

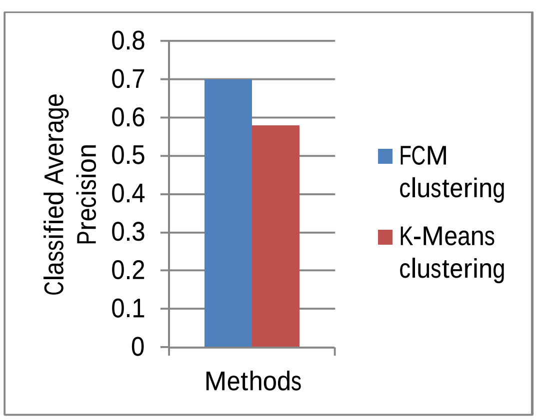

Figure 3: Classified Average Precision (AP) comparison

Risk

VAP is still an unsatisfactory criterion. Taking into consideration an extreme case, if every URL in the click session is categorized into one class, VAP will forever be the highest value that is 1 no matter whether user contain so many investigate goals or not. Consequently present be supposed to be a risk to avoid classify exploration results into too many classes by error. The risk proposed as follows:

$$Risk = \frac{\sum_{i,j=1(i<j)}^{m} d_{ij}}{C_m^2}$$

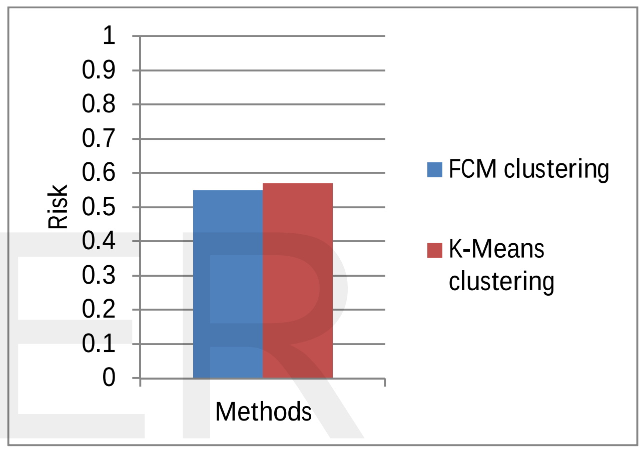

Figure 4: Risk comparison

6.Conclusion

In this paper FCM approach with annotation has been proposed to infer user search goals for a query by clustering its feedback sessions represented by pseudo-documents. Primarily we introduce feedback sessions to be analyzed to infer user search goals rather than search results or clicked URLs. Next, we map feedback sessions to pseudo documents to approximate goal texts in user minds with similarity based measures and then pseudo-documents can supplement the URLs with additional textual contents including the titles and snippets. It cluster the different pseudo documents of the user search goals with feedback session and those pseudo-documents, user search goals can then be exposed and depict with a number of keywords. Finally FCM clustering approach with annotation of the result and Cosine similarity based K means approach were measured the presentation on new criterion CAP, AP, VAP and Risk is formulate of user search goal inference. Experimental results on client click-through logs from a commercial search engine reveal the efficiency of our proposed methods.

References

[1] X. Wang and C.-X Zhai, "Learn from Web Search Logs to Organize Search Results," *Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '07),* pp. 87-94, 2007

[2] H.-J Zeng, Q.-C He, Z. Chen, W.-Y Ma, and J. Ma, "Learning to Cluster Web Search Results," *Proc. 27th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '04)*, pp. 210-217, 2004.

[3] Zamir O., Etzioni O. Grouper: A Dynamic Clustering Interface to Web Search Results. *In Proceedings of the Eighth International World Wide Web Conference (WWW8),* Toronto, Canada, May 1999.

[4] Spink, B. J. Jansen, and H. C. Ozmultu." Use of query reformulation and relevance feedback by Excite users" *Internet Research: Electronic Networking Applications and Policy,* 10(4):317–328, 2000.

[5] R. Jones and K.L. Klinkner, "Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs," *Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08),* pp. 699-708, 2008.

[6] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," *Proc. 14th Int'l Conf. World Wide Web (WWW '05),* pp. 391-400, 2005.

[7] R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query Recommendation Using Query Logs in Search Engines," *Proc. Int'l Conf. Current Trends in Database Technology (EDBT '04)*, pp. 588-596, 2004.

[8] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-Aware Query Suggestion by Mining Click-Through," *Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '08),* pp. 875-883, 2008.

[9] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs," *J. Am. Soc. for Information Science and Technology,* vol. 54, no. 7, pp. 638-649, 2003.

[10] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay, "Accurately Interpreting Clickthrough Data as Implicit Feedback," *Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05),* pp. 154-161, 2005.

[11] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," *Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds.*, pp. 79-96, Physica/Springer Verlag, 2003.

[12] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. Eighth ACM SIGKDD *Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02),* pp. 133-142, 2002.

[13] H. Chen and S. Dumais, "Bringing Order to the Web: Automatically Categorizing Search Results," *Proc. SIGCHI Conf. Human Factors in Computing Systems (SIGCHI '00)*, pp. 145-152, 2000.

Authors

[1]K.Kalaivani, PG Scholar, Department of Computer Science and Engineering, Hindusthan Institute of Technology, Coimbatotr, Tamilnadu, India.

[2]K.Lavanya, PG Scholar, Department of Computer Science and Engineering, Hindusthan Institute of Technology, Coimbatotr, Tamilnadu, India.

[3]Dr.S.Uma, Professor and Head, PG Department of Computer Science and Engineering, Hindusthan Institute of Technology, Coimbatore, India.