International Journal of Smart Home
Vol.8, No.5 (2014), pp.25-32
http://dx.doi.org/10.14257/ijsh.2014.8.5.03

Automatic Deep Web Query Results User Satisfaction Evaluation with Click-through Data Analysis

Zhen Liu[1,2,3], Yong Feng[1, 2] and Huijuan Wang[1, 2]

[1] *College of Computer Science, Chongqing University, Chongqing 400030, China*
[2] *Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education, Chongqing University, Chongqing 400030, China*
[3] *Chongqing Medicine Exchange, Chongqing 400010, China*

Abstract

We browse through hundreds of Deep web pages everyday to find information of interest. We feel happy when Deep web browsing operations provide us with necessary information; otherwise, we feel bitter. Now, the measurement of this user satisfaction has become a hot research topic. In this paper, we propose a click-through-data-based and unsupervised user satisfaction evaluation system, CNEITE, to evaluate the user satisfaction of Deep Web query result pages. It applies query type classifying, navigational query evaluating, informational/transactional query evaluating to solve the challenging tasks. We evaluated our CNEITE system on the AOL data sets, experimental results show that CNEITE achieves higher classify precision than a widely used classify method , Dtree, and higher annotate answer accuracy than method proposed in [17].

Keywords*: Automatic user satisfaction evaluation, Deep Web, click-through data analysis*

1. Introduction

With the advent of information technology, a user is able to obtain relevant information from the World Wide Web which contains a huge amount of information simply and quickly by entering search queries. In response to the queries, the database servers generate the information and deliver it directly to the user. And now more and more information generated from the Deep Web.

Deep Web refers to Web data sources that provide a considerable amount of information with backend databases that are not indexed by general search engines [1]. A survey [2] published in 2004 estimated that there were 450,000 Deep Web data-sources.

We browse through hundreds of Deep web pages everyday to find information of interest. We feel happy when Deep web browsing operations provide us with necessary information; otherwise, we feel bitter. Now, the measurement of this user satisfaction has become a hot research topic. In the past few years, many approaches for measuring user satisfaction for E-commerce and special Websites have been reported in the literature [3-9]. In this paper, we propose a click-through-data-based and unsupervised user satisfaction evaluation system, CNEITE (Query type Classification, Navigational query evaluation, Informational/transactional query evaluation), to evaluate the user satisfaction of Deep Web query result pages.

2. CNEITE Framework

Definition 1(User Satisfaction): In an early attempt to define "user satisfaction" as a concept, Tessier, Crouch, and Atherton (1977) stated that satisfaction was "ultimately a state experienced inside the user's head" (p. 383) and therefore was a response that "may be both intellectual and emotional" (p. 384).

ISSN: 1975-4094 IJSH

Definition 2(Deep Web Query Results User Satisfaction): In the process of evaluating the user satisfaction of Deep Web query results, we hope that we can obtain a rating value that can represent the user satisfaction by mining the user click through data; this satisfaction value can be used to improve the search strategy and the performance of website.

CNEITE framework has three parts as shown in Figure 1, in the first part, we classify the queries into navigational queries and Informational/transactional queries by using a decision tree based classification algorithm, and then in the second and third part, we proposed an automatic satisfaction evaluation method respectively.

With analysis into search engine user behavior, Broder [10] and Rose [11] independently found that search goals behind user queries can be informational, navigational or transactional. Further experiment results in TREC [12-13] showed that informational and navigational search results benefit from different kinds of evidences.

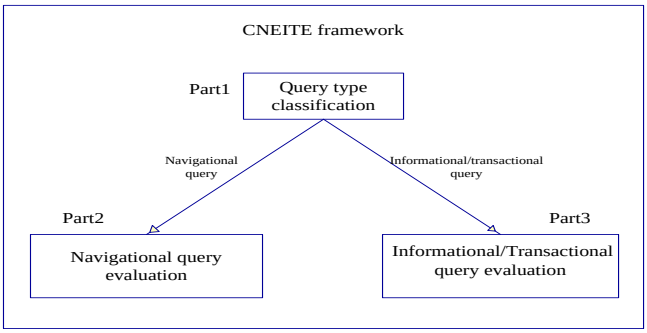

Figure 1. CNEITE Framework

3. Query Type Classification Using Click through Data

In order to verify reliability and scalability of our classification method, we obtained part of query logs from AOL [14] data sets. Using click through data to classify user queries could better understand what users want so that more accurate classification results can be expected. In this section, we propose a novel evidences extracted from click through data: Key-URL Similarity (KUS). It can be used as a feature in our query type identification algorithm.

In order to find the differences between navigational and informational / transactional type queries, we developed a training set of queries which contains 146 navigational queries and 91 informational queries. These queries are randomly selected from query logs and manually classified by 4 assessors using voting to decide queries' categories.

3.1. Key-URL Similarity (KUS) Evidence

Key-URL Similarity (KUS) evidence is extracted from click through data, it is proposed for the AOL data set. In the AOL data set, the queries are all English words; thus, we can calculate the similarity of queries and URLs. It is based on the following assumption:

Assumption (Similarity Assumption): While performing a navigational type search request, the similarity of query and URLs user clicked is relatively high.

For instance, one web search user has a navigational goal, the query he submitted is "Google", he has a fixed search target in mind and would like to find the target URL (www.google.com). We can see that the similarity of "Google" and "google" (the URL after processed) is very high.

According to the Similarity Assumption, we can judge a query type by the KUS. KUS feature is defined as follows:

URL1=URL removes “http://www.”; (1)

Key1=Key removes “http://www.” If it has one; (2)

URL2 = URL1 removes suffix S; (3)

Key2=Key1 removes S If it has one; (4)

KUS = 1－ LD(Key2,URL2) / maxLength (5)

In the formula (5), LD (Levenshtein Distance) is a text comparison algorithm; supposing lengthK = the length of Key2, lengthU = the length of URL2, maxLength = lengthK if lengthK > lengthU, else, maxLength = lengthU.

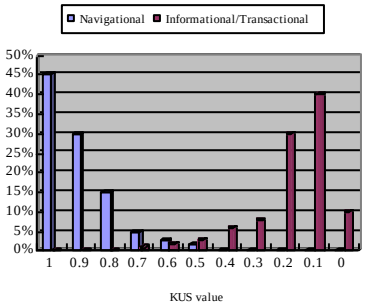

Figure 2. KUS Feature Distribution in the Training Set

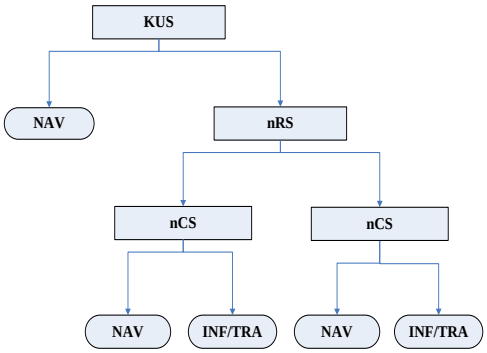

Figure 3. A Query Classification Decision Tree composed of nRS and nCS Features

According to Figure 2, navigational type queries have larger KUS than informational /transactional ones. Most navigational queries have a KUS larger than 0.7 while 90% informational /transactional queries’ KUS is less than 0.5. It means this feature can separate a large part of navigational queries.

3.2. A Learning Based Classification Algorithm

Based on the new feature proposed in Section 3.1, we can separate informational/transactional queries from navigational ones. Besides this feature, n Clicks Satisfied (nCS) and top n Results Satisfied (nRS) proposed by Liu [15] are also believed to be able to identify web search queries.

In order to combine these 3 features: KUS, nRS and nCS to finish the query type classification task, we adopted a typical decision tree algorithm. It is a method for approximating discrete-valued functions that is robust of noisy data and capable of learning disjunctive expressions. We choose decision tree because it is usually the most effective and efficient classifier when we have a small number (3 features here) of features.

We used standard C4.5 algorithm to combine these 3features and get the following decision tree shown in Figure 3. According to C4.5 algorithm, the effectiveness of features can be estimated by the distance away form the root. We can see that KUS is more effective in classification than nRS and nCS. The new feature proposed is more reliable here according to the metric of information ratio in C4.5 algorithm.

4 Automatic Navigational Query Satisfaction Evaluation with Click through Data Analysis

Evaluation is one of the key questions in information retrieval (IR) research. Traditional evaluation methods rely on much human efforts and are therefore quite time-consuming. Several recent attempts have been made towards automatically evaluation in order to tackle the difficulties related to manual assessment. The evaluation of the performance of information retrieval system is a part of the user satisfaction evaluation.

In this section, we propose a fully automatic approach that evaluates user satisfaction of navigational queries based on click through data. Instead of building a query set and annotating relevant documents manually, we annotate answers automatically by analyzing user's query log and click through data. The correctness of the automatically annotated answers is compared with manual ones to verify the reliability of this approach.

Click distribution (CD) is a feature proposed by Lee *et al.* in [16]. It was used by Liu *et al.* in automatic answer annotation in [17]. In the literature [17], CD of a query q is defined as:

$$CD(Query\ \ q) = \frac{\#(Session\ \ of\ \ q\ \ that\ \ involves\ \ clicks\ \ on\ \ R_{most})}{\#(Session\ \ of\ \ q)} \tag{6}$$

For a navigational type query q, Rmost is defined as the URL which is clicked by the most Web search users who are querying q. Users who propose a navigational type query will click a certain result because they consider this result as their search target. Hence Rmost is likely to be the correct answer for q as long as search engines can return the answer as a relative front position so that the users can find and click it. Then the annotation process can be described as a process to locate Rmost for each Query q. However, if a URL is clicked intentionally many times by user, this URL will be the Rmost, but in the fact that this URL isn't the correct answer, which will lead to lower precision.

Aim at improving the precision of automatic answer annotation; we propose a novel automatic annotation approach. For a navigational type query q, suppose that the total number of URLs are clicked is m; the URLs are URL1, URL2, URL3……URLm, each URL might be clicked multiple times; the number of clicks of each URL is respective count1, count2, count3……countm; we use KUS(x,y) to represent the similarity of x and y, x represent the URL and y represent the query q. Given the definitions:

$$KUS_i = KUS(URL_i, q) \quad (1 \le i \le m) \tag{7}$$

$$RKUS_i = KUS_i * count_i \quad (1 \le i \le m) \tag{8}$$

Supposing that RKUSj is the maximum value in RKUSi ($1 \le i \le m$), then we take URLj as the correct answer for query q.

We can see from the above discussion, if we only consider the Rmost, the precision will reduce when some people repeatedly clicking some unrelated URLs deliberately. However, if we consider both the similarity of URL and query q and the click count of URL, a URL is more likely to be the correct answer if the click count is relative high, at the same time, the similarity of URL and query q is very high. Compared with the approach proposed in [17], RKUS has higher precision and better against cheat ability.

With the query set and the answer set annotated above, we use traditional metrics MRR[1] to represent the user satisfaction, the details of user satisfaction are shown in experiments section.

[1] Mean Reciprocal Rank (MRR) is a metric in navigational type evaluation. RR equals to the reciprocal of the correct answer's ranking in the result list and MRR is the mean of the topics' RRs.

5 Automatic Informational/Transactional Query Satisfaction Evaluation with Click through Data Analysis

For an informational/transactional type query q, there is no single correct answer; users may need to click multiple relational URLs to obtain the information they need. In this case, we have to consider the correlation of URLs that users clicked and information users desired as well as the location of URLs.

In this section, we propose a fully automatic approach that evaluates user satisfaction of informational/transactional queries based on click through data. We also annotate the correlation of URLs that users clicked and information users desired automatically by analyzing user's query log and click through data.

5.1. Automatically Correlation Annotation

For the same informational/transactional type query q, different users may need different information, and each user may need to click multiple relational URLs to obtain the information he need. In this case, we believe that the more clicks on the URL the stronger correlation the URL has with the information users desired. Hence, we propose a correlation function CorrUI (URL) to represent the correlation of URL and the information user need.

For a informational/transactional type query q, suppose that the total number of URLs are clicked is m; the URLs are URL1, URL2, URL3……URLm, each URL might be clicked multiple times; the number of clicks of each URL is respective count1, count2, count3……countm; supposing that countj is the maximum value in counti ($1 \le i \le m$), CorrUI(URL) is defined as follows:

$$CorrUI\quad(URL_i) = \frac{count_i}{count_j} \qquad (1 \le i \le m) \tag{9}$$

We can see from the equation (9) that the correlation of URL and the information is associated with the count of URL is clicked.

5.2. Automatically User Satisfaction Evaluation

When we evaluate the user satisfaction of informational/transactional type queries, we have to consider not only the correlation of URL and information, but also the URL position. Here, we use the RUR to represent the location information of the URL, RUR equals to the reciprocal of the URL's ranking in the result list.

We define a satisfaction function Satisfaction (q). For an informational/transactional type query q, suppose that the total number of URLs are clicked is n; the URLs are URL1, URL2, URL3……URLn, CorrUI (URLi) ($1 \le i \le n$) represents that the correlation of URLi and information; RURi represents the reciprocal of the URLi's ranking in the result list; then we have the definition as follow:

$$Satisfaction\ (q) = \frac{\sum_{i=1}^{n} CorrUI\quad(URL_i) \times RUR_i}{n} \tag{10}$$

Apparently, we obtain the average satisfaction of a query here; similarly, we also can access a user's average satisfaction and the average satisfaction of all users over a period of time.

6 Experiments and Discussions

All the experiments are based on the click through data come form AOL data set. The logs are collected from March 1st to March 31th in the year 2006.

6.1. Query Type Classification Experiment

We developed a test set to verify the effectiveness of our classification algorithm. This test set is composed of 113 informational/transactional type queries and 186 navigational queries. These queries are all come from AOL data set.

We use traditional precision/recall framework to judge the effectiveness of the query type classification task. Precision and Recall values are calculated separately for two kinds of queries. They are also combined to F-measure value to judge the overall performance. Experiment results are shown in Table 1.

Table 1. Query Type Classification Experimental Results

	Training set		Test set	
	INF/TRA	NAV	INF/TRA	NAV
Precision	80.00%	93.67%	78.45%	88.67%
Recall	77.84%	92.25%	79.02%	87.54%
F-measure	0.79	0.93	0.79	0.88

According to the experimental results in Table 1, precision and recall values over 80% are achieved to classify queries. It shows that most queries are successfully classified with the help of click through information.

6.2. Navigational Query Correct Answer Annotation Experiment

We annotated three groups of queries using click through logs during different time periods. About 5% of the annotated queries are picked up randomly and manually checked for correctness.

Table 2. Size of the Annotated Query Set and Accuracy of the Annotated Answers

	#(Annotated queries)	#(Checked sample set)	Accuracy
March.1 - March.10	6,540	327	98.17%
March.11 - March.20	6,892	344	97.58%
March.21 - March.31	6,267	313	98.36%

According to the results shown in Table 2, we can see that during each time period, over six thousand queries are successfully annotated and over 97% of the sampled annotated answers are correct.

We checked the wrongly-annotated answers in the sample set and found that these answers are usually similar with the correct answers. For example, when users propose the query of “aaroncarter”, more users may like to click the URL http://aaroncarter.ca instead of its homepage http://www.aaroncarter.com; although the KUS (http://aaroncarter.ca, aaroncarter) is the same with KUS (http://www.aaroncarter.com, aaroncarter), RKUS (http://aaroncarter.ca, aaroncarter) > RKUS (http://www.aaroncarter.com, aaroncarter); and then the automatically approach annotate http://aaroncarter.ca as the answer instead of the homepage (http://www.aaroncarter.com).

6.3. User Satisfaction Evaluation Experiment

MRR is used in our experiments for evaluating the user satisfaction of navigational type queries and the results are shown in Figure 4.

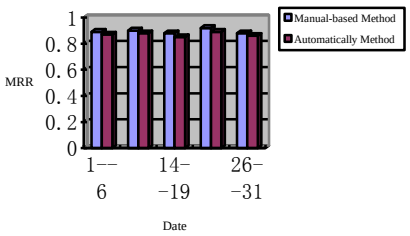

Figure 4. Comparison in Different Time Periods' Evaluation Results between Manual-based Method and Automatically Method

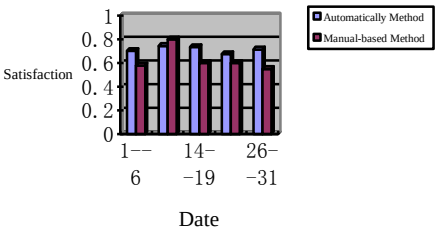

Figure 5. The User Satisfaction of Informational/transactional Type Queries during Different Time Periods

In manual-based methods, a set of 180 navigational queries are selected randomly from the query logs. The answers are annotated by 4 assessors. According to Figure 4, we found that the automatically evaluation result has the same performance ranking as the manual one. The correlation value between MRRs of the two methods is 0.945, which indicates the two evaluation results are quite similar. The user satisfaction of informational/transactional type queries are shown in Figure 5.

In manual-based methods, a set of 156 informational/transactional queries are selected randomly from the query logs. These queries are given a satisfaction value by 4 assessors. The correlation value between satisfaction values of the two methods is 0.635, which because that the satisfaction values given by assessors may not be very accurate. However, according to Figure 5, we can found that our automatically method can reflect users satisfaction to some extent.

By observing Figure 4 and Figure 5 we can found that, the user satisfaction values of informational /transactional queries are lower than that of navigational queries, which verifies the fact that the performance of navigational queries is better than that of informational/transactional queries for search engines.

7. Conclusions and Future Work

In this paper, we propose a click-through-data-based and unsupervised user satisfaction evaluation system, CNEITE, to evaluate the user satisfaction of Deep Web query result pages. It applies query type classifying, navigational query evaluating, Informational/transactional query evaluating to solve the challenging tasks. We can conclude from the experiments that CNEITE achieves higher classify precision than Dtree, and higher annotate answer accuracy on AOL data sets.

Future study will focus on the following aspects: How to combine the retention time that users access the page and the times that users submit queries with the quality of result pages to evaluate the user satisfaction.

Acknowledgment

The authors acknowledge the support of the National Nature Science Foundation of China (No. 61103114), Fundamental Research Funds for the Central Universities (No. CDJZR185502), National Key Technology R & D Program (No. 2012BAH19F00), and the Key Project Chongqing Postgraduate Education Reform (No. yjg132027).

References

[1] J. Cafarella Michael, H. Alon and M. Jayant, "Structured data on the web", Communications of the ACM, vol. 54, no. 2, **(2011)**, pp. 72-79.
[2] T. M. Ghanem and W. G. Aref, "Databases deepen the web", Computer, vol. 37, no. 1, (**2004**), pp. 116-117.
[3] N. Gudigantala, J. Song and D. Jones, "User satisfaction with Web-based DSS: The role of cognitive antecedents", International Journal of Information Management, vol. 31, no. 4, **(2011)**, pp. 327-338.
[4] P. Verdegem and G. Verleye, "User-centered E-Government in practice: A comprehensive model for measuring user satisfaction", Information Quarterly, vol. 26, no. 3, **(2009)** July, pp. 487-497.
[5] Y.-S. Wang and Y.-W. Liao, "The conceptualization and measurement of m-commerce user satisfaction", Computers in Human Behavior, vol. 23, no. 1, **(2007)** January, pp. 381-398.
[6] B. John, S. Khaddaj and A. Hoppe, "Accelerating User Satisfaction in Semantic E-Commerce", 2nd International Conference on the Applications of Digital Information and Web Technologies, **(2009)**, pp. 849–851.
[7] C.-Y. Dai, M.-T. Kao, C.-T. Harn, Y.-H. Yuan and W.-F. Chen, "The Research on User Satisfaction of Easy Teaching Web of Taipei Assessed via Information Quality,System Quality, and Technology Acceptance Model", Computer Science & Education (ICCSE), 2011 6th International Conference , **(2011)**, pp. 758–762.
[8] P. Seongwon, O. Duckshin and L. Bong Gyou, "Analyzing User Satisfaction Factors for Instant Messenger-Based Mobile SNS*", Communications in Computer and Information Science, vol. 185, **(2011)**, pp. 280-287.
[9] M. Niamh and K. Jurek, "Measuring user-satisfaction with electronic consumer products: The Consumer Products Questionnaire, International Journal of Human Computer Studies, vol. 69, no. 6, **(2011)** June, pp. 375-386.
[10] A. Broder, "A taxonomy of web search", SIGIR Forum, vol. 36, **(2002)**.
[11] D. E. Rose and D. Levinson, "Understanding User Goals in Web Search", In proceedings of the 13th World-Wide Web Conference, **(2004)**.
[12] N. Craswell and D. Hawking, "Overview of the TREC-2002 web track", In the eleventh Text Retrieval Conference (TREC-2002), NIST, **(2003)**.
[13] N. Craswell and D. Hawking, "Overview of the TREC-2003 web track", In the twelfth Text Retrieval Conference (TREC-2003), NIST, **(2004)**.
[14] G. Pass, A. Chowdhury and C. Torgeson, "A Picture of Search", The First International Conference on Scalable Information Systems, Hong Kong, **(2006)** June.
[15] Y. Liu, M. Zhang, L. Ru and S. Ma, "Automatic Query Type Identification Based on Click Through Information", Lecture Notes in Computer Science, vol. 4182, **(2006)**, pp. 593-600.
[16] U. Lee, Z. Liu and J. Cho, "Automatic Identification of User Goals in Web Search", in the 14th WWW Conference, **(2005)**.
[17] L. Yiqun, F. Yupeng, Z. Min, M. Shaoping and R. Liyun, "Automatic Search Engine Performance Evaluation with Click-through Data Analysis*", 16th International World Wide Web Conference, **(2007)**, pp. 1133-1134.

