Information Sciences 430–431 (2018) 554–566

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A multi-view semi-supervised approach for task-level web search success evaluation

Alin Fan, Ling Chen*, Gencai Chen

College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

A R T I C L E I N F O

Article history:
Received 19 September 2016
Revised 11 October 2017
Accepted 4 December 2017
Available online 6 December 2017

Keywords:
Disagreement-based semi-supervised learning
Multi-view mechanism
Web search success evaluation

A B S T R A C T

Web search success evaluation is an effective way to evaluate the performance of search engines from the perspective of search experience. Many research efforts have been made to evaluate web search success via modeling user search behavior by analyzing search logs. Most of these studies consider the web search success evaluation as a binary classification problem, and use supervised learning approaches to evaluate whether a search experience is successful or not, which often require a large number of labeled data to learn accurately. Since unlabeled data of user search behaviors are easily obtainable in search logs, semi-supervised approaches have been exploited to improve the performance of web search success evaluation via combining labeled data and unlabeled data. However, the existing semi-supervised web search success evaluation approach would suffer from the model assumption violation, i.e., when the assumption of the model is not correct, training the web search success evaluation model with large amounts of unlabeled data would hurt the evaluation performance. In order to address this problem, a Multi-view Semi-Supervised web Search Success Evaluation (M4SE) approach is proposed, which exploits a multi-view mechanism during the semi-supervised learning process of web search success evaluation. M4SE considers the transitions between any two actions as the action view, and treats the dwell time between contiguous actions as the time view. M4SE uses the strategy of different parameter configurations on action and time views to generate diverse classifiers. Experiments on different search log datasets show that the proposed approach achieves better performance than the state-of-the-art semi-supervised approach for evaluating web search success.

1. Introduction

In recent years, web search success has become an important indicator to measure the performance of search engines from the perspective of search experience. By modeling user search behavior using search logs, the quality of a search experience could be evaluated (i.e., to evaluate whether a search experience is successful or not) [1,14,20,25]. In comparison with traditional single-query metrics (e.g., Mean Average Precision (MAP), Precision @k, and Discounted Cumulative Gain (DCG) [26]), web search success evaluation approaches require neither relevance judgments nor explicit feedback; they only depend on the user search behaviors that naturally consist in search logs. These advantages make web search success evaluation approaches useful for search providers.

* Corresponding author.
E-mail addresses: fanalin@zju.edu.cn (A. Fan), lingchen@zju.edu.cn (L. Chen), chengc@zju.edu.cn (G. Chen).

https://doi.org/10.1016/j.ins.2017.12.003
0020-0255/

Table 1
A list of action types.

Action	Description
Q	Issue a query to a search engine
SR	Click on a returned document in search result page
HL	Click on a hyperlink in the current clicked document (not in a search result page)
E	End the search task (a dummy action)

Generally, the aim of the web search success evaluation approach is to classify each search task into one of two categories (success or failure). Although the search log can be segmented at query, task, or session level, Liao et al. [29] showed that evaluating web search success at task level is more accurate than evaluating at query level and session level. Hence most of the existing researches consider web search success at task level, and use supervised learning approaches to assess whether a search task is successful or not [1,15,17,18,37]. A major concern with these supervised learning approaches is that they often require large quantities of labeled data to learn accurately; nevertheless, collecting a large number of labeled data for web search success evaluation is expensive and time consuming.

Since unlabeled data of user search behaviors are easily obtainable in search logs, researcher starts to use both labeled and unlabeled data to evaluate web search success more effectively, e.g., Hassan [19] proposed a generative-model based semi-supervised approach to further improve the performance of web search success evaluation by using labeled and unlabeled data simultaneously, which adopted Expectation–Maximization (EM) [12] to conduct maximum likelihood estimation of the generative-model parameters. Although the generative-model based semi-supervised approach has made some achievements in web search success evaluation, there exist two drawbacks: First, generative-model based semi-supervised approach concatenated potential sub-views into one single view, which is not physically meaningful, because each sub-view has a specific statistical property. Second, the EM algorithm essentially is a solution seeking local maxima; if the initial classifier is not good enough, fitting the model using a large number of unlabeled data would cause performance degradation [9,10,45].

To address these drawbacks, we propose the Multi-view Semi-Supervised web Search Success Evaluation (M4SE) approach. M4SE is a disagreement-based semi-supervised learning approach that constructs multiple diverse classifiers, and makes them cooperate to utilize unlabeled data [43]. In M4SE, two views are tailored to the problem of web search success evaluation, i.e., action view (the transitions between any two actions) and time view (the dwell time between contiguous actions), and different parameter configurations are used in action and time views to generate diverse classifiers. The two classifiers trained from action and time views would be integrated through inferring labels to each other by turns during the training process of EM, and then the performance degradation of generative-model based semi-supervised approach caused by local maxima could be alleviated through jointly optimizing the performance of all classifiers with a large diversity.

Our contributions can be summarized as follows:

(1) Combine labeled and unlabeled data to further improve the performance of web search success evaluation by exploiting a disagreement-based semi-supervised learning approach with multi-view mechanism;
(2) Exploit the strategy of different parameter configurations on action and time views to generate diverse classifiers;
(3) Perform empirical experiments on different search log datasets and the experimental results show that the proposed approach achieves better performance than the state-of-the-art semi-supervised approach for evaluating web search success.

The remainder of this paper is organized as follows. Section 2 gives the problem definition. We describe the M4SE approach in Section 3. Section 4 presents the experiments. We discuss the related work in Section 5. Finally, Section 6 gives the conclusions and future work.

2. Problem definition

Since evaluating web search success at task level is more accurate than evaluating at other levels, web search success evaluation in this paper especially refers to evaluate whether a search task extracted from search logs is successful or not.

Definition 1 (*Action*)**.** An action, denoted by a, is an interaction that a user performs with a search engine, e.g., submitting a query string, clicking an interested URL, and clicking the spelling correction link.

Actions include all types of interactions performed by users while using the search engine. Following the work of Ageev et al. [1], we use the set of actions described in Table 1.

Definition 2 (*Transition time*)**.** A transition time t refers to the time between two contiguous actions.

Definition 3 (*Search task*)**.** A search task S is an atomic information need [17,21], which could be treated as an ordered sequence of actions along with the time between these actions, i.e., given a search task with n actions, it could be expressed as $S = \langle a_1, t_1, a_2, t_2, \ldots, t_{n-1}, a_n \rangle$, where a_n is always the dummy action "*E*" that indicates the end of a search task.