The Electronic Library

Automatic prediction of news intent for search queries: An exploration of contextual and temporal features

Xiaojuan Zhang, Shuguang Han, Wei Lu,

Received 26 June 2017
Revised 27 November 2017
31 January 2018
20 April 2018
Accepted 10 June 2018

Automatic prediction of news intent for search queries

An exploration of contextual and temporal features

Xiaojuan Zhang
Department of Computer and Information Science, Southwest University, Chongqing, China

Shuguang Han
Department of Information Science, University of Pittsburgh, Pittsburgh, PA, USA, and

Wei Lu
Department of Information Management, Wuhan University, Hubei, China

Abstract

Purpose – The purpose of this paper is to predict news intent by exploring contextual and temporal features directly mined from a general search engine query log.

Design/methodology/approach – First, a ground-truth data set with correctly marked news and non-news queries was built. Second, a detailed analysis of the search goals and topics distribution of news/non-news queries was conducted. Third, three news features, that is, the relationship between entity and contextual words extended from query sessions, topical similarity among clicked results and temporal burst point were obtained. Finally, to understand the utilities of the new features and prior features, extensive prediction experiments on SogouQ (a Chinese search engine query log) were conducted.

Findings – News intent can be predicted with high accuracy by using the proposed contextual and temporal features, and the macro average F1 of classification is around 0.8677. Contextual features are more effective than temporal features. All the three new features are useful and significant in improving the accuracy of news intent prediction.

Originality/value – This paper provides a new and different perspective in recognizing queries with news intent without use of such large corpora as social media (e.g. Wikipedia, Twitter and blogs) and news data sets. The research will be helpful for general-purpose search engines to address search intents for news events. In addition, the authors believe that the approaches described here in this paper are general enough to apply to other verticals with dynamic content and interest, such as blog or financial data.

Keywords Query classification, News intent, News queries, Query intent

Paper type Research paper

Introduction

It has been reported that around 10 per cent of web search queries are related to current news events (Bar-Ilan *et al.*, 2009). To better support the news information needs, modern search engines such as Google and Bing have aggregated the news contents into their search

This research is supported by National Social Science Foundation of China under grant no. 15CTQ019 and National Nature Science Foundation of China under grant no. 71173164.

The Electronic Library
Vol. 36 No. 5, 2018
pp. 938-958
© Emerald Publishing Limited
0264-0473
DOI 10.1108/EL-06-2017-0134

results. News results occupy the space of "regular" search results so that the misunderstanding of news intent for search queries would hurt users' search experiences. It is therefore important to predict the news intent of search queries which provide guidance to a general search engine in deciding whether to aggregate news content into the search results or only provide non-news results. Note that, we name the queries with news intent as news queries, otherwise non-news queries. However, automatic prediction of news intent of search queries is a challenge because of at least two reasons. First, limited amount of information can be directly derived from both the query strings and the users who issue the queries. Second, the prediction should ideally be in near real time. The most common method is to treat it as a classification problem, where a list of features was extracted and selected for better prediction accuracy. Consequently, feature extraction plays an important role in news intent prediction. Existing approaches have tried to extract the classification features from social media (e.g. Wikipedia, Twitter and blogs) or news data sets, and at least two of these data resources are used each time. Although being able to produce a reasonable prediction performance, it requires too many computation resources for crawling, parsing and integrating these resources and what is especially important is that it is hard to get these data resources. Hence, we would like to explore new feature(s) that can be quickly computed. Users' query logs and their behaviors on search engine result pages can provide useful resources to extract such features, guiding us to extract features directly from search engines. What is more, we believe that the news intent will be classified more precisely because user logs from the same search engine under the same search scenario are supported, and some hints for his or her intent at that time will be obtained by analyzing the user's query click record and query reformulating history. To justify our hypothesis, we conducted a series of experiments on real-world search logs, and particularly, we focused on the Chinese search engines because of the availability of query logs.

News intent of a search query means that the query is related to a currently newsworthy event, and news intent prediction is to predict whether or not a query is related to an ongoing or recent event (Louis *et al.*, 2011). Here, the event refers to "something happening in a specific place at a specific time, and tagged with a specific term" (Ruocco and Ramampiaro, 2012). Thus, if a query has the news intent, then this query may be related to certain factors of an event, such as named entities (e.g. person, place and organization), topical words (specific terms assisting in describing the topic of an event) and temporal information. We summarize these factors into two types, contextual-based and temporal-based features. The former one is based on the query strings and textual context (e.g. the named entities contained in the query string and the clicked results) in which the query keywords occur. The latter one tracks the corpus frequency (the number of times queries occur in the query logs over time) and quantifies the temporal distribution of clicked results. Moreover, we are interested in studying the effectiveness of utilizing these features for news intent prediction.

In summary, we address the issue of news intent prediction by proposing proper methods to extract a different set of contextual and temporal features directly from a general search engine query log without use of such large corpora as social media (e.g. Wikipedia, Twitter and blogs) and news data sets in this paper. Importantly, the main contributions of this paper lie in the following three aspects:

(1) We build a ground-truth data set with correctly marked news and non-news queries, and this data set can be reused in similar tasks. And we conduct a detailed analysis of the search topic distribution of news/non-news queries, which has not been thoroughly studied.

(2) We propose two new contextual features and one temporal feature for news intent classification, including the co-occurrence between a named entity (we use the

word “entity” and “named entity” interchangeably in this paper to refer to the same concept) and a topical word in query sessions, topical similarity among clicked results and temporal burst point extracted from query logs.

(3) We perform a thorough analysis to study the effectiveness of our proposed contextual and temporal features combined with prior proposed features, through which we find that news intent is predicted with high accuracy and all the proposed features have positive effects on deciding news intent of a query.

Related work

The research on the news intent prediction includes the following three aspects:

(1) news intent prediction for search queries;

(2) regency queries identification; and

(3) event detection based on contextual and temporal features.

News intent prediction for search queries

Prior work on news intent prediction has introduced a number of features, and these features are mainly mined from social media (e.g. Wikipedia, Twitter and blogs) or news data sets, and at least two of these data resources are used each time. For example, Diaz (2009) proposed a machine learning approach for news intent prediction. This approach trains an initial model using features extracted from news articles and additional query features from past web and news vertical query logs, and they further use a user’s subsequent clicks to enhance the model over time. They showed that the search query logs when combining with user feedback were effective resources for determining the news relatedness of a query. Arguello *et al.* (2009) expanded upon the approach by Daiz for multiple vertical intent by extracting features from each vertical considered to build a classification model. Konig *et al.* (2009) also examined a machine learned query classifier for determining news intent. They proposed the use of additional features from the query (i.e. query length and query terms), in addition to features describing the distribution of the query terms in blog, newswire and Wikipedia corpora. Louis *et al.* (2011) explored the ways of utilizing query similarity to improve the news intent prediction. They calculated the similarity between queries based on the URLs, titles and abstracts of clicked news and blog pages. McCreadie *et al.* (2013) performed a user study to investigate when and where to integrate news-related content from newswire, blogs and Twitter and Wikipedia sources. Their results further showed the potential of using social media to serve news queries. However, because of the high computation costs and the difficulty of obtaining resources, it is hard to use the social media and news corpora on a large scale.

To the best of our knowledge, only little work has addressed the news intent identification just using query log but without social media and news data set. For example, McCreadie *et al.* (2010) studied the generation and validation of a news query classification data set comprised of labels crowdsourced from Amazon’s Mechanical Turk and detailed insights were gained. However, they did not explore this dataset to realize the automatic classification of news queries. Hassan *et al.* (2009) used geographic features extracted from a query log of a general search engine to identify news queries. The information they considered included cues derived from the location of the user, the IP address, the location relevant to the query and the relation between the two locations. They built a classifier that used geographical cues to predict news intent of a query. However, for protecting user privacy, it is difficult to get the IP address of users from query logs in most cases. Nevertheless, this method initially validated the feasibility of using a general search engine query log to identify news intent without any social media and news data set. However, this study did not consider the contextual-based and temporal-based features.

Recency queries identification

It should be noted that recency queries identification is also related to our work. Particularly, recency queries are those occurring right after breaking news or more recent events and they are defined as those occurring right after breaking news or more recent events (Moulahi *et al.*, 2016; Joho *et al.*, 2016); therefore, the work of distinguishing recency queries from non-recency queries resembles to the news intent identification in a sense. Recently, the NTCIR Temporialia task pushes further this study and proposes whether a given query is related to *past, recency, future* or *atemporal*. Within this context, the most performing system is based on machine learning approach. For example, Sakaguchi and Kurohashi (2016) took a supervised machine learning approach, using features of bag of words, POS and word vectors mined from *New York Times* corpus, and they also incorporated knowledge about temporal and holiday expressions to classify the queries. Li *et al.* (2016) proposed four groups of features including trigger word, word POS, explicit time gap and temporal probability of word. Gui and Lu (2016) proposed 19 features in total from query itself. These approaches did not consider user information because it is hard to get the corresponding user logs, and thus, they did not exploit the data which can express a user's real intent. Accordingly, the accuracy of recency queries identification is affected.

Event detection based on contextual and temporal features

The detection of events from web document streams and databases has been treated extensively in the literature (Allan *et al.*, 1998; Brants *et al.*, 2003; Ruocco and Ramampiaro, 2012). As each event is usually characterized by a set of name entities (such as person, place and organization), topic words (referring to the specific term describing the topic of an event) and temporal information (Ruocco and Ramampiaro, 2012), such information has been widely used for event detection. For example, Kumaran and Allan (2004, 2005) studied the ways of detecting stories that are reported on new events through calculating the similarity between the new article and other old articles. The weights of name entities were boosted to the overall similarity score, performing better than not using the name entities. Fu *et al.* (2010) proposed a new event detection algorithm with more emphasis on the elements of news (e.g. person and place). Vavliakis *et al.* (2013) integrated the named entity recognition, dynamic topic map discovery and topic clustering into one unified framework. Sun and Hu (2011) proposed to study query-guided event detection from two parallel document streams (i.e. news and blog). They grouped user queries, news articles and blog posts into events to which they are related. Wei *et al.* (2014) developed an event episode discovery mechanism to organize news documents pertaining to an event of interest. In particular, they proposed two novel time-based metrics that could be successfully used for feature selection and document representation.

The above-mentioned methods tried to detect events from the textual information of a web page. Besides, existing studies have also explored the web search logs for event detection. For example, Zhao *et al.* (2006) proposed a two-phase clustering method, in which the semantic and temporal similarity were used to group similar query-page pairs that are corresponding to real life events. Chen *et al.* (2008) explored the way of using query session as the indicator of an event. They first mapped each query session to a popular space on the basis of contextual and temporal similarity, and then the query sessions were grouped based on both similarities to represent an event. Parikh and Sundaresan (2008) studied the role of query in burst event detection by using daily query streams from a large-scale e-commerce system. Zhang *et al.* (2010) developed features like query frequency, click information and user intent dynamics within a search session based on historical query logs and trained classifier to identify recurrent event queries. Gu *et al.* (2010) first divided the whole query log data into topics for efficiency consideration and then incorporated link information, temporal information and query content to ensure the quality of detected events.

These studies have demonstrated the effectiveness of using entities, topical words and temporal information from query logs in predicting events, motivating us to further study on this topic. To our best knowledge, the most related work to our study is from Ghoreishi and Sun (2013), who identified 20 features including both contextual (including the named entities and topical words) and temporal features from a small set of search results of a query to predict its event-relatedness. These features are derived from the content of a query and its search engine results. Based on this study, we find that the contextual and temporal features are both important to decide the event-relatedness of a given query. This motivates us to utilize the features proposed in the study conducted by Ghoreishi and Sun (2013), while our work also differs from the study (Ghoreishi and Sun, 2013) in three aspects:

(1) we build a ground-truth dataset with correctly marked news and non-news queries;

(2) we utilize the clicked documents instead of all the documents from the search engine result page; and

(3) we propose three new features.

The construction of data collection

To conduct news intent prediction, we need to build a ground-truth data collection. We obtain our experimental data sets from a general search engine and manually labeled a subset of them.

Query log

As far as we know, there are three suitable publicly available query logs to support experiments with query intent identification: the AOL data set (Zamora *et al.*, 2014), the MSN data set (Brenes *et al.*, 2009) and the SogouQ data set (Liu *et al.*, 2006). Our experimental data came from SogouQ[1] data set and it is mainly caused by the following reason. Manual labeling of queries is a very important task in query classification and the quality of labeling data set often directly affects the final experimental results. What is more, the quality of data annotation to a great extent depends on the degree of labelers' familiarity with news events. Taking the geographical location into consideration, it is more convenient for us to ask Chinese to finish our labeling work. As Chinese labelers are more familiar with Chinese news events than foreign new events and most queries issued into Sogou search are related with events having happened in China, the quality of labeled results from Chinese labelers on SogouQ are more likely to be superior to those on MSN or AOL query log.

SogouQ was extracted from June 1 to June 30 in 2008 including 51,537,393 web queries1, and private information has been removed as well as illegal query sessions. As shown in Table I, each record contains the following fields:

- QueryTime: when the query was issued;
- UserID: an anonymous user-ID;
- Query: the actual query terms;

- Item-Rank: the rank of clicked result;
- Click Sequence: the number of the click in the sequence of clicks in a search session; and
- Clicked URL: the result the user clicked.

The same pre-processing work for query log as Cai *et al.* (2014) work was performed. We used time-based method to segment our session and set time-out threshold to15 minute (He and Goker, 2000) in our experiment.

To create our query sets, we need to sample representative and unbiased queries from the query log. We extracted queries of intermediate time period of June 11, 2008 to June 20, 2008 so as to analyze characteristics of news queries issued before and after the query-related news event broke out. To get our data set, we used the Poisson sampling strategy proposed in the study of McCreadie *et al.* (2010). Considering that some queries with little click information would bring the problem of data sparseness, we removed those queries containing less than 20 unique clicks. In total, we got 3,075 unique sample queries.

Manual labeling of queries

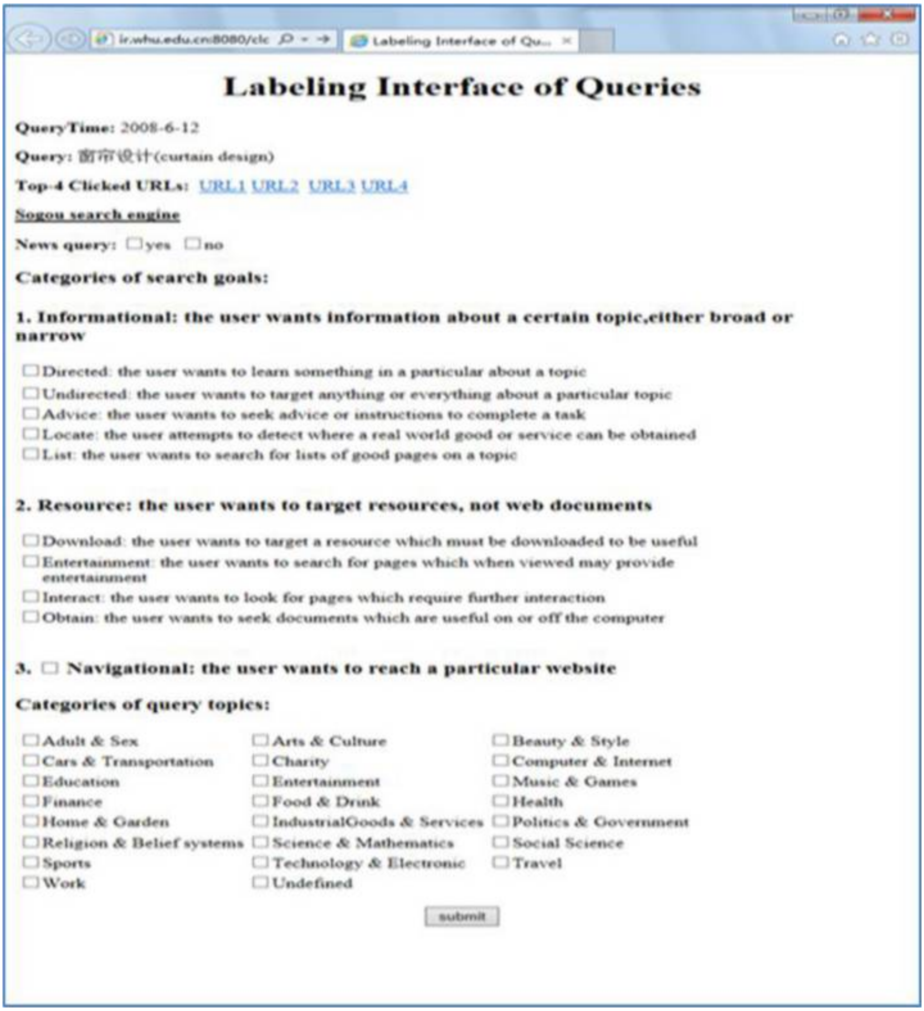
In this paper, we will use automatic classification to predict query intent. As automatic classification is a machine learning method, which requires manually mining new training data and retrained a classification model with mined data. Thus, the manual labeling of queries is the first step in the query classification. Our labeling tasks were fulfilled by 10 graduate students. We developed an annotation interface (in Figure 1) similar to McCreadie *et al.* (2010) work, and we integrated the top four clicked news pages into the annotation interface. Most importantly, we explicitly provided the information about whether or not the query words appeared on the title, abstract or body field of a news page. If this page was created on the day of query time, then it could be labeled as a news query. If the labelers still had confusions, then they could use “Sogou search engine” (a Chinese search engine) links in the interface to get more information in the Sogou search results.

In addition, we also asked the labelers to annotate the “goal” (i.e. a goal a user wants to achieve through searching) and “topic” (i.e. the information a user wants to find) of each query. It is quite common for a query to have multiple goals. For instance, given the query “MP3”, the user might be interested in downloading MP3 files or MP3 related topics. In such cases, we asked the labelers to assign each query to the most expected goal category. The goal hierarchy used in this paper was proposed by Rose and Levinson (2011). The lists of topics (Gonzalez-Caro *et al.*, 2011) were built from the first level of categories used by ODP, Yahoo! and Wikipedia. Similar to search goal, we also demanded the labelers to classify each query to the most appropriate topic category.

To calculate the inter-work agreement for labeling news/non-news queries, we also had 400 queries annotated by three labelers. The Kappa (Cohen, 1960) of average pair-wise annotator agreement among the three labelers was 0.82, suggesting the perfect inter-work agreement (Landis and Koch, 1977). A high level of agreement indicated that the resulting labels were of good quality; hence, the labeling method of our work was suitable. In total, we

Table I. Examples of the Sogou query log data sets

Query time	User ID	Query (transition into English)	Item-rank	Click sequence	Clicked URL
00:00:03	8234353	电影下载(Movie download)	9	6	www.tvbsale.com
00:00:04	720986435	李荷娜(You HaNa)	2	3	club.koook.com



Labeling Interface of Queries

QueryTime: 2008-6-12

Query: 窗帘设计(curtain design)

Top-4 Clicked URLs: URL1 URL2 URL3 URL4

Sogou search engine

News query: □yes □no

Categories of search goals:

1. Informational: the user wants information about a certain topic,either broad or narrow

- □Directed: the user wants to learn something in a particular about a topic
- □Undirected: the user wants to target anything or everything about a particular topic
- □Advice: the user wants to seek advice or instructions to complete a task
- □Locate: the user attempts to detect where a real world good or service can be obtained
- □List: the user wants to search for lists of good pages on a topic

2. Resource: the user wants to target resources, not web documents

- □Download: the user wants to target a resource which must be downloaded to be useful
- □Entertainment: the user wants to search for pages which when viewed may provide entertainment
- □Interact: the user wants to look for pages which require further interaction
- □Obtain: the user wants to seek documents which are useful on or off the computer

3. □ Navigational: the user wants to reach a particular website

Categories of query topics:

□Adult & Sex
□Arts & Culture
□Beauty & Style
□Cars & Transportation
□Charity
□Computer & Internet
□Education
□Entertainment
□Music & Games
□Finance
□Food & Drink
□Health
□Home & Garden
□IndustrialGoods & Services
□Politics & Government
□Religion & Belief systems
□Science & Mathematics
□Social Science
□Sports
□Technology & Electronic
□Travel
□Work
□Undefined

submit

Figure 1.
Labeling Interface of queries

got 317 news-related queries, accounting for about 10.3 per cent of the total number of sample queries. This result is roughly in line with the finding reported in the study (Bar-Ilan *et al.*, 2009).

An analysis of labeled queries

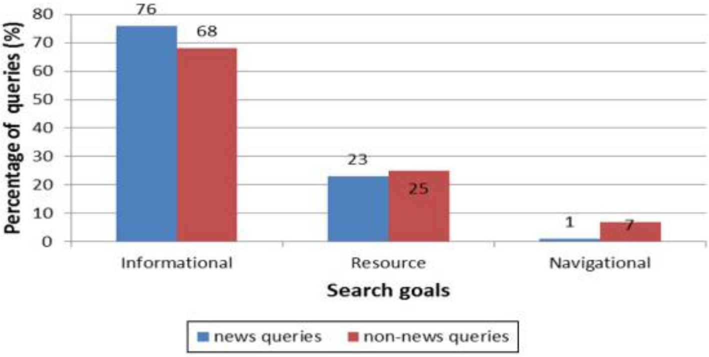
In this section, we conduct an analysis to depict the distributions of search goals as well as topics for our labeled news and non-news queries, respectively. As shown in Figure 2, there is no distinct difference in the distributions of search goals (e.g. informational, resource and navigational) between news and non-news queries. Figure 3 provides a more detailed proportion of the search goals about 10 sub goal categories. The “Directed” (i.e. to get something in about a particular topic) and “Entertainment” (i.e. to search for pages which can provide entertainment when viewed) take up the most part of news queries, which suggests that goals of news queries are supposed to obtain information for a particular topic or for some entertainment purposes. In addition to "Directed", the goals of non-news query are also more likely to lie in the categories of "Advices" (i.e. the user wants to get advice, ideas, suggestions or instructions) and "Undirected" (i.e. the user wants to learn anything/everything about a topic).

Figure 4 shows the search topic distribution in the news and non-news queries, indicating that search topics of news queries tend to be entertainment, economy, politics and sports, whereas those of non-news queries tend to be industrial goods and services, music and games and information about jobs.

The above results in Figures 3 and 4 give us some inspiration when extracting some classification features for news intent identification.

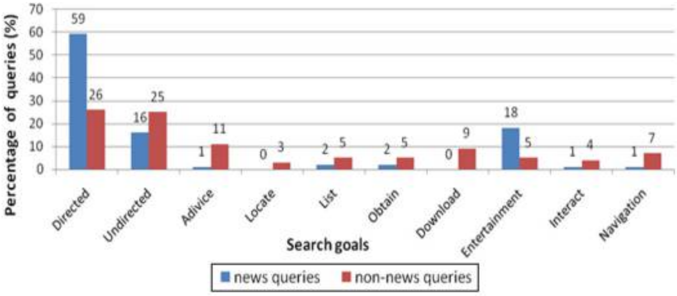
Extracting features for news intent prediction

Determining which features should be used to represent a query is a key decision in query classification, and the features we adopted in this study belongs to two sets, namely, contextual features and temporal features. Specifically, 3 new and 20 prior proposed features



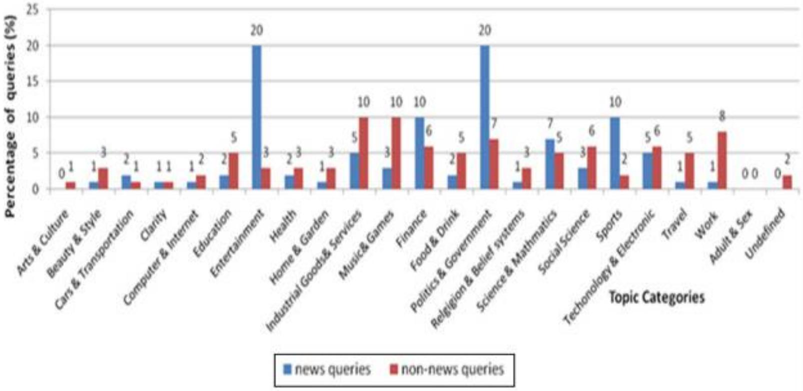
Note: Note that the bars in each color sum up to a total of 1.0

Figure 2. Distribution of search goals with respect to the taxonomy of Broder (2002) in the news and non-news queries



Note: Note that the bars in each color sum up to a total of 1.0

Figure 3. Distribution of search goals with respect to the taxonomy of Rose and Levinson (2011) in the news and non-news queries



Note: Note that the bars in each color sum up to a total of 1.0

Figure 4.
Distribution of query topics in the news and non-news queries

are included. We will first describe how to extract the new features in detail and then give a brief introduction to existing features in this section.

Newly proposed features

We propose two contextual features (i.e. the co-occurrence between entity and topical word in query sessions and topical similarity among clicked results) and one temporal feature (i.e. temporal burst point) for news intent prediction.

Co-occurrence between entity and topical word in query session. We believe that the content storyline of an event can be characterized by a set of entities and a set of topical terms (Kumaran and Allan, 2005). Thus, the content of an event-related query can be divided into two parts: entities and other topical words (referring to the non-entity words). However, most queries are usually very short (Cui *et al.*, 2002), which may miss certain content aspects. For example, after the event "Wenchuan earthquake" broke out, some users would use short queries, such as "Wenchuan" or "earthquake" to express their news intent related to this event. Because of the lack of sufficient information in short queries, it is difficult for search engines to identify the news event relatedness. As explored in He *et al.* (2013), a user usually submits a series of queries during one episode of interactions with the web search engine for single information need to get what they really need. Accordingly, it is a good way to expand corresponding entities or topical words for implicit queries from query sessions in the past three days. As indicated in Figure 3, the goal of news queries is to obtain information for a particular topic, so we believe that users always type queries containing specific combination of one or several named entities and topical words to get news-related information. More importantly, in one case, when a query contains one or more name entities, we got all the topical words from query sessions that this query exists. And we assume that if this named entity or one of named entities in this query always co-occurs with a topical word, we may judge that this query has news intent. In another case, for a query not containing any named entity, we get all the named entities (e.g. person, location, organization) occurring in the query session that is present. We also assume that if one or more topical words in this query usually co-occur with a named entity, we may judge that this query has news intent. We used mutual information (MI) shown in Equation (1) to compute the co-occurrence relationship between the topical word and named entity:

$$I(e,w) = \sum_{X_e, X_w} P(X_e, X_w) log \frac{P(X_e X_w)}{P(X_e)P(X_w)} \tag{1}$$

Where X_e and X_w are two binary random variables corresponding to the presence/absence of named entity e and topical word w in each query session. P($X_e X_w$) is the joint probability distribution function of X_e and X_w. For example, P(X_e = 1, X_w = 1) can be calculated as the proportion of the user sessions in which e and w are both present. P(X_e) and P(X_w) are the marginal probability distribution functions of X_e and X_w, respectively. For example, P(X_e = 1) is defined as the proportion of queries in which e is present, and P(X_e = 0) is defined as the proportion of user queries in which e is absent. Similarly, P(w) has the same meaning as P(X_e). Generally speaking, the greater is the value of I(e,w), the greater probability that a query containing e or w has news intent. We took the top 10 values of I(e,w) as the weight of this feature for each query. To analyze queries in depth, we adopted ICTCLAS[2] to segment query words or terms and to recognize person, organization, as well as place entities automatically. In addition, we removed the stop words for each query.

Topical similarity among clicked results. In most instances, a user clicks a result page only when he or she thinks this page is related to his/her intent; thus, the clicked results are a significant source for understanding query intent. On the basis of conclusion (i.e. news queries are supposed to obtain information for a particular topic) drawn from Figure 3 and the findings (i.e. several news articles wrote the same story during the duration of a news event) indicated by the Claypool *et al.* (2001), we assume that if a query contains news intent, then there may exist a topical or theme similarity among contents of the clicked results to a certain degree. In view of this, we tried to crawl top 20 clicked result pages in the past three days for each given query to verify our assumption. Having collected a large sum of web pages, data processing jobs like web page cleanup, text extraction, Chinese word segmentation, stop word removal and entity identification were performed. In the process of entity identification, we considered four major types of entities: person, organization, location and time entities. Hence, the ICTCLAS was used to recognize the former three and the last one was identified by regular expression.

In this study, the topical similarity between two web pages is in Equation (2). sim (a_1,a_2) denotes the content similarity between two web pages. tf(t,a_2) is the value of term frequency when term t is in the web page a_2, calculated in Equation (3). count(t in a_2) refers to the occurrence frequency of t in text a_2 and size(a_2) means to the total number of terms in text a_2. The calculation of idf(t) (inverse document frequency) is shown in Equation (4), where count(a_2 has t) refers to the number of documents containing t. Intuitively, two documents would share both named entities as well as topic terms if they are on the same topic (Kumaran and Allan, 2004). Therefore, when calculating the value of sim(a_1, a_2), the word representing an entity is more important than other words. Accordingly, boot(t) in Equation (2) is a boost constant to increase the importance of entity, and it equals to 3 when a term is an entity (e.g. person, organization, location or time named entities), otherwise equals to 1. The higher the value of *sim*(a_1,a_2) is, the more likely the query has news intent:

$$\mathrm{sim}(a_1, a_2) = \sum_{t \in a_1} [tf(t, a_2) \times idf(t) \times boost(t)] \tag{2}$$

$$tf(t, a_2) = \frac{count(t\ in\ a_2\ )}{size(a_2)} \tag{3}$$

$$idf(t) = 1 + \frac{\log N}{count\ (a_2\ has\ t)} \tag{4}$$

Temporal burstpoint. When a news event happens, it is possible that many users search for related information which makes the news-related queries popular (Sun and Hu, 2011). Consequently, we exploited the query popularity over time (Kulkarni *et al.*, 2011) to get classification feature for predicting news intent. Figure 5 shows the probability distribution over time of news query "陆幽(Lu You)" and non-news query "乒乓球视频(Ping-Pong video)". In Figure 4, horizontal axis represents the specific day in June 2008, and vertical axis represents the corresponding probability value, which is calculated by the ratio of the query frequency in one day and in a month. "陆幽 (Actor Lu You)" (there was some gossips news about the actor of Lu You at that time) is news queries, while "乒乓球视频 (Ping-Pong video)" is non-news queries in the labeled data set. As presented in this figure, the curve of news queries probability distribution fluctuates acutely. For news query "陆幽 (Actor Lu You)", the probability distribution curve from June 1 to June 11 is smooth but become a burst point on June 12, leading a violet fluctuation. Within the next few days, there exists a peak and then restores stable in the following days. It is not surprising to find this trend because news does not fade away quickly. In this study, the peak phase duration of a query is referred as the duration time of news event. For discovering regions of burst point in a sequence, our approach is based on the computation of the Moving Average (MA) (Vlachos *et al.* (2004). Through statistical analysis, we find that the average duration time of news event is five days. Specifically, given a query, we assume that if there is a burst point during the first four days before the query time, then the query may have news intent. We used Boolean value to represent this feature value, namely, if a query has burst point, then this feature value is 1; otherwise, it is 0.

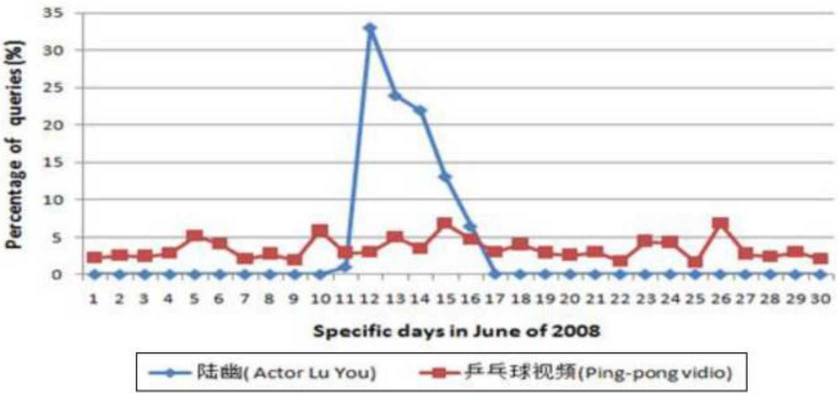


Figure 5.
The probability distribution over time of news query "陆幽 (Lu You)"and non-news query "乒乓球视频(Ping-Pong video)"

Applicability of existing features

We implemented features utilized by Ghoreishi and Sun (2013) which are mined from the current search results return by search engines. Because of the clicked document recorded in the query logs can often express what users really want after issuing a query at that time, we think the clicked documents can be more relevant to users' intent compared to current search results in our work. Consequently, we used top 20 clicked results instead of search results returned by search engines, that is, to construct query profile of query in this paper. In the following, the possible contextual and temporal features are derived from the query string and query profile.

Contextual features. Topical Specificity and Cohesiveness. This set measures the topic specificity and cohesiveness of the clicked results. (1) Topic specificity (*ts*): It is quantified using query profile clarity. It is the Kullback-Leibler (KL) divergence [shown in Equation (5)] between the word distribution estimated from the query profile and the word distribution of the document stream, where P(w|D_q) is estimated by the relative document frequency of word *w* in query profile D_q, similarly for P(w|S). (2) Topic cohesiveness: It is evaluated using two features: centroid-based cohesion (*tcc*) that is computed by the averaged cosine similarity between a document w∈$\mathrm{D_q}$ and the centroid of D_q and pairwise similarity (*tcp*) that is the average Jaccard coefficient of the set of words contained in each pair of documents.

$$\text{Clarity}\left(D_q\right) = \sum_{w \in D_q} P\left(w|D_q\right) log_2 \frac{P\left(w|D_q\right)}{P(w|S)} \tag{5}$$

Named Entities and Newsworthiness. A news event is something that happens in a certain place at certain time. Documents clicked for a news-related query are highly probable to have keywords related to location, organization or person. Therefore, we consider the following features: number of entities under each category and all three categories, namely, number of person named entity (*npe*), number of location named entity (*nle*), number of organization named entity (*noe*) and number of all the three types of named entity (*nae*). Number of distinct person named entity (*ndpe*), number of distinct location named entity (*ndle*), number of distinct organization named entity (*ndoe*) and number of distinct all the three types of named entity (*ndae*).

String's newsworthiness. The newsworthiness of a string *s* is the probability of *s* appears in the clicked news pages that contain the string s. We computed the newsworthiness of the NEs in the three categories (i.e. person, location and organization) respectively as three features. Namely, newsworthiness of person named entity (*nwpe*), newsworthiness of location named entity (*nwle*), newsworthiness of organization named entity (*nwoe*) and newsworthiness of all three types of named entity (*nwae*).

Recurrent Event Seed Words. Two features are derived based on the seed words. Query string recurrence (*qsr*) is a binary feature indicating whether any token in a query string. Query profile recurrence (*qpr*) is the total number of appearances of seed words in a query profile. The query logs we use are derived from a Chinese search engine, and most queries typed are related to some event in China; thus, the seed words (referring to events in American) listed in the study (Ghoreishi and Sun, 2013) are not applicable in this experiment. In our study, we labeled 80 seed words, and each seed word was a token that is generalized and highly meaningful to be recurrent over time, not particularly for specific places or groups of people, location and nation. Table II lists 30 of the 80 seed words in our experiments. And these seed words mainly cover some topics which news queries tend to be as listed in Figure 4, such as sports, entertainment, politic, education, economic, etc.

Query String Frequency (qsf). The frequency of a string in the past three days.

Temporal features

Temporal Features. Two features were considered to describe the temporal characteristics of a query profile, name recency and temporal clarity. *Recency(tr)*. It reflects the freshness of the clicked results query profile D_a with respect to the query time q.t. Recency measure is the median time difference between the query time and the time stamp of the documents in a query profile. *TemporalClarity (tc)*. As shown in Equation (6), it measures the difference between the temporal distribution of clicked documents for a query and the temporal distribution of documents in the document stream. In this equation, P(t|D_q) is the relative document frequency of documents created in time window t, which is set to be one day in our experiment:

$$\text{TClarity}\left(D_q\right) = \sum_{\min(d,t-q.t)}^{\max(d,t-q.t)} P\left(t\ |D_q\right) log_2 \frac{P\left(t|D_q\right)}{P\left(t|S\right)} \tag{6}$$

In total, we get 20 contextual features and three temporal features. What is more, what we should point out here is that the three features proposed by us, that is, co-occurrence between entity and a topical word in query sessions, topical similarity among clicked results and temporal burst point feature are abbreviated as *qe, cs* and *qsb*, respectively, in the following section.

Experiments and results analysis

In the previous section, we have extracted several features for predicting news intent. In this section, we evaluate the effectiveness of these features by using our benchmark query sets.

Evaluation methods

We used three evaluation metrics including precision, recall and F_1 measures to evaluate the effectiveness of the query type identification task. These metrics were calculated separately for two kinds of queries (news and non-news queries) and then were averaged (i.e. we used macro-average). We used a tenfold cross-validation in our experiments. Particularly, the annotated queries are randomly partitioned into ten equal size subsamples, and a single subsample is retrained as the validation data for testing model, and the remaining nine subsamples are used as training data. The cross-validation process is then repeated k times,

Table II.
News event seed words

结婚 (marriage)	事件 (events)	去世 (death)	离婚 (divorcement)	高考 (national entrance examination)
地震(earthquake)	火炬(torch)	案发(incidents)	阅卷(scoring test papers)	绯闻(gossip)
战争(war)	火灾(file)	会议(meeting)	台风(hurricane)	冲突(conflict)
打架(flight	发布(release)	报告(report)	改革(reform)	凶犯(perpetrators)
袭击(attack)	国务院(State Council)	教育部(Ministry of Education)	央视(CCTV)	演唱会(concert)
楼盘(houses)	房交会(Housing fair)	NBA	爆炸(explosion)	油价(petroleum price)

with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged to produce a single estimation and presented the classification accuracy obtained for taxonomies. C_{news}, C_{nonews}, C_{all}, which implied three different tasks: distinguish news from non-news queries, non-news from news queries and identify news intent between the two possibilities. To design good classifier for query intent prediction, we investigated which classifier is suitable our application. More specifically, we used the SVM^{light} implementation[3] of Support Vector Machines, the Weka[4] implementations of Bayesian Logistic Regression (Bayesian LR), Multinomial Logistic Regression (Multinomial LR) and Naïve Bayes classifiers (NB). For SVM, Radial basis function kernels were used for its better accuracy over other kernels. Default settings were adopted for other parameters.

Experimental results

After generating all the above features for our benchmark queries, we first applied contextual and temporal features separately to determine their impacts. We then studied the impact of each of them by removing it from the whole set of features.

Feature analysis

First, classification performances obtained using different classifiers are provided in Table III. From these results, we can observe that the general performances of the classifiers are good with utilization of the proposed set of features. It suggests that the news intent of the query is predicted with high accuracy. Specifically, the best classification precision and recall is 0.8621 and 0.8680, respectively. Furthermore, these results also suggest the feasibility of mining features for news intent prediction just from query logs. What is more, the experimental results of four classifiers are different. Overall, the best result was gained by SVM classifier, with macro-average F1 of around 0.8677, and it significantly outperforms results obtained by other classifiers. This result is consistent with the findings reported in previous work (Baeza-Yates and Calder¨on-Benavides, 2006; Lu *et al.*, 2006) that SVM Function shows better performance in query categorization than other classifiers. The performance of classification algorithms is affected by the randomness and sparseness of the actual data set, the size of the data set and the number of independent features. Moreover, these factors can also explain why our result is different from in the work of Ghoreishi and Sun (2013), which showed that the Multinomial LR and Bayesian LR achieved better accuracy than other two classifiers. In particular, the classifiers with default parameters and no tuning played an important role. Since we focus on the ability to treat the classifier as a "black-box", squeezing out every bit of performance by tweaking the

Table III. Classification performances obtained using different classifiers

Classification		F_1			Precision			Recall	
method	C_{all}	C_{news}	C_{nonews}	C_{all}	C_{news}	C_{nonews}	C_{all}	C_{news}	C_{nonews}
Bayesian LR	0.8271	0.8241	0.8301	0.8142	0.8014	0.8233	0.8392	0.8481	0.8370
Multinomial LR	0.8426	0.8399	0.8453	0.8355	0.8452	0.8236	0.8478	0.8347	0.8682
Naïve Bayes	0.8467	0.8367	0.8390	0.8261	0.8270	0.8343	0.8433	0.8466	0.8438
SVM^{Light}	0.8677*#Δ	*0.8692*	*0.8661*	*0.8621*	*0.8730*	*0.8549*	*0.8680*	*0.8654*	*0.8776*

Notes: The best scores in each column are type-set boldface. The symbols "*", "#" and "Δ" in the superscript in the table indicates statistical significance (two-tailed t-test, $p < 0.05$) with respect to the Bayesian LR, Multinomial LR, NB, for the macro-average F_1 scores, respectively classifiers' parameters is beyond the scope of this work. In conclusion, we choose SVM as the base classifier for our framework.

Table IV represents the values of news intent prediction by applying different set of features in isolation in data sets. Notice that, the larger the accuracy obtained by using a set of features, the greater possibility they are good for news intent prediction. Overall, among all features evaluated contextual features are significantly more effective than temporal features, and this result is in agreement with that obtained in the work of Ghoreishi and Sun (2013). Specifically, these results may be due to three factors that may affect the performance of classification algorithms. First, the contextual set includes more features than the temporal set. Second, because of the characteristic vocabulary, contextual features can describe users' real intent more directly than temporal features. Third, contextual set considers the features derived from the entities but the temporal set does not. Some entities such as famous people and organization are often involved in emerging news events, and they are always related with news intent.

Meanwhile, Table V indicates the classification accuracy obtained by removing different individual features using SVM$^{\text{Light}}$. Note that, in these cases, the smaller the accuracy obtained after removing a feature, the greater its capability to provide correct decisions on whether a query has news intent. From Table V, we can see that the impact of the feature "temporal burst point (*qsb*)" is the most significant. This result suggests that temporality is an important characteristic of news queries. Moreover, *qe is* the second most effective feature, indicating that query session is an important source for extracting feature of news intent. What is more, this feature is more important than the features extracted through computing string's newsworthiness (e.g. *nwpe, nwle, nwoe* and *nwae*) and the features

Table IV. The macro-averaged values of news intent prediction with different sets of features

Feature set	SVM$^{\text{Light}}$	BayesianLR	Multinomial LR	NB
Temporal features	0.6698	0.6621	0.6213	0.6001
Contextual features	0.7691*	0.7491*	0.7131*	0.6913*

Notes: The symbol "*" in the superscript in the table indicates statistical significance (two-tailed t-test, $p < 0.05$) with respect to the temporal features

Table V. Classification accuracy obtained by removing different individual features using SVM$^{\text{Light}}$

Feature removed	Macro-averaged value	Feature removed	Macro-averaged value
None	0.8677	nwpe	0.8567
ts	0.8597	nwle	0.8551
tcc	0.8589	nwoe	0.8565
tcp	0.8600	nwae	0.8543
npe	0.8673	qsr	0.8512
nle	0.8674	qpr	0.8606
noe	0.8673	qsf	0.8514
nae	0.8600	tr	0.8498
ndpe	0.8664	tc	0.8502
ndle	0.8668	qe	0.8434
ndoe	0.8664	qsb	*0.8379*
Ndae	0.8569	rs	0.8412

Notes: The line none represents the combination of all features. The best score is type-set italic explored by judging the content contain of corresponding event seed words (e.g. *qsr*). These results prove that query sessions can be a more useful source for mining features of news intent prediction, because users' real intent is always recorded truthfully in query log sessions.

In addition, the feature *rs* is more useful than other features measuring the similarity between the clicked documents – i.e. *ts, tcc* and *tcp.* The possible reason is that *rs* treats the named entity word differently from the non-named entity words when calculating similarity among clicked documents, while the other three features treat named entity and non-named entity words equally. From these results, we can draw the conclusion that named entity can play a more important role than other words in describing a news event.

Comparison with previous research results

We also carried out a comparative study on the performance of news intent identification. Table VIII-VI indicates the comparison results among previous research features. The "*Baseline*" stands for the method that use the prior features listed in the previous section, and "*Baseline* + *qe*", "Baseline + *cs*", "*Baseline* + *bst*" and "*Baseline* + *qe* + *cs* + *bst*" refer to the methods that combine the prior features with new feature "*qe*", "*cs*", "*bst*" and all three of them. From these results, the methods "*Baseline* + *qe*", "*Baseline* + *cs*" and "*Baseline* +*bst*" significantly outperform the method "*Baseline*". These results indicate that any of three features can improve the classification accuracy of news intent prediction. In addition, the method "*Baseline* + *qe* + *cs* + *bst*" achieves the best performance, with macro-average F1 of around 0.8677 and significantly outperforms all other methods.

The results suggest that the method using all of the three proposed new features and prior features simultaneously is better than the method just using previous features or combining only one new feature with prior features. Our proposed features are useful and significant in improving the accuracy of news intent prediction.

The effect of each our proposed features on other prior features

As shown in Table VI, the three new features are all effective in the classification task studied here. It is also interesting to verify whether each of these three features can increase the discriminatory nature of other features. To better understand how a new feature impacts other features, we combined each one of other features with this new feature. The results are presented in Tables VII-IX. It is worth noticing that the results in these tables are obtained over using features isolation; more specifically, they equal to the method combing the new

Research method	F_1	Precision	Recall
Baseline	0.8042	0.7853	0.8241
Baseline + qe	0.8220*	0.8034	0.8427
Baseline + cs	0.8234*	0.8070	0.8410
Baseline + bst	0.8356*	0.8295	0.8424
Baseline + cs + qe + bst	*0.8677*$^{*\#\Delta\dagger}$	0.8642	0.8680

Notes: The best scores in each column are type-set boldface. The symbols "*", "#"," Δ" and "†" in the superscript in the table indicates statistical significance (two-tailed *t*-test, $p < 0.05$) with respect to the "Baseline", "*Baseline* + *qe*", "*Baseline* + *cs*", "*Baseline* + *bst*" and "*Baseline* + *cs* + *qe* + *bst*", respectively

Table VI. Comparison among previous research methods using $\mathrm{SVM}^{\mathrm{Light}}$

Table VII.
How the feature "*qe*" impact other features

Feature	Gain (%)	Feature	Gain (%)
Ts	2.5	nwpe	*19.4*
tcc	1.7	nwle	*20.4*
tcp	2.2	nwoe	*21.4*
npe	7.6	nwae	*29.7*
nle	6.5	qsr	*23.3*
noe	10.4	qpr	12.1
nae	8.0	qsf	12.3
ndpe	10.8	tr	*11.6*
ndle	8.8	tc	4.6
ndoe	7.6	qsb	13.4
ndae	5.5	rs	12.6

Notes: Gains are calculated over using the features in isolation and utilizing taxonomy as well as classifier SVM$^{\text{Light}}$. The five best scores are type-set italic

feature with another feature against the method just using the feature alone. It should be noted that the greater value of results in Tables VII-IX, the more significantly that the new feature impact another feature.

As shown in Table VII, the five most improved features (i.e. *nwpe, nwle, nwoe, nwae* and *qsr*) after the combination with *qe* are those features extracted from query expression, indicating that the combination of features extracted from the sessions can improve the accuracy the features extracted from the query when identifying news intent. Table VIII shows that the five features (i.e. *qsf, tf, tc, qe* and *rs*) most affected by the feature *qsb* are related to time; hence, the feature *qsb* can improve the accuracy of features involving query frequencies or temporal property. In addition, Table IX indicates that the feature "rs" affects accuracy of the five features (i.e. *ts, tcc, tcp, tr* and *tc*) more than other features, representing that the feature *rs* can improve the accuracy of features used to calculate the similarity among clicked pages.

Table VIII.
How the feature "*qsb*" impact other features

Feature	Gain (%)	Feature	Gain (%)
ts	10.5	nwpe	1.40
tcc	8.7	nwle	13.4
tcp	7.2	nwoe	13.4
npe	6.6	nwae	12.3
nle	9.5	qsr	11.3
noe	10.4	qpr	12.1
nae	10.0	qsf	*31.7*
ndpe	11.8	tf	*21.6*
ndle	12.8	tc	*23.6*
ndoe	13.6	qe	*24.4*
ndae	10.5	rs	*15.6*

Notes: Gains are calculated over by using features in isolation and utilizing taxonomy *as well as classifier* SVM$^{\text{Light}}$. The best five scores are type-set italic

Feature	Gain (%)	Feature	Gain (%)
ts	*22.5*	nwpe	6.4
tcc	*18.8*	nwle	1.3
tcp	*19.6*	nwoe	3.2
npe	10.4	nwae	2.3
nle	2.3	qsr	11.2
noe	1.2	qpr	12.1
nae	1.0	qsf	1.9
ndpe	5.8	tr	*15.3*
ndle	2.4	tc	*14.2*
ndoe	2.3	qe	13.3
ndae	2.3	qsb	2.3

Notes: Gains are calculated over by using features in isolation and utilizing taxonomy as well as SVM$^{\text{Light}}$. The best five scores are type-set italic

Table IX. How the feature "*rs*" impact other features

Conclusions

In this study, we have addressed the issue of automatic identification of news intent by exploiting three new classification features. We try to verify the effectiveness of predicting news intent of a query using contextual and temporal-based features derived from a general search engine query log without using any social media and news data sets. We first annotated sampled queries, and then goals and topics of news intent are analyzed on the basis of a human-annotated collection. Three news features, that is, the relationship between entity and contextual words extended from query sessions, topical similarity among clicked results and temporal burst point are obtained. We train four classifiers (i.e. Support Vector Machine, Naive Bayes, Multinomial Logistic Regression and Bayesian Logistic Regression) by combining our features with features proposed in previous work to predict news intent of a query. The results of experiments indicate that three newly proposed features are effective in identifying queries with news intent, being supported by promising results with macro average F_1 of 0.8677.

This study can be extended in a variety of directions. Some future work is proposed:

- we intend to improve the accuracy of recognizing named entities in queries, as the segmentation tools we used were effective for long text (e.g. document) than for a short text as a search query with little contextual information;
- it is worthwhile to try to explore the query words distribution over time;
- to improve the accuracy of the labeling work, it is be necessary to use the idea of crowdsourcing in the future work; and
- it is essential to evaluate the effectiveness of our methods in other search logs such as AOL and MSN query log.

Notes

1. www.sogou.com/labs/dl/q.html
2. www.ictclas.org/
3. www.svmlight.joachims.org/
4. www.cs.waikato.ac.nz/ml/weka/

References

Allan, J., Papka, R. and Lavrenk, V. (1998), "On-line new event detection and tracking", *Proceedings of 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY*.pp. 37-45.

Arguello, J., Diaz, F., Callan, J. and Crespo, J.F. (2009), "Sources of evidence for vertical selection", *International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY*, pp. 315-322.

Baeza-Yates, R. and Calder'on-Benavides, L. (2006), "The intention behind web queries", *Proceedings of the 13th international conference on String Processing and Information Retrieval, Springer-Verlag, Berlin*, pp. 98-109.

Bar-Ilan, J., Zhu, Z. and Levene, M. (2009), "Topic-specific analysis of search queries", *Proceedings of the 2009 Workshop on Web Search Click Data, ACM, New York, NY*, pp. 35-42.

Brants, T., Chen, F. and Farahat, A. (2003), "A system for new event detection", *Proceedings of 26th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY*, pp. 330-337.

Brenes, D.J., GayoAvello, D. and Perez-Gonzalez, K. (2009), "Survey and evaluation of query intent detection methods", *Proceedings of the Workshop on Web Search Click Data, ACM, New York, NY*, pp. 1-7.

Broder, A. (2002), "Taxonomy of web search", *ACM SIGIR Forum*, Vol. 36 No. 2, pp. 3-10.

Cai, F., Liang, S. and de Rijke, M. (2014), "Time-sensitive personalized query auto completion", *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, ACM, New York, NY*, pp. 1599-1608.

Chen, L., Hu, Y. and Nejdl, W. (2008), "Deck: detecting events from web click-through data", *Presented at the Proceedings of 8th IEEE International Conference on Data Mining*, pp. 123-132, available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4781107 (accessed 10 May 2017).

Claypool, M., Brown, D., Le, P. and Waseda, M. (2001), "Inferring user interest", *IEEE Internet Computing*, Vol. 5 No. 6, pp. 32-39.

Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 37-46.

Cui, H., Wen, J.R., Nie, J.Y. and Ma, W.Y. (2002), "Probabilistic query expansion using query logs", *Proceedings of the 11th International Conference on World Wide Web, ACM, New York, NY*, pp. 325-332.

Diaz, F. (2009), "Integration of news content into web results", *Proceedings of the Second ACM International Conference on Web Search and Data Mining, ACM, New York, NY*, pp. 182-191.

Fu, Y., Zhou, M.Q., Wng, X.S. and Luan, H. (2010), "On-line event detection from web news stream", *Presented at the Proceedings of the 5th Pervasive Computing and Applications*, pp. 105-110, available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5704083 (accessed 10 May 2017)

Ghoreishi, S.N. and Sun, A. (2013), "Predicting event-relatedness of popular queries", *Proceedings of the 22nd ACM International on Information and Knowledge Management, ACM, New York, NY*, pp. 1193-1196.

Gonzalez-Caro, C., Calderon-Benavides, L. and Baeza-Yates, R. (2011), "Web queries: the tip of the iceberg of the user's intent", *Proceedings of the 34th International Conference on Web Search and Web Data Mining, ACM, New York, NY*, pp. 282-291.

Gu, Y., Cui, J., Liu, H., Jiang, X., He, J., Du, X. and Li, Z. (2010), "Detecting hot events from web search logs", *Proceedings of the 11th International Conference on Web-age Information Management, Springer-Verlag, Berlin*, pp. 417-428.

Gui, S. and Lu, W. (2016), "WHUIR at the NTCIR-12 temporal intent disambiguation task", *Proceedings of the 12th NTCIR Conference, National Institute of Informatics, Tokyo*.

Hassan, A., Jones, R. and Diaz, F. (2009), "A case study of using geographic cues to predict query news intent", *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, New York, NY*, pp. 33-41.

He, D.Q. and Goker, A. (2000), "Detecting session boundaries from web user logs", available at: www.sis.pitt.edu/~daqing/docs/he00detecting.pdf (accessed 10 May 2017).

He, R., Wang, J., Tian, J., Chu, C., Mauney, B. and Perisc, L. (2013), "Session analysis of people search within a professional social network", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 5, pp. 929-950.

Joho, H., Jatowt, A., Blanco, R., Yu, H. and Yamamoto, S. (2016), "Overview of NTCIR-12 temporal information access (temporalia-2) task", *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Institute of Informatics*, *Tokyo*, pp. 217-224.

Konig, A.C., Gamon, M. and Wu, Q. (2009), "Click-through prediction for news queries", *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY*, pp. 347-354.

Kulkarni, A., Teevan, J., Svore, K. and Dumains, S.T. (2011), "Understanding temporal query dynamic", *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, ACM, New York, NY*, pp. 167-176.

Kumaran, G. and Allan, J. (2005), "Using names and topics for new event detection", *Proceedings of HLT 2005, Association for Computational Linguistics*, *Stroudsburg, PA*, pp. 121-128.

Kumaran, G. and Allan, J. (2004), "Text classification and named entities for new event detection", *Proceedings of 26th annual international ACM SIGIR conference, ACM*, *New York, NY*, pp. 297-304.

Landis, J.R. and Koch, G.G. (1977), "The measurement of observer agreement for categorical data", *Biometrics*, Vol. 33 No. 1, pp. 159-174.

Li, D., Liu, B., Zhang, Y., Huang, D. and Cao, J. (2016), "Using time-Series for temporal intent disambiguation in NTCIR-12 temporalia", *Proceedings of the 12th NTCIR Conference, National Institute of Informatics*, *Tokyo*, pp. 267-271.

Liu, Y., Zhang, M., Ru, L. and Ma, S. (2006), "Automatic query type identification based on click through information", *Lecture Notes in Computer Science*, Springer, Berline, Vol. 4182, pp. 593-600.

Louis, A., Crestan, E., Billawala, Y., Shen, R., Diaz, F. and Crespo, J. (2011), "Use of query similarity for improving presentation of news verticals", *presented at the Proceedings of Very Large Data Search*, pp. 62-67, available at: www.google.co.jp/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjniNqAgKPJAhVGUZQKHRsyDQ4QFgghMAA&url=http%3A%2F%2Fceur-ws.org%2FVol-880%2FVLDS-p62-Louis.pdf&usg=AFQjCNGhTTYqozHyLtxxVtCvT3jtVMHptA (accessed 10 May 2017).

Lu, Y., Peng, F., Li, X. and Ahmed, N. (2006), "Coupling feature selection and machine learning methods for navigational query identification", *Proceedings of the 15th International Conference on Information and Knowledge Management, ACM*, *New York, NY*, pp. 682-689.

McCreadie, R., Macdonald, C. and Ounis, I. (2010), "Crowdsourcing a news query classification dataset", *Proceedings of the 33 th ACM SIGIR Workshop on Crowdsourcing for Search Evaluation, ACM, New York, NY*, pp. 31-38.

McCreadie, R., Macdonald, C. and Ounis, I. (2013), "News vertical search: when and what to display to users", *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY*, pp. 253-262.

Moulahi, B., Tamine, L. and Yahia, S.B. (2016), "When time meets information retrieval: past proposals, current plans and future trends", *Journal of Information Science*, Vol. 42 No. 6, pp. 725-747.

Parikh, N. and Sundaresan, N. (2008), "Scalable and near real-time burst detection from e-commerce queries", *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY*, pp. 972-980.

Rose, D.E. and Levinson, D. (2011), "Understanding user goals in web search", *Proceedings of the 13th international conference on World Wide Web, ACM, New York, NY*, pp. 13-19.

Ruocco, M. and Ramampiaro, H. (2012), "Exploratory analysis on heterogeneous tag-point patterns for ranking and extracting hot-spot related tags", *Proceedings of the SIGSPATIAL LBSN, ACM, New York, NY*, pp. 16-23.

Sakaguchi, T. and Kurohashi, S. (2016), "Kyoto at the NTCIR-12 temporalia task: Machine learning approach for temporal intent disambiguation subtask", *Proceedings of the 12th NTCIR Conference, National Institute of Informatics*, *Tokyo*, pp, 288-292.

Sun, A. and Hu, M. (2011), "Query-guided event detection from news and blog streams", *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol. 41 No. 5, pp. 834-839. Part A,

Vavliakis, K., Symeonidis, A.J. and Mitka, P.A. (2013), "Event identification in web social media through named entity recognition and topic modeling", *Journal of Data& Knowledge Engineering*, Vol. 88 No. 2013, pp. 1-24.

Vlachos, M., Meek, C. and Vagena, Z. (2004), "Identifying similarities, periodicities and bursts for online search queries", *Proceedings of International Conference on Management of Data and Symposium on Principles Database and Systems, ACM, New York, NY*, pp. 131-142.

Wei, C., Lee, Y.H., Chiang, Y.S., Chen, C.T. and Yang, C.C.C. (2014), "Exploiting temporal characteristics of features for effectively discovering event episodes from news corpora", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 3, pp. 621-634.

Zamora, J., Mendoza, M. and Allende, E. (2014), "Query intent detection based on query log mining", *Journal of Web Engineering*, Vol. 13 Nos 1/2, pp. 024-052.

Zhang, R., Konda, Y., Dong, A., Kolari, P., Chang, Y. and Zheng, Z. (2010), "Learning recurrent event queries for web search", *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, MIT, MA*, pp. 1129-1139.

Zhao, Q., Liu, T.Y., Bhowmick, S.S. and Ma, W. (2006), "Event detection from evolution of click-through data", *Proceedings of the 12th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY*, pp. 484-493.

Further reading

Kanhabua, N., Ngoc Nguyen, T. and Nejdl, W. (2015), "Learning to detect event-related queries for web search", *Proceedings of the 24th International Conference on World Wide Web, ACM, New York, NY*, pp. 1339-1344.

Salton, G. and Buckley, C. (1988), "Term-weighting approaches in automatic text retrieval", *Information Processing and Management*, Vol. 24 No. 5, pp. 513-523.

Corresponding author

Wei Lu can be contacted at: reedwhu@gmail.com