Journal of Network and Computer Applications 43 (2014) 42–55

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Generating temporal semantic context of concepts using web search engines

Zheng Xu [a,b,*], Yunhuai Liu [a], Lin Mei [a], Chuanping Hu [a], Lan Chen [a]

[a] The Third Research Institute of the Ministry of Public Security, 339 Bisheng Road, Shanghai 201142, China
[b] Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 25 February 2013
Received in revised form
5 March 2014
Accepted 4 April 2014
Available online 5 May 2014

Keywords:
Temporal semantic context
Semantic annotation
Content analysis
Web mining

ABSTRACT

In this paper, the problem of generating temporal semantic context for concepts is studied. The goal of the proposed problem is to annotate a concept with temporal, concise, and structured information, which can reflect the explicit and faceted meanings of the concept. The temporal semantic context can help users learn and understand unfamiliar or newly emerged concepts. The proposed temporal semantic context structure integrates the features from dictionary, Wikipedia, and LinkedIn web sites. A general method to generate temporal semantic context of a concept by constructing its associated words, associated concepts, context sentences, context graph, and context communities is proposed. Empirical experiments on three different datasets including Q–A dataset, LinkedIn dataset, and Wikipedia dataset show that the proposed algorithm is effective and accurate. Different from manually generated context repositories such as LinkedIn and Wikipedia, the proposed method can automatically generate the context and does not need any prior knowledge such as ontology or a hierarchical knowledge base. The proposed method is used on some applications such as trend analysis, faceted exploration, and query suggestion. These applications prove the effectiveness of the proposed temporal semantic context problem in many web mining tasks.

1. Introduction

With the high speed development of the internet, search has emerged as a key technology to facilitate access to information for users. Millions of users submit millions of queries to web search engines such as Google[1] and Yahoo[2]. Web search engines allow users to browse on the web, find related information, or as a starting point for entertainment.

Given a new concept to the user, she/he may use the web search engines to index the web pages, which may help users to learn the concept conveniently. With sophisticated algorithms, web search engines have made accessing information easy. Some researchers (Sparrow et al., 2011) suggest that when faced with difficult questions, people are primed to think about using computers. When people expect to have future access to information, they have lower rates of recall of the information itself and enhanced recall instead for where to access it. In other words, when faced with a new concept, users prefer to use search engines rather than learn it through their own prior knowledge.

Though web search engines have become a major intermediary for seeking information, finding relevant information satisfying a user's needs based on the user's initial search queries has become an increasingly difficult task (Leung et al., 2003), which makes users.

(1) **Make costly efforts to find useful information.** In White and Drucker (2007), the authors give some statistics: about 29% of users will modify their original queries in a search task session; the average number of re-visit web pages in a search task session is 5; about 21% of operations in a search task session are backward (users revisit the web pages).
(2) **Face high cognitive burden.** The average steps of a search task session are 17.7 (White and Drucker, 2007), which means that the users browse 17.7 web pages before finishing the task.
(3) **Become lost in the search task.** The average branches of a search task session are 4.1 (White and Drucker, 2007). This is the number of times a user revisited a previous page on the trail and then proceeds forward to view another page.

Modified original queries, so many backward operations, too many browsed web pages, and inappropriate branches – such obstacles make it difficult to get the relevant and correct information. In our view, the following causes contribute to difficulties in a search task session:

* Corresponding author at: The Third Research Institute of the Ministry of Public Security, 339 Bisheng Road, Shanghai 201142, China.
E-mail address: xuzheng@shu.edu.cn (Z. Xu).

[1] www.google.com.
[2] www.yahoo.com.

http://dx.doi.org/10.1016/j.jnca.2014.04.002
1084-8045/

(1) **Caused by users.** A study by Jansen et al. (1998) proposes that the average length of a query submitted to popular web search engines is only 2.35 terms. The short queries may not be able to describe users' real information needs. Thus, the short queries given by users can hardly bring good search results. The reasons for users submitting short queries are the low length queries lessen user's cognitive burden; users face an open-ended search task; and users have difficulty in formatting proper queries (Pandit and Olston, 2007).
(2) **Caused by search engines.** Web search engines generally adopt a "one-size-fits-all" approach for search results presentation, which does not consider the personal need of the users. Different users need different information from the same queries.
(3) **Caused by concepts.** The short queries submitted by users are usually represented by ambiguous concepts. For example, the user may submit "apple" when she/he wants to buy a product produced by Apple Computer Company. But the concept usually has different meanings. For example, the concept "apple" can be fruit or a computer company. Besides the diverse semantics of the concept, new meanings of the old concept and newly emerged concepts often lessen the accuracy of the search results.

In order to provide an accurate annotation for a concept, the problem of automatically generating temporal semantic context (TSC) for concepts is studied. Explicit and concise information for a concept is provided, which indicates the semantics and hidden meanings of the concept. What is a good semantic context of a concept? Let's see two examples from the Cambridge Advanced Learner's Dictionary and Wikipedia,[3] which are shown in Fig. 1. In Fig. 1, the semantic context of a concept is structured as follows. First, the definition of the concept is presented. Second, some example sentences are given to show the usage of the concept. In addition, a visual thesaurus graph of the concept is given to show some related concepts. Wikipedia provides the disambiguated meaning of the concept besides the definition of the concept. For example, Wikipedia gives the link to the Apple Inc. which is another meaning of the concept "apple".

Analogously, if the structured and semantic related information of a concept are provided, it will be very helpful for her/him to understand and further explore it. Of course, when a concept is temporally changed, the new meaning may add to the concept. For example, "Gangnam" is a region in Seoul, Korea. But with the popularity of the song "Gangnam Style" recently, the concept "Gangnam" may be related to the popular music. Thus, the temporal feature to the semantic context must be added. So, what is a good temporal semantic context? Let's see another example from LinkedIn,[4] which is shown in Fig. 2. In Fig. 2, the experience of "Barack Obama" is listed by time sequence. In different time intervals, the concept "Barack Obama" has different semantic context. Thus, inspired by the annotations from the dictionary, Wikipedia, and LinkedIn, the temporal semantic context should include:

(1) **Example sentences**. Given a concept, the example sentence can help the users understand the context of the concept. Moreover, the example sentences can help users apply the concepts in a real context. This factor can be found from the dictionary.
(2) **Diverse meanings.** Given a concept, the different meanings should be given, which can help users learn and explore the concept. This factor can be found from Wikipedia.
(3) **Semantically related concepts.** Similar to the synonyms or thesauri in a dictionary, related concepts should be added to the temporal semantic contexts.
(4) **Temporal annotations.** In different time intervals, the concept may have different meanings. The appropriate semantic context in different time intervals should be mined. This factor can be found from LinkedIn.

To the best of our knowledge, the temporal semantic context of concepts has not been well addressed in existing work. The detailed analysis of the existing work will be given in the next section. The major contributions of this paper are as follows.

(1) In this work, the problem of generating temporal semantic context of concepts is proposed. A general method to automatically generate structured temporal contexts of a concept including semantically related words, example sentences, diverse meanings, and temporal annotations is given. The proposed TSC structure integrates the features from dictionary, Wikipedia, and LinkedIn web sites, which is helpful for users to understand and explore the concept.
(2) Empirical experiments on three different datasets including Q–A dataset, LinkedIn dataset, and Wikipedia dataset show that the proposed algorithm is effective and accurate. Different from manually generated context repositories such as LinkedIn and Wikipedia, the proposed method can automatically generate the context. Moreover, the proposed method does not need any prior knowledge such as ontology or a hierarchical knowledge base such as WordNet.[5]
(3) Some applications using the proposed TSC method are given. The proposed method can be used on trend analysis, faceted exploration, and query suggestion. These applications prove the importance of the proposed TSC problem in many web mining tasks.

The rest of the paper is organized as follows. The related work is given in Section 2. In Section 3, the problem of TSC is formally defined and a series of definitions is given. In Section 4, how to generate the TSC of a concept by web search engines is introduced. Our experiments and results are discussed in Section 5. In Section 6, three applications using the proposed TSC method are introduced. Finally, some conclusions are given.

2. Related work

To the best of our knowledge, the problem of temporal semantic context has not been well studied in existing work. In this section, the related work of the proposed method is given: the recent work of semantic annotations and temporal context.

In the semantic annotation field, with the explosion of community contributed multimedia content available online, many social media repositories (e.g., Flickr,[6] YouTube, and Zooomr[7]) allow users to upload media data and annotate content with descriptive keywords which are called social tags. These tags can be seen as a type of semantic context of the objects such as images or videos. Considering usage patterns and semantic values of social tags, Golder (Golder and Huberman, 2006) mined usage patterns of social tags based on the delicious dataset.[8] Davis (Al-Khalifa and Davis, 2006) concluded that social tags were semantically richer than automatically extracted keywords. Suchanek (Suchanek et al., 2008) used YAGO and WordNet to check

[3] en.wikipedia.org/wiki.
[4] www.linkedin.com.
[5] wordnet.princeton.edu.
[6] www.flickr.com.
[7] www.zooomr.com.
[8] www.delicious.com.

Fig. 1. The annotation from the dictionary.

Fig. 2. The annotation from LinkedIn.

the meaning of social tags and concluded that top tags were usually meaningful. Halpin (Halpin et al., 2007) examined why and how the power law distribution of tag usage frequency was formed in a mature social tagging system over time. These research efforts focus on mining the usage patterns from social tags. Different from these works, the proposed method aims at generating appropriate annotations of concepts. Recently, many researchers investigated the applications of social tags in information retrieval and ranking. In Ramage et al. (2009), the authors empirically study the potential value of social annotations for web search. Zhou (Zhou et al., 2008) proposed a model using latent allocation, which incorporates the topical background of documents and social tags. Xu (Xu et al., 2007) developed a language model for information retrieval based on the metadata property of social tags and their relationships to annotated documents. Bao (Bao et al., 2007) introduced two ranking methods: (1) SocialSimRank, which ranked pages based on the semantic similarity between tags and pages, and (2) SocialPageRank, which ranked returned pages based on their popularity. Schenkel (Schenkel et al., 2008) developed a top-k algorithm which ranked search results based on the tags shared by the user who issued the query and the users who annotated the returned documents with the query tags. Radinsky (Radinsky et al., 2011) proposed to use temporal information as a complementary source of signal to detect semantic relatedness of words. The temporal semantic relatedness of two words can be computed in the different time intervals. The above annotation methods are entered manually or semi-automatically. In this paper, the TSC of the objects are generated by web search engines, which are automatically and temporally changed.

In the temporal context field, to the best of our knowledge, there is no similar work in the related literature. In the computer vision field, temporal context is used for video annotation (Li et al., 2010; Jannach and Leopold, 2007), object recognition (Liu et al., 2011), and target tracking (Nguyen et al., 2007). These works use temporal or spatial context on image or videos, which is different from the proposed work. The proposed works focus on the concept instead of image or videos. In the text mining field, temporal context is used for text processing (Weng et al., 2009; Choi et al., in press), text segmentation (Dumont and Quenot, 2012), and text representation (Glaser and Zelnik-Manor, 2011). These works aim at using temporal context on text instead of concept. Manica (Manica et al., 2012) presented a comprehensive study that describes the evolution of search engines on the exploitation of temporal information. That work mentioned the importance of the temporal feature of given information. In this paper, besides the temporal feature, the faceted feature of a given concept is also considered. Milea (Milea et al., 2008) introduced a new OWL-based temporal formalism for the representation of time, change, and state transitions. Based hereon Milea presented a financial Web-based application centered on the aggregation of stock recommendations and financial data. That work integrates the temporal feature into the semantic web (Berners-Lee et al., 2001; Zhuge, 2011; Luo et al., 2011) field, which is different from the research scope of the proposed work. In the cognitive science (Howard et al., 2011), a predictive temporal context model used the definition of temporal context to generate semantic memory representations. In the web search field, Boughareb (Boughareb and Farah, 2011) used the search log data as the temporal context for identifying the user search needs. Hwang (Hwang et al., 2011) proposed an enrichment of semantic relation network and used it for word sense disambiguation. Leung (Leung and Ng, 2008) proposed a content-based personalized search method. That method builds a graph structure of the search session. Different from these methods, the proposed method considers both faceted and temporal features, and does not need any additional data such as search log or user need.

3. Problem formulation

In this section, the basic definitions of the problem of temporal semantic context (TSC) are formally given.

Let$C=\{w_1,w_2,...,w_{|C|}\}$be a concept containing a set of words, which can be a query to the web search engine. Let$S=\{s_1,s_2,...,s_{|S|}\}$be a set of documents related to the concept. Based on these two sets, five basic definitions for semantic context of concept C are formulated.

Definition 1. Associated Word (AW): An associated word co-occurs frequently with the concept C, which appears frequently in the documents set S. The set of associated words of the concept C is denoted as

$$AW_C=\{aw_1,aw_2,...,aw_{|AW_C|}\}. \quad (1)$$

Definition 2. Associated Concept (AC): An associated concept co-occurs frequently with the concept C, which appears frequently in the documents set S. The associated concepts are composed of some associated words. The set of associated concepts of the concept C is denoted as

$$AC_C=\{ac_1,ac_2,...,ac_{|AC_C|}\}. \quad (2)$$

The difference between the associated words and the associated concepts is that the concepts are a sequence of associated words. For example, the word “iPad” is an associated word of the concept “apple”. The concept “iPad mini” is an associated concept including the words “iPad” and “mini”. In this paper, an associated concept contains more than one word.

Definitions 1 and 2 are similar to the thesaurus of the dictionary in Fig. 1.

Definition 3. Context Sentence (CS): A context sentence contains concept C and a sequence of words, which appears in documents set S. The set of context sentences of the concept C is denoted as

$$CS_C=\{cs_1,cs_2,...,cs_{|CS_C|}\}. \quad (3)$$

Definition 3 is similar to the example sentence of the dictionary.

Definition 4. Context Graph (CG): A context graph is a data structure of the associated words, which reflects an associated relation between associated words (e.g., iPhone is associated with Apple). The nodes N in the context graph are the associated words of the concept C. The edges E in the context graph are the relations between the associated words. The context graph of the concept C can be denoted as

$$CG_C=\{N,E\}$$
$$N=AW_C$$
$$E=\{e_1,e_2,...,e_{|E|}\}. \quad (4)$$

The edge e_k can be denoted as a triple$e_k=\langle aw_i,aw_j,\lambda\rangle$, which means the edge e_k is from the node$aw_i$to$aw_j$with the weightλ. In the Section 4.5, the computation method ofλwill be introduced in detail.

Definition 4 is similar to the visual thesaurus of the dictionary.

Definition 5. Context Community (CC): A context community is a subgraph of the context graph, which reflects a part of context of the concept C. The set of context communities of the concept C is denoted as

$$CC_C=\{cc_1,cc_2,...,cc_{|CC_C|}\}$$
$$\forall aw_i\in cc_i\wedge\forall aw_j\in cc_j\rightarrow aw_i\neq aw_j. \quad (5)$$

The set of context communities is a segmentation of the context graph. Each context community is a part of the context graph, but with no common associated words from another context community.

Definition 5 is similar to the disambiguation meaning of the Wikipedia.

These five basic definitions form the semantic context of a concept C, which can be used for annotation. The following example is a semantic context of the concept “apple”, which is generated from the top ten snippets by the web search engine Google.[9]

Example 1. A semantic context of the concept “apple”

Associated Words:
AW={iPhone, iPad, Mac, tree, fruit, records,…}
Associated concepts:
AC={iPhone 5, iPad mini, apple tree, apple records,…}
Context sentences:
CS_1={Apple designs and creates iPod and iTunes}
CS_2={the apple is the pomaceous fruit of the apple tree}
Context graph:
N={iPhone, iPad, Mac, tree, fruit, records,…}
$E=\{<\text{iPhone, iPad},\lambda_1>, <\text{tree, fruit},\lambda_2>,…\}$
Context communities:
CC_1={iPhone, iPad, Mac, iPod, computer,…}
CC_2={fruit, tree, rose,…}
CC_3={records, sound, Beatles,…}

Example 1 integrates the annotation features from dictionary and Wikipedia. Figure 3 gives the context graph generated by the proposed method for Example 1. Different colors of nodes mean the different communities. With the five basic definitions, the temporal feature to the semantic context of the concept C is added. Definition 6 is given as

[9] The searching results are obtained in the data 10/29/2012.

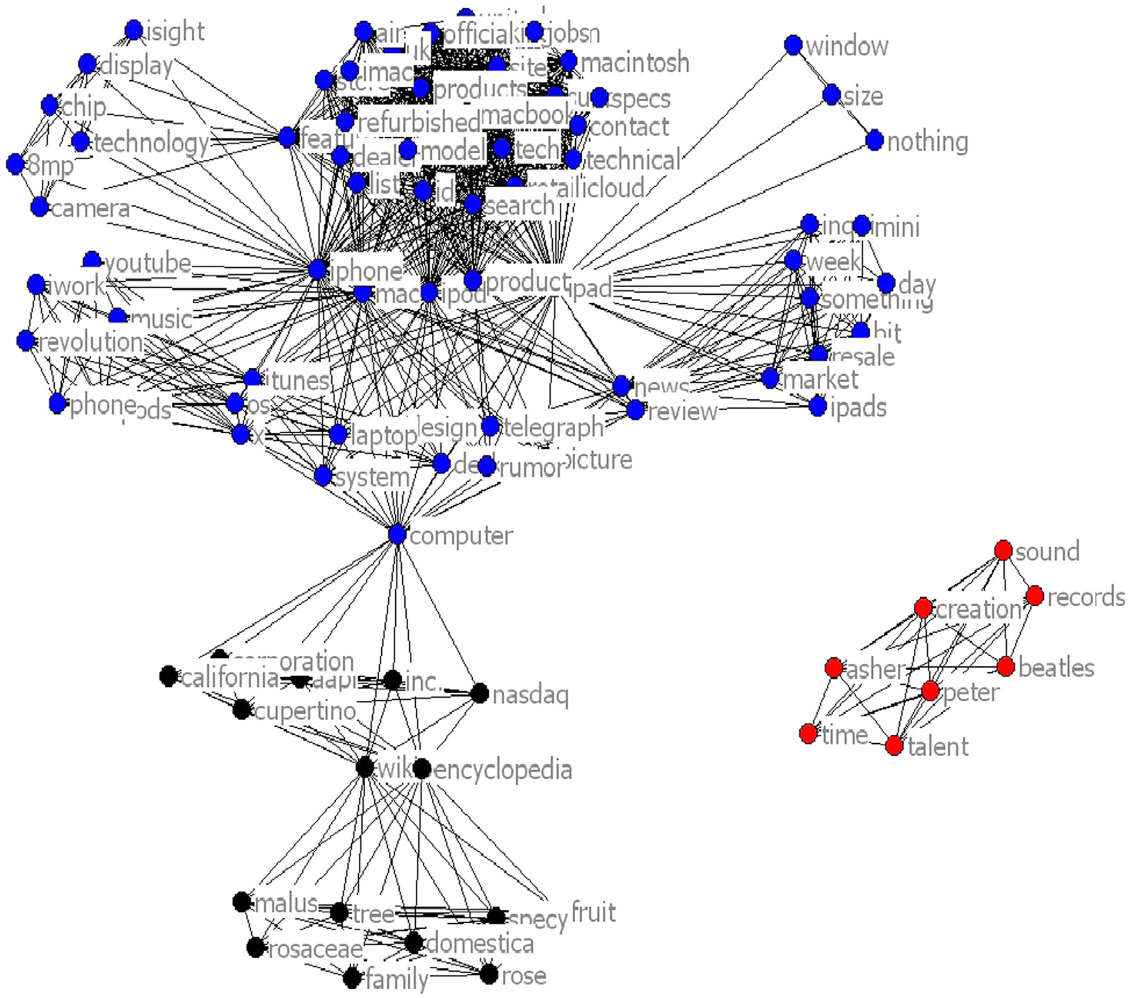


Fig. 3. The context graph of "apple".

Definition 6. Time Interval (TI): A time interval is a range from the starting timestamp t_s to the ending timestamp t_e, which is denoted as $\langle t_s, t_e \rangle$.

With the definition of time interval $\langle t_s, t_e \rangle$, the temporal feature to the semantic context of the concept C can be added. For example, the associated words of time interval $\langle t_s, t_e \rangle$ can be revised as $AW_C(t_s, t_e)$. The starting timestamp and the ending timestamp can be added to Definitions 1–5. Example 2 gives a temporal semantic context of the concept "apple" in the year 2006.

Example 2. A temporal semantic context of "apple"

Associated words:

AW(2006)={Mac, technology, media, cook,…}

Associated concepts:

AC(2006)={Apple technology, Apple media tool,…}

Context sentences:

CS (2006)={The apple media tool was a multimedia authoring tool…}

Context graph and context communities:

Illustrated in Fig. 4

Example 2 integrates the annotation features from LinkedIn, which reflect the temporal feature of the concept. With the definitions above, the problem of generating **Temporal Semantic Context (TSC)** is defined as

Input: Given a concept C and the time interval $\langle t_s, t_e \rangle$

Output: The related semantic context of the given concept C, including the associated words, the associated concepts, the context sentences, the context graph, and the context communities from the starting timestamp t_s to the ending timestamp t_e.

Of course, the proposed TSC problem is not easy, which includes three challenges:

(1) It is not easy to find the appropriate knowledge base for generating temporal semantic context. The background, the prior knowledge, or the ontology is not clear before the users provide the concept.
(2) It is not immediately clear how to select an appropriate computation method for ranking the temporal semantic context. Since the temporal semantic context may be large for a given concept, it is necessary to select the best related context for the users.
(3) It is not easy to add the time information to the semantic context. The meanings of the given concept change with time; it is necessary to select the correct context for the different time intervals.

On the other hand, the above challenges also show the advantages of the proposed method. The proposed method integrates the

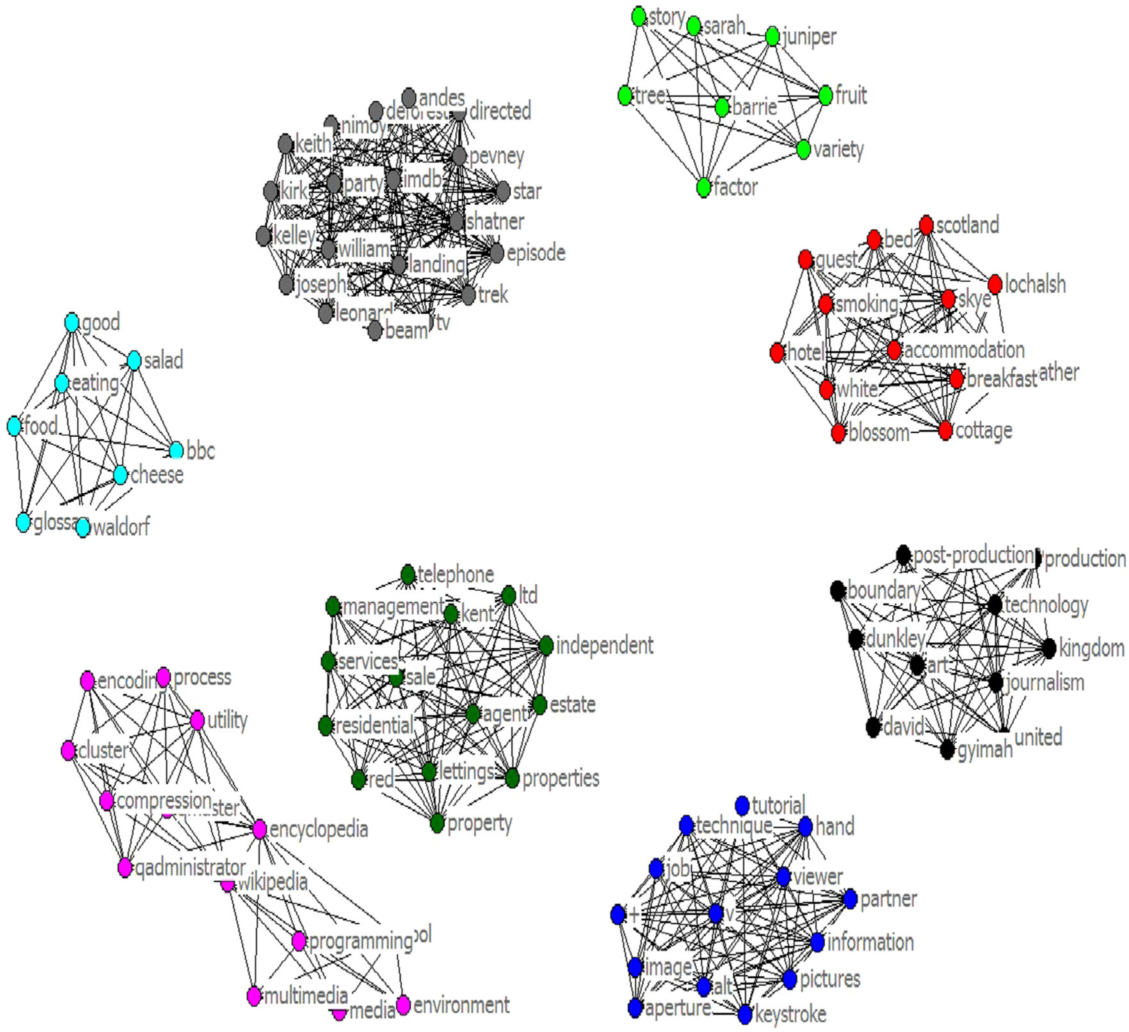


Fig. 4. The context graph of "apple" in the year 2006.

temporal feature and the semantic context of the concept and does not need any domain knowledge or manual ontology. In the following section, the method for generating temporal semantic context of the concept is given.

4. Generating temporal semantic context

In this section, the method for generating TSC of a concept is presented. First, the feasible repository for collecting contexts is introduced. Second, the methods for obtaining five basic elements of the semantic context are given. Last, the ranking mechanism for selecting an appropriate TSC is proposed.

4.1. Temporal semantic context repository

In order to generate TSC of a concept, it is necessary to choose an appropriate repository. The reasons are as follows.

(1) **Updating information rapidly.** The meaning of a concept refreshes over time. For example, the concept "US President" may be related to "George Bush" in 2007. But in 2008, "Barack Obama" was related to the concept "US President". Besides the temporal change of the concept, a new concept may appear. For example, the concept "iPad mini" occurred in October of 2012. Web search engines can keep up with the change and evolution of the concepts. This feature of web search engines can ensure the objectivity and completeness of the proposed method. Usually, the most related concepts and words are listed in the top results of web search engines.

(2) **Convenient interface for temporal features.** It is important to obtain the related information of a concept in a given time interval. Web search engines such as Google provide the interface for searching web pages in a given time interval. The temporal feature of a web search engine can ensure the proposed method will catch the new concepts.

Besides the convenient interface and appropriate updating speed, web search engines provide two important resources for generating TSC:

(1) **Page count**. Page count of a concept is the number of the web pages containing the query words. For example, the page count of "apple" is 22,900,000 on the date 10/1/2012. The page count reflects the popularity of a concept in a time interval. In the following section, the page count of a concept is denoted as P_C.

(2) **Snippet**. Snippet is a brief window of text extracted by a search engine around the query words in a web page. Processing a snippet is also efficient as it releases the trouble of downloading web pages, which might be time consuming and make parsing infeasible. In the following section, the snippets of a concept are denoted as $S_C = \{s_1, s_2, ..., s_{|S_C|}\}$.

The steps for obtaining the resources from the repository (in this paper, Google is chosen as the TSC repository) are as follows:

(1) Issue the concept C and time interval $\langle t_s, t_e \rangle$ as the query to Google.
(2) Get the page counts of the query in the time interval, denoted as $P_C(t_s, t_e)$.
(3) Get the snippets of the query in the time interval, denoted as $S_C(t_s, t_e)$.

A parsing tool for extracting the page count and snippets from the search result html pages is developed. In this paper, the top 20 Google search results[10] of the concept are parsed.

4.2. The generation of associated words

In this section, the generation of associated words is presented. In Section 4.1, the snippets of a concept are collected from the context repository. Extracting appropriate words from the snippets is needed. A weight of each word is given, which reflects the importance of the words against the concept.

Since the snippets may contain stop words such as "in", "a", the Stanford tagger[11] is used to preserve the noun words in the snippets. The noun words have real and clear meanings, which can reflect the real context of a concept. Suppose the associated words extracted from the snippets set S as

$$AW = \{aw_1, aw, \ldots, aw_m\} \tag{6}$$

where m is the number of associated words. Thus, the appearance of the associated words in snippets can be represented as a matrix

$$AWS = \begin{pmatrix} a_{11} & \ldots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}, \tag{7}$$

where n is the number of snippets.$a_{nm}$is 1 if the mth word appears in the nth snippet. The AWS matrix is a 1-0 matrix.

Besides the extraction of associated words, the weight of each word should be given to reflect which is the most associated to the concept. Some existing work such as term frequency-inverse document frequency (tf-idf) (Salton et al., 1975) is used to give a weight to the word. In our work, the snippet frequency (SF) of each word is used to rank its importance against the concept. The snippet frequency means the appearance frequency of each word in snippets. For example, if a word appears in 10 snippets from a total of 100 snippets, the snippet frequency of it is 0.1. Thus, the weight of each word is computed as

$$SF(aw_i) = \sum_{j=1}^{n} a_{ji}/n, a_{ji} \in AWS. \tag{8}$$

Using snippet frequency as the ranking schema is low complexity. The snippet frequency of each word can be computed by summing the elements in AWS matrix of its row. The top searching results are usually related to the query. The word with high snippet frequency is usually more related to the query than that of low snippet frequency.

4.3. The generation of associated concepts

In this section, the methods for generating associated concepts are presented. Different from the associated words, an associated concept is a sequence of words. Generally speaking, associated concepts are with the exact meanings. For example, "iPad" and "mini" are two associated words. In fact, "iPad mini" is an accurate concept integrating "iPad" and "mini". Besides mining associated concepts from the snippets, the weight of each associated concept is also employed.

In some research work, such as data mining, frequent patterns are itemsets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold (Han et al., 2007). A subsequence, such as first buying milk, and then an apple, if it occurs frequently in a shopping history database, is a frequent sequential pattern. In our work, an associated concept is thought of as a frequent sequential pattern. Herein, the frequent sequential pattern mining method (Han et al., 2007) is used for generating associated concepts.

Suppose the snippets set S is the transactions set of the frequent sequence patterns mining, then the associated concepts can be extracted from the AWS matrix. Besides mining the associated concepts, the ranking schema of associated concepts also should be given. Support (Agrawal and Srikant, 1994) is used to select interesting frequent patterns in the data mining area. The support of each associated concept (SUP) is employed as its weight, which is computed as

$$SUP(ac_i) = |ac_i|/n, \tag{9}$$

where $|ac_i|$ means the number of associated concepts appearing in AWS matrix. In fact, the support of an associated concept is similar to the snippet frequency. Both of them compute the appearance frequency in the snippets set.

4.4. The generation of context sentences

In this section, the method for generating context sentences of the concept is presented. Comparing with the associated words and associated concepts, the context sentence can be understood by the users more easily and clearly. The context sentence is composed of the associated words, which can reflect the relations of the words. For example, the sentence "Apple designs and creates iPod and iTunes" reflects the relation between the associated words "iPod" and the concept "apple". Users can understand the relation between the associated words and the concept by the context sentence. Of course, mining the context sentence is easier than mining the associated concepts. The string segmentation algorithm is used when scanning the end of a string. In our work, ending strings of a sentence are '.', '?', '!', and '…'. If the string segmentation algorithm scans the ending string, then the scanned string can be seen as a sentence. A context sentence can appear in different snippets, and a snippet can provide different context sentences. The weight of context sentences (WCS) can be computed by the sum of snippet frequencies of their words:

$$WCS(cs_i) = \sum_j SF(aw_j), aw_j \in cs_i. \tag{10}$$

For example, the sentence "Apple designs and creates iPod and iTunes" contains two associated words "iPod" and "iTunes". The weight of the sentence is the sum of the snippet frequency of these two words. The basic idea of Eq. 10 is that the high weight sentences should contain high weight associated words. The long sentences may have the higher weight of support. In the proposed method, the long sentences are considered more important since they provide more effective information.

4.5. The generation of context graph

The context sentences can reflect the relations of the associated words, which can help the user to understand the meaning of the concept. Unfortunately, the context sentence can only reflect the

[10] The low number of search results may lessen the recall and the high number of search results may lessen the precision.

[11] nlp.stanford.edu.

relation of a few words since the length of a sentence is limited. Therefore, in this section, the context graph is proposed, which can help the users understand all aspects of the concept.

In Definition 4, the context graph is denoted as the associated words and the edges between them. The nodes of the context graph can be obtained from the associated words set. In our previous work (Xu et al., 2011), the pointwise mutual information (PMI) (Bollegala et al., 2007) is suitable for computing the semantic similarities between a given pair of concepts. In this paper, inspired by the previous work, the weight of the edge between two associated words is computed by

$$PMI(aw_i, aw_j) = \log\left(\frac{SF(aw_i, aw_j)}{SF(aw_i)\,SF(aw_j)}\right) / \log(n), \tag{11}$$

where n is the number of the snippets. $SF(aw_i, aw_j)$ means the number of snippets containing both aw_iandaw_j. It is noted that the PMI equation is symmetrical, which means that $PMI(aw_i, aw_j) = PMI(aw_j, aw_i)$.

From Figs. 3 and 4, the nodes with high number of edges can be seen as those with the high snippet frequencies. This feature is easily understood. The associated words with high snippet frequencies appear in many snippets, which may co-occur with many different associated words. In other words, the associated words with high snippet frequency are central nodes of the context graph. Of course, the context graph can be optimized to help users understand the concept more clearly. The edges with low weight can be deleted in the context graph. Generally, the low weight edges are removed as much as possible under the condition of preserving the connectivity of the context graph. The spanning tree algorithm is used to discover the context graph. In our work, the algorithm is revised as the maximum spanning tree, which preserves the high weight edges as much as possible. Figure 5 shows the spanning tree of the context graph of Fig. 3. The tree structure is clearer than the original context graph, which can reflect the relations between the words succinctly.

4.6. The generation of context community

Besides the tree structure of a concept, the semantics of a concept may differ. For example, the concept “apple” can be seen as a fruit, as a computer company, or even as a daily newspaper. So the different aspects of a concept must be mined to reflect the different meanings.

In this paper, the Girvan and Newman (GN) algorithm is used to detect the context community of the context graph. Girvan and Newman, 2002 proposed an algorithm based on the iterative removal of edges with high “betweenness” scores that appears to identify such structures with some sensitivity. The GN algorithm has been employed by a number of authors in the study of such diverse systems as networks of email messages, social networks of animals, collaborations of jazz musicians, metabolic networks, and gene networks (Newman, 2004). In our work, the GN algorithm is used to cluster words with the same semantics. The words make up a set representing the same meanings of a context. In order to detect the communities efficiently, the GN algorithm is employed on the spanning tree of the context graph. Figure 5 shows the communities detection results of the concept “apple”. From Fig. 5, each different community reflects a different meaning of the concept “apple”. For example, the associated words “iPhone”, “iPad”, “Mac” in the community with blue nodes reflects that “apple” is a computer company. The associated words “fruit”, “tree”, “rose” in the community with black nodes reflects that “apple” is a kind of fruit. In Section 4.3, the weight of context sentences are computed by the sum of snippet frequencies of their words. Since different sentences may reflect the different meanings of a concept, the context sentences are also divided into different communities. Figure 6 shows the overall model of generating the temporal semantic context of a concept.

5. Experiments and results

In this section, experimental results are presented on three different datasets to show the effectiveness of the proposed temporal semantic context technology for various real-world tasks.

5.1. Question–answer dataset

The first dataset is a set of forty questions. These questions are mainly about four concepts including: CEO, president, champion, and winner. For example, considering the concept “CEO”, ten questions are set such as “who is the CEO of apple”, “who is the CEO of Microsoft.” Ten companies are chosen, such as “XXX CEO”, where XXX is the name of company. Like the concept “CEO”, “president of country”, “champion of match”, and “winner of prize” are set. The queries are listed in Table 1. Our experiments are designed as follows.

(1) Issue the data in Table 1 as the queries to Google, for example, the query “President of US” by Google is searched.
(2) Select a different time interval of the concept. The time interval is set as one year. The starting year is 2003 and the ending year is 2012. For example, the query “President of US” in the different years are searched from 2003 to 2012.
(3) Generate the temporal semantic context of each query in different years. The number of snippets is set as 20 (Xu et al., 2011).

Two data of the temporal semantic context of each query are considered: the associated concepts and the associated words. The associated concepts or words may reflect the related concepts of the query, which can be seen as the answer of the question. For example, the associated concept with top snippet frequency of the query “president of US” in 2005 is “George Bush”. But in 2009, the associated concept changes to “Barack Obama”. The goal of experiments on this dataset is to show the effectiveness of the TSC to generate an accurate context. Table 2 shows some selected results of the queries. In total, the forty queries in the years from 2003 to 2012 can generate 400 different semantic contexts. If the associated concepts or words with the top three ranks are the right answer, then the semantic context is correct. For example, if the associated concept in top three ranks of the query “Apple CEO” in 2004 is “Steve Jobs”, then semantic context is thought right. If the associated concept in top three ranks of the query “Microsoft CEO” in 2010 is “Bill Gates”, then that semantic context is thought wrong (the right answer is “Steve Ballmer”).

The accuracy of the 400 temporal semantic contexts is 92%, which means that the most of the questions get the right answer in the proposed TSC method. These questions are also queried in the Q–A website[12]. The accuracy is lower than 10%. The existing Q–A system does not support the temporal function. For example, it is hard for the existing Q–A system to answer the question such as “Who is the CEO of IBM in 1996?” The experiments in questions dataset show the accuracy and effectiveness of the proposed TSC method. The generated semantic context is related to the query, which captures the meanings of the given concept.

[12] www.answers.com.

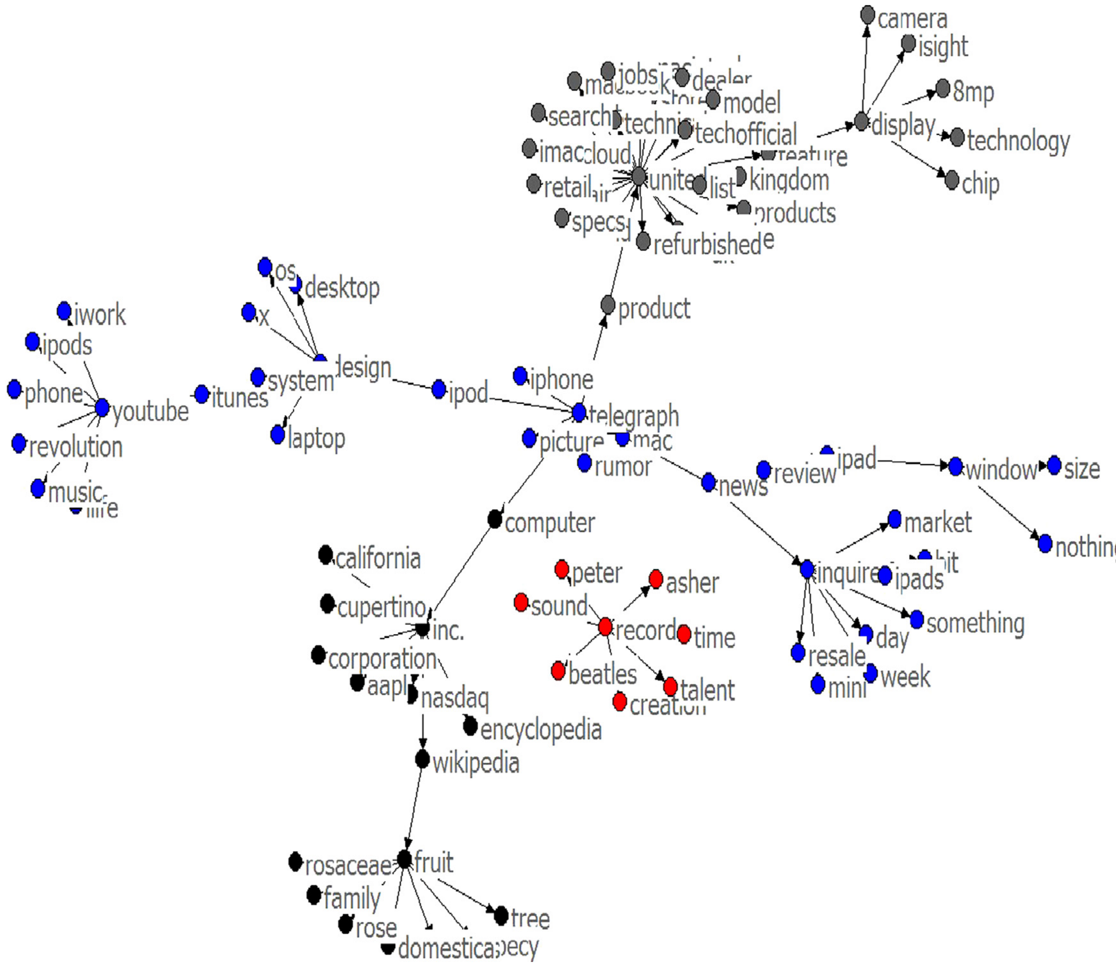


Fig. 5. The maximum spanning tree of Fig. 3.

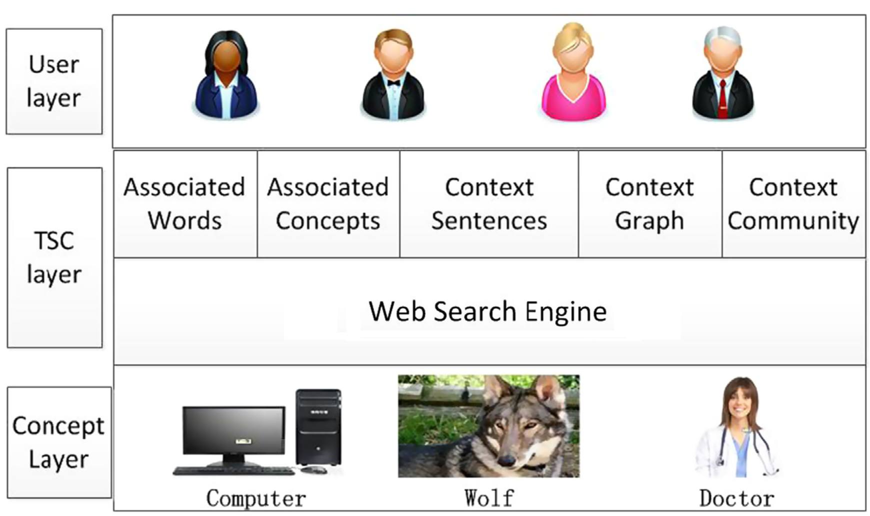


Fig. 6. The model of generating the temporal semantic context.

5.2. LinkedIn dataset

LinkedIn is a social networking website for people in professional occupations. As of June 2012, LinkedIn reported more than 175 million registered users in more than 200 countries and territories. Figure 2 gives a LinkedIn page of "Barack Obama", who is the president of United States. He has four experiences including the president, US Senator, State Senator, and Senior lecturer in law. The different experiences are associated with the different time intervals. For example, he was a state senator of Illinois from 1997 to 2004. In this experiment, the TSC of the people in different time intervals is generated according to his/her experiences in LinkedIn. If the TSC includes the experience in the correct time interval, the proposed method performs well in temporal aspect. 100 persons who are chairs, members, or keynote speakers of the WWW, KDD, CIKM, ICDM, and SIGIR from 2009 to 2013 are selected. All of these persons have

Table 1
The Q&A dataset in proposed experiments.

Concept	1	2	3	4	5	6	7	8	9	10
CEO	Apple	Microsoft	IBM	Yahoo	Facebook	Twitter	GE	Google	Baidu	HP
President	US	UK	China	Germany	France	Russia	Italy	Japan	Brazil	India
Champion	NBA	NFL	US Open	France Open	Wimbledon Open	Australia Open	Formula 1	Seria A	Liga	NHL
Winner	Best leading actor of Oscar	Best supporting actor of Oscar	Best leading actress of Oscar	Best supporting actress of Oscar	Best director of Oscar	Best picture of Oscar	Noble prize in Physics	Noble prize in Chemistry	Noble prize in Literature	Noble peace prize

Table 2
The selected results of the experiments on the Q&A dataset.

Query	Answers extracted from associated concepts			
	2004	2006	2008	2010
Yahoo CEO	Terry Semel	Terry Semel	Jerry Yang	Carol Bartz
Japanese prime minister	Junichiro Koizumi	Junichiro Koizumi	Yasuo Fukuda	Yukio Hatoyama
NBA Champion	Detroit Pistons	Miami Heat	Boston Celtics	Los Angeles Lakers
Nobel peace prize winners	Shirin Ebadi	Yunus	Ahtisaari	Liu xiaobo

Table 3
The selected persons from www2009 to www2013.

www2013	Tanya Berger-Wolf	Virgilio Almeida	Daniel Schwabe	Luis Von Ahn
www2012	Wolfgang Nejdl	Deepak Agarwal	Evgeniy Gabrilovich	Sharad Goel
www2011	Elisa Bertino	Panos Ipeirotis	Andrew Tomkins	ChengXiang Zhai
www2010	Vint Cerf	Danah Boyd	Juliana Freire	Elizabeth Churchill
www2009	Juan Quemada	Ziv Bar-Yossef	Steffen Staab	Lada Adamic

LinkedIn pages and complete experiences. Table 3 gives the name of the International World Wide Web Conference. The complete names of these people are set as the query to Google. The experiments are designed as follows.

(1) Issue the name of the authors as queries to Google. For example, the query “Virgilio Almeida” is searched by Google;
(2) Generate the temporal semantic context of each query in each year from 2003 to 2012.

The context sentences of the temporal semantic context of each query are considered. The context sentences reflect the exact meaning of the query, which can be seen as a background of the query. For example, the context sentence of the query ‘Virgilio Almeida’ in 2011 is ‘Virgilio Almeida is the IT policy secretary at the Brazilian Ministry of Science and Technology (MCT)’. But in 2009, the context sentence is ‘Virgilio Almeida is a professor of computer science at the Universidade Federal de Minas Gerais (UFMG)’. The context sentence of the different time intervals can be matched to the experience from LinkedIn. The goal of experiments on the LinkedIn dataset is to test the effectiveness of the TSC to generate an accurate context sentence. Totally, the 100 queries generate 511 experiences (2003–2012). Since some web pages of early stage may not be contained by Google, the experiences are only selected from the years 2003 to 2012. If the set of the context sentences contains the right experience, then the semantic context is considered correct.

Table 4 shows some selected results extracted from the context sentence of the queries. The precision of the experiment is 0.95, which means that about 95% sentences contain the right experiences of the person. The recall of the experiment is 0.82, which means about 419 right experiences are mined from the context sentences. The *F*-measure is 0.88. The missed and incorrect experiences are usually caused by the recall of the returned search results by Google and the same name of the other people. The experiments in the LinkedIn dataset show the accuracy of the proposed TSC method. The generated semantic context can catch the temporal feature of the concept.

5.3. Wikipedia dataset

The Q&A datasets prove the accuracy of the generated associated words and associated concepts. The LinkedIn dataset shows the effectiveness of the generated context sentences. In this section, we test the accuracy of the context community. Five concepts are selected from Wikipedia including: apple, java, china, mercury, and transformer. All of them have diverse meanings. Table 5 gives the concepts and their diverse context sentences. The concepts are issued as queries to Google, and generate the context communities. If the meaning of the context community appears in the Wikipedia, then the temporal semantic context is considered correct. For example, one context community about the concept “Java” is a programming language, and another context community is about a famous touring site. These two meanings can be matched to the diverse definitions of “Java” in Wikipedia.

In total, the five concepts generate 13 context sentences from different context communities. The different context sentences reflect the different aspects of the concept. Meanwhile, the different context sentences reflect the different relations between associated words.

6. Applications

In this section, some applications are employed using the proposed TSC method including query suggestion, faceted exploration, and

Table 4
The selected results of the experiments on the LinkedIn dataset.

Person	Experiences extracted from context sentences
Tanya Berger-Wolf	University of New Mexico (2003–2004), DIMACS (2003–2004), University of Illinois at Chicago (2005–2007, 2010–now), University of Illinois (2008–2009)
Evgeniy Gabrilovich	Israel Institute of Technology (2003, 2005–2006), Microsoft Research (2004), Yahoo (2007–2012), Google (2012–now)
Andrew Tomkins	IBM Almaden Research (2003–2005), IBM Research (2004), Yahoo Research (2005–2009), Google (2010–now)
Elizabeth Churchill	FX Palo Alto Laboratory (2003), Palo Alto Research Center (2004), Yahoo Research (2005–now)
Steffen Staab	University of Karlsruhe (2003–2004), University of Koblenz-Landau (2004–now), Institute For Web Science and Technologies (2010–now)

Table 5
The selected results of the experiments on the Wikipedia dataset.

Concepts	Context sentences
Apple	Apple designs and creates iPod and iTunes, Mac laptop and desktop computers, the OS X operating system, and the revolutionary iPhone and iPad. The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). Peter Asher recounts the creation of The Beatles' Apple Records which brought together some of the most eclectic sounds and talents of the time. Make the most of the British apple season with our delicious apple recipes.
Java	Java is a set of several computer software products and specifications from Sun Microsystems (which has since merged with Oracle Corporation). Java (Indonesian: Jawa) is an island of Indonesia.
China	China officially the People's Republic of China (PRC), is a country in East Asia. Chinese cuisine is any of several styles originating from regions of China, some of which have become increasingly popular in other parts of the world
Mercury	Mercury is the innermost planet in the Solar System. Mercury is a chemical element with the symbol Hg and atomic number 80. Alt-J win the Mercury prize 2012 after a rather subdued ceremony at the Roundhouse in London.
Transformer	A transformer is a power converter that transfers electrical energy from one circuit to another through inductively coupled conductors. Transformer is a 2007 American science fiction action film based on the Transformer toy line.

trend analysis. These applications show the importance of the proposed TSC problem in many web mining tasks.

6.1. Query suggestion

Queries to web search engines are usually short and ambiguous, providing insufficient information to determine needs of users for effectively retrieving relevant Web pages (Xu et al., 2011a,b). In order to address this problem, query suggestion is implemented by most search engines (e.g., Google's 'Search related to' and Yahoo's 'Also Try'). In this section, the proposed TSC method is applied to the query suggestion task. The steps for using TSC for query suggestion are as follow.

(1) Suppose the input query concept as q.
(2) Generate the TSC of q using the proposed method.
(3) Rank the associated words of the TSC by their snippet frequencies.

According to the above three steps, the high ranking associated words can be recommended for the candidate query suggestion. In order to evaluate the accuracy of using TSC on query suggestions, some concepts are used in the test query suggestion task. Table 6 shows the queries used in our suggestion task, which are the top 10 searched queries from years 2010 to 2011 in Google Zeitgeist. Each query is given 5 query suggestions. Two evaluation methods including objective and subjective methods are conducted. In the subjective evaluation method, 50 randomly selected human raters are invited to rank the query suggestion results by TSC, Google 'search related to', and Yahoo 'also try'. These human raters are university students. Each human rater evaluated the query suggestion on a 5-point scale (the higher the rating, the better the method is). Figure 7 gives the average point results. The proposed method performs better than Google and Yahoo, which indicates that the suggested concept is basically related to the query.

Table 6
The top 10 search queries of Google Zeitgeist.

2011		2010	
Rebecca black	iPhone 5	Chatroulette	Myxer
Google+	Adele	iPad	Katy Perry
Ryan Dunn	东京 电力	Justin Bieber	Twitter
Casey Anthony	Steve Jobs	Nicki Minai	Gamezer
Battlefield 3	iPad2	Friv	Facebook

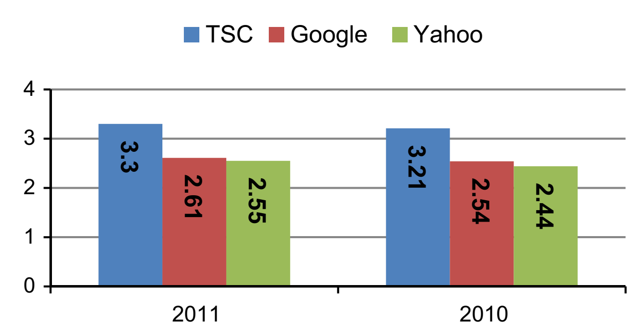


Fig. 7. The average points result of three different methods.

The objective evaluation methodology means that the average semantic similarity between the query and its query suggestion results obtained by Google, Yahoo, and our method are computed, respectively. The average semantic similarity means the average value of the semantic similarities between the suggested queries and the search query. The higher the average semantic similarity is, the better the method is. In Fig. 8, the objective comparison results of our method, Google and Yahoo are given. As seen in Fig. 8, the average semantic similarity of our method is higher than Google and Yahoo in each year's queries. Since the invited human raters are all students, the query suggestion results may have limitations. Moreover, the temporal feature can be added to the query suggestion, which can provide the query suggestion results in different time intervals.

6.2. Faceted exploration

Faceted exploration of search results is widely used in search interfaces for structured databases such as shopping catalogs, job listings, and house search (van Zwol and Sigurbjornsson, 2011). The proposed TSC method is used on faceted exploration, which can help users explore the different aspects of the query. The steps for using TSC for faceted exploration are as follow.

(1) Suppose the input query concept as q.
(2) Generate the TSC of q using the proposed method.
(3) Rank the associated words of the TSC by their snippet frequencies.
(4) Select the associated words with high ranking of the different context communities.

The associated words from different context communities can provide the different meanings of the concept, which can support the faceted exploration task. For example, given the query "apple", the associated words "iPhone", "fruit", and "Beatles" can show the different aspects to the users. In order to evaluate the accuracy of using TSC on faceted exploration, given a query, the different images of different communities are provided. Figure 9 gives the image based faceted exploration results using the proposed TSC method of the five queries "buy Cartier", "buy Chanel", "buy Gucci", "buy Hermes", and "buy Louis Vuitton". In Fig. 9, the TSC of each query is generated. The top associated word for each of the three context communities is used for the faceted exploration. These words are set as the query to the Google image search. The top returned image is recommended to the users. Thus, users can browse different images in different aspects.

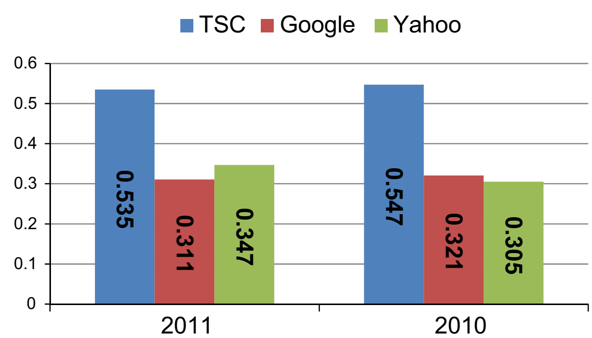


Fig. 8. The average semantic similarity result of three different methods.

6.3. Trend analysis

The meanings of concepts change with time. The new meaning emerges and the old meaning diminishes in usage. The proposed TSC method is employed on trend analysis of a concept. The steps for using TSC for trend analysis are as follow.

(1) Suppose the input query concept as q.
(2) Generate the page counts plot in a given time interval.
(3) Generate the TSC of q using the proposed method in the different time intervals.
(4) Rank the example sentences of TSC.

Given a concept, the top context sentences of the nodes with high page counts are given. Figure 10 gives an example of the concept "iPhone". The number of news stories is found from 2008.1 to 2011.12 per month from Google. Five nodes with the local maximum page counts are selected. The criterion for selecting these five nodes is inspired by Google Trends. Google Trends provides the milestone points in the time plot. The top five milestones from Google Trends are selected, which can be seen as milestone nodes in their own time interval. The common feature of these five nodes is that the immediately preceding and immediately succeeding nodes have lower page counts.

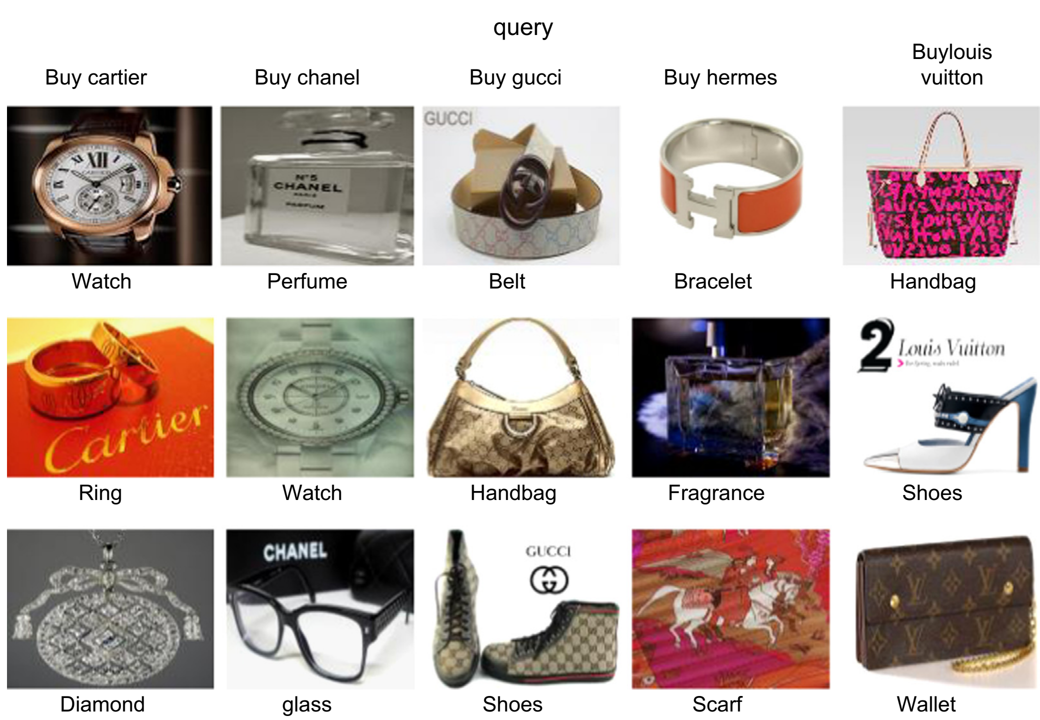
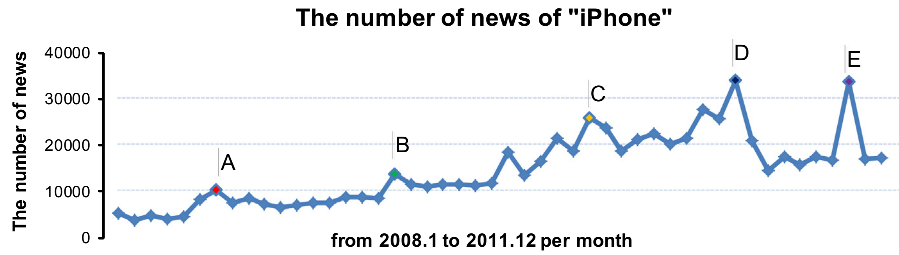


Fig. 9. The faceted exploration using TSC.



Context Sentence A: FAQ: iPhone 3G launch day is Friday (July, 2008)
Context Sentence B: Apple launched iPhone 3GS at WWDC (June, 2009)
Context Sentence C: iPhone 4 launchday: the ultimate survival guide (June, 2010)
Context Sentence D: Apple ships 100 million iPhones (March, 2011)
Context Sentence E: Apple unveils iPhone4S, no iPhone 5 in sight (October, 2011)

Fig. 10. The trend analysis using the proposed method.

The context sentences of these five nodes are given. The context sentences in the time interval are all important news. Four context sentences are about Apple computer company issued new iPhone, and another shows the huge success of iPhone sales. The context sentences can help users analyze the trend of a concept. The important news stories cause high page counts, which raise the popularity of the concept.

7. Conclusion

Existing web search engines usually generate "one-size-fits-all" search results to the users, which cannot provide appropriate semantic context. In this paper, considering the temporal and semantic features of context, the problem of temporal semantic context (TSC) is studied. Compared to existing methods, the proposed method does not need any prior knowledge such as ontology or a hierarchical knowledge base.

A temporal semantic context consists of a set of context indicators including associated words, associated concepts, context sentences, context graph, and context communities is defined. A general approach to automatically generate structured contexts of a concept including semantically related words, example sentences, diverse meanings, and temporal annotations are proposed. The proposed TSC structure integrates the features from dictionary, Wikipedia, and LinkedIn web sites, which is helpful for users to understand and explore the concept. Empirical experiments on three different datasets including Q–A dataset, LinkedIn dataset, and Wikipedia dataset show that the proposed algorithm is effective and accurate for generating TSC. Different from manually generated context repositories such as LinkedIn and Wikipedia, the proposed method can automatically generate the context. Some applications are demonstrated employing the proposed TSC method. The proposed method is applied to trend analysis, faceted exploration, and query suggestion. These applications prove the importance of the proposed TSC problem in many web mining tasks.

Acknowledgment

This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grants 2012AA011504, 2013AA014601, 2013AA014603, in part by National Key Technology Support Program under Grant 2012BAH07B01, in part by the National Natural Science Foundation of China under Grant 61300202, and in part by the Science Foundation of Shanghai under Grant 13ZR1452900.

References

Agrawal R Srikant R. Fast algorithms for mining association rules. In: Proceedings of the 1994 international conference on very large data bases; 1994. p. 487–99.
Al-Khalifa H, Davis H. Measuring the semantic value of folksonomies. In: Proceedings of the 2nd IEEE international conference on innovation information technology; 2006. p. 1–5.
Bao S, Wu X, Fei B, Xue G, Su Z, Yu Y. Optimizing Web search using social annotations. In: Proceedings of the 16th international world wide web conference; 2007. p. 501–510.
Berners-Lee T, Hendler J, Lassila. O. The semantic web. Sci Am 2001;284(5):34–43.
Bollegala D, Matsuo Y, Ishizuka M. Measuring semantic similarity between words using web search engines. In: Proceedings of the 16th international world wide web conference; 2007. p. 757–766.
Boughareb D, Farah N. Toward a web search personalization approach based on temporal context. Commun Comput Inf Sci 2011;166:33–44.
Choi D, Ko B, Kim H, Kim P. Text analysis for detecting terrorism related articles on the web. J Netw Comput Appl 2014 (in press).
Dumont E, Quenot G. A local temporal context-based approach for TV news story segmentation. In: Proceedings of the 2012 IEEE international conference on multimedia and expo; 2012. p. 973–8.
Girvan M, Newman M. Community structure in social and biological networks. In: Proceedings of the national academy of science of the United States of America; 2002. p. 7821–6.
Glaser T, Zelnik-Manor L. Incorporating temporal context in Bag-of-Words models. In: Proceedings of the 2011 IEEE international conference on computer vision workshops; 2011. p.1562–9.
Golder S, Huberman B. Usage patterns of collaborative tagging systems. J Inf Sci 2006;32(2):198–208 (2).
Halpin H, Robu V, Shepherd H. The complex dynamics of collaborative tagging. In: Proceedings of the 16th international world wide web conference; 2007. p. 211–20.
Han J, Chen H, Xin D, Yan X. Frequent pattern mining: current status and future directions. Data Min Knowl Discov 2007;15:55–86.
Howard M, Shankar K, Jagadisan U. Constructing semantic representations from a gradually changing representation of temporal context. Top Cognitive Sci 2011;3:48–73.
Hwang M, et al. Automatic enrichment of semantic relation network and its application to word sense disambiguation. IEEE Trans Knowl Data Eng 2011;23 (6):845–58.
Jannach D, Leopold K. Knowledge-based multimedia adaptation for ubiquitous multimedia consumption. J Netw Comput Appl 2007;30(3):958–82.
Jansen M, Spink A, Bateman J, Saracevic T.. Real life information retrieval: a study of user queries on the Web. In: Proceedings of ACM SIGIR forum, vol. 32; 1998. p. 5–17.
Leung K, Ng W. Personalized concept-based clustering of search engine queries. IEEE Trans Knowl Data Eng 2008;20(11):1505–18.
Leung K, Ng W, Lee D. Personalized word-based clustering of search engine queries. IEEE Trans Knowl Data Eng 2003;20(11):1505–18.
Li Y, Tian Y, Duan L, Yang J, Huang T, Gao W. Sequence multi-labeling: a unified video annotation scheme with spatial and temporal context. IEEE Trans Multimed 2010;12(8):814–28.
Liu X, Lin L, Yan S, Jin H, Tao W. Integrating spatio-temporal context with multiview representation for object recognition in visual surveillance. IEEE Trans Circuits Syst Video Technol 2011;21(4):393–407.
Luo X, Xu Zheng, Yu J, Chen X. Building association link network for semantic link on web resources. IEEE Trans Autom Sci Eng 2011;8(3):482–94.
Manica E, Dorneles C, Galante R. Handling temporal information in web search engines. ACM SIGMOD Rec 2012;41(3):15–23.
Milea V, Frasincar F, Kaymak U. Knowledge engineering in a temporal semantic web context. In: Proceedings of the eighth international conference on web engineering; 2008. p. 65–74.
Newman M. Fast algorithm for detecting community structure in networks. Phys Rev E 2004;69(6):1–5.

Nguyen H, Ji Q, Smeulders. A. Spatio-temporal context for robust multitarget tracking. IEEE Trans Pattern Anal Mach Intell 2007;29(1):52–64.
Pandit S, Olston C. Navigation-aided retrieval. In: Proceedings of the 16th international conference on world wide web; 2007. p. 319–400.
Radinsky K, Agichteiny E, Gabrilovichz E, Markovitch. S. A word at a time: computing word relatedness using temporal semantic analysis. In: Proceedings of the 20th international world wide web conference; 2011. p. 337–46.
Ramage D, Heymann P, Manning C, Garcia-Molina H.. Clustering the tagged Web. In: Proceedings of the 2nd ACM international conference on web search and data mining; 2009. p. 54–63.
Salton G, Wong A, Yang. C. A vector space model for automatic indexing. Commun ACM 1975;18(11):613–20.
Schenkel R, Crecelius T, Kacimi M, Michel S, Neumann T., Parreira J, et al. . Efficient top-k querying over social tagging networks. In: Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval; 2008. p. 523–30.
Sparrow B, Liu J, Wegner D. Google effects on memory: cognitive consequences of having information at our fingertips. Science 2011;333:776–8.
Suchanek F, Vojnovic M, Gunawardena D. Social tags: meaning and suggestions. In Proceedings of the 17th ACM international conference on information knowledge management; 2008. p. 223–32.
Weng J, Zhang Q, Chi M, Xue X. Complex text processing by the temporal context machines. In: Proceedings of IEEE 8th international conference on development and learning; 2009. p. 1–8.
White R, Drucker S.. Investigating behavioral variability in web search. In: Proceedings of the 16th international world wide web conference; 2007. p. 21–30.
Xu S, Bao S, Cao Y, Yu Y.. Using social annotations to improve language model for information retrieval. In Proceedings of the 15th ACM international conference on information and knowledge management; 2007. p. 1003–6.
Xu Z, Luo X, Yu J, Xu W. Measuring semantic similarity between words by removing noise and redundancy in web snippets. Concurr Comput Pract Exp 2011;23(18):2496–510.
Xu Z, Luo X, Yu J, Xu W. Mining web search engines for query suggestion. Concurr Comput Pract Exp 2011;23(10):1101–13.
Zhou D, Bian J, Zheng S, Zha H, Giles C. Exploring social annotations for information retrieval. In: Proceedings of the 17th international world wide web conference; 2008. p. 715–24.
Zhuge H. Semantic linking through spaces for cyber-physical-socio intelligence: a methodology. Artif Intell 2011;175:988–1019.
van Zwol R, Sigurbjornsson B. Faceted exploration of image search results. In Proceedings of the 19th international world wide web conference; 2011. p. 1–10.