Special Issue Papers and Technical Correspondence

Mining Evolutionary Topic Patterns in Community Question Answering Systems

Zhongfeng Zhang, Qiudan Li, and Daniel Zeng

***Abstract*—Community Question Answering (CQA) is becoming a popular Web 2.0 application. By analyzing evolutionary topic patterns from CQA applications, one can gain insights into user interests and user responses to external events. This paper proposes a novel evolutionary topic pattern mining approach. This approach consists of three components: 1) extraction of the topics being discussed through a temporal analysis; 2) discovery of topic evolutions and construction of evolutionary graphs of extracted topics; and 3) life cycle modeling of the extracted topics. We show empirically the effectiveness of our approach using two real-world data sets.**

***Index Terms*—Community Question Answering (CQA), evolutionary topic patterns, life cycle.**

I. Introduction

Community question answering (CQA) has become a novel social media platform where users can ask and answer questions on any topic of interest. On a CQA site, e.g., Yahoo! Answers and Baidu Zhidao, users post specific questions and other users may help answer them. Both questions and answers are stored in a searchable format for future uses.

In CQA, the set of the questions and related answers about a given topic can be viewed as a text stream with time stamps. Considering the temporal dimension of questioning and answering is essential for an in-depth understanding of the topic under discussion, and can help track topic evolution and dissemination. In the meanwhile, by examining the question–answer pairs posted, one can gain an overview of the topic evolution, and better understand community user behavior and interests.

In this paper, we study the problem of evolutionary topic pattern (ETP) discovery in CQA. We propose a three-step approach: 1) extracting question topics from CQA question–answer pairs through a temporal approach; 2) discovering topic transitions and constructing the corresponding evolutionary graphs; 3) modeling topic "strength" over time and analyzing life cycles of the extracted topics. The proposed ETP mining approach is unique in that it focuses on the topic level representation and abstraction of questions and answers, as opposed to individual questions or users as studied in the previous work. This approach also provides a higher-level view of group-based discussions, enabling better understanding of community user behavior. We have empirically evaluated the proposed approach using two real-world CQA data sets, one on Infectious Disease and the other on Economics. The results show that our approach can discover interesting ETPs.

Manuscript received August 1, 2009; revised March 29, 2010; accepted July 17, 2010. Date of publication June 23, 2011; date of current version August 23, 2011. This research is partly supported by the National High Technology Research and Development Program (863) of China under Grant 2006AA010106, the National Basic Research Program (973) of China under Grant 2007CB311007, the Chinese Academy of Sciences project under Grant 2F07C01, and the National Natural Science Foundation of China projects under Grant 60703085 and 90924302. This paper was recommended by Associate Editor C. Yang.

Z. Zhang and Q. Li are with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: sonfon@gmail.com; qiudan.li@ia.ac.cn).

D. Zeng is with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Department of Management Information Systems, University of Arizona, Tucson, AZ 85721-0108 USA (e-mail: zengdaniel@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2011.2157131

II. Related Work

CQA research has become popular very recently. As the quality of the submitted questions and answers varies widely, researchers proposed content quality measures [1]. Jurczyk and Agichtein [2] analyzed link structures in CQA to detect authoritative users. These previous studies have mainly focused on individual users or contents at the individual question or answer level. In this paper, we instead focus on a macro perspective to analyze the topics and their evolution.

Recently, ETP discovery has drawn a lot of attention. The key idea is to cluster documents into different topics and track the changes in these topics over time. Das *et al.* [12] presented a differential evolution-based strategy for automatic clustering of large real-word data sets. Carvalho *et al.* [13] studied the dynamic clustering of interval-valued data based on adaptive quadratic distances. Mei and Zhai [3] proposed probabilistic approaches to mine evolutionary theme patterns automatically. In their work, the pLSA model was used to model documents as expressing several themes, and K-L divergence was used to measure theme cohesion over time. Their work was extended to analyze spatiotemporal theme patterns from weblogs in [5], by considering locations of blog authors.

Wang *et al.* [6] presented a non-Markov continuous time model based on the Latent Dirichlet Allocation (LDA) model, in which each topic is associated with a continuous distribution over time stamps. While the time stamps were treated as observations of latent topics, the topic mixture proportions of each document were assumed to be dependent on previous topic mixture proportions [7]. Yang *et al.* proposed an event evolution pattern discovery technique from news corpora [11]. More recently, Liu *et al.* [4] proposed a sentence-level probabilistic model for discovering evolutionary theme patterns, in which themes are represented as lists of named entities. Our work reported in this paper follows in general this stream of research. We have chosen the LDA model for temporal topic extraction since this model has been shown to deliver good performance in related text analysis tasks. The widely-used cosine similarity measure [12] is used to explore the evolutionary relationships among temporal topics.

III. Evolutionary Topic Pattern Discovery

A. Problem Formulation

In CQA, a question document is generated by assembling a question and its associated answers together. Let $Q = \{q_1, q_2, \ldots, q_T\}$ be the collection of generated question documents, where q_t refers to a question document at time stamp t. Given this collection Q, *evolutionary topic pattern (ETP)* discovery aims to extract the temporal topic evolution patterns automatically. Another aspect of ETP is to analyze the "strength" of a topic throughout its life cycle.

We now introduce formally the related definitions.

Definition 1 (Temporal Topic): A temporal topic z refers to a topic with a starting time $s(z)$ and an ending time $e(z)$. It is represented as a probabilistic distribution over words, written as $p(w|z)$.

High-probability words typically suggest what the particular topic is about. For instance, in a topic about swine flu, words such as "swine," "flu," "infectious" would appear with high frequencies.

Definition 2 (Topic Transition): Given two temporal topics z_1 and z_2, with z_1 starting earlier than z_2, i.e., $s(z_1) < s(z_2)$, a topic transition occurs if $\text{sim}(z_1, z_2) > \varepsilon$, where $\text{sim}(z_1, z_2)$ is the similarity score between z_1 and z_2, and ε is a predetermined threshold. In this case, z_2 is said to be evolved from z_1, represented by $z_1 \Rightarrow z_2$.

This concept of topic transition is crucial for describing the relationships between topics evolving over time and across different time spans. We are able to mine the ETPs by assembling the discovered topic transitions along the time line.

Definition 3 (Topic Life Cycle): Given a collection of question documents, the topic life cycle of a topic is defined as "the strength distribution of the topic over the entire time line" [3].

This concept of topic life cycle enables us to gain a summary view of user participation in the topic during different time periods.

B. Temporal Topic Extraction

We first partition the question documents into sub-collections according to time intervals, i.e., $Q = Q_1 \cup Q_2, \ldots, \cup Q_n$. For each sub-collection Q_i, we extract temporal topics using the LDA model [9]. Each question document is modeled as a mixture of different topics, and each topic is represented by a probabilistic distribution over words.

Let $z_1, \ldots, z_k$ be the K topics to be extracted, V the number of unique words in the domain vocabulary, and D the number of question documents in Q_i. The generative process of extracting temporal topics from Q_i is given below [8]:

---

1) Draw K multinomials ϕ_{z_k} from a Dirichlet prior β, one for each temporal topic z_k;
2) Draw D multinomials θ_q from a Dirichlet prior α, one for each question document q;
3) For each word token j in the question document q:
   a) Sample a topic z_j from multinomial θ_q; $p(z_j|\alpha)$;
   b) Sample a word w_j from multinomial ϕ_{z_j}; $(p(w_j|z_j, \beta))$.

The topic–word distribution ϕ and the topic–question document distribution θ are two main variables of interest. They are estimated using the Gibbs sampling method [9]. The probability of assigning the current word token w_j to each temporal topic z_j, conditioned on the topic assignment of all other word tokens [9], is calculated as

$$p(z_j = m|z_{-j}, w_j, q_j, \cdot) \propto \frac{C^{VK}_{w_j m} + \beta}{\sum_{w=1}^{V} C^{VK}_{wm} + V\beta} \frac{C^{DK}_{q_j m} + \alpha}{\sum_{k=1}^{K} C^{DK}_{qk} + K\alpha} \quad (1)$$

where C^{VK} is the word–topic matrix of counts with dimensions $V \times K$, and C^{VK}_{wm} is the number of times word w being assigned to topic m, excluding the current token j. Similarly, C^{DK} is the question document-topic matrix of counts with dimensions $D \times K$, where C^{DK}_{qk} is the number of times topic z_k being assigned to some word token in question document q, excluding the current token j. z_{-j} represents the topic assignments of all other word tokens. $z_j = m$ indicates that token j is assigned to topic m. After a number of iterations, the approximated posterior in (1) can be used to estimate ϕ and θ, by examining the counts of word assignments to topics and topic occurrences in question documents.

By applying the LDA model to all the question documents in sub-collection Q_i, we can obtain the K temporal topics and the topic structure of each question document. The same model can be applied to the entire collection Q to extract trans-collection topics. These topics will be used for life cycle modeling in the subsequent procedure.

C. Topic Transition and Evolution Graphs

Let z_1 and z_2 be two temporal topics, where $s(z_1) < s(z_2)$, the cosine similarity between them is calculated as

$$\text{sim}(z_1, z_2) = \frac{\sum_{j=1}^{|V|} p(w_j|z_1) \times p(w_j|z_2)}{\sqrt{\sum_{j=1}^{|V|} p^2(w_j|z_1)} \times \sqrt{\sum_{j=1}^{|V|} p^2(w_j|z_2)}}. \quad (2)$$

A topic transition occurs between z_1 and z_2, if $\text{sim}(z_1, z_2)$ is above a threshold ε. By calculating the pair-wise similarities of temporal topics between two sub-collections Q_i and Q_{i+1}, we can explore all the topic transitions between Q_i and Q_{i+1}. The topic evolution graph can be constructed by assembling all the identified topic transitions along the time line.

D. Topic Life Cycle Modeling

Life cycle theory has been proposed to study news topics in the previous work [10]. Topics in CQA have "the life cycle with the stages of birth, growth, decay and death" [10] as well. Topics have different life spans, depending on the degree of user participation. An energy function has been used to measure the "strength" of a topic throughout its life cycle [10]. The energy of a topic increases if more question documents are discussing this topic.

Given the collection of question documents $Q = \{q_1, q_2, \ldots, q_T\}$, we use a fixed-size sliding time window to measure the energy of each topic at a given time. Let $\{q_{t,1}, \ldots, q_{t,n}\}$ be the set of question documents during the time slice $[t - (W/2), t + (W/2)]$, where W is the width of the sliding time window. The absolute energy of topic z at time t is calculated as

$$A\text{Energy}(z, t) = \sum_{j=1}^{n} p(z|q_{t,j}) \quad (3)$$

where $p(z|q_{t,j})$ is the number of times topic z being assigned to question document $q_{t,j}$.

The life cycle of a topic can be modeled as the changes of its energy over time.

Algorithm 1 provides the pseudo-code for our three-step approach.

Algorithm 1: Evolutionary Topic Pattern (ETP) Discovery

Input: A collection of question documents $Q = \{q_1, q_2, \ldots, q_T\}$, where q_t refers to a question document received at time stamp t.

Output: 1) topics extracted for each timestep, with each topic represented by a probabilistic distribution over words; 2) evolutionary topic patterns; 3) absolute energy levels for topics extracted.

Algorithmic Steps:

1) Partition the question documents into sub-collections according to time intervals, i.e., $Q = Q_1 \cup Q_2, \ldots, \cup Q_n$, and apply a series of preprocessing procedures to each question document, including tokenization, stop word removal, and stemming.
2) For each sub-collection Q_i, extract the temporal topic sets $Z_i = \{z_{i,1}, \ldots, z_{i,k}\}$ using the LDA topic model.
3) Determine whether a topic transition occurs between topics in two sub-collections Q_i and Q_{i+1}, and construct the corresponding evolution graph.
4) Extract the topic set from the entire set of question documents Q, and analyze the life cycle of each identified topic.

IV. Experiments and Findings

In this section, we report our experiments and key findings.

A. Performance Evaluation

Perplexity has been adopted to evaluate the performance of topic models [8]. It measures the ability of a probabilistic model to generalize to hold-out data. Lower perplexity indicates better generalization performance to unseen data. For a hold-out test set with D question documents, perplexity is calculated as

$$\text{perplexity} = \exp\left\{-\frac{\sum_{d=1}^{D}\sum_{w=1}^{V}\hat{n}_{d,w}\log\left(\sum_{k=1}^{K}\theta_{d,k}\varphi_{k,w}\right)}{\sum_{d=1}^{D}\sum_{w=1}^{V}\hat{n}_{d,w}}\right\} \quad (4)$$

where $\hat{n}_{d,w}$ is the number of times word w has been observed in question document d.

In our experiments, the perplexity measures were calculated by averaging over 10-fold cross validation. For each run, we randomly split the data set into a training set (90%) for model construction and a hold-out set (10%) for testing purposes.

B. Data Preparation

We have constructed two data sets, one concerning Infectious Disease and the other Economics, by collecting solved questions and selected best answers from Yahoo! Answers. Although many of the answers that are not deemed "best" may contain valuable information and represent topic shifts and user involvement, the quality of these answers varies widely. On the other hand, the best answers, selected by the users who posted the initial questions or voted by other community participants, are generally high-quality in terms of contents. The posting time of the question is used as the time stamp of the related question document. The key statistics concerning these two data sets are summarized in Table I.

TABLE I
Basic Data Set Information

Data Set	# of question documents	Time Span
Infectious Disease	11390	Apr. 6th 2009- May 18th 2009
Economics	13762	Dec. 2008 – Apr. 2009

For temporal topic detection, we set the LDA parameters $\beta = 0.01$, $\alpha = 50/\mathrm{K}$, where K is the number of topics to be extracted. Gibbs sampling was repeated for 300 iterations. These parameter settings have been widely adopted in the literature.

C. Experiments on the Infectious Disease Data Set

We first evaluated the proposed approach using the Infectious Disease data set. We split the collected question documents into five time intervals, each spanning about two weeks and adjacent intervals overlapping with each other (the length of the overlapped period is one week). The choice of timestep length is domain dependent. For emergency events such as swine flu outbreaks, the related topics are time-sensitive and change rapidly as these events develop. A short time interval would better capture the topic structure and its evolution. Computationally, partitioning the question documents into overlapping collections seems to be a reasonable approach for such events.

We have examined the perplexity value for each timestep on different settings of topic number as shown in Fig. 1(a). We computed the curvature variation of the perplexity curve at each point and set the number of topics such that the corresponding curvature variation is less than 2 and (near) minimum perplexity can be achieved. For convenience, we have fixed the topic number to be 60 for each timestep and analyzed the ETPs.

Fig. 2 plots several ETPs discovered by our method, with each topic represented by a list of top-ranked keywords. The weight associated with an edge between time-stamped topics indicates the cosine similarity between the topic at $t-1$ and that at t. Each ETP is represented by a thread labeled with one lower-case letter. We have verified these ETPs by manually checking the topic keywords and representative question documents for each topic.

Consider, for instance, thread a, which is about the immune system. It started with malaria-related questions and answers during I1, shifted to antibody during I2, health condition during I3, the concerned age and health problems during I4, and finally the mechanisms of the immune system during I5. Thread b is concerned with the vaccine problems. It began with the reaction and dose problems during I1 and I2, the manufacturing problems during I3, the time table for vaccine during I4, and evolved into the discussion about the injection reactions of children during I5. Thread c is about swine flu, which had spread around the world since mid April in 2009. The two topics during I2 were about the outbreaks of swine flu, and the drugs for swine flu (e.g., tamiflu, relenza), respectively. Later, the cure and symptoms of swine flu, and the reactions of people and governments dominated the discussion during I3. The dangers of swine flu, as well as potential problems with eating pork were discussed during I4. Finally, the death and panic caused by swine flu were talked over during I5.

The next step is to build life cycle models for trans-collection topics. For this task, we used the entire Infectious Disease data set. Table II lists five extracted topics, with each topic represented by top 10 keywords. Topic T20 seems to be about death and panic caused by swine flu. T22 discusses the outbreak and pandemic of swine flu throughout the world. T23 is concerned with the immune system. T40

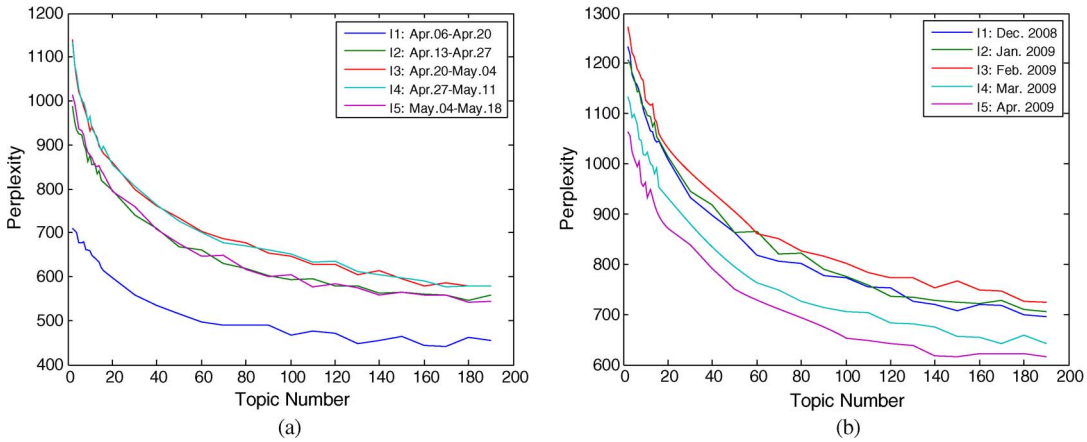

Fig. 1. Perplexity values for (a) the Infectious Disease data set and (b) the Economics data set with different topic number settings.

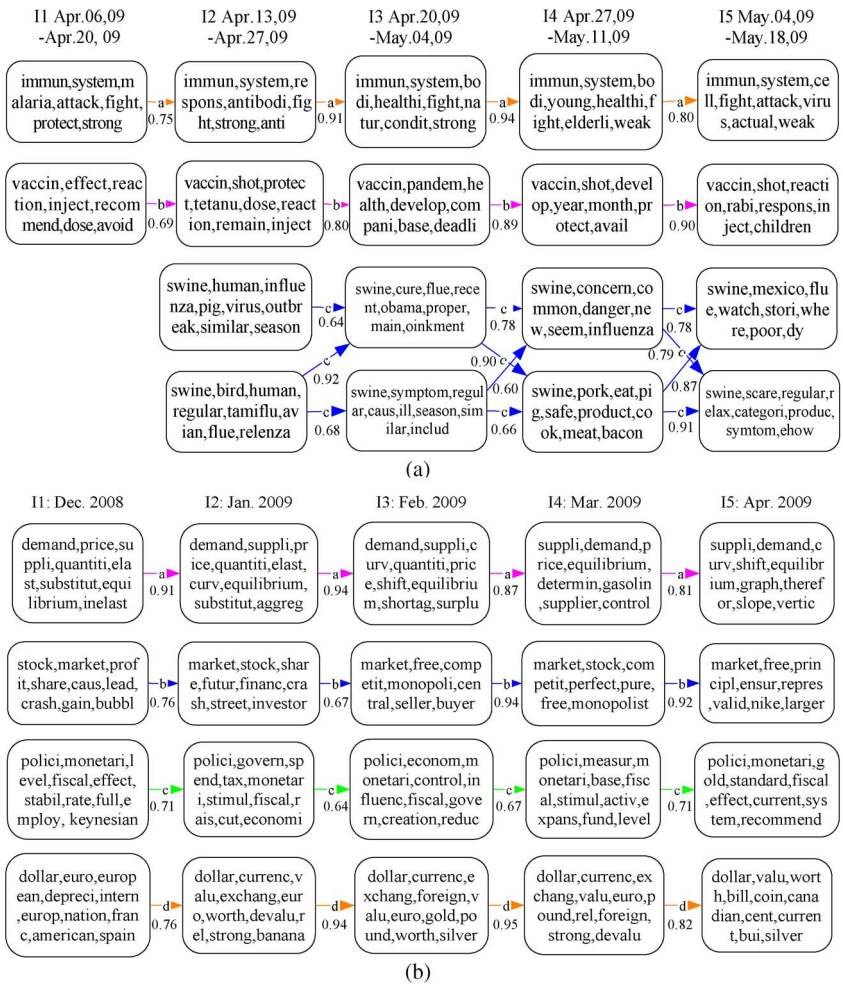

Fig. 2. A partial topic evolution graph for (a) the Infectious Disease dataset and (b) the Economics dataset.

TABLE II
FIVE TOPICS DISCOVERED FROM THE INFECTIOUS DISEASE DATA SET

ID	Topic Words
T20	peopl, regular,freak,deal,panic,swine,death,new,dy
T22	world,pandem,health,million,epidem,outbreak,govern,control,worldwide
T23	immun,system,bodi,healthi,viru,young,fight,attack,elderli,weak
T40	cough,nose,symptom,runni,sneez,doctor,throat,fever,sore,headach
T51	vaccin,cell,produc,effect,protein,respons,antibodi,studi,develop,children

discusses the symptom of cough and sore throat. The vaccine problems are discussed in T51.

Fig. 3(a) shows the absolute energy levels of these five topics over time. We can see that during the period of 10 to 15 days after Apr. 6, the absolute energy of all these five topics was increasing slowly. Swine flu mainly spread in Mexico during this time period. As such, it had not attracted broad discussion in the Yahoo! Answers community. As swine flu began to spread in America and throughout the world after Apr. 21, there was a sharp increase in the energy levels of these topics,

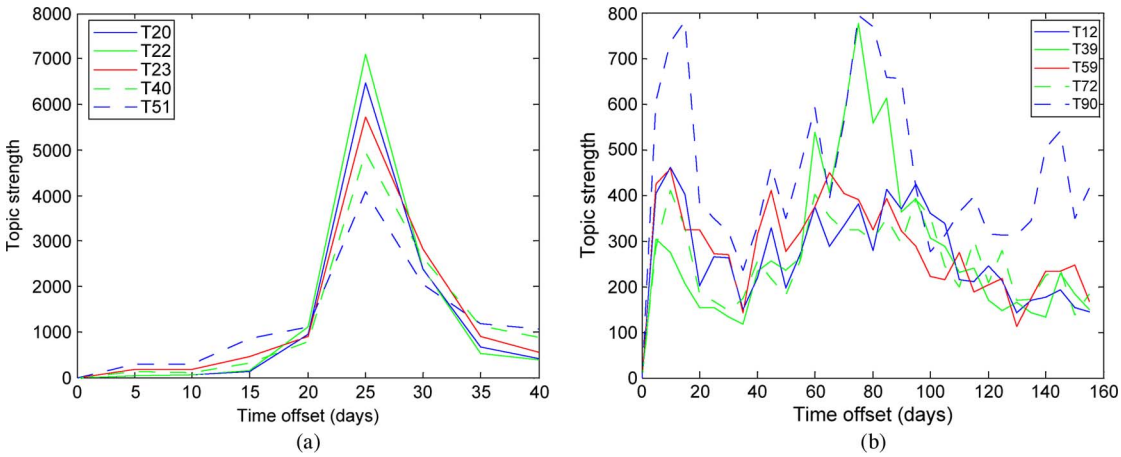

Fig. 3. Absolute energy levels for topics in (a) the Infectious Disease dataset and (b) the Economics dataset.

TABLE III
FIVE TOPICS DISCOVERED FROM THE ECONOMICS DATA SET

ID	Topic Words
T12	market, stock, futur, share, crash, loss, happen, lead, specul
T39	economi,plan,stimulu,obama,spend,go,packag,recoveri,stimul,check
T59	china,india,chines,countri,america,indian,south,cheap,outsourc,north
T72	unemploy,rate,employ,full,forc, peopl,economi, number
T90	demand,price,suppli,quantiti,elast,chang,inelast,equilibrium,substitute

which reached their peak values in just 10 days at around May 1. Thereafter, the energy levels began to decrease and stabilized after May 11. There are two possible reasons for this downward trend: 1) Yahoo! Answers encourages users to search for similar historical questions before submitting new ones. As a large number of questions about swine flu were already solved, duplicated submissions of similar questions did not make it into the system; 2) with the rapid spread of swine flu, WHO and various governments took significant preventative efforts and conducted information campaigns to calm the general public.

D. Experiments on the Economics Data Set

In the case of the Economics data set, as the changes in topics took place much slower, a longer time interval, one month, was sufficient to capture the topic structure and its evolution. In addition, no overlaps between time intervals were needed. The perplexity values varying with different topic number settings are shown in Fig. 1(b). We can see that the optimal performance is achieved at around 100 topics for each timestep. In our experiments, we fixed the topic number to be 100 for each timestep to analyze the ETPs. A subset of discovered ETPs is shown in Fig. 2(b).

Thread a is about the supply-demand relationship, a common topic in Economics. Some minor differences showed up as time progresses. The elastic and inelastic demands were discussed during I1. The equilibrium price was discussed during I2 and I3. During I4, the gasoline supplement received much attention. Finally, the supply and demand curves were discussed during I5. Thread b is about the stock market, also of great concern in Economics, covering a range of topics including the market incentives, market bubbles, financial policies and regulations, etc. Thread c is about the monetary policy. Thread d is concerned with the currency policy of different countries coping with the economic crisis.

We now analyze the life cycle of the Economics topics. Five extracted topics and associated top-ranked words are listed in Table III. T12 is about the stock market, T39 the Obama stimulus plan, T59 economics in developing countries, such as China, India and South American, and T72 the unemployment problems, and T90 the supply-demand relationship. Fig. 3(b) shows the absolute energy levels of these topics. It can be seen that the energy curves of these topics are relatively smooth.

E. Sensitivity to Time Interval

In this subsection, we investigate empirically the impact of choosing different time intervals through additional experiments. We set the time interval to be three weeks and two weeks for the Infectious Disease data set and the Economics data set, respectively, and 50% overlapping was chosen between adjacent time windows. Fig. 4 shows perplexity values varying with different topic numbers. For the Economics data set, only five curves are displayed for clarity. It can be seen that the performance becomes stable after 80 topics and 100 topics, respectively, for these two domains.

We set the topic number to be 100 for both data sets. Fig. 5 plots several ETPs that were discovered. For Economics, thread a is about the supply-demand relationship, while thread b is about the stock market. For Infectious Disease, thread c and thread d are about the immune system and the pandemic of swine flu from Mexico throughout the world, respectively. Comparing the experimental findings with those reported in previous sections, we note that the main ETPs could be discovered under a range of time interval settings. For emergency events such as the swine flu outbreak, the drift of topics may be dramatic if a longer time interval is selected. For the Economics domain in which topics evolve slowly, a small time interval may cause the topics to be split into unnecessarily detailed and insignificant subgroups.

V. CONCLUSION AND DISCUSSIONS

The rapid adoption of Web2.0 technologies and platforms, including CQA, has made available an expanding and enriched set of channels for information seeking and sharing. In CQA applications, the participative and iterative nature of interactions among individual contributors enables potentially accurate reflection of the "pulse" of the society and the general public's reaction to a range of events and policies. In this paper, we propose a three-step approach to discover ETPs automatically from question document collections. The experimental findings based on two data sets collected from Yahoo! Answers show that this approach can be used to discover interesting ETPs from the CQA community. It not only reveals the hidden topic structures, but also reflects the users' interests changing over time. The application of this type of approach is potentially broad. For

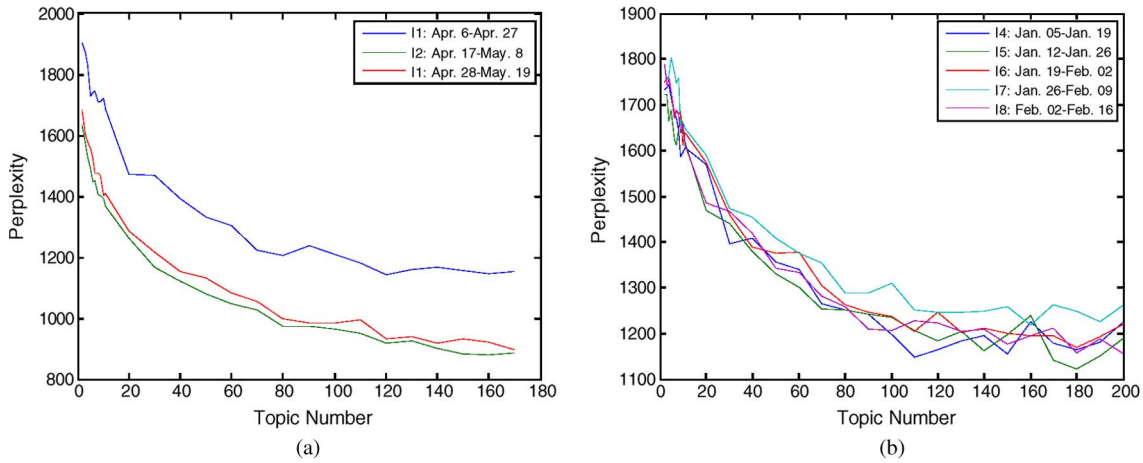

Fig. 4. Perplexity values for (a) the Infectious Disease dataset and (b) the Economics dataset with different time intervals.

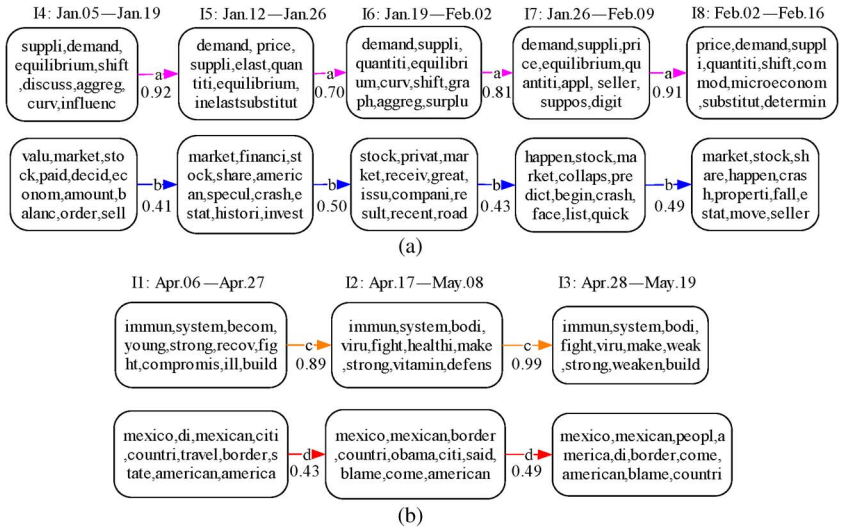

Fig. 5. A partial topic evolution graph for (a) the Economics dataset and (b) the Infectious Disease dataset with different time intervals.

instance, in emergency situations, CQA might be used as an alternative information and feedback channel to learn about people's interests and concerns, and help track people's reaction to both events and policy decisions. Such knowledge can lead to better-informed decisions and more effective policy implication. Discovering ETPs can also be used as a knowledge retrieval and learning tool. For instance, for users who want to learn about the monetary policies, questions in thread c as extracted from the Economics data set would be of interest. ETPs mined from CQA can also be readily used for content recommendation purposes.

Research reported in this paper represents one of the first studies on ETP mining in CQA. As part of our ongoing research, we are evaluating the proposed approach on other data sets and conducting comparative studies, comparing our methods with other approaches.

REFERENCES

[1] X. Wang, X. Tu, D. Feng, and L. Zhang, "Ranking community answers by modeling question–answer relationships via analogical reasoning," in *Proc. SIGIR*, 2009, pp. 179–186.

[2] P. Jurczyk and E. Agichtein, "Discovering authorities in question answer communities using link analysis," in *Proc. CIKM*, 2007, pp. 919–922.

[3] Q. Mei and C. Zhai, "Discovering evolutionary theme patterns from text—An exploration of temporal text mining," in *Proc. SIGKDD*, 2005, pp. 198–207.

[4] S. Liu, Y. Merhav, W. Yee, N. Goharian, and O. Frieder, "A sentence level probabilistic model for evolutionary theme pattern mining from news corpora," in *Proc. SAC*, 2009, pp. 1742–1747.

[5] Q. Mei, C. Liu, H. Su, and C. Zhai, "A probabilistic approach to spatio-temporal theme pattern mining on weblogs," in *Proc. WWW*, 2006, pp. 533–542.

[6] X. Wang and A. McCallum, "Topics over time: A non-Markov continuous-time model of topical trends," in *Proc. SIGKDD*, 2006, pp. 424–433.

[7] X. Wei, J. Sun, and X. Wang, "Dynamic mixture models for multiple time series," in *Proc. IJCAI*, 2007, pp. 2909–2914.

[8] L. AlSumait, D. Barbará, and C. Domeniconi, "On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking," in *Proc. ICDM*, 2008, pp. 3–12.

[9] M. Steyvers and T. Griffiths, "Probabilistic topic models," in *Latent Semantic Analysis: A Road to Meaning*, T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, Eds. Mahwah, NJ: Laurence Erlbaum, 2007.

[10] C. C. Chen, Y. Chen, and M. C. Chen, "An aging theory for event life-cycle modeling," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 2, pp. 237–248, Mar. 2007.

[11] C. C. Yang, X. D. Shi, and C. Wei, "Discovering event evolution graphs from news corpora," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 4, pp. 850–863, Jul. 2009.

[12] S. Das, A. Abraham, and A. Konar, "Automatic clustering using an improved differential evolution algorithm," *IEEE Trans. Syst. Man, Cybern. A, Syst. Humans*, vol. 38, no. 1, pp. 218–237, Jan. 2008.

[13] F. D. T. de Carvalho and Y. Lechevallier, "Dynamic clustering of interval-valued data based on adaptive quadratic distances," *IEEE Trans. Syst. Man, Cybern. A, Syst. Humans*, vol. 39, no. 6, pp. 1295–1306, Nov. 2009.