



Multi-Topic Tracking Model for dynamic social network



Yuhua Li^{a,*}, Changzheng Liu^a, Ming Zhao^b, Ruixuan Li^a, Hailing Xiao^a,
Kai Wang^a, Jun Zhang^a

^a Intelligent and Distributed Computing Lab, School of Computer Science and Technology,
Huazhong University of Science and Technology, Wuhan 430074, PR China

^b School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, United States

HIGHLIGHTS

- The selection behavior of users is considered in our topic tracking model.
- Selection model is combined with influence model to model dynamics of social links.
- A novel statistical model MTTD is proposed.
- MTTD can track multi-topic of dynamic social networks simultaneously.

ARTICLE INFO

Article history:

Received 4 December 2015

Available online 20 February 2016

Keywords:

Multi-Topic Tracking Model

Dynamic social network

Influence phenomenon

Selection phenomenon

ABSTRACT

The topic tracking problem has attracted much attention in the last decades. However, existing approaches rarely consider network structures and textual topics together. In this paper, we propose a novel statistical model based on dynamic bayesian network, namely Multi-Topic Tracking Model for Dynamic Social Network (MTTD). It takes influence phenomenon, selection phenomenon, document generative process and the evolution of textual topics into account. Specifically, in our MTTD model, Gibbs Random Field is defined to model the influence of historical status of users in the network and the interdependency between them in order to consider the influence phenomenon. To address the selection phenomenon, a stochastic block model is used to model the link generation process based on the users' interests to topics. Probabilistic Latent Semantic Analysis (PLSA) is used to describe the document generative process according to the users' interests. Finally, the dependence on the historical topic status is also considered to ensure the continuity of the topic itself in topic evolution model. Expectation Maximization (EM) algorithm is utilized to estimate parameters in the proposed MTTD model. Empirical experiments on real datasets show that the MTTD model performs better than Popular Event Tracking (PET) and Dynamic Topic Model (DTM) in generalization performance, topic interpretability performance, topic content evolution and topic popularity evolution performance.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Topic tracking is to track the trend of topics which people are interested in [1]. It is a useful method in information retrieval, event tracking and trend evolution analysis. Multi-topic tracking is used to track the trends of multiple topics and

* Corresponding author.

E-mail addresses: ldcliuhua@hust.edu.cn (Y. Li), liucz@sina.cn (C. Liu), mingzhao@asu.edu (M. Zhao), rxli@hust.edu.cn (R. Li), harryshayne@gmail.com (H. Xiao), wk0719@gmail.com (K. Wang), zhangjun9204@gmail.com (J. Zhang).

<http://dx.doi.org/10.1016/j.physa.2016.02.038>

0378-4371/© 2016 Elsevier B.V. All rights reserved.

help users understand the trend of popularity. Besides, it can identify interesting information and address changes in user interests over time [2]. Furthermore, it can analyze the relationship between topics and events and monitor event progress and influence factors in event tracking. For example, it can be used to follow research trend changes in science fields and help researchers identify areas that are interesting. An event in the reality can involve more than one topic. Thus, when you conduct multi-topic tracking, you can understand the whole event better not just only one topic or one respect of the whole event.

With the popularity of online social networks [3], it is appealing to develop a system that tracks the diffusion and evolution of trending topics in these platforms. Topic tracking allows us to infer interesting topics, origin of interests, and the evolution and spread of topics in the networks. However, topics are influenced by various factors including the change of network structures, users' interests, network information diffusion and selection phenomenon in dynamic social networks. In particular, the selection phenomenon characterizes the homophily, which suggests that users with similar interests are more willing to make connections among each other. These complex factors involved in dynamic social networks make topic tracking a difficult problem to solve. There are several methods studied in the related work to address this problem, but their effectiveness is limited. The Dynamic Topic Model (DTM) [4] omits network structures and considers only textual information. Although the Popular Event Tracking (PET) [5] takes social network factors into account, it cannot track multiple topics simultaneously and does not consider the selection phenomenon, which is natural in social networks and valuable to topic tracking.

Taking all the aforementioned factors into consideration, we propose a unified model, Multi-Topic Tracking Model for Dynamic Social Network (MTTD). It integrates textual information and network structures and can track multiple topics simultaneously. It provides a unified probabilistic model that considers (1) the network effect in information diffusion; (2) the selection phenomenon; (3) the burstiness of user interest; and (4) the evolution of textual topics. To the best of our knowledge, our MTTD model is the first to consider all four factors in a unified way. Although in this paper we focus on topic tracking, considering only primitive topics as the input, our model can be combined with existing topic models that can automatically extract topics.

To summarize, the contributions of this paper are as follows.

- (1) The selection behavior of users in social networks is considered in our topic tracking model. It is rarely taken into account in the existing tracking models. A link generative process is leveraged to model users' selection behavior according to their interests over topics.
- (2) The selection model is combined with well studied influence model to model the dynamics of social links. Specifically, the topic interests of users are affected by both the link structure of previous time slice and that of current time slice. Experimental results show that the model benefits from this combination.
- (3) The document generative model detects multi-topic and MTTD can track multi-topic of dynamic social networks simultaneously.
- (4) We incorporate influence model, selection model, document generative model and topic evolution model into a unified probabilistic model. We also present the detailed parameter estimation algorithm of the model based on Expectation Maximization (EM) framework [30]. Our extensive real dataset based experiments show that the MTTD is more effective than the existing tracking models in generalization performance, topic interpretability performance, topic evolution and topic popularity evolution performance.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 formally defines the problem. Section 4 describes the four sub-models of MTTD model and its parameter estimation algorithm. Section 5 presents the experimental results and Section 6 concludes the paper.

2. Related work

There is a good body of related work on topic tracking in the literature. In order to solve the topic tracking problem, we often have to utilize the topic model to obtain topics first, whereas some topic models can be directly applied to topic tracking. Thus, in this section we introduce the related work from these two areas, topic models and topic tracking.

Topic model. Topic modeling approaches have been applied to many textual mining applications. Two of the most well-known models are Probabilistic Latent Semantic Analysis (PLSA) [6] proposed by Hofmann in 2001 and Latent Dirichlet Allocation (LDA) [7] proposed by David M. Blei in 2003. Many approaches integrate other factors to extend these classical topic models, mainly from four different aspects. First, dynamic topic model, such as DTM [4] and Continuous Time Dynamic Topic Model (CTDTM) [8] consider time as a factor. Both of them can be directly applied into topic tracking and DTM is considered as a baseline model in our experiments. Second, incorporating network structure into the topic model has been studied in related work [9,10]. Third, because topics are often correlated in reality, their correlation is considered in the Correlated Topic Model (CTM) [11] proposed by David M. Blei and the Relational Topic Model (RTM) [12] proposed by Jonathan Chang. Finally, Author-Topic models [13,14] are proposed to extract the topic distribution of authors and associated documents simultaneously. There also exist other topic models considering other factors. For example, RankTopic [15] includes a mutual enhancement framework between ranking and topic modeling. The work by Hong et al. [16] incorporates geographical factor into topic model. These different topic models discussed above can be utilized by our proposed topic tracking approach to obtain the topic distribution at the beginning of topic tracking.

Download English Version:

<https://daneshyari.com/en/article/976529>

Download Persian Version:

<https://daneshyari.com/article/976529>

[Daneshyari.com](https://daneshyari.com)