2015 IEEE International Conference on Data Mining

Modeling Emerging, Evolving and Fading Topics using Dynamic Soft Orthogonal NMF with Sparse Representation

Yong Chen[1,2,a], Hui Zhang[1,2,a], Junjie Wu[3,b,*], Xingguang Wang[1,b], Rui Liu[1,b], and Mengxiang Lin[a]
[1]School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]National Science and Technology Resources Sharing Services Engineering Research Center, Beijing 100191, China
[3]School of Economics and Management, Beihang University, Beijing 100191, China
Email: [a]{chenyong, hzhang, mxlin}@nlsde.buaa.edu.cn, [b]{wujj, wangxggc, lr}@buaa.edu.cn
[*]*Corresponding Author*

***Abstract*—Dynamic topic models (DTM) are of great use to analyze the evolution of unobserved topics of a text collection over time. Recent years have witnessed the explosive growth of streaming text data emerging from online media, which creates an unprecedented need for DTMs for timely event analysis. While there have been some matrix factorization methods in the literature for dynamic topic modeling, further study is still in great need to model emerging, evolving and fading topics in a more natural and effective way. In light of this, we first propose a matrix factorization model called SONMFSR (Soft Orthogonal NMF with Sparse Representation), which makes full use of soft orthogonal and sparsity constraints for static topic modeling. Furthermore, by introducing the constraints of emerging, evolving and fading topics to SONMFSR, we easily obtain a novel DTM called SONMFSR$_d$ for dynamic event analysis. Extensive experiments on two public corpora demonstrate the superiority of SONMFSR$_d$ to some state-of-the-art DTMs in both topic detection and tracking. In particular, SONMFSR$_d$ shows great potential in real-world applications, where popular topics in Two Sessions 2015 are captured and traced dynamically for possible insights.**

Keywords*—Dynamic Topic Model (DTM); Non-negative Matrix Factorization (NMF); Soft Orthogonality; Sparse Representation; Topic Detection and Tracking (TDT)

I. Introduction

Since the last decade, topic model has attracted increasing interests in machine learning, information retrieval, natural language processing, machine translation and other similar fields. It can be roughly divided into two categories: probabilistic approaches such as probabilistic Latent Semantic Analysis (PLSA) [1] and Latent Dirichlet Allocation (LDA) [2], and matrix factorization approaches such as Latent Semantic Indexing (LSI) [3], Regularized LSI [4] and Non-negative Matrix Factorization [6]. Classic topic models are essentially static topic models (STM for short) that are concerned with mining latent topics from a whole collection without time constraints.

With the rapid development of Web2.0 applications in recent years, like micro-blogging, micro-messaging, and social networking, tremendous amounts of online streaming data are emerging, which brings unprecedented demand for dynamic topic models (DTM for short). For example, finding topics from news or online reviews in a dynamic manner could facilitate users' understanding of hot topics. Moreover, mining dynamic topics from abundant tweets could help detect unexpected events (riots, fires, etc.) so as to save public losses. In the area of sociology, dynamic topic models have been leveraged to find ordinary people's life routines and the shifts of preferences.

In the literature, LDA [2] and NMF [6] are two widely adopted STMs and naturally become the bases of most DTMs. However, topics mined by LDA usually share more words than that by NMF, which could bring troubles to dynamic topic tracking, where less overlapped topics are particularly sought. Classic NMF has been reported as good candidate for local structures mining from images [11], but its suitability for dynamic event analysis on streaming text data remains an open question. Therefore, in this paper, we aim to improve classic NMF by introducing more effective constraints so as to make it suitable for dynamic topic modeling.

Dynamic topic model essentially introduces time information over static topic model, and therefore have three basic states: emerging, evolving and fading. While there have been some matrix factorization DTMs proposed in the literature, some [7], [8], [9], [10] are just concerned with emerging and evolving topics but totoally ignore fading topics, and still some [7], [11] run with fixed number of topics without any adjustment over time, which is likely to lose some emerging topics. A possible way to model fading topics is to build a mapping matrix between adjacent topic (term-topic) matrices [12], which however is relatively complex and hard to handle in practice. From this perspective, there is still a great need in building DTMs for mining emerging, evolving and fading topics in a flexible and effective way.

In light of these, in this paper, we build a novel topic model called SONMFSR and extend it to SONMFSR$_d$ for dynamic topic modeling. Our main contributions are summarized as follows:

1) We propose a model called SONMFSR (Soft Orthogonal Non-negative Matrix Factorization with Sparse Representation) for static topic mining. By introducing soft orthogonality and sparsity constraints, SONMFSR is capable of detecting topics with less overlapping.

2) We extend SONMFSR to SONMFSR$_d$, which can model emerging, evolving and fading topics in a flexible and intuitive

1550-4786/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDM.2015.96

way. In particular, the fading topics are modeled by using the sparse representation of documents, which has seldom been touched in previous studies.

3) We conduct extensive experiments to demonstrate the effectiveness of SONMFSR$_d$ in dynamic topic detection and tracking. A case study on dynamic event analysis further illustrates the usefulness of SONMFSR$_d$ in real-world applications.

The remainder of this paper is organized as follows. In Sect. 2, we introduce related work. In Sect. 3, we define our problem and give necessary notations. In Sect. 4, we propose SONMFSR and extend it to SONMFSR$_d$ to model emerging, evolving and fading topics. In Sect. 5, we present an approximate learning process for matrix factorization and derive the optimization algorithms. We conduct experiments and a case study to evaluate the proposed models in Sect. 6, and finally draw conclusions in Sect. 7.

II. Related Work

Classic/Static Topic models (STM) can be divided into two kinds: probabilistic approaches and matrix factorization approaches. In the probabilistic approaches, a topic is defined as a probability distribution over terms and a document is defined as data generated from mixtures of topics. PLSA [1] and LDA [2] are such widely used generative models. While in the matrix factorization approaches, term-document matrix can be projected into a low dimensional (K dimensions) subspace which is also referred as latent topic space through linear or nonlinear transformation. In the topic space, each coordinate axis corresponds to a topic, and then each document can be represented as a linear combination of these K topics. LSI [3] and NMF [6] are such matrix factorization models.

Dynamic topic models throw time information on topics other than static; the topics in a document collection evolve over time, and it's of great interest to explicitly model the dynamics (emerging, evolving and fading) of the underlying topics. Specifically, collected documents can be divided into sequential parts according to timeslots and document exchangeability will not affect the dynamic topics at the same timeslot. Since STM has two categories, so does DTM: dynamic probabilistic approaches and dynamic matrix factorization approaches.

Probabilistic DTMs embed independent LDA in the timeline through bridging the evolving between adjacent topics. Blei [7] built the evolving relationship using Gaussian fluctuation with respect to the same topic over time and proposed the first dynamic topic model. Chong et al. [13] developed a continuous time dynamic topic model with Brownian motion extending the discrete dynamic topic model [7] for fine-grained discretization. Tomoharu Iwata [14] proposed online multiscale DTM under the assumption that current topic-specific distributions over words are generated based on the multiscale word distributions of the previous epoch and designed an efficient stochastic EM algorithm for fast detecting and tracking dynamic topics. However, such LDA-based DTM always choose fixed number of topics during the evolving span and then identify emerging, evolving and fading (often ignored in many research) topics according to the term-topic matrix. This approach is clearly not flexible and may lose some important topics. Therefore, in this paper part of new topics will be introduced at each new timeslot to capture the emerging topics adaptively.

Matrix factorization based DTMs mainly characterize dynamic topics in the NMF framework [6], [15], [17], [18], [19], [20] and many variants were proposed recently. Cao et al. [8] modeled emerging and evolving topics by injecting a small number of new topics at each new timeslot and proposed a framework for detecting and tracking the latent factors in data streams by extending the NMF with orthogonal constraints. Ankan Saha et al. [9] also only concentrated on the emerging and evolving topics and proposed an online NMF framework to capture these two dynamics in unstructured text under temporal regularizations. Carmen Vaca et al. [12] provided a simple way to discover trends: emerging, evolving and fading topics by introducing collective time-based factorization algorithm for topic detecting and tracking of evolving input streams. However, such works either ignored the fading topics, or choose the fixed number of topics, or use mapping matrix for identifying dynamic topics for simplifications. Therefore, this paper tries to model these three dynamic topics in a more simple and intuitive way.

What's more, we will make full use of topic-document matrix to remove fading topics by the sparsity and ratio of each topic in an intuitive idea while other DTMs may not focus on this point. All in all, we model the three kind of dynamic topics in the SONMFSR framework and propose a simple and intuitive DTM named Modeling Emerging, Evolving and Fading Topics using Dynamic Soft Orthogonal NMF with Sparse Representation (SONMFSR$_d$) in this paper. A series of experiments in section 6 would show that SONMFSR$_d$ has a good performance on topic detecting and tracking.

III. Problem Statement

Given a set of documents D with size N, we can construct a vocabulary with size M after data processing and feature extraction. Then a document can be simply represented as an M-dimensional vector d, where the m-th entry denotes the score of the m-th term. The N documents in D are then represented as an $M \times N$ term-document matrix $D = [d_1, d_2, \cdots d_N]$, in which each row corresponds to a term and each column corresponds to a document.

A topic is defined over terms in the vocabulary and is also represented as an M-dimensional vector u, where the m-th entry denotes the weight of the m-th term in the topic. Intuitively, the term with larger weight is more relevant to the topic. Suppose that there are K topics in the datasets; then the K topics can be summarized into an $M \times K$ term-topic matrix $U = [u_1, u_2, \cdots u_K]$, in which each column corresponds to a topic.

Matrix factorization is essentially discovering the latent topics in the datasets as well as modeling the documents by representing them as mixtures of the topics. More precisely, given topics $u_1, u_2, \cdots u_K$, document d_n is represented as $d_n = \sum_{k=1}^{K} v_{kn} u_k = U v_n$, where v_{kn} denotes the weight of the k-th topic in document d_n. Let $V = [v_1, v_2, \cdots v_N]$ be the topic-document matrix, where column vector v_n stands for the representation of the n-th document in the latent topic space.

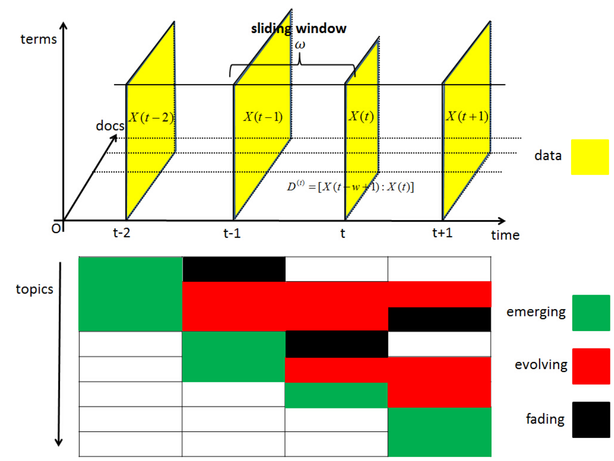


Fig. 1. Illustration of dynamic topic models: (1) The upper figure illustrates the evolving of collected data (term-document matrix) over different time slots, where $X(t)$ denotes the documents collected from timestamp t-1 to timestamp t; note that different data share the same vocabulary but different documents; (2) The lower figure shows various types of dynamic topics over different time slots: emerging (green), evolving (red) and fading (black).

DTM is of great interest to explicitly model the dynamic topics: emerging, evolving and fading. The big difference between DTM and STM seems to be the time span of topics. Formally DTM is just marked with time tag while STM is not. Let $X(t) \in R^{M^{(t)} \times N^{(t)}}$ be the collected data from the end of time $(t-1)$ to the start of time t, where $M^{(t)}$ denotes the size of vocabulary and $N^{(t)}$ is the number of collected documents at time t. Assume a vocabulary is shared through all the evolving time, and then $M^{(t)} = M$, where M represents the number of terms in the vocabulary. Let $D^{(t)} \in R^{M \times N^{(t)}}$ be the collected data within the sliding window at time t, specifically $D^{(t)} = [X(t-w+1) : X(t)]$, where ω is the size of the sliding window i.e. from the end of time $(t-w)$ to the start of time t, which is illustrated in the Fig. 1.

Therefore, we can perform topic model on collected documents at a specific time t and get the term-topic matrix $U^{(t)} \in R^{M \times K^{(t)}}$ and topic-document matrix $V^{(t)} \in R^{K^{(t)} \times N^{(t)}}$, where $K^{(t)}$ denotes the number of topics at time t. The dynamics of underlying topics are elaborated as follows: (1) at time t there are parts of topics evolving from the previous time $(t-1)$; (2) some topics at time $(t-1)$ may disappear at time t; (3) and some new topics would emerge at time t as illustrated in the Fig. 1. Thus $K^{(t)}$ can be divided into two parts: one is the number of evolving topics from the previous time $(t-1)$ denoted by $K^{(t)}_{evolve}$, and the other is the number of the newly emerging topics denoted by $K^{(t)}_{emerge}$ and there holds $K^{(t)} = K^{(t)}_{evolve} + K^{(t)}_{emerge}$. It is worth mentioning that the $U^{(t-1)}_{-fading}$ is formed from the $U^{(t-1)} \in R^{M \times K^{(t-1)}}$ with fading topics removed at time $(t-1)$. Obviously $U^{(t)}_{evolve}$ just evolves from $U^{(t-1)}_{-fading}$ with small fluctuations. Table I gives a summary of math notations and explanations.

TABLE I. MATH NOTATIONS

Notations	Explanations
$M^{(t)} = M$	Size of vocabulary at time t; and it does not vary over time;
$N^{(t)}$	Number of documents at time t;
$K^{(t)}$	Number of topics at time t, and $K^{(t)} = K^{(t)}_{evolve} + K^{(t)}_{emerge}$;
$K^{(t)}_{evolve}$	Number of evolving topics from the previous time $(t-1)$ at time t;
$K^{(t)}_{emerge}$	Number of emerging topics at time t;
$D^{(t)} \in R^{M \times N^{(t)}}$	Term-document matrix collected within the sliding window at time t; $D^{(t)} = [d^{(t)}_1, d^{(t)}_2, \cdots, d^{(t)}_{N^{(t)}}]$
$d^{(t)}_n$	The n-th document represented in the term space at time t;
$d^{(t)}_{mn}$	Weight of the m-th term in the n-th document at time t;
$U^{(t)} \in R^{M \times K^{(t)}}$	Term-topic matrix at time t, where $U^{(t)} = [U^{(t)}_{evolve}, U^{(t)}_{emerge}]$; $U^{(t)} = [u^{(t)}_1, u^{(t)}_2, \cdots, u^{(t)}_{K^{(t)}}]$
$U^{(t)}_{evolve} \in R^{M \times K^{(t)}_{evolve}}$	The evolving topics from previous time $(t-1)$;
$U^{(t)}_{emerge} \in R^{M \times K^{(t)}_{emerge}}$	The emerging topics at time t;
$U^{(t-1)}_{-fading} \in R^{M \times K^{(t)}_{evolve}}$	The term-topic matrix $U^{(t-1)}$ with fading topics removed at time $(t-1)$ as the next time t's evolving topics;
$u^{(t)}_k$	The k-th topic represented in the term space at time t;
$u^{(t)}_{mk}$	Weight of the m-th term in the k-th topic at time t;
$V^{(t)} \in R^{K^{(t)} \times N^{(t)}}$	Topic-document matrix at time t, where $V^{(t)} = \begin{bmatrix} V^{(t)}_{evolve} \\ V^{(t)}_{emerge} \end{bmatrix}$, $V^{(t)}_{evolve}$ and $V^{(t)}_{emerge}$ correspond to evolving and emerging topics respectively; $V^{(t)} = [v^{(t)}_1, v^{(t)}_2, \cdots, v^{(t)}_{N^{(t)}}]$
$v^{(t)}_n$	The n-th document represented in the topic space at time t;
$v^{(t)}_{kn}$	Weight of the k-th topic in the n-th document at time t;

IV. Modeling Emerging, Evolving and Fading Topics using Dynamic Soft Orthogonal NMF with Sparse Representation

A. The SONMFSR Model

Different matrix factorization models choose different schemes to model matrices U and V and impose different constraints on them. For example, in LSI [3], $u_1, u_2, \cdots u_K$ are orthogonal between each other and then can be solved with SVD [21], [22]. But the obtained matrices U and V can be negative or non-negative, and could not make a good explanation on the practical problems such as image understanding, topics mining, signal processing, genetic fingerprint extraction and so on. Then in 1999 D.D.Lee and H.S.Seung proposed Non-negative Matrix Factorization [6] and successfully solved this problem. NMF owns the ability to explain the actual data and surprisingly mine structures as human cognition, so it is widely used in machine learning, data mining, biomedicine and other related fields.

However we found that NMF usually could not get fully localized structures which were often mixed with noises in practice [23]. In order to handle such problem, we introduce soft orthogonal constraints (i.e. any two different topics are close to orthogonal, which can be formalized as $||U^T U - I_K||^2_F \le \upsilon$, $(\upsilon \ge 0)$) on term-topic matrix and $L1$ regularization on topic-document matrix for better characterizing sparse data in large scale dataset, then propose Soft Orthogonal Non-negative Matrix Factorization with Sparse Representation framework.

Taking into account the above description and analysis, our SONMFSR framework is to solve the following optimization problem:

$$\min_{U,V} \|D-UV\|_F^2 + \lambda \sum_{n=1}^{N} \|v_n\|_1^1 \quad (1)$$

$$s.t. \begin{cases} ||U^TU - I_K||_F^2 \leq \upsilon, \ (\upsilon \geq 0) \\ u_{mk} \geq 0, \ m = 1, 2, \cdots, M; \\ v_{kn} \geq 0, \ n = 1, 2, \cdots, N; \ k = 1, 2, \cdots K, \end{cases}$$

where $|| * ||_F$ is the Frobenius norm, υ and λ are two hyper parameters with non-negative values, I_K is a $K \times K$ diagonal matrix. The much smaller υ is, the much closer to orthogonal U is; while the larger λ is, the sparser v_n will be.

For the simplicity of designing algorithms, the optimization problem (1) is equivalently converted to solve the following optimization problem:

$$\min_{U,V} \|D-UV\|_F^2 + \lambda \sum_{n=1}^{N} \|v_n\|_1^1 + \xi \|U^TU - I_K\|_F^2 \quad (2)$$

$$s.t. \begin{cases} u_{mk} \geq 0, m = 1, 2, \ldots, M; \\ v_{kn} \geq 0, n = 1, 2, \ldots, N; k = 1, 2, \ldots, K, \end{cases}$$

where ξ is a parameter with non-negative values, and the larger ξ is, the closer to orthogonal U will be.

B. SONMFSR$_d$: Dynamic Topic Model based on SONMFSR

The main tasks of our DTM are to model the emerging, evolving and fading topics over time. These three dynamics can be summarized as follows:

(1) From time $(t-1)$ to time t, there are two dynamics: evolving and fading, i.e. part of the topics continue to evolve with a smooth fluctuation; the others would disappear in the next timeslot.

(2) At time t, there would probably be new emerging topics.

All these three dynamics of topics can be analyzed and modeled in a very simple and intuitive way.

(1) Modeling Emerging Topics: At time t, there are some topics evolving from the previous time $(t-1)$ and some new emerging topics. With respect to the emerging ones, we can dynamically add a possible number of topics $U_{emerge}^{(t)}$ to the evolving parts $U_{evolve}^{(t)}$.

(2) Modeling Evolving Topics: Topic evolving is a gradual process, and there will be some topics evolving from time $(t-1)$ to time t. We assume the evolving topics change over time with small-scale fluctuations for the same topic. Therefore, $U_{evolve}^{(t)}$ should correspond to $U_{-fading}^{(t-1)}$ with the following constraint:

$$||U_{evolve}^{(t)} - U_{-fading}^{(t-1)}||_F^2 \leq \delta, \quad (3)$$

where δ is a threshold to characterize the fluctuations. The much smaller δ is, the less fluctuation the topic would be. The greatest benefit of this model is that the same column of $U_{evolve}^{(t)}$ and $U_{-fading}^{(t-1)}$ naturally represents the same topic at different times.

(3) Modeling Fading Topics: At time t we can get the term-topic matrix $U^{(t)} \geq 0$ and topic-document matrix $V^{(t)} \geq 0$ through topic models. Here, more attention would be paid to the topic-document matrix with $L1$ norm regularizer for sparsity which is import for our fading strategies. With respect to some specific topic at time $(t-1)$, if few documents contain the topic (the corresponding row of the topic is very sparse in $V^{(t-1)}$), then we could remove it in the next time. More specifically, if

$$\frac{\sum_{n=1}^{N^{(t-1)}} I\{v_{kn}^{(t-1)} = 0\}}{N^{(t-1)}} > \tau, \quad (4)$$

where τ is a threshold of sparsity, then the k-th topic will be removed at time t. $I\{\bullet\}$ is an indicator function. If the condition inside the braces is satisfied, then the function value is equal to 1; otherwise 0.

In addition, if the ratio of the topic is very low, we can also identify it as the fading topic. More specifically, if

$$\frac{\sum_{n=1}^{N^{(t-1)}} v_{kn}^{(t-1)}}{\max\{\sum_{n=1}^{N^{(t-1)}} v_{kn}^{(t-1)} : k = 1, 2, \cdots, K^{(t-1)}\}} < \sigma, \quad (5)$$

where σ is the ratio threshold, then the k-th topic will be identified as a fading topic.

Then we can build our DTM based on SONMFSR. Specifically, at time t, there holds the following optimization problem:

$$\min \ ||D^{(t)} - U^{(t)}V^{(t)}||_F^2 + \lambda \sum_{n=1}^{N^{(t)}} ||v_n^{(t)}||_1^1 \quad (6)$$

$$s.t. \begin{cases} U^{(t)} \geq 0, \ V^{(t)} \geq 0 \\ U^{(t)} = [U_{evolve}^{(t)}, U_{emerge}^{(t)}] \\ ||U_{evolve}^{(t)} - U_{-fading}^{(t-1)}||_F^2 \leq \delta \\ ||[U^{(t)}]^T[U^{(t)}] - I_{K^{(t)}}|| \leq \upsilon, \end{cases}$$

where $U^{(t)} \geq 0$, $V^{(t)} \geq 0$ denote all the elements non-negative. The optimization (6) can be transformed into the equivalent optimization as follows:

$$\begin{aligned} \min \ & ||D^{(t)} - U^{(t)}V^{(t)}||_F^2 + \xi||[U^{(t)}]^T[U^{(t)}] - I_{K^{(t)}}||_F^2 \\ & + \eta||U_{evolve}^{(t)} - U_{-fading}^{(t-1)}||_F^2 + \lambda \sum_{n=1}^{N^{(t)}} ||v_n^{(t)}||_1^1 \end{aligned} \quad (7)$$

$$s.t. \begin{cases} U^{(t)} \geq 0, \ V^{(t)} \geq 0 \\ U^{(t)} = [U_{evolve}^{(t)}, U_{emerge}^{(t)}], \end{cases}$$

where ξ, η and λ are non-negative hyper parameters, and the larger η is, the less fluctuation the evolving topics will be.

V. ALGORITHM DERIVATIONS

For a specific time t, we can easily obtain $D^{(t)}$ and $U^{(t-1)}_{-fading}$, and then the solution of optimization (7) is to find the optimal $U^{(t)}$ and $V^{(t)}$. Here the time stamps are omitted to simplify the algorithm derivations. Let $U_{pre} = U^{(t-1)}_{-fading}$, then the optimization (7) is converted into the following expressions:

$$\begin{aligned} \min\ & ||D - UV||_F^2 + \xi||[U]^T[U] - I_K||_F^2 \\ & + \eta||U_{evolve} - U_{pre}||_F^2 + \lambda \sum_{n=1}^{N} ||v_n||_1^1 \\ s.t.\ & \begin{cases} U \geq 0,\ V \geq 0 \\ U = [U_{evolve}, U_{emerge}]. \end{cases} \end{aligned} \tag{8}$$

The optimization problem (8) is convex with respect to U when V is fixed and convex with respect to V when U is fixed. However, it is not convex with respect to both of them. Following the practice in Sparse Coding [24], [25], [26], [27], we optimize the objective function in (8) by alternately minimizing it with respect to term-topic matrix U and topic-document matrix V. This procedure is summarized as Algorithm 1 in Table II.

TABLE II. APPROXIMATE LEARNING PROCESS

Algorithm 1: Our DTM (SONMFSR$_d$) based on SONMFSR
Input: $D \in R^{M \times N}$ and $U_{pre} \in R^{M \times K_{evolve}}$
Output: $U \in R^{M \times K}$ and $V \in R^{K \times N}$

$V^{(0)} \leftarrow random(0, 1)$
for s=1:S do
 $U^{(s)} \leftarrow UpdateU(D, U_{pre}, V^{(s-1)})$
 $V^{(s)} \leftarrow UpdateV(D, U^{(s)})$
end for
$return\ U^{(S)}, V^{(S)}$

A. Update of Matrix U

Holding matrix V fixed, the update of U amounts to solve the following optimization problem:

$$\begin{aligned} \min_U\ & ||D - UV||_F^2 + \xi||U^TU - I_K||_F^2 \\ & + \eta||U_{evolve} - U_{pre}||_F^2 \end{aligned} \tag{9}$$
$$s.t.\ u_{mk} \geq 0,\ m = 1, 2, \cdots, M;\ k = 1, 2, \cdots K.$$

For any matrix $A \in R^{M \times N}$, there is $tr(A^TA) = \|A\|_F^2$ where $tr(\bullet)$ denotes the trace. Then let the objective function of the optimization problem (9) be L as follows:

$$\begin{aligned} L &= ||D - UV||_F^2 + \xi||U^TU - I_K||_F^2 \\ &+ \eta||U_{evolve} - U_{pre}||_F^2 \\ &= tr((D - UV)^T(D - UV)) + \xi tr((U^TU - I_K)^T(U^TU \\ &- I_K)) + \eta tr((U_{evolve} - U_{pre})^T(U_{evolve} - U_{pre})). \end{aligned} \tag{10}$$

The derivation of the objective function L over matrix U can be derived as the following equation:

$$\begin{aligned} \frac{\partial L}{\partial U} &= 2(UVV^T - DV^T) + \xi\left(4UU^TU - 4U\right) \\ &+ \eta([2U_{evolve} - 2U_{pre}, O]). \end{aligned} \tag{11}$$

Then we can solve this optimization problem with projected gradient descent method. Specific procedures are elaborated as follows:

(step-1): $U_{(s+1)} = U_{(s)} - \Delta\frac{\partial L}{\partial U_{(s)}}$, where $\Delta > 0$ is the step size, specifically $\Delta = f(C, s)$ (which can be set $\Delta = \frac{C}{\sqrt{s++}}$ or $\Delta = C$), and C is a pre-defined constant, s is the iterative index;

(step-2): $U_{(s+1)} = \mathrm{P}_\Omega[U_{(s+1)}]$, where $\mathrm{P}_\Omega[\bullet]$ is the projection operator, i.e. $X = \mathrm{P}_\Omega[X] = \arg\inf_{Y \in \Omega} ||X - Y||$, where X can be any matrix and Ω is the projection domain.

(step-3): Iterate step-1 and step-2 until convergence.

Thus the procedure for updating U is summarized as Algorithm 2 in Table III.

TABLE III. UPDATE OF MATRIX U

Algorithm 2: UpdateU
Input: $D \in R^{M \times N}, U_{pre} \in R^{M \times K_{evolve}}, V \in R^{K \times N}$
$iterMax \in R, Eps \in R, C \in R$
Output: $U \in R^{M \times K}$

$U_{(1)} \in R^{M \times K} \leftarrow random(0, 1)$
$s \leftarrow 1$
$repeat$
 $L \in R^{M \times K} \leftarrow 2U_{(s)}VV^T - 2DV^T$
 $R \in R^{M \times K} \leftarrow 4\xi(U_{(s)}U_{(s)}^TU_{(s)} - U_{(s)})$
 $Q \in R^{M \times K} \leftarrow 2\eta([(U_{evolve})_{(s)} - U_{pre}, O])$
 $G \in R^{M \times K} \leftarrow L + R + Q$
 $\triangle \leftarrow C/\sqrt{s++}$
 $U_{(s+1)} \leftarrow \mathrm{P}_\Omega[U_{(s)} - G \times \triangle], \Omega = \{U \geq 0\}$
 $s \leftarrow s + 1$
$until\ s > iterMax\ or\ \|U_{(s+1)} - U_{(s)}\|_F \leq Eps$
$return\ U_{(s)}$

B. Update of Matrix V

Holding matrix U fixed, the Update of matrix V is equivalent to solve the following optimization problem:

$$\begin{aligned} & \min_V \|D - UV\|_F^2 + \lambda\sum_{n=1}^{N}\|v_n\|_1^1 \\ \Leftrightarrow & \min_{\{v_n\}} \sum_{n=1}^{N}\|d_n - Uv_n\|_2^2 + \lambda\sum_{n=1}^{N}\|v_n\|_1^1 \\ & s.t.\ v_{kn} \geq 0, n = 1, 2, \ldots, N; k = 1, 2, \ldots, K, \end{aligned} \tag{12}$$

where v_{kn} represents the (kn)-th element of V. Because v_n is independent on each other, then the optimization problem (12) can be decomposed into N sub-optimizations, with each sub-optimization corresponding to one v_n and can be solved in parallel:

$$\begin{aligned} & \min_{v_n}\|d_n - Uv_n\|_2^2 + \lambda\|v_n\|_1^1 \\ & s.t. v_{kn} \geq 0, k = 1, 2, \ldots, K; (n = 1, 2, \ldots, N). \end{aligned} \tag{13}$$

The optimization problem (13) is a least square problem with non-negative constraints and $L1$ regularization. Here we can optimize the problem with the thought of coordinate descent. Let $U_{\setminus k}$ denote the matrix of U with the k-th column removed, $v_{n\setminus k}$ denote the vector of v_n with the k-th entry removed, and $const$ be a constant with respect to v_{kn}, we have

$$
\begin{aligned}
L &= \|d_n - Uv_n\|_2^2 + \lambda\|v_n\|_1^1 \qquad (14)\\
&= \|d_n - (U_{\setminus k}v_{n\setminus k} + u_k v_{kn})\|_2^2 + \lambda\sum_{k=1}^{K} v_{kn}\\
&= \lambda\sum_{k=1}^{K} v_{kn} + [(d_n - U_{\setminus k}v_{n\setminus k}) - u_k v_{kn}]^T\\
&\qquad \times [(d_n - U_{\setminus k}v_{n\setminus k}) - u_k v_{kn}]\\
&= (d_n - U_{\setminus k}v_{n\setminus k})^T(d_n - U_{\setminus k}v_{n\setminus k}) + \lambda\sum_{k=1}^{K} v_{kn}\\
&\qquad - 2v_{kn}(u_k)^T(d_n - U_{\setminus k}v_{n\setminus k}) + \|u_k\|_2^2(v_{kn})^2\\
&= \|u_k\|_2^2(v_{kn})^2 - 2v_{kn}(u_k)^T(d_n - U_{\setminus k}v_{n\setminus k})\\
&\qquad + \lambda v_{kn} + const.
\end{aligned}
$$

Then the optimization problem (13) is transformed to solve K sub-optimization problems:

$$
\begin{aligned}
\min\, (u_k)^T u_k(v_{kn})^2 - 2v_{kn}(u_k)^T(d_n - U_{\setminus k}v_{n\setminus k}) \qquad (15)\\
+ \lambda v_{kn} + const\\
s.t. v_{kn} \geq 0; (k = 1, 2, \ldots, K).
\end{aligned}
$$

Let S_{ij} and R_{ij} are the $(ij)^{th}$ entries of $K \times K$ matrix $S = U^T U$ and $K \times N$ matrix $R = U^T D$, respectively. Therefore the optimization problem (15) is just a simple optimization problem with non-negative constraints which can easily be solved as follows:

$$
v_{kn} = \max\left\{0, \frac{R_{kn} - \sum_{1\leq l\leq K, l\neq k} S_{kl} \times v_{ln} - \frac{\lambda}{2}}{S_{kk}}\right\}. \qquad (16)
$$

Finally, the algorithm for updating of matrix V with coordinate descent can be described as Algorithm 3 in Table IV:

TABLE IV. UPDATE OF MATRIX V

Algorithm 3: UpdateV
Input: $D \in R^{M\times N}, U \in R^{M\times K}$
Output: $V \in R^{K\times N}$
$S \leftarrow U^T U$
$R \leftarrow U^T D$
$for\ n = 1 : N\ do$
$\quad v_n \leftarrow random(0, 1)$
$\quad repeat$
$\qquad for\ k = 1 : K\ do$
$\qquad\quad v_{kn} = \max\left\{0, \frac{R_{kn} - \sum_{1\leq l\leq K, l\neq k} S_{kl} \times v_{ln} - \frac{\lambda}{2}}{S_{kk}}\right\}$
$\qquad end\ for$
$\quad until\ convergance$
$end\ for$
$return\ V$

VI. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed SONMFSR model on various public text corpora for topic detection in both static and dynamic environments. A case study on the focus event in China: *Two Sessions (of the NPC and the CPPCC) 2015* further illustrates the effectiveness of our model.

A. Experimental Setup

1) Data Sets: Two publicly available news corpora are adopted for comparative studies on competing models, whose documents are labeled with news categories. Details are given as follows.

Sougou corpus. Sougou text corpus contains tens of thousands documents from the Sohu news website, most of which are edited and classified manually. It consists of a classification system of dozens of classified nodes. The broad categories are automobile, finance, IT, health and other categories. Users with interests can refer to [31] for more details. We here randomly select 500 documents from the Finance, Tourism, Health, IT and Sports categories, respectively, and thus construct a corpus with 2500 news docoments. For the experiments on dynamic topic detection, we further partition the whole corpus randomly into 10 parts as 10 timeslots in sequence, and each timeslot contains 250 news documents.

20Newsgroup. The 20Newsgroups data set is a collection of approximately 20000 documents, partitioned (nearly) evenly across 20 different newsgroups, each corresponding to a different news category. Users with interests can refer to [32] for more details. Here, we randomly select 250 documents from each of the 10 categories and construct a corpus with 2500 documents. Similar to the Sougou corpus, we also construct a synthetic dynamic news corpus over 10 timeslots, each with 250 documents.

2) Baseline Models: Various state-of-the-art topic models are adopted here for comparative study.

The baselines for static topic modeling include: (1) FSTM: Fully Sparse Topic Model [29]; (2) NMF: Non-negative Matrix Factorization [6], [19]; (3) SNMF: Sparse NMF [20], [24]; (4) ONMF: Orthogonal NMF [8], [11]. We set the number of topics to 20, 40, 60, 80, and 100, respectively. The parameters of SONMFSR are set as follows: $\xi = 0.7$, $\lambda = 0.1$, the step $C = 0.02$, and $\sigma = 0.1$ ($\tau = 0.95$ optional).

The baselines for dynamic topic modeling include: (1) T-model [12], [20]; (2) Online NMF [8]; (3) Dynamic NMF [9]. Since the combination of parameter configurations is huge for dynamic experiments, we take the following two as examples: (1) the initial number of topics is 8 (i.e., $K^{(1)} = 8$) and each increases 8 (i.e., $K^{(t)}_{emerge} = 8$, $\forall\ t > 1$) at a new timeslot; (2) the initial number of topics is 26 (i.e., $K^{(1)} = 26$) and each increases 6 (i.e., $K^{(t)}_{emerge} = 6$, $\forall\ t > 1$) at a new timeslot. In both cases, we can get 80 topics in total at the last timeslot. Moreover, after conducting cross-validations, we set $\eta = 1$ or $\eta = 10$ for better results.

3) Evaluation Measures: We here briefly introduce the two measures, i.e., micro-averaged $F1$ and intensity, which are employed for the evaluations of topic detecting and tracking, respectively.

Since document labels are available for the data sets, we use the widely adopted external measure: *micro-averaged* $F1$ [9], [35] ($F1$ for short) to evaluate the effectiveness of topic detecting. Now that the documents in Sougou and 20Newsgroup are labeled with classes (e.g. sports, economy, health, tourism etc.), then we can call these labeled classes as "true topics" or "ground truth". Assume that the topic model

TABLE V. STATIC TOPIC MINING PERFORMANCE: BY MICRO-AVERAGED $F1$

		T=20			T=40			T=60			T=80			T=100		
		K=1	K=2	K=3	K=1	K=2	K=3	K=1	K=2	K=3	K=1	K=2	K=3	K=1	K=2	K=3
Sougou	NMF	0.474	0.445	0.419	0.471	0.447	0.44	0.466	0.463	0.437	0.48	0.495	0.48	0.464	0.486	0.464
	ONMF	0.454	0.432	0.411	0.468	0.444	0.427	0.48	0.448	0.422	0.467	0.456	0.446	0.472	0.476	0.453
	SNMF	0.523	0.49	0.448	0.534	0.515	0.493	0.505	0.521	0.505	0.501	0.536	0.531	0.493	0.497	0.523
	FSTM	0.624	**0.714**	**0.684**	0.514	**0.676**	**0.717**	0.403	0.551	0.646	0.364	0.522	0.601	0.351	0.503	0.6
	SONMFSR	**0.719**	<u>0.658</u>	<u>0.608</u>	**0.574**	<u>0.606</u>	<u>0.629</u>	**0.654**	**0.724**	**0.756**	**0.524**	**0.622**	**0.663**	**0.565**	**0.615**	**0.666**
20Newsgroup	NMF	0.303	0.247	0.223	0.333	0.279	0.251	0.361	0.298	0.262	0.38	0.318	0.281	0.37	0.333	0.293
	ONMF	0.282	0.236	0.214	0.313	0.26	0.235	0.34	0.276	0.245	0.361	0.296	0.257	0.348	0.293	0.257
	SNMF	0.323	0.265	0.23	0.367	0.31	0.268	0.421	0.347	0.293	0.41	0.377	0.332	0.348	0.293	0.257
	FSTM	**0.512**	**0.472**	**0.414**	0.477	**0.539**	**0.521**	0.378	0.479	**0.489**	0.364	0.447	0.481	0.33	0.422	0.447
	SONMFSR	<u>0.409</u>	<u>0.328</u>	<u>0.284</u>	**0.562**	<u>0.501</u>	<u>0.453</u>	**0.555**	**0.496**	<u>0.479</u>	**0.567**	**0.55**	**0.507**	**0.527**	**0.549**	**0.514**

Note: The best results are in bold, and the second-best results are underlined.

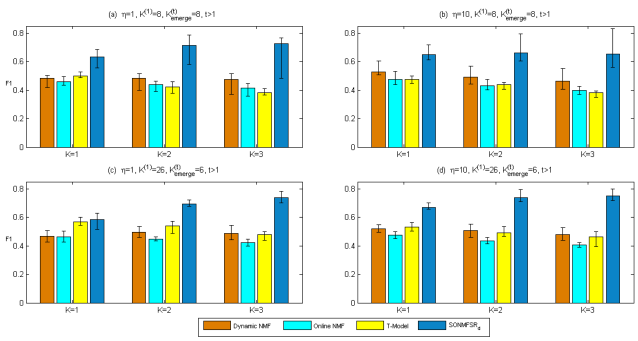


Fig. 2. Topic detecting performance on Sougou corpus (by $F1$).

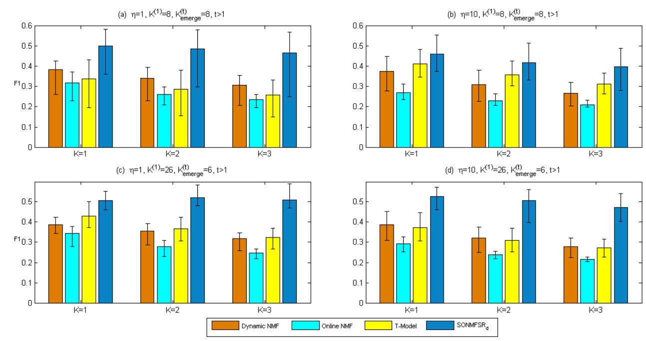


Fig. 3. Topic detecting performance on 20Newsgroup corpus (by $F1$).

generates T topics labeled $(1, 2, \ldots, T)$ where in general $T \neq L$ (L is the number of class labels (C_1, C_2, ..., C_L) of collected documents) i.e., the topic model is allowed to generate any number of topics. Then we construct the $L \times T$ confusion matrix CM between classes and topics, i.e., $CM(c, t)$ is the number of documents that were tagged C_l by the ground truth and tagged t by the topic model. From this matrix, for each label C_l, we identify $top_K(C_l)$ as the set of top-K most frequently co-occuring topics generated by topic model. Then we can compute the microaveraged $F1$ measure as follows:

$$Precision_K = \frac{\sum_{l=1}^{L} |D_{true}(C_l) \cap D_{model}(C_l)|}{\sum_{l=1}^{L} |D_{model}(C_l)|}, \tag{17}$$

$$Recall_K = \frac{\sum_{l=1}^{L} |D_{true}(C_l) \cap D_{model}(C_l)|}{\sum_{l=1}^{L} |D_{true}(C_l)|}, \tag{18}$$

$$F1_K = \frac{2 \times Precision_K \times Recall_K}{Precision_K + Recall_K}, \tag{19}$$

where $D_{true}(C_l)$ is the set of documents labeled C_l by the ground truth and $D_{model}(C_l)$ is the set of documents topic-model-tagged with a topic in the $top_K(C_l)$, $|\bullet|$ denotes the cardinality and $\cap$ represents intersection operations on sets.

As to topic tracking, we use about intensity. Since dynamic topics evolve with small fluctuations over different times, then we choose the top-10 terms of the k-th dynamic topic at its emerging as a query denoted q_k and compute the similarities with the documents within the corresponding timeslot t, then the averaged similarity could be treated as the intensity of the ground truth's topic denoted $Intensity_{true}(t, k)$ [12], [9]. More specifically,

$$Intensity_{true}(t, k) = \frac{\sum_{n=1}^{N^{(t)}} sim(q_k, d_n^{(t)})}{N^{(t)}}, \tag{20}$$

where $sim(q_k, d_n^{(t)})$ is the cosine value of query q_k and document $d_n^{(t)}$, i.e., $sim(q_k, d_n^{(t)}) = \frac{<q_k, d_n^{(t)}>}{|q_k| \times |d_n^{(t)}|}$.

While the intensity of the k-th dynamic topic at timeslot t of our proposed DTM denoted $Intensity_{SONMFSR_d}(t, k)$ is the average value of the k-th row vector in the topic-document matrix at timeslot t. More specifically,

$$Intensity_{SONMFSR_d}(t, k) = \frac{\sum_{n=1}^{N^{(t)}} v_{kn}^{(t)}}{N^{(t)}}. \tag{21}$$

Naturally, each timeslot will get one intensity and then the intensity trending of the ground truth and our DTM are formed respectively.

B. Static Topic Mining from Sougou and 20Newsgroup

Given the two text corpora, we do word segmentation, filter stop words, and conduct other related treatments as done in [30], [33]. The data sets are finally represented as a term-document matrix suitable for topic modeling.

We first evaluate the performance of SONMFSR in static topic mining. We compare SONMFSR with some state-of-the-art topic models: NMF, SNMF, ONMF and FSTM. Table V lists the $F1$ values of the results generated by competing models. As can be seen from Table V, according to the $F1$ measure, SONMFSR shows overwhelming advantages over

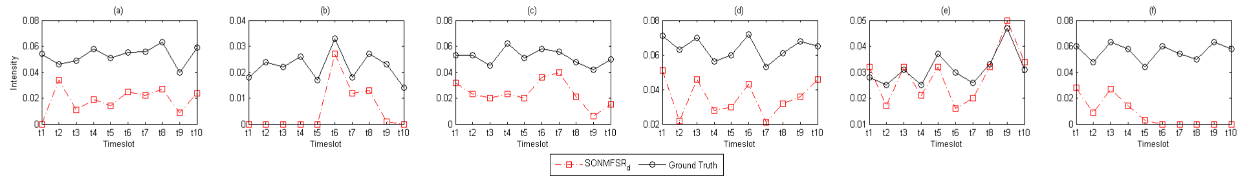


Fig. 4. Topic tracking performance on Sougou corpus (by *intensity*).

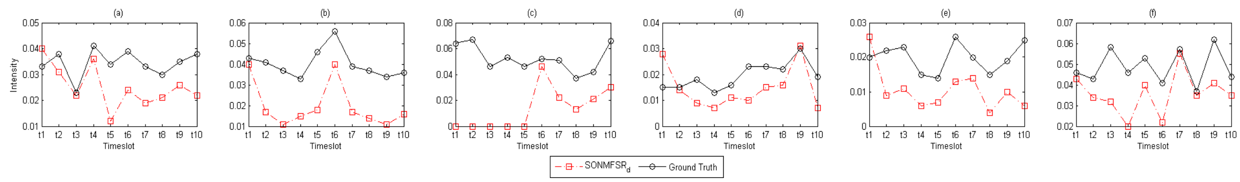


Fig. 5. Topic tracking performance on 20Newsgroup corpus (by *intensity*).

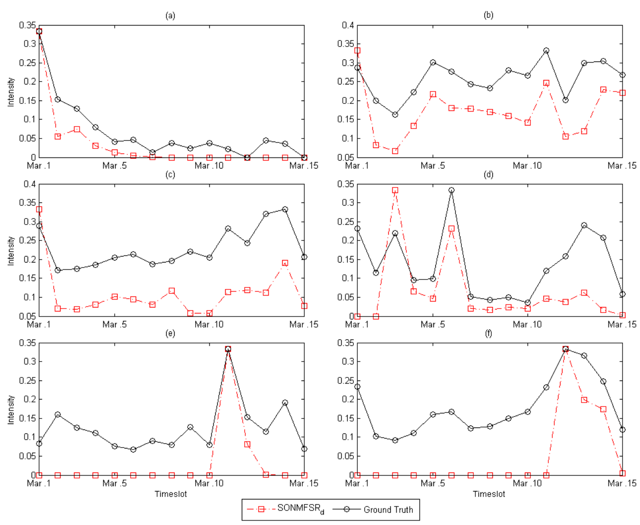


Fig. 6. Topic intensity tracking on Two Sessions 2015 by SONMFSR$_d$.

NMF, SNMF and ONMF, and performs better than FSTM in most cases. These well demonstrate that the combined effect of soft orthogonality and sparse representation is critically important for the success of SONMFSR. It is also interesting to see that compared with competing models, SONMFSR performs particularly well when setting a relatively large topic number, as indicated by the $F1$ values in bold.

C. Dynamic Topic Mining from Sougou and 20Newsgroup

We here compare our SONMFSR$_d$ model with some state-of-the-art dynamic topic models: T-model, Online NMF and dynamic NMF, in dynamic topic detecting and tracking.

Fig. 2 and Fig. 3 show the $F1$ values of the topics detected from the two data sets, respectively, where the bar plots indicate the average results over 10 timeslots, and the box plots indicate the variations. According to the figures, our SONMFSR$_d$ model shows significant superiority to all the baseline methods no matter under what parameter configurations. Moreover, the relatively stable results under different configurations of parameters imply that SONMFSR$_d$ is quite robust to the setting of parameters, which is particularly important for real-life practice.

To further illustrate the performance of SONMFSR$_d$ in topic tracking, we randomly select six topics and track their evolutions along ten timeslots. Fig. 4 and Fig. 5 show the shapes of topic evolutions for the two data sets, respectively. As can be seen from the figures, generally speaking the topic trending captured by SONMFSR$_d$ coincides with the trending of ground truth quite well. For instance, SONMFSR$_d$ shows excellent performances from Fig. 4(b) to Fig. 4(e), whose correlation coefficient statistics between the two trends are all beyond 0.6. The best tracking performances for the 20Newsgroup data set are highlighted by Fig. 5(b) and Fig. 5(f), although the overall performance is less competent than the Sougou corpus. These results well domonstrate the advantage of SONMFSR$_d$ for topic tracking.

D. Case Study on Two Sessions 2015

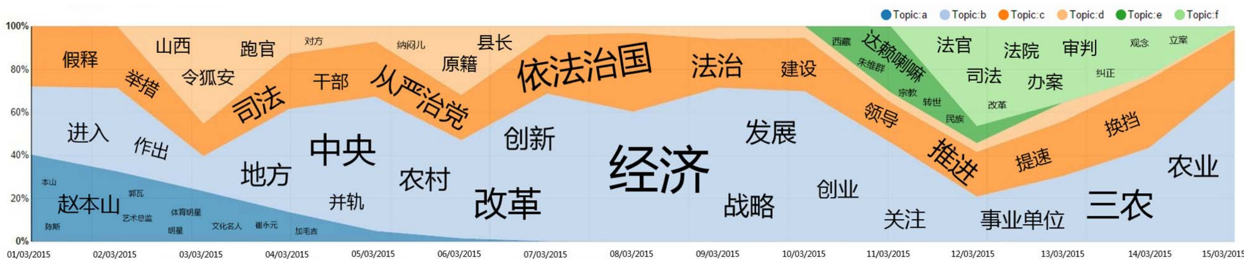
The *Two Sessions (of the NPC and the CPPCC)* is the annual key event in China, with plenty of hot topics concerned mostly by the public. We therefore apply our SONMFSR$_d$ model for dynamic topic detecting and tracking of Two Sessions 2015.

We used a Web-crawler to collect the related news released in the period from March 1 to March 15, 2015 on the Sina website [34]. Our data set comprises of 2034 news documents spreading over 15 days, and the number of news documents in each day ranges from 50 to 260. We used NLPIR tool [30] to split the Chinese news and obtained 10007 terms after removing the ones with very low term frequencies, say less than 5. We then apply SONMFSR$_d$ for dynamic topic modeling, with parameters set as follows: $\xi = 0.7$, $\lambda = 0.1$, $\eta = 10$, $\sigma = 0.02$, the initial number of topics $K^{(1)} = 12$, and the number of emerging topics in a new day $K^{(t)}_{emerge} = 2$, $\forall\ t > 1$.

Fig. 6 reports the intensity tracking of six hot topics in Two Sessions 2015, with the corresponding top-10 topic terms listed in Table VI. From Fig. 6, we can see that SONMFSR$_d$ well captures the evolutions of these topics, especially the peaks and valleys in the trending lines. We further relate the trends

TABLE VI. TOP-10 TERMS FOR TOPICS IN FIG. 6

#Topic	Topic	Top-10 Terms
a	Stars in sports & entertainment	ZHAO Benshan, GUO Wa, star, sports star, JIA Maoji, art director, CUI Yongyuan, cultural celebrities, CHEN Si, Benshan
b	Economics & developments	economy, reform, rural, merger, local, innovation, central, enter, development, strategy
c	Rule of law	The rule of law, law, justice, cadres, tightening party discipline, promoting,leadership, speed, construction, act
d	Anti-corruption	Shanxi, LINGHU An, official positions, the opposite side, puzzled, origin, relatives, trustee, county magistrate, report group
e	Darai Lama & religion issues	The Darai Lama, Tibet, ZHU Weiqun, reincarnation, religion, nationality, abandon, leaders, self-immolation, foreign
f	Justice & miscarriages	judge, justice, court, case, trial, ideas, filing, reform, correct, degree



in Fig. 6 with the topic semantics in Table VI to gain insights from topic evolution. According to Table VI and Fig. 6, we can roughly divide the six topics into three categories. The first category is about routine discussions on long-standing economics and politics developments, i.e., Topics b and c in Table VI, which exhibit relatively stable trends in Fig. 6. The second category is about long-term hot issues such as anti-corruption and unjust miscarriages, corresponding to Topics d and f in Table VI, which show various peaks in trending lines in Fig. 6, implying continuous attention from the public. The last category is about short-term hot concerns on for example entertainment and political stars, correspoinding to Topics a and e, which as shown in Fig. 6 attract attention easily by some special events but then trend downwards quickly.

In summary, the above results demonstrate the effectiveness of our SONMFSR$_d$ model for dynamic analysis of real-world events. In fact, we have taken a further step in building a demo system for dynamic event analysis based on SONMFSR$_d$. Fig. 7 exhibits some screenshots of our implemented system, which we plan to make public after some time.

VII. CONCLUSIONS

In this paper, we propose SONMFSR$_d$, a novel dynamic topic detection and tracking model with soft orthogonal constraint, to capture emerging, evolving and fading topics in a flexible and intuitive way. Extensive experiments on two public corpora demonstrate the superiority of SONMFSR$_d$ to some state-of-the-art DTMs in both topic detection and tracking. In particular, SONMFSR$_d$ shows great potential in real-world applications, where popular topics in Two Sessions 2015 are captured and traced dynamically for possible insights.

ACKNOWLEDGMENT

This research was partially supported by National High Technology Research and Development Program (863 Program of China) under Grant 2014AA021504 (Research and Development on the Key Technologies of Microbial Digital Information System Integration) and NSFC (M1450009). Junjie Wu was supported in part by National High Technology Research and Development Program of China (SS2014AA012303), NSFC (71322104, 71171007, 71531001, 71471009), National Center for International Joint Research on E-Business Information Processing (2013B01035), Foundation for the Author of National Excellent Doctoral Dissertation of PR China (201189), and the Fundamental Research Funds for the Central Universities. We also thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] T. Hofmann, *Probabilistic latent semantic indexing*. Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1999: 50-57.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent dirichlet allocation[J]*.the Journal of machine Learning research, 2003, 3: 993-1022.

[3] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas and R. A. Harshman, *Indexing by latent semantic analysis[J]*. JAsIs, 1990, 41(6): 391-407.

[4] Wang Q, Xu J, Li H, et al, *Regularized latent semantic indexing[C]*. Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011: 685-694.

[5] Wang Q, Cao Z, Xu J, et al, *Group matrix factorization for scalable topic modeling[C]*. Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012: 375-384.

[6] D. D. Lee, H. S. Seung, *Learning the parts of objects by non-negative matrix factorization[J]*. Nature, 1999, 401(6755): 788-791.

[7] D. M. Blei, J. D. Lafferty, *Dynamic topic models[C]*. Proceedings of the 23rd international conference on Machine learning. ACM, 2006: 113-120.

[8] Cao B, Shen D, Sun J T, et al, *Detect and Track Latent Factors with Online Nonnegative Matrix Factorization[C]*. IJCAI, 2007, 7: 2689-2694.

[9] Saha A, Sindhwani V, *Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization[C]*. Pro-

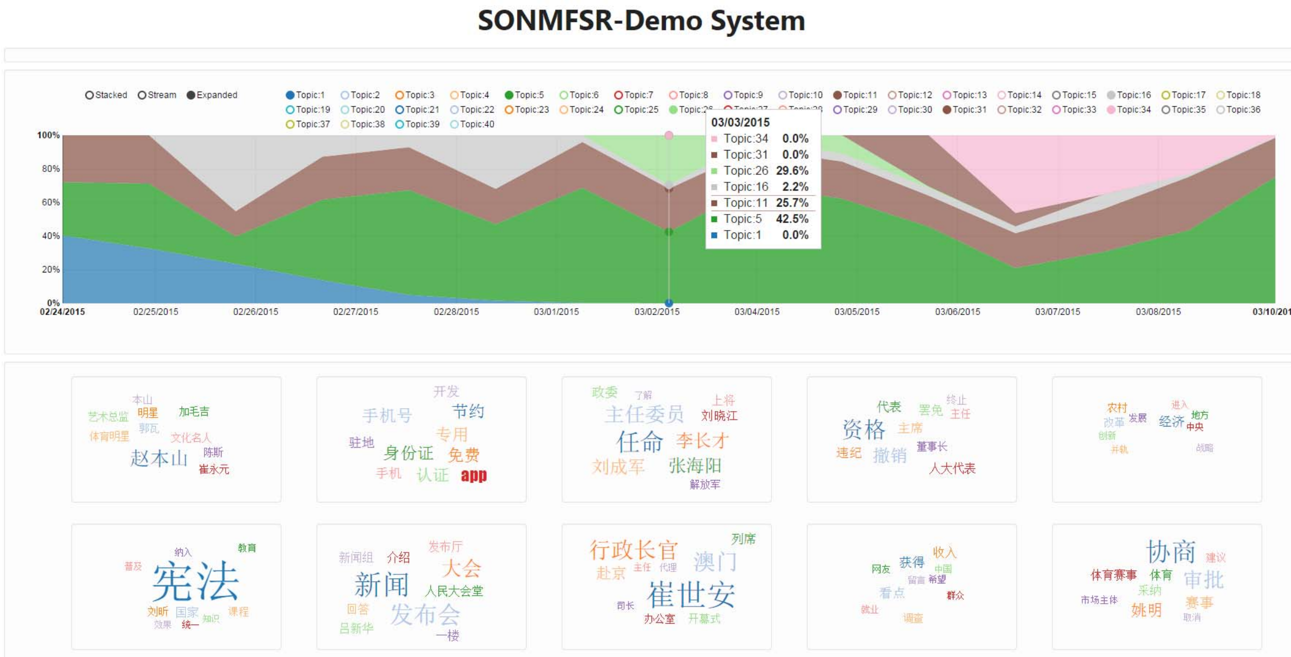


Fig. 7. Demo system for dynamic topic modeling: (1) The upper figure is for intensity tracking of dynamic topics for given events; (2) The lower figures are word clouds of the corresponding topics.

ceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012: 693-702.

[10] Derntl M, Gnnemann N, Tillmann A, et al, *Building and Exploring Dynamic Topic Models on the Web[C]*. Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, 2014: 2012-2014.

[11] Cichoki A, Zdunek R, Phan A H, et al, *Nonnegative Matrix and Tensor Factorization [Book]*. 2009.

[12] Vaca C K, Mantrach A, Jaimes A, et al, *A time-based collective factorization for topic discovery and monitoring in news[C]*. Proceedings of the 23rd international conference on World wide web. International World Wide Web Conferences Steering Committee, 2014: 527-538.

[13] Chong Wang, David M. Blei, David Heckerman, *Continuous Time Dynamic Topic Models*. UAI 2008: 579-586.

[14] Iwata T, Yamada T, Sakurai Y, et al, *Online multiscale dynamic topic models[C]*. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010: 663-672.

[15] Jiho Yoo, Seungjin Choi, *Nonnegative Matrix Factorization with Orthogonality Constraints*. JCSE 4(2): 97-109 (2010)

[16] Chua F C T, Oentaryo R J, Lim E P. *Modeling temporal adoptions using dynamic matrix factorization[C]* Data Mining (ICDM), 2013 IEEE 13th International Conference on. IEEE, 2013: 91-100.

[17] Li X, Cheung W K W, Liu J, et al, *A novel orthogonal NMF-based belief compression for POMDPs[C]*. Proceedings of the 24th international conference on Machine learning. ACM, 2007: 537-544.

[18] Lyu S, Wang X, *On algorithms for sparse multi-factor NMF[C]*. Advances in Neural Information Processing Systems. 2013: 602-610.

[19] Kim J, He Y, Park H, *Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework[J]*. Journal of Global Optimization, 2014, 58(2): 285-319.

[20] Cichocki A, Zdunek R, Phan A H, et al, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation[M]*. John Wiley & Sons, 2009.

[21] Gerbrands J J, *On the relationships between SVD, KLT and PCA[J]*. Pattern recognition, 1981, 14(1): 375-381.

[22] Navas M, Ordonez C, *Efficient computation of PCA with SVD in SQL[C]*. Proceedings of the 2nd Workshop on Data Mining using Matrices and Tensors. ACM, 2009: 5.

[23] Ding C, Li T, Peng W, et al, *Orthogonal nonnegative matrix t-factorizations for clustering[C]*. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006: 126-135.

[24] Seungyeon Kim, Joonseok Lee, Guy Lebanon, Haesun Park, *Local Context Sparse Coding*. AAAI 2015: 2260-2266.

[25] Abdul-Saboor Sheikh, Jacquelyn A. Shelton, Jorg Lcke, *A truncated EM approach for spike-and-slab sparse coding*. Journal of Machine Learning Research 15(1): 2653-2687 (2014).

[26] Gabriel Leivas Oliveira, Erickson R. Nascimento, Antonio Wilson Vieira, Mario Fernando Montenegro Campos, *Sparse Spatial Coding: A Novel Approach to Visual Recognition*. IEEE Transactions on Image Processing 23(6): 2719-2731 (2014).

[27] Jim Jing-Yan Wang, Xin Gao, *Semi-supervised sparse coding*. IJCNN 2014: 1630-1637.

[28] Newman D, Lau J H, Grieser K, et al, *Automatic evaluation of topic coherence[C]*. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010: 100-108.

[29] Than K, Ho T B, *Fully sparse topic models [M]*. Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2012: 490-505.

[30] *http://ictclas.nlpir.org/*, NLPIR2014.

[31] *http://download.labs.sogou.com/resources.html*, Sougou Corpus.

[32] *http://qwone.com/~jason/20Newsgroups/*, 20Newsgroup.

[33] *http://nlp.stanford.edu/software/corenlp.shtml*, Stanford CoreNLP.

[34] *http://news.sina.com.cn/c/z/2015qglh/*, 2015 NPC & CPPCC news.

[35] Chou T C, Chen M C., *Using incremental PLSI for threshold-resilient online event analysis[J]*. Knowledge and Data Engineering, IEEE Transactions on, 2008, 20(3): 289-299.