

A Joint Model for Topic-Sentiment Evolution over Time

Mohamed Dermouche, Julien Velcin, Leila Khouas, Sabine Loudcher

▶ **To cite this version:**

Mohamed Dermouche, Julien Velcin, Leila Khouas, Sabine Loudcher. A Joint Model for Topic-Sentiment Evolution over Time. 2014 IEEE International Conference on Data Mining (ICDM), Dec 2014, Shenzhen, China. 10.1109/ICDM.2014.82 . hal-01762995

HAL Id: hal-01762995
https://hal.science/hal-01762995

Submitted on 10 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Joint Model for Topic-Sentiment Evolution over Time

Mohamed Dermouche[1,2], Julien Velcin[1], Leila Khouas[2], and Sabine Loudcher[1]
[1]Université de Lyon (ERIC LYON 2)
Email: {julien.velcin, sabine.loudcher}@univ-lyon2.fr
[2]AMI Software R&D
Email: {mde, lkh}@amisw.com

***Abstract*—Most existing topic models focus either on extracting static topic-sentiment conjunctions or topic-wise evolution over time leaving out topic-sentiment dynamics and missing the opportunity to provide a more in-depth analysis of textual data. In this paper, we propose an LDA-based topic model for analyzing topic-sentiment evolution over time by modeling time jointly with topics and sentiments. We derive inference algorithm based on Gibbs Sampling process. Finally, we present results on reviews and news datasets showing interpretable trends and strong correlation with ground truth in particular for topic-sentiment evolution over time.**

***Keywords*—joint topic sentiment models; time series; trend analysis; topic models; sentiment analysis; opinion mining.**

I. Introduction

Topic modeling and sentiment analysis are two popular tasks that deal with textual data. The former deals with extracting topics (what is it about?) while the latter is about sentiment and opinion classification (what is the underlying opinion?). These two tasks are complementary to the extent that sentiments are usually issued about topics and topics are often the basis of subjective positions. This is why topics and sentiments should be jointly extracted and analyzed. In the recent years, joint topic-sentiment modeling has emerged as a separate text-mining task. Some useful work has been done in this line [1], [2], [3], [4], [5] but most of this work extract sentiments about topics in a static way ignoring the dynamic nature of textual data. Another bulk of work [6], [7], [8] focus on analyzing content evolution only at the topic level leaving out topic-sentiment correlations. Motivated by this observation, we propose a novel topic-model-based approach to extract topic-sentiment associations from text as well as their evolution over time.

The proposed model results in a 3-level output: topics, topic's sentiments, and topic-sentiment evolution over time. It first serves as a traditional topic-discovering model able to extract the hidden topical structures from a document collection. Second, it models the association between topics and sentiments (the overall sentiment towards each of the extracted topics). Finally, it provides an efficient tool for tracking and visualizing the strength of topic-sentiment association over time. All of these information are extracted simultaneously and jointly without any post-processing.

Our approach has three main features that are not jointly addressed by other models of the literature. First, time is jointly modeled with topics and sentiments, which allows to capture the evolution of topic's sentiment over time. Second, topic-specific sentiments are extracted for the whole data at once and not for each single document, providing an overall view of topic-sentiment correlations. Finally, no post-processing is needed to match similar topics under different sentiment polarities.

Based on a ground-truth evaluation framework, we compare our model to other state-of-the-art topic-sentiment models, JST and ASUM. We demonstrate the efficiency of our model in extracting accurate topic-sentiment-time associations on two different Web datasets including product reviews and news articles.

The rest of this paper is organized as follows: An overview of related work is given in Section II. Our approach and evaluation framework are presented in Sections III and IV. Experiments, results, and discussion are given in Sections V and VI. Finally, the paper is concluded in Section VII.

II. Related Work

A. Joint Topic-Sentiment Modeling

Topic and sentiment modeling tasks are quite related to the extent that sentiments are usually issued about a topic at hand. In order to model topic-sentiment conjunction, a lot of work use topic models: statistical models for discovering low-dimensional structures (topics) from text based on word co-occurrence patterns. These patterns are captured using the so-called latent or hidden variables. Earlier topic models, like LDA [9] and PLSA [10], have mainly focused on extracting homogeneous topics but more recently, these models have been extended to capture other aspects of text, like sentiment. As an example, Joint Sentiment Topic model (JST) [2], [11] has been developed for topic extraction under different sentiment labels. This is performed by extending LDA with a new sentiment layer inserted before the topic layer. Thus, to generate a word for a document, a sentiment label s is drawn first, then a topic is drawn conditioned on s. Reverse-JST [2] is a variant of JST where the order of sentiment and topic layers is inverted.

Slightly different models exist like Topic Sentiment Mixture (TSM) [3], Sentiment-LDA [1], Aspect-and-Sentiment Unification Model (ASUM) [4] and Sentiment-Topic model with Decomposed Prior (STDP) [5]. All these models are based on LDA except TSM that is based on PLSA.

B. Topic Evolution over Time

Documents are usually collected over time (online discussions, news, emails etc.) and consequently their content may evolve and change over time. Here, we focus on the quantitative evolution, i.e., the "amount" of data discussing some topic at some timestamp t. Topics over Time (TOT) [8] is an LDA-based model for quantitative topic evolution. While the meaning of a particular topic is assumed constant, its size is supposed evolving and is captured using a Beta distribution. In [7], LDA model has been used to capture quantitative topic evolution over time by counting the number of documents associated to each topic at each timestamp.

On the other hand, qualitative evolution focuses on all other aspects of a topic (word distribution, inter-topic correlation, vocabulary etc.). In this line, Dynamic Topic Model (DTM) [6] has been proposed to model the change of topic-word distributions over time. Documents are first grouped by timestamp, then a separate LDA model is fitted for each group by evolving the model from the previous timestamp.

C. Joint Topic-Sentiment Evolution over Time

The issue of modeling topic-sentiment evolution is relatively recent and it is less covered in the literature. Mei et al. [3] were among the first to deal with this problem. The previously proposed TSM model has been used to extract topic-sentiment associations. Then, the quantitative evolution has been characterized by the number of words assigned to a topic and a sentiment label in the same timestamp. In another work [12], TSM has been extended with a new layer to capture the temporal dimension.

The works in [3], [12] are the closest to ours. However, our model is different from [3] in that it does not require a post-processing to deduce time evolution. The model in [12] is based on PLSA, which is known to have many deficits like overfitting the learning data and high inference complexity due to the large number of learnt parameters [9]. Our model is based on the well-known LDA model and consequently shares the same strengths, in particular the explicit use of Dirichlet hyperparameters to smooth the multinomial distributions. These hyperparameters have been shown to be specifically useful for topic-sentiment models to guide the joint topic-sentiment discovery [2], [4], [5].

Our proposal is also different from a number of topic models that are based on model adaptation (qualitative evolution). In this line, He et al. [13], [14] introduced Dynamic-JST on top of the previously proposed JST model to capture the qualitative topic evolution over time. Their approach is similar to DTM [6] where the model at each timestamp is derived from the model at the previous one.

III. Time-aware Topic-Sentiment (TTS) Model

In this section, we describe Time-aware Topic-Sentiment (TTS) model, our approach to modeling topic-sentiment associations as well as their quantitative evolution over time. Our proposal is motivated by the following observations made on the previously proposed models:

- Time is not jointly modeled with topics and sentiments [1], [2], [4], [5], [13], [14].

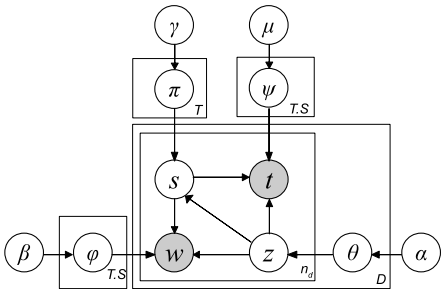


Fig. 1. Time-aware Topic-Sentiment (TTS) graphical model.

- Topic-specific sentiments are estimated for each document separately [1], [2], [4], [5].
- Similar topics from different sentiment polarities are not automatically matched [1], [2], [3], [4], [5].

To address these issues, we propose a novel topic model based on three main features: first, time is jointly modeled with topics and sentiments providing a quantitative analysis of topic-sentiment evolution over time. Second, topic-specific sentiments are extracted for the whole data and not for each single document, providing an overall view of topic-sentiment correlations. Finally, no post-processing is needed to match topics under different sentiment polarities because the same topic has multiple distributions over words, one for each sentiment polarity.

We extend LDA model by adding two new layers, s and t, in order to capture sentiment and time respectively (Figure 1). Our proposal is built on the traditional topic modeling assumptions: each document in the learning collection is a mixture of topics (multinomial distribution over topics). In addition, we assume that each topic has multiple facets, one for each sentiment polarity, and consequently multiple multinomial disributions over words. Moreover, we assume that the "strength" of topic-sentiment association would evolve over time.

Documents from the learning data must be annotated with time (e.g., creation date). Time is first discretized and each document receives a discrete timestamp label (e.g. day, month, year). In the learning step, time modality is captured using the variable t, and consequently topic-sentiment evolution is captured using a multinomial distribution over timestamps ψ. The notation used in the rest of this paper is given in Table I.

A. Generative Process

TTS is a fully-generative model of words, sentiments and timestamps. Its generative process is as follows:

1) Draw $T \times S$ multinomials $\varphi_{z,s} \sim \text{Dir}(\beta)$
2) Draw $T \times S$ multinomials $\psi_{z,s} \sim \text{Dir}(\mu)$
3) Draw T multinomials $\pi_z \sim \text{Dir}(\gamma)$
4) For each document d, draw a multinomial $\theta_d \sim \text{Dir}(\alpha)$, then for each word w_i in d:
   a) Draw a topic $z_i \sim \theta_d$
   b) Draw a sentiment label $s_i \sim \pi_{z_i}$
   c) Draw a word $w_i \sim \varphi_{z_i,s_i}$
   d) Draw a timestamp $t_i \sim \psi_{z_i,s_i}$

TABLE I. NOTATION.

D	Number of documents
V	Vocabulary size
T	Number of topics
S	Number of sentiment labels
H	Number of timestamps
Θ	$[\theta_d]$: $D \times T$ matrix of document-specific distributions over topics
Φ	$[\varphi_{z,s}]$: $T \times S \times V$ matrix of topic-sentiment-specific distributions over words
Π	$[\pi_z]$: $T \times S$ matrix of topic-specific distribution over sentiments
Ψ	$[\psi_{z,s}]$: $T \times S \times H$ matrix of time distributions specific to topic-sentiment pairs
n_d	Number of words in document d
$n_{d,j}$	Number of words in document d affected to topic j
n_j	Number of words affected to topic j
$n_{j,k}$	Number of words affected to topic j and sentiment k
$n_{i,j,k}$	Number of times a word i is affected to topic j and sentiment k
$n_{j,k,h}$	Number of times a word with timestamp h is affected to topic j and sentiment k
n^{-p}	Count variable excluding word at position p of the current document

By examining the graphical model and the generative process of TTS, one can notice that different timestamps may be generated for different words in the same document. Yet, all words in a document should have the same timestamp. In practice, this is not a real problem because TTS is still efficient in modeling topic-sentiment dynamics. However, as time modality is involved in topic discovery, this may impact the homogeneity of topics because time modality is assumed having the same "weight" as word modality while actually it is not (one time modality vs. n_d word modalities in a document d). To address this issue, we adopt the same strategy as in TOT model [8] and Group-Topic model [15] where a balancing hyperparameter is introduced in order to balance word and time contributions in topic discovery. A natural setting would be to use the inverse of the number of words n_d as a balancing hyperparameter. This hyperparameter is taken into account when calculating posterior distribution.

B. Inference

Gibbs sampling is a popular approach to parameter estimation (inference) in topic models [5], [2], [8]. We adopt this approach because it often yields relatively simple algorithms. Due to space limitation, we only give the final formulas. A detailed derivation of inference on LDA via Gibbs sampling is provided in [16]. Derivation on TTS is performed in the same way.

a) Joint distribution.: Using Bayes conditional independence rule, the joint probability of words, topics, sentiments and timestamps can be factored as follows:

$$\begin{aligned} & p(\mathbf{w}, \mathbf{t}, \mathbf{s}, \mathbf{z} | \alpha, \beta, \gamma, \mu) \\ = \; & p(\mathbf{w}|\mathbf{s}, \mathbf{z}, \beta) \cdot p(\mathbf{t}|\mathbf{s}, \mathbf{z}, \mu) \cdot p(\mathbf{s}|\mathbf{z}, \gamma) \cdot p(\mathbf{z}|\alpha). \end{aligned} \quad (1)$$

The first term is obtained by integrating over φ.

$$p(\mathbf{w}|\mathbf{s}, \mathbf{z}, \beta) = \left(\frac{\Gamma(V\beta)}{\Gamma(\beta)^V} \right)^{T \cdot S} \prod_j \prod_k \frac{\prod_i \Gamma(n_{i,j,k} + \beta)}{\Gamma(n_{j,k} + V\beta)}, \quad (2)$$

Where Γ denotes Gamma function. Subscripts i, j, k, h are used to loop over words, topics, sentiments and timestamps respectively. The second term of Equation 1 is obtained by integrating over ψ.

$$p(\mathbf{t}|\mathbf{s}, \mathbf{z}, \mu) = \left(\frac{\Gamma(H\mu)}{\Gamma(\mu)^H} \right)^{T \cdot S} \prod_j \prod_k \frac{\prod_h \Gamma(n_{j,k,h} + \mu)}{\Gamma(n_{j,k} + H\mu)}, \quad (3)$$

The remaining terms of Equation 1 are obtained in the same way by integrating over π, θ respectively.

$$p(\mathbf{s}|\mathbf{z}, \gamma) = \left(\frac{\Gamma(\sum_k \gamma_k)}{\prod_k \Gamma(\gamma_k)} \right)^T \prod_j \frac{\prod_k \Gamma(n_{j,k} + \gamma_k)}{\Gamma(n_j + \sum_k \gamma_k)}, \quad (4)$$

$$p(\mathbf{z}|\alpha) = \left(\frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \right)^D \prod_d \frac{\prod_j \Gamma(n_{d,j} + \alpha_j)}{\Gamma(n_d + \sum_j \alpha_j)}, \quad (5)$$

b) Posterior distribution.: Posterior distribution is estimated by sampling the variables z, s given all other variables. We use the superscript $-p$ to denote the quantity of data that excludes the word at position p of the current document d. Posterior probability can be derived from joint probability as follows:

$$\begin{aligned} & p(s_p = k, z_p = j | \mathbf{w}, \mathbf{t}, \mathbf{s}^{-p}, \mathbf{z}^{-p}, \alpha, \beta, \gamma, \mu) \\ \propto \; & \frac{n_{d,j}^{-p} + \alpha_j}{n_d^{-p} + \sum_j \alpha_j} \cdot \frac{n_{w_p,j,k}^{-p} + \beta}{n_{j,k}^{-p} + V\beta} \cdot \frac{n_{j,k}^{-p} + \gamma_k}{n_j^{-p} + \sum_k \gamma_k} \cdot \frac{n_{j,k,t_p}^{-p} + \mu}{n_{j,k}^{-p} + H\mu}. \end{aligned} \quad (6)$$

The balancing hyperparameter $\frac{1}{n_d}$ is introduced as an exponential power of the last term of Equation 6. Samples obtained from the Markov chain are then used to estimate the distributions φ, θ, π and ψ as follows:

$$\begin{aligned} \varphi_{j,k,i} &= \frac{n_{i,j,k} + \beta}{n_{j,k} + V\beta}, \quad \theta_{d,j} = \frac{n_{d,j} + \alpha_j}{n_d + \sum_j \alpha_j}, \\ \pi_{j,k} &= \frac{n_{j,k} + \gamma_k}{n_j + \sum_k \gamma_k}, \quad \psi_{j,k,h} = \frac{n_{j,k,h} + \mu}{n_{j,k} + H\mu}. \end{aligned} \quad (7)$$

c) Gibbs sampling algorithm.: A complete overview of Gibbs sampling procedure is given in Algorithm 1.

Algorithm 1 Inference on TTS

Require: $\alpha, \beta, \gamma, \mu, T$
1: Initialize matrices Φ, Θ, Π, Ψ.
2: **for** iteration c = 1 **to** *nbGibbsIterations* **do**
3: **for** document d = 1 **to** D **do**
4: **for** p = 1 **to** n_d **do**
5: Exclude word w_p from d and update count variables
6: Sample a topic and a sentiment label for word w_p using Equ. 6
7: Update count variables with new topic and sentiment label
8: **end for**
9: **end for**
10: **end for**
11: Update matrices Φ, Θ, Π, Ψ with posterior estimates using Equ. 7

C. Incorporating Prior Knowledge

We use prior knowledge, represented in the form of a sentiment lexicon (a list of words annotated with prior sentiment labels), to guide sentiment discovery. Prior knowledge is incorporated when sampling a sentiment for a word (line 6 of Algorithm 1). Thus, if the word is in the lexicon, it

is affected to its corresponding sentiment label taken from the lexicon. Otherwise, the sentiment label is generated using Equation 6. This strategy is also adopted in [1], [2], [5]. In all our experiments, we use a subset of MPQA subjectivity lexicon[1] (words tagged with "strong subjectivity"), to which we add a number of words, n-grams and slang words that we have collected and tagged manually. The final stemmed lexicon comprises 1502 positive and 2541 negative stems.

IV. Evaluation Framework

In the original papers, JST and ASUM have been evaluated using techniques from supervised learning (sentiment prediction at the document level) but this is not the initial purpose of these models, neither of TTS. Evaluating TTS model involves at least two aspects: topic-sentiment association and topic-sentiment evolution over time. We propose to evaluate these aspects by comparing model results to a ground truth. To this end, we rely on a dataset in which each document is annotated with topic, sentiment and time. Then, for each topic-sentiment pair, we calculate the "real" (observed) distribution over words $p(w|s,z)$ by merging all the documents from topic z that are annotated with sentiment s. We calculate for each topic the real distribution over sentiments $p(s|z)$ by counting the number of documents from topic z annotated with sentiment s. Finally, we calculate the real distribution of topic-sentiment pairs over time $p(t|s,z)$ in the same way.

Based on the ground truth, we define two separate evaluation measures: topic-sentiment association accuracy Q_s and topic-sentiment evolution accuracy Q_t. These measures are based on the calculation of a distance between "estimation" and "reality". The general approach is performed in two steps: topic matching and evaluation.

A. Topic Matching

For simplicity, we suppose a binary sentiment modality, positive and negative. Let r, e be real, respectively estimated topics. Each topic r is matched to a topic e based on the calculation of KL divergence between topic distributions over vocabulary φ_r and φ_e. As KL divergence is not a distance measure, we use KL distance (KLD) instead. It can be calculated for two multinomial distributions P and Q as follows [17][2]:

$$\text{KLD}(P,Q) = \text{KL}(P||Q) + \text{KL}(Q||P) \quad (8)$$

$$= \sum_i \left[(P(i) - Q(i)) \cdot \log \frac{P(i)}{Q(i)} \right] \quad (9)$$

The matching procedure is realized by taking up the pair of topics with the lowest KLD value iteratively. The general procedure is performed in two steps: First, real and estimated topics are matched separately under positive and negative polarity. Second, each topic e_p from estimated positive topics is matched to a topic e_n from estimated negative ones if e_p and e_n were matched with the same real topic in the previous step. This double matching is only necessary for JST and ASUM models. In TTS, it is automatically provided by the model.

[1]http://mpqa.cs.pitt.edu

[2]For ease of reproducibility, zero values are replaced by 10^{-12}.

B. Evaluation Measures

Let $\mathcal{M}$ be the result of the previous step ($\mathcal{M}$ contains the pairs of matched topics regardless of polarity). Each pair of topics $(r,e) \in \mathcal{M}$ is characterized by a distribution over sentiments. The calculation of this distribution is specific to each model. For TTS, it is directly produced by the model (distribution π). For JST and ASUM, $p(s|z)$ is obtained in a way similar to the calculation of the real distribution but with the new (estimated) annotations. Each document d is re-annotated with the sentiment and the topic maximizing probability θ_d.

The first evaluation measure Q_s (topic-sentiment association accuracy) is the average KL distance between real and estimated π distributions[3] of matched pairs of topics:

$$Q_s = \frac{1}{T} \cdot \sum_{(r,e)\in\mathcal{M}} \text{KLD}(\pi_r, \pi_e) \quad (10)$$

The second evaluation measure Q_t is based on the calculation of estimated topic-sentiment distributions over time (ψ). This information is directly produced by TTS. For JST and ASUM, we use the real timestamps associated to documents in order to estimate ψ distributions. Finally, topic-sentiment-time association accuracy, Q_t, is the average KL distance between real and estimated ψ distributions of matched pairs of topics:

$$Q_t = \frac{1}{T} \cdot \sum_{(r,e)\in\mathcal{M}} \text{KLD}(\psi_r, \psi_e) \quad (11)$$

V. Experiment

We use two datasets in English: MDS (a subset of Multi-Domain Sentiment dataset [18]) and NYSK (New York v. Strauss-Kahn) dataset. MDS consist of reviews for different types of products from Amazon[4]. Documents are annotated with 24 different topics (books, apparel, software, kitchen etc.), sentiment (positive or negative) and creation date ranging from year 1996 to 2007.

TABLE II. Statistics on the used datasets.

Dataset	Type	D	V	Annotation	Timestamps
MDS	reviews	29379	43834	topic, sent., time	years
NYSK	news	10421	51188	time	days

NYSK dataset is about the case relating to allegations of sexual assault against the former IMF (International Monetary Fund) director Dominique Strauss-Kahn (DSK) during May 2011[5]. We have created this dataset by crawling leading english-speaking online newspapers and news agencies using a meta-search engine with the following query: "dsk" OR "strauss-kahn" OR "strauss-khan". Documents of NYSK dataset are annotated with creation date ranging from 05/17/2011 to 05/26/2011. We have made this dataset publicly available on UCI Machine Learning Repository[6].

[3]Here, π refers to the distribution $p(s|z)$ which is different from π distribution in JST and ASUM models.

[4]http://www.amazon.com

[5]A chronology of the story is available at: http://www.nbcnewyork.com/news/local/DSK-Dominique-Strauss-Kahn-Case-Timeline-124854459.html

[6]http://archive.ics.uci.edu/ml/datasets/NYSK

TABLE III. TOP WORDS OF SELECTED TOPICS FROM MDS (LEFT) AND NYSK (RIGHT) DATASETS. THE LAST ROW REPRESENTS THE OVERALL SENTIMENT PROBABILITY SPECIFIC TO TOPIC ($\pi_z(s)$). POSITIVE AND NEGATIVE WORDS FROM THE LEXICON ARE REPRESENTED IN GREEN (UNDERLINED) AND RED (ITALIC) RESPECTIVELY.

z_1: **computer & video games**		z_2: **beauty**		z_3: **software**		z_4: **gourmet food**		z_5: **allegation**		z_6: **investigation**		z_7: **Christine Lagarde**		z_8: **French elections**	
positive	*negative*	*positive*	*negative*	*positive*	*negative*	*positive*	*negative*	*positive*	*negative*	*positive*	*negative*	*positive*	*negative*	*positive*	*negative*
game	way	hair	smell	use	comput	tast	tea	*alleg*	*accus*	evid	strauss	lagard	strauss	french	strauss
play	player	product	scent	software	xp	good	coffe	strauss	*guilti*	told	kahn	minist	next	polit	sarkozy
one	*bore*	great	shaver	work	upgrad	flavor	drink	kahn	*crimin*	time	hotel	imf	lagard	socialist	presid
fun	*puzzl*	shave	*thick*	support	internet	love	milk	charge	*deni*	court	dna	candid	nation	parti	franc
good	run	dri	eye	file	crash	try	brand	victim	media	*assault*	new	french	crisi	leader	*assault*
graphic	*long*	feel	*iron*	system	manual	best	fat	arrest	french	investig	bail	economi	european	dsk	parti
level	*bad*	recommend	*disappoint*	featur	*slow*	chocol	textur	polic	kahn	consensu	investig	support	possibl	newspap	imag
better	*hard*	clean	*irrit*	easi	connect	sweet	treat	sexual	hotel	believ	maid	lead	member	candid	aubri
0.59	0.41	0.36	0.64	0.49	0.51	0.69	0.31	0.87	0.13	0.16	0.83	0.47	0.53	0.53	0.47

Both datasets have been stemmed and lowercased. We have also removed stopwords, numbers and punctuation marks. Statistics on the preprocessed data are given in Table II.

We have implemented TTS model based on the code from GibbsLDA++[7]. For our experiments, we consider two sentiment polarities: positive and negative. For both datasets, we set the number of topics T to 24. The symmetric hyperparameters α, β and μ are set to $\frac{50}{T}$, 0.04 and 0.01 respectively. Experiments showed that TTS is not sensitive to parameter μ and time sparsity is not a big problem even with very low values of μ. For the evaluation purpose, the hyperparameter $\gamma_{pos.}$ is set to 1 while $\gamma_{neg.}$ is variable. Experiments showed that topic-sentiment models TTS, JST and ASUM are not sensitive to the values of these parameters but to their ratio $\frac{\gamma_{neg.}}{\gamma_{pos.}}$, here denoted γ_{ratio}. All results are obtained at the 500th iteration of Gibbs sampler.

VI. RESULTS AND ANALYSIS

A. Topic-sentiment extraction

The first goal of TTS is extracting topic-sentiment associations. Figure 2 represents a quantitative evaluation of TTS, JST and ASUM models using topic-sentiment accuracy measure Q_s (cf. Section IV). This result is obtained on MDS dataset by varying γ_{ratio}. In terms of Q_s, ASUM gives the best result ($Q_s = 0.2$ for $\gamma_{ratio} = 20$), followed by TTS ($Q_s = 1.39$) then JST ($Q_s = 2.26$). This experiment reveals that topic-sentiment models are more efficient when dealing with sentences as coherent units (as in ASUM) rather than words (as in TTS and JST). It also reveals that TTS and ASUM are less sensitive to the initialization step (random assignment of words to topics and sentiments) than JST.

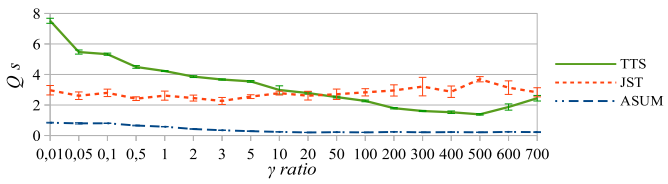


Fig. 2. Topic-sentiment association accuracy Q_s on MDS dataset (the lower, the better). Error bars give standard deviation based on 5 random initializations.

[7]http://www.gibbslda.sourceorge.net

However, TTS is still efficient for extracting topics and topic's sentiments. Table III shows selected examples of the extracted topics from each dataset, MDS and NYSK, using TTS model. γ_{ratio} is set to 250 for both datasets. A topic is represented by an ordered list of most likely words under the two sentiment polarities. As it can be seen from the table, the extracted topics are clearly opinionated. Under each sentiment label, the most likely words are quite coherent and sentiment-bearing. For example, topic z_1 (computer and video games) is described positively under the positive polarity ("good", "better", "enjoy" etc.). The same topic under negative polarity is described rather with negative words ("bored", "bad" etc.).

B. Topic-sentiment evolution over time

The second goal of TTS is modeling topic-sentiment evolution over time. Figure 3 shows the variation of Q_t measure w.r.t. γ_{ratio}. Q_t measures how the model is able to capture accurate topic-sentiment associations over time. Figure 3 shows that TTS model significantly outperforms JST and ASUM in terms of Q_t measure. The best result reached by TTS is $Q_t = 1.31$ while it is 3.61 for JST and 3.22 for ASUM. This experiment shows that incorporating time information into the modeling process, as in TTS, is useful and allows extracting more accurate topic-sentiment-time associations.

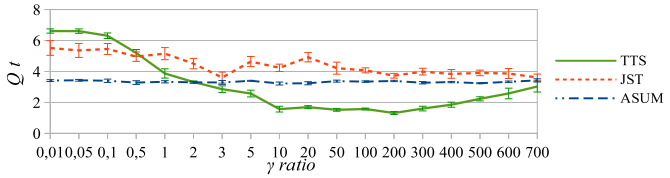


Fig. 3. Topic-sentiment-time association accuracy Q_t on MDS dataset (the lower, the better). Error bars give standard deviation based on 5 random initializations.

In addition to this quantitative evaluation, we present a case study of TTS model on NYSK dataset. For ease of interpretation, we choose to measure quantitative topic-sentiment evolution of a topic z and a sentiment label s by the number of documents belonging to topic z and sentiment s simultaneously. The number of documents at each timestamp t, denoted $\text{nbDocs}_{z,s}(t)$ is calculated as follows:

$$\text{nbDocs}_{z,s}(t) = \psi_{z,s,t} \cdot \pi_{z,s} \cdot \text{topicSize}(z) \qquad (12)$$

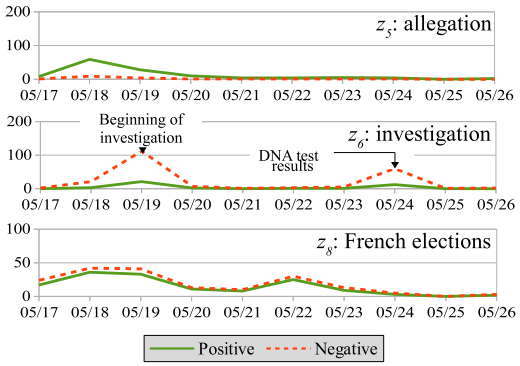


Fig. 4. Estimated topic-sentiment evolution (number of documents) over time on NYSK dataset using TTS model.

Where topicSize(z) is the number of documents assigned to topic z using maximum probability.

Figure 4 shows the estimated evolution for a set of topics from NYSK dataset. A deep examination of topic-sentiment evolution in Figure 4 leads to the following piece of information:

- Topic z_5 is referring to the very first allegations against DSK, it is clearly situated in the beginning of the timeline. At this period, one can notice the use of words like "victim", "accusation", "guilty", "deny" etc (cf. Table III). The overall sentiment is strongly positive reflecting a relatively strong support to DSK in the first days. This also reveals that media was taking this topic with a pinch of salt, but not for long.
- Topic z_6 relates the investigations made to ensure the allegation's credibility. This topic is located mainly on May 19, the date of first investigations and May 24, the date on which DNA test results have been reported by media. The overall sentiment is strongly negative, which is consistent with DNA test results.
- Finally, the 2012 French presidency where DSK was considered to be a leading candidate is a topic that has been invoked throughout the timeline (z_8). The overall mixed sentiment reveals that the consequences of DSK case on the elections was not that catastrophic. This result is consistent with an opinion poll realized by the French institute CSA[8] on May 16 where it has been shown that "victory of socialists was always possible even without DSK".

VII. Conclusion

In this paper, we have discussed TTS: a novel topic-model-based approach to jointly model topic and sentiment dynamics. Through an evaluation framework based on ground truth, we have demonstrated that TTS outperforms two other state-of-the-art models in extracting accurate topic-sentiment-time associations. We have illustrated the efficiency of TTS model for modeling homogeneous topics, associated sentiments and their evolution over time using both academic and real-world datasets collected on the Web. Also, analyzing topic-sentiment evolution on real-world data helped us discover and gain insight into hidden phenomena impossible with previous models, which generates a wide range of promising applications.

As a future direction, it would be interesting to work on hyperparameter setting for TTS model. The results shown in Figures 2 and 3 reveals that TTS, compared to other topic-sentiment models, is rather sensitive to the topic-sentiment prior γ, usually fixed empirically after several experiments. A work is already ongoing to examine the effectiveness of the methods usually used in the literature, in particular those based on Maximum-Likelihood estimation [19]. The first results showed that these methods perform well for estimating α and β hyperparameters but poorly for estimating γ. We believe that the automation of the topic-sentiment analysis process with TTS should be based on user's interaction to guide the analysis process.

[8]http://www.csa.eu/multimedia/data/sondages/data2011/opi20110516-les-premieres-consequences-politiques-de-l-affaire-dsk.pdf

References

[1] F. Li, M. Huang, and X. Zhu, "Sentiment analysis with global topics and local dependency," *AAAI'10*, pp. 1371–1376, 2010.

[2] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly Supervised Joint Sentiment-Topic Detection from Text," *TKDE*, vol. 24, no. 6, pp. 1134–1145, Jun. 2012.

[3] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: modeling facets and opinions in weblogs," in *WWW'07*. Banff, Canada: ACM, 2007, pp. 171–180.

[4] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," in *WSDM'11*. Hong Kong, China: ACM, 2011, pp. 815–824.

[5] C. Li, J. Zhang, J.-t. Sun, and Z. Chen, "Sentiment Topic Model with Decomposed Prior," in *SDM'13*. Austin, TX, USA: SIAM, 2013, pp. 767–776.

[6] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *ICML'06*. Pittsburgh, PA, USA: ACM, 2006, pp. 113–120.

[7] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 5228–5235, Apr. 2004.

[8] X. Wang and A. McCallum, "Topics over time: a non-Markov continuous-time model of topical trends," in *KDD'06*. Philadelphia, PA, USA: ACM, 2006, pp. 424–433.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *JMLR*, vol. 3, pp. 993–1022, 2003.

[10] T. Hofmann, "Probabilistic latent semantic indexing," in *SIGIR'99*. Berkeley, CA, USA: ACM, 1999, pp. 50–57.

[11] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *CIKM'09*. Hong Kong, China: ACM, 2009, pp. 375–384.

[12] M. Zheng, C. Wu, Y. Liu, X. Liao, and G. Chen, "Topic Sentiment Trend Model: Modeling facets and sentiment dynamics," in *CSAE'12*, no. 1. Zhangjiajie, China: IEEE, 2012, pp. 651–657.

[13] Y. He, C. Lin, W. Gao, and K. F. Wong, "Dynamic Joint Sentiment-Topic model," *TIST*, vol. 9, no. 4, 2014.

[14] Y. He, C. Lin, W. Gao, and K.-F. Wong, "Tracking Sentiment and Topic Dynamics from Social Media," in *ICWSM'12*. Dublin, Ireland: AAAI, 2012, pp. 483–486.

[15] X. Wang, N. Mohanty, and A. McCallum, "Group and topic discovery from relations and text," in *LinkKDD'05*. Chicago, IL, USA: ACM, 2005, pp. 28–35.

[16] G. Heinrich, "Parameter estimation for text analysis," Tech. Rep., 2005.

[17] B. Bigi, "Using Kullback-Leibler Distance for Text Categorization," in *ECIR'03*. Pisa, Italy: Springer-Verlag, 2003, pp. 305–319.

[18] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boomboxes and Blenders: Domain Adaptation for Sentiment Classification," in *ACL'07*. Prague, Czech Republic: ACL, 2007, pp. 440–447.

[19] T. P. Minka, "Estimating a Dirichlet distribution," MIT, Tech. Rep. 8, 2003.