Time-aware multi-viewpoint summarization of multilingual social text streams

Ren, Z.; Inel, O.; Aroyo, L.; de Rijke, M.

DOI
10.1145/2983323.2983710

Publication date
2016

Document Version
Final published version

Published in
CIKM'16

Time-aware Multi-Viewpoint Summarization of Multilingual Social Text Streams

Zhaochun Ren
University of Amsterdam
Amsterdam, The Netherlands
z.ren@uva.nl

Oana Inel
VU University Amsterdam
Amsterdam, The Netherlands
oana.inel@vu.nl

Lora Aroyo
VU University Amsterdam
Amsterdam, The Netherlands
lora.aroyo@vu.nl

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

ABSTRACT

A viewpoint is a triple consisting of an entity, a topic related to this entity and sentiment towards this topic. In time-aware multi-viewpoint summarization one monitors viewpoints for a running topic and selects a small set of informative documents. In this paper, we focus on time-aware multi-viewpoint summarization of multilingual social text streams. Viewpoint drift, ambiguous entities and multilingual text make this a challenging task. Our approach includes three core ingredients: dynamic viewpoint modeling, cross-language viewpoint alignment, and, finally, multi-viewpoint summarization. Specifically, we propose a dynamic latent factor model to explicitly characterize a set of viewpoints through which entities, topics and sentiment labels during a time interval are derived jointly; we connect viewpoints in different languages by using an entity-based semantic similarity measure; and we employ an update viewpoint summarization strategy to generate a time-aware summary to reflect viewpoints. Experiments conducted on a real-world dataset demonstrate the effectiveness of our proposed method for time-aware multi-viewpoint summarization of multilingual social text streams.

Keywords

Multi-viewpoint summarization; Dynamic viewpoint modeling; Topic modeling; Multilingual social text streams

1. INTRODUCTION

Focused on an entity [35], a *viewpoint* refers to a topic with a specific sentiment label. As an example, consider the entity "Japan" within the topic "#Whale hunting," with a negative sentiment. With the development of social media, we have witnessed a growth in the number of social media posts that express dynamically changing viewpoints in different languages around the same topic [38]. Unlike viewpoints in stationary documents, time-aware viewpoints of social text streams are dynamic, volatile and cross-linguistic [15].

CIKM'16 , October 24 - 28, 2016, Indianapolis, IN, USA

ACM ISBN 978-1-4503-4073-1/16/10. . . $15.00
DOI: http://dx.doi.org/10.1145/2983323.2983710

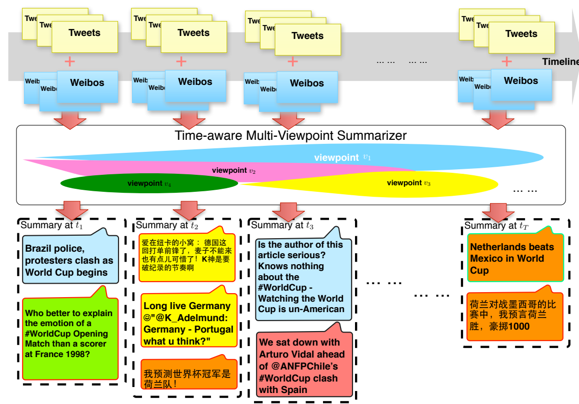


Figure 1: An example of time-aware multi-viewpoint summarization of multilingual social text streams about #FIFA World-cup 2014. The timeline at the top is divided into multiple time periods. The social text stream is composed of English language tweets and Chinese language weibos, which are shown at the top as yellow and blue rectangles, respectively. The time-aware multi-viewpoint summarizer detects temporal viewpoints by analyzing social text and generating an update summary at each period to reflect salient viewpoints. The summarization results are shown as colored round rectangles.

The task we address is *time-aware multi-viewpoint summarization of multilingual social text streams*: we extract a set of informative social text documents to highlight the generation, propagation and drift process of viewpoints in a given social text stream. Fig. 1 shows an example of our task's output for the topic "#FIFA World-Cup 2014."

The growth in the volume of social text streams motivates the development of methods that facilitate the understanding of those viewpoints. Their multi-lingual character is currently motivating an increasing volume of information retrieval research on multi-lingual social text streams, in areas as diverse as reputation polarity estimation [38] and entity-driven content exploration [43]. Recent work confirms that viewpoint summarization is an effective way of assisting users to understand viewpoints in stationary documents [17, 19, 26, 29, 30, 34, 46]—but viewpoint summarization for multilingual social text streams has not been addressed yet.

The most closely related work to time-aware viewpoint summarization is the viewpoint summarization of stationary documents [37], in which a sentence ranking algorithm is used to summarize contrastive viewpoints based on a topic-aspect model [36]. Compared with viewpoint summarization in stationary documents, the task of time-aware multi-viewpoint summarization of social text streams faces four challenges: (1) the ambiguity of entities in social text streams; (2) viewpoint drift, so that a viewpoint's statistical properties change over time; (3) multi-linguality, and (4) the shortness of social text streams. Therefore, existing approaches to viewpoint summarization cannot be directly applied to time-aware viewpoint summarization of social text streams.

We propose a method to tackle the above challenges: (1) We employ a state-of-the-art entity linking method to identify candidate entities from social text; (2) We represent a viewpoint as a tuple of an entity, a topic and a sentiment label, and propose a dynamic latent factor model, called the viewpoint tweet topic model (VTTM), to discover life cycles of a viewpoint. Unlike most existing topic models, VTTM jointly tracks dynamic viewpoints and any viewpoint drift arising with the passing of time. VTTM employs Markov chains to capture the sentiment dependency between two adjacent words. At each time period, VTTM detects viewpoints by jointly generating entities, topics and sentiment labels in social text streams. Gibbs sampling is applied to approximate the posterior probability distribution. (3) Focusing on multi-linguality, we employ an entity-based viewpoint alignment method to match viewpoints in multiple languages by calculating semantic similarities between viewpoints. (4) Lastly, we present a random walk strategy to extract update summaries to reflect viewpoints.

To evaluate our proposed strategy to summarizing dynamic viewpoints in multilingual social text streams, we collect multilingual microblog posts for 6 well-known topics from 2014. Based on both online and offline human annotations, the evaluation of our proposed method for time-aware viewpoint summarization is shown to be effective.

Our main contributions are: (1) We propose the task of time-aware multi-viewpoint summarization of multilingual social text streams; (2) We propose a viewpoint tweet topic model (VTTM) to track dynamic viewpoints from text streams; (3) We align multilingual viewpoints by calculating semantic similarities via an entity-based viewpoint alignment method; (4) We present a Markov random walk strategy to summarize viewpoints from multilingual social text streams, which is shown to be effective in experiments using a real-world dataset.

2. RELATED WORK

We divide related work into three parts: opinion summarization; update summarization, and topic modeling for social text streams.

2.1 Opinion summarization

In recent years, significant progress has been made on the opinion summarization task [22]. Opinion summarization generates structured [22, 28, 31, 34] or semi-structured [18, 23] summaries for a set of opinionated documents. Given opinionated documents, a structured opinion summary shows positive/negative opinion polarities. Semi-structured opinion summarization extracts sentences to describe opinion polarities. Hu and Liu [23] apply a sentence ranking approach based on the dominant sentiment according to polarity. Kim et al. [27] propose a method to extract explanatory sentences as opinion summary. Ganesan et al. [18] propose an unsupervised method to generate a concise summary to reflect opinions. Recently, several methods have been proposed for integrating sentence ranking and extraction. For instance, Paul et al. [37] propose a topic model to distinguish topics into contrastive categories. Because of the ambiguity of entities, viewpoint drift and multi-linguality, their approach is a poor fit for our task.

Compared with previous work, our task is different: (1) our work is to extract and summarize dynamic viewpoints from multilingual social text streams; (2) our strategy is based on a latent factor model that focuses on dynamic viewpoints on the timeline; (3) our task pays attention to methods on cross-language processing and update summarization.

2.2 Update summarization

Traditional document summarization is retrospective in nature. Update summarization extracts and synthesizes novel information in a collection of documents [11, 33]. Given a *base* collection that users have already read and another *update* collection of recent documents, the goal of update summarization is to generate an update summary by analyzing the novelty, contrast and prevalence. An intuitive solution to update summarization is to remove redundancy from the output generated by a multi-document summarizer [10, 16]. Wan [44] propose a co-ranking algorithm to optimize a trade-off strategy between novelty and relevance metrics. McCreadie et al. [33] propose a pair-wise learning to rank algorithm to produce an update summary. They also train a regression model to predict the novelty of the given documents in each time period.

As far as we know, no previous work has addressed time-aware multi-viewpoint summarization.

2.3 Topic modeling for social text streams

Early research on topic modeling addressed the topic detection and tracking (TDT) task, where one needs to find and follow topics and events in a stream of broadcast news stories [3, 4]. With the development of social media, topic modeling for social text streams has received increased attention [2, 9, 32, 41]. Yang et al. [48] propose a large-scale topic modeling system that infers topics of tweets over an ontology of hundreds of topics in real-time. Focusing on sparsity and drift, Albakour et al. [2] propose a query expansion method to tackle real-time filtering in microblogs. To help users understand events and topics in social text streams, tweets summarization has also received attention [9, 41, 42].

Topic models have been successfully applied to topic modeling of social text streams [12, 39, 41]. In topic models [7, 21], each document is represented as a finite mixture of topics, whereas each topic is represented as a mixture of words. Beyond the insufficient "bag of words" representation, dynamic topic models are proposed to analyze topic evolution in streaming documents, such as the Dynamic Topic Model [6], Dynamic Mixture Models [47] and the Topic Tracking Model [25]. To describe the whole life cycle of a topic, Ahmed and Xing [1] propose an infinite dynamic topic model on temporal documents. Instead of assuming that a vocabulary is known *a priori*, Zhai and Boyd-Graber [49] propose an extension of the Dirichlet process to add and delete terms over time. Topic models have also been applied to explore personalized topics and time-aware events in social text streams [12, 41].

To the best of our knowledge, there is little previous work on summarizing multiple dynamic viewpoints. By jointly modeling temporal topics, sentiment labels and entities in multilingual social text streams, we propose a cross-language strategy to tackle the viewpoint summarization task for multilingual social text streams.

3. PROBLEM FORMULATION

In this section, we introduce key concepts about time-aware multi-viewpoint summarization. First of all, Table 1 lists the notation we use in this paper.

Given a social text stream $\mathcal{D}$ including T time periods, we define $\mathcal{D}_t \subset \mathcal{D}$ to be the set of documents published during the t-th period. We suppose there are two different languages used in $\mathcal{D}$; we

Table 1: Glossary.

Symbol	Description
$\mathcal{D}$	all documents
$\mathcal{W}$	vocabulary of documents $\mathcal{D}$
$\mathcal{E}$	entities set in $\mathcal{D}$
$\mathcal{L}$	sentiments in $\mathcal{D}$
$\mathcal{Z}$	topics in $\mathcal{D}$
$\mathcal{V}$	viewpoints in $\mathcal{D}$
K	the number of topics, i.e., $\lvert\mathcal{Z}\rvert$
E	number of entities
$\mathcal{D}_t$	documents posted at t
$\mathcal{D}_t^{(A)}$	documents posted in language A at t
$\mathcal{N}_d$	words in document d
D_t	number of documents posted at t, i.e., $\lvert\mathcal{D}_t\rvert$
N_d	number of words in document d, i.e., $\lvert\mathcal{N}_d\rvert$
d_t	a document in $\mathcal{D}_t$ posted at t
v_d	a viewpoint in document d, $v \in \mathcal{V}$
e_d	a entity present in document d, $e \in \mathcal{E}$
w_i	the i-th word present in document, $w \in \mathcal{W}$
z_i	a topic present in word w_i, $z \in \mathcal{Z}$
l_i	a sentiment label present in word w_i
π_t	distribution of viewpoint at t
θ_t	distribution of entity over viewpoint at t
μ_t	distribution of topics over viewpoint at t
$\phi_{v,z,l,t}$	distribution of words over v, z and l at t
$\mathcal{S}_t$	time-aware multi-viewpoint summary at t

divide $\mathcal{D}_t = \{d_1, d_2, \dots, d_{D_t}\}$ into $\mathcal{D}_t^{(A)} \cup \mathcal{D}_t^{(B)}$, where $\mathcal{D}_t^{(A)}$ and $\mathcal{D}_t^{(B)}$ indicates the set of documents written in language A and B respectively.

We begin by defining the notions of topic, entity and sentiment in our work. Following the definition of topic models [7, 8], we define a *topic*, denoted as z, as a probabilistic distribution over words. Assuming K topics exist in the social text streams on which we focus, we set $z \in \{1, 2, \dots, K\}$. We define an *entity*, denoted as e, as a rigid designator of a concept around a *topic*, e.g., "China" with "disputed islands between China and Japan". Using a state-of-the-art entity linking method [35], for each document we find an associated entity $e_d \in \mathcal{E}$. *Sentiment* is defined as a probability distribution over sentiment labels positive, negative, and neutral. A *sentiment label* $l_i \in \mathcal{L}$ is attached with each word w_i. Following [28], we assume that the sentiment label l_i for a word w_i depends on the sentiment label for its previous word w_{i-1} and the topic z_i simultaneously. Specifically, we set $l_i = -1$ when word w_i is "negative", whereas $l_i = 1$ when w_i is "positive."

Given a topic z, sentiment label l and entity e, we define a *viewpoint* to be a finite mixture over the *sentiment*, *entity* and *topic*, i.e., a tuple $v = \langle z, l, e\rangle$. Unlike previous work that considers viewpoints to be stationary [18, 37, 46], we assume that each viewpoint is also changing over time, which effects topics, sentiments and entities at each time interval. Thus for each viewpoint at time t, we represent it as a tuple $v = \langle z, l, e, t\rangle$. Given documents $\mathcal{D}_t$, because documents in social text streams are short, we assume that in each document $d \in \mathcal{D}_t$ only one viewpoint v_d exists. We further assume that there exist a probability distribution of viewpoints at each time period.

At time t, we set π_t to be a probability distribution of viewpoints at t, μ_t a probability distribution of topics over viewpoints at t, and θ_t a probability distribution of entities over viewpoints t. In social text streams, the statistical properties of viewpoints change over time. Thus we assume that the probability distribution of viewpoints π_t at time t is derived from a Dirichlet distribution over π_{t-1}. Assuming that the distribution of topics and sentiments also drifts over time, we set ϕ_t to be a probability distribution of words in topics and sentiment labels at time t, which is derived from a Dirichlet distribution over ϕ_{t-1} at the previous time $t-1$.

Finally, we define the task of time-aware multi-viewpoint summarization of multilingual social streams. Let multilingual social text streams $\mathcal{D}$ posted in T time periods be given. Then,

- at time period $t = 1$, the target of time-aware multi-viewpoint summarization of multilingual social text streams is to select a set of relevant documents as $\mathcal{S}_1$ as a summary of viewpoints $\mathcal{V}_1$;
- at a time period $t, 1 < t \leq T$, the target is to select a set of both relevant and novel documents, to summarize both the content of viewpoints $\mathcal{V}_t$ at time period t and the difference between $\mathcal{V}_t$ and viewpoints $\mathcal{V}_{t-1}$.

4. METHOD

4.1 Overview

Before providing the details of our proposed method for time-aware viewpoint summarization, we first give an overview in Fig. 2. We divide our method in 3 phases: (A) dynamic viewpoint modeling; (B) cross-language viewpoint alignment; and (C) multi-viewpoint summarization. Given a multilingual social text stream $\mathcal{D}_t = \{d_1, d_2, \dots, d_{D_t}\}$ published at time t, in phase A we propose a dynamic viewpoint model to draw viewpoints for each document. Using a set of viewpoints $\mathcal{V}_t$ extracted from phase A, in phase B we use cross-language viewpoint alignment to link similar viewpoints in different languages by computing the similarity between two entities. Phase C then summarizes documents according to viewpoint distributions using a co-ranking based strategy. In the end we get a time-aware multi-viewpoint summary $\mathcal{S}_t$ at time t.

4.2 (A) Dynamic viewpoint modeling

At time period t, given documents $\mathcal{D}_t$ in two different languages, our task during phase A is to detect dynamic viewpoints from the documents in $\mathcal{D}_t$. Using an extension of dynamic topic models [6], we propose a dynamic latent factor model, the *viewpoint tweets topic model* (VTTM), that jointly models viewpoints, topics, entities and sentiment labels in $\mathcal{D}_t$ at each time interval t.

Using a state-of-the-art entity linking method for social media [35], for each document d at t, we discover entities by calculating the *COMMONNESS* value of the document. We assume that there are, in total, V viewpoints and K topics in social text steams. For each document d, there are an entity e_d and N_d words; for each word $w_i \in d$, there is a topic z_i and a sentiment label l_i. We assume that the viewpoint v_d in d is derived via a multinomial distribution over a random variable π_t that indicates a probability distribution over viewpoints at t; each topic z, each sentiment label l and each entity e in document d is derived from the viewpoint v_d. The probability distribution π_t is derived from a Dirichlet mixture over the viewpoint distribution π_{t-1} at the previous period.

In VTTM we consider the sentiment dependency between two adjacent words. That is, a Markov chain is formed to represent the dependency relation between the sentiment labels of two adjacent words. Given a word w_i, the sentiment label l_i is selected depending on its previous one. The transition probability distribution is derived from the sentiment label of l_{i-1} and a transition variable x_i. The transition variable $x \in \mathcal{X}$ determines where the corresponding sentiment label comes from. If $x = 1$, then the sentiment label l_i of w_i is identical to the sentiment label l_{i-1} of word w_{i-1}; whereas if $x_i = -1$, the sentiment label l_i is opposite to l_{i-1},

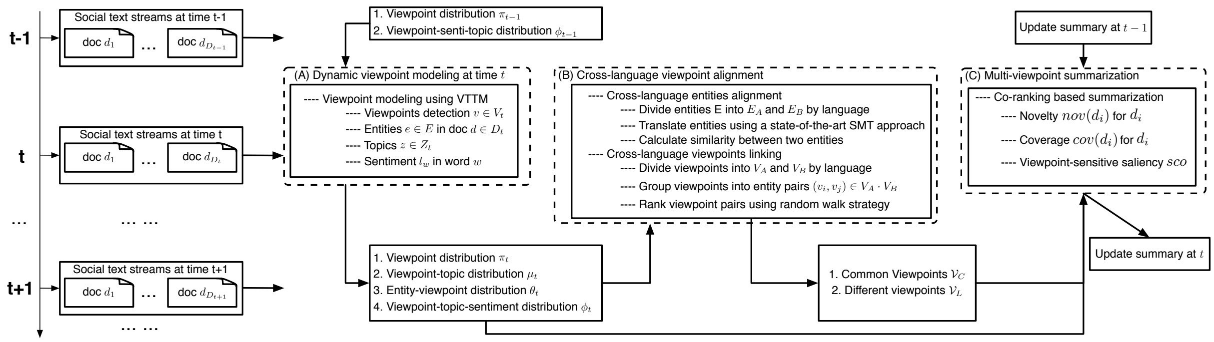


Figure 2: Overview of our approach to dynamic viewpoint summarization in social text streams. (A) dynamic viewpoint modeling; (B) cross-language viewpoint alignment; (C) multi-viewpoint summarization and generation of the update summary.

- For each topic $z \in \mathcal{Z}$ and sentiment l at time t:
  - Draw $\phi_{z,l,t} \sim Dir(\phi_{z,l,t-1} \cdot \beta_t)$;
- For each viewpoint $v \in \mathcal{V}$:
  - Draw $\pi_{v,t} \sim Dir(\alpha \cdot \pi_{v,t-1})$;
  - Draw $\mu_{v,t} \sim Dir(\chi)$; $\theta_{v,t} \sim Dir(\delta)$
  - For each topic z, draw $\rho_{v,z} \sim Beta(\eta)$;
- For each document $d \in \mathcal{D}_t$:
  - Draw a viewpoint $v_d \sim Multi(\pi_t)$;
  - Draw an entity $e_d \sim Multi(\theta_{v_d,t})$;
  - Draw $\sigma \sim Dir(\tau)$;
  - For each word $w_i \in \mathcal{N}_d$, $0 < i < N_d$:
    * Draw a topic $z_i \sim Multi(\mu_{v_d,t})$;
    * Draw $x_i \sim Multi(\sigma)$;
    * If $x_i = 1$, draw $l_i \sim l_{i-1}$
    * If $x_i = -1$, draw $l_i \sim (-1) \cdot l_{i-1}$;
    * If $x_i = 0$, draw $l_i \sim Bern(\rho_{v_d,z_i})$;
    * Draw word $w_i \sim Multi(\phi_{z_i,l_i,t})$:

Figure 3: Generative process in VTTM at time period t.

which shows that the sentiment label changes form one polarity to the other. Thus, we set the transition variable $x_i = 1$ when w_i and w_{i-1} are connected by a correlative conjunction, such as "and" and "both"; and we set $x_i = -1$ when w_i and w_{i-1} are connected by an adversative conjunction, such as "but" and "whereas"; we set $x_i = 0$ for other kinds of conjunctions. The generative process of VTTM is shown in Fig. 3.

Similar to other topic models [6, 7, 25, 45], it is intractable to derive the explicit posterior distribution of viewpoint $v_{d,t}$ at time period t. We apply a Gibbs sampling method [12] for sampling from the posterior distribution over viewpoints, entities, topics and sentiment labels. The sampling algorithm provides a method for exploring the implicit topic for each word and the particular viewpoint for each document.

At time period t, given document d, the target of our sampling is to approximate the posterior distribution $p(v_d, \vec{z}_d, \vec{l}_d, \vec{x}_d \mid \mathcal{W}, \mathcal{Z}, \mathcal{V}, \mathcal{E}, t)$, where $\vec{z}_d$, $\vec{l}_d$ and $\vec{x}_d$ indicate document d's topic vector, sentiment labels, and transition vector, respectively. Conceptually, we divide our sampling procedure into two parts. First, we sample the conditional probability of viewpoint v_d in each document $d \in \mathcal{D}_t$ given the values of inferred topics and sentiment labels, i.e., $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z})$. Second, given the current state of viewpoints, we sample the conditional probability of topic z_i with sentiment label l_i for word w_i, i.e., $P(z_i = k, l_i = l, x_i = x \mid \mathcal{X}_{-i}, \mathcal{L}_{-i}, \mathcal{Z}_{-i}, \mathcal{W}, v_d)$.

As the first step in our sampling procedure, for each document $d \in \mathcal{D}_t$, to calculate the probability of viewpoint v_d by sampling $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z})$, we have:

$$P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}) \propto \frac{n_{v,t}^{-d} + \alpha \cdot \pi_{t-1}}{n_t^{-d} + 1} \cdot \prod_{e \in \mathcal{E}} \frac{n_{v,e,t}^{-d} + \delta}{n_{v,t}^{-d} + E\delta} \cdot \tag{1}$$
$$\prod_{z \in \mathcal{Z}} \frac{n_{v,z,t}^{-d} + \chi_{z,t}}{n_{z,t}^{-d} + \sum_{z \in \mathcal{Z}} \chi_{z,t}} \cdot \prod_{l \in \mathcal{L}} \prod_{w \in \mathcal{N}_d} \frac{n_{z,l,v,t}^{w} + \beta_t \cdot \phi_{t-1,w}^{z,l,v}}{n_{z,l,v,t}^{-i} + \sum_{w \in \mathcal{N}} \beta_t \cdot \phi_{t-1,w}^{z,l,v}},$$

where $n_{v,t}^{-d}$ indicates the number of times that documents have been assigned to viewpoint v at t, except for document d; $n_{v,e,t}^{-d}$ indicates the number of times that entity e has been assigned to viewpoint v at t, excluding d; $n_{v,z,t}^{-d}$ indicates the number of times that topic z, at time t, has been assigned to viewpoint v, except for topic z in d; $n_{z,l,v,t}^{w}$ indicates the number of times that word w has been assigned to z, l and v jointly at t; $\phi_{t-1,w}^{z,l,v}$ is the probability of word w given v, z and l at $t-1$.

As the second step in our sampling procedure, given the viewpoint v_d sampled from document d, when $x_i \neq 0$ and $x_{i+1} \neq 0$ we sample the ith word w_i's topic z_i and sentiment label l_i using the probability in Eq. 2:

$$\begin{aligned} & P(z_i = k, l_i = l, x_i = x \mid \mathcal{X}_{-i}, \mathcal{L}_{-i}, \mathcal{Z}_{-i}, \mathcal{W}, v_d) \propto \\ & \frac{n_{v_d,k,t}^{-i} + \chi_{k,t}}{n_{v_d,t}^{-i} + \sum_{z \in \mathcal{Z}} \chi_{z,t}} \cdot \frac{n_{k,l,v_d,t}^{w_i,-i} + \beta_t \cdot \phi_{t-1,w_i}^{k,l,v_d}}{n_{k,l,v_d,t}^{-i} + \sum_{w \in \mathcal{N}} \beta_t \cdot \phi_{t-1,w}^{k,l,v_d}} \cdot \\ & \frac{n_{-i,x}^{w_i} + \tau_x}{n_{-i}^{w_i} + \sum_{x \in \mathcal{X}} \tau_x} \cdot \frac{n_{-(i+1),x_{i+1}}^{w_{i+1}} + I(x_{i+1} = x_i) + \tau_x}{n_{-(i+1)}^{w_{i+1}} + 1 + \sum_{x \in \mathcal{X}} \tau_x} \end{aligned} \tag{2}$$

where $n_{v_d,k,t}^{-i}$ indicates the number of times that a word with viewpoint v_d has been assigned to a topic k at time period t, except for the ith word; $n_{-i,t}^{d}$ indicates the number of words in document d, except for the ith word; $n_{-i,k,l}$ indicates the number of times that a word has been assigned to topic z and sentiment l synchronously, excluding the ith word; ϕ_{t-1}^{k,l,w_i} is the probability of word w_i given z and l at $t-1$; $n_{-i,x}^{w_i}$ indicates the number of times that w_i has been assigned to x, excluding the current one; and $I(x_{i+1} = x_i)$ gets the value 1 if $x_{i+1} = x_i$, and 0 otherwise. When $x_i = 0$, w_i's sentiment label l_i is derived from a Bernoulli distribution ρ_{v_d,z_i}, thus the last part in Eq. 2 is replaced by a posterior distribution over η, i.e., $(n_{z,l,v,t}^{-i} + \eta_l)/(n_{v,z,t}^{-i} + \sum_{l \in \mathcal{L}} \eta_l)$.

After sampling the probability for each viewpoint v, topic z and sentiment label l, at time period t we approximate the random variable ϕ_t that indicates the probability distribution over viewpoints, topics and sentiments labels, a viewpoint distribution π_t, a topic distribution μ_t over viewpoints, and entity distribution θ_t over viewpoints, similar to Iwata et al. [25].

4.3 (B) Cross-language viewpoint alignment

Using VTTM, we extract viewpoints from multi-lingual social text streams. Multi-linguality may make the viewpoint set $\mathcal{V}$ redundant and ambiguous. To address this, we present a cross-language viewpoint alignment strategy to connect the same viewpoint across languages. Shortness and sparseness hinder statistical machine translation in social text streams. We consider entities, i.e., concepts that can be linked to a specific Wikipedia document, as a means to connect viewpoints by comparing the similarity between two linked Wikipedia documents. We divide viewpoints $\mathcal{V}$ extracted from VTTM into $\mathcal{V}_A$ and $\mathcal{V}_B$ according to their languages L_A and L_B. Similarly, we divide entities $\mathcal{E}$ into $\mathcal{E}_A$ and $\mathcal{E}_B$ according to their languages.

Given viewpoint $v_A \in \mathcal{V}_A$, at time period t we extract the most relevant entity $e_i \in \mathcal{E}_A$ that has the highest $\theta_{v,e_i,t}$, i.e., $P(e_i \mid v, t)$. The same procedure is adapted to obtain $e_j \in \mathcal{E}_B$ for another viewpoint $v_B \in \mathcal{V}_B$. We compute the similarity between v_A and v_B by comparing the similarity between two entities e_i and e_j, shown in Eq. 3:

$$sim_t(v_A, v_B \mid t) = sim(e_i, e_j) \cdot \theta_{v_A,e_i,t} \cdot \theta_{v_B,e_j,t}, \tag{3}$$

where $sim(e_i, e_j)$ is the similarity between e_i and e_j in two languages. To compute $sim(e_i, e_j)$, we compute the similarity between two linked Wikipedia documents. Using links to English Wikipedia documents on Wikipedia pages, we translate a non-English Wikipedia document to an English Wikipedia document, i.e., a corresponding English Wikipedia document $\widehat{\mathcal{W}}_{e_j}$ for document $\mathcal{W}_{e_j}$. We use LDA [7] to represent each Wikipedia document $\mathcal{W}$ as a K-dimension topic vector $\vec{\varphi}_{\mathcal{W}}$. Then $sim(e_i, e_j)$ is computed proportionally to the inner product of the two vectors:

$$sim(e_i, e_j) = \frac{|\vec{\varphi}_{\mathcal{W}_{e_i}} \cdot \vec{\varphi}_{\widehat{\mathcal{W}}_{e_j}}|}{|\vec{\varphi}_{\mathcal{W}_{e_i}}| \cdot |\vec{\varphi}_{\mathcal{W}_{e_j}}|}, \tag{4}$$

where $\vec{\varphi}_{\mathcal{W}_{e_i}}$ indicates the topic vector for entity e_i's Wikipedia document, and $\vec{\varphi}_{\widehat{\mathcal{W}}_{e_j}}$ indicates the topic vector for entity e_j's translated Wikipedia document. We sum up the similarities between v_A and v_B at all time periods to obtain the similarity between v_A and v_B: $sim(v_A, v_B) = \sum_t sim_t(v_A, v_B)$. Thus, for each viewpoint $v_A \in \mathcal{V}_A$, we find the most similar viewpoint $v_B \in \mathcal{V}_B$ to match with the highest $sim(v_A, v_B)$. By generating such viewpoint pairs, we extract a set of viewpoint pairs $\mathcal{V}_s$ from $\mathcal{V}$. To remove redundant viewpoint pairs from $\mathcal{V}_s$, we employ a random walk-based ranking strategy [14] to rank $\mathcal{V}_s$ iteratively, in which each viewpoint pair's score, sa, receives votes from other pairs. As shown in Eq. 5, we use the similarity between two viewpoint pairs as the transition probability from one to another:

$$tr((v_A, v_B), (v'_A, v'_B)) = \frac{|sim(v'_A, v_B) \cdot sim(v_A, v'_B)|}{|sim(v_A, v_B)| \cdot |sim(v'_A, v'_B)|}. \tag{5}$$

At the beginning of the iterative process, an initial score for each pair is set to $1/|\mathcal{V}_s|$, and at the c-th iteration, the score of a viewpoint pair i is computed in Eq. 6:

$$sa(i)^{(c)} = \mu \sum_{i \neq j} \frac{tr\,(i, j)}{\sum\limits_{j' \in \mathcal{V}_s} tr(i, j')} \cdot sa(j)^{(c-1)} + \frac{(1-\mu)}{|\mathcal{V}_s|}, \tag{6}$$

where $|\mathcal{V}_s|$ equals the number of viewpoint pairs; μ denotes a decay parameter that is usually set to 0.85. The iterative process will stop when it convergences. Then we extract the top $|\mathcal{V}_C|$ viewpoint pairs from the ranked list, and merge two viewpoints in a pair into a single viewpoint. Below, we write $\mathcal{V}_C$ to denote $|\mathcal{V}_C|$ common viewpoints shared by both $\mathcal{V}_A$ and $\mathcal{V}_B$, and $\mathcal{V}_L = (\mathcal{V}_A \cup \mathcal{V}_B, v) \backslash \mathcal{V}_C$ to denote viewpoints $v \notin \mathcal{V}_C$.

4.4 (C) Multi-viewpoint summarization

The last step of our method, after cross-language viewpoint alignment is time-aware multi-viewpoint summarization of social text streams. Following [11, 16, 33], we propose a time-aware multi-viewpoint summarization method to summarize time series viewpoints by extracting a set of documents at each time period.

Suppose a set of viewpoint summaries $\{\mathcal{S}_s\}_{s=1}^{t-1}$ has been generated and read during the previous $t-1$ time periods. Based on viewpoint pairs $\mathcal{V}_s$ and viewpoint distributions inferred via VTTM, our target is to generate an update summary $\mathcal{S}_t$ to reflect the distribution of viewpoints at time period t. Inspired by Wan [44], we employ a co-ranking based algorithm to calculate the saliency of each tweet by considering both novelty and coverage. *Novelty* concerns the semantic divergence of viewpoint probabilities between a candidate document $d_i \in \mathcal{D}_t$ and previous summaries $\{\mathcal{S}_s\}$. *Coverage* concerns the relevance of a candidate document $d_i \in \mathcal{D}_t$ to a given viewpoint. Each document d_i's total saliency score $sco(d_i)$ is composed of a novelty score $nov(d_i)$ and a coverage score $cov(d_i)$.

As in co-ranking, Markov random walks are employed to iteratively optimize the ranked list. Three matrices are constructed to capture the transmission probability between two documents. Given a viewpoint $v \in \mathcal{V}_C \cup \mathcal{V}_L$, item $M^A_{i,j}$ in matrix M^A is about the similarity between two candidate documents d_i and d_j in $\mathcal{D}_t$:

$$M^A_{i,j} = \frac{\sum\limits_{e,e'} sim(e, e') \cdot \sum\limits_{z \in \mathcal{Z}} \sum\limits_{l \in \mathcal{L}} \phi^{z,l,v}_{d_i,t} \cdot \phi^{z,l,v}_{d_j,t}}{\|\Phi^v_{d_i,t}\| \cdot \|\Phi^v_{d_j,t}\|}, \tag{7}$$

where entity e and e' belong to $\mathcal{E}_{d_i}$ and $\mathcal{E}_{d_j}$, respectively; $\Phi^v_{d_i,t}$ is a matrix over topics and sentiment labels; each item for z, l, i.e., $\phi^{z,l,v}_{d_i,t}$ in Eq. 7, is calculated by averaging the value of $\phi^{z,l,v}_{t,w}$ of all words $w \in d_i$. Since the transmission matrix must be a stochastic matrix [13], we normalize M^A to $\widehat{M}^A$ by making the sum of each row equal to 1.

Similarly, we use $\widehat{M}^B$ to represent the transmission matrix among summaries during the previous $t-1$ time periods; we use M^{AB} to represent the similarity between $\mathcal{D}_t$ and $\{\mathcal{S}_s\}_{s=1}^{t-1}$. We normalize M^{AB} to $\widehat{M}^{AB}$ by making the sum of each row equal to 1.

The third and last matrix, W^{AB}, is about the divergence between $\mathcal{D}_t$ and $\{\mathcal{S}_s\}_{s=1}^{t-1}$; given a viewpoint v, we calculate each item $W^{AB}_{i,j}$ in $\widehat{W}^{AB}$ using Eq. 8:

$$W^{AB}_{i,j} = \frac{|t-s| \cdot |\pi_{v,t} - \pi_{v,s}| \cdot \|\Phi^v_{d_i,t} - \Phi^v_{d_j,t}\|}{\|\Phi^v_{d_i,t}\| \cdot \|\Phi^v_{d_j,t}\|}, \tag{8}$$

After row-normalization, we obtain $\widehat{W}^{AB}$ from W^{AB}. Using a co-ranking based update summarization algorithm [44], given a viewpoint v, for each iteration we use two column vectors $nov(d) = [nov(d_i)]_{i \in \mathcal{D}_t}$ and $cov(d) = [cov(d_i)]_{i \in \mathcal{D}_t}$ to denote the novelty scores and coverage scores of the documents in $\mathcal{D}_t$, respectively. In order to compute the viewpoint-biased scores of the documents, we use column vectors $\kappa_{d,v} = [\kappa_{d_i,v}]_{i \in \mathcal{D}_t}$ to reflect the relevance of the documents to the viewpoint v, where each entry in $\kappa_{d,v}$ corresponds to the conditional probability of the given viewpoint in documents, i.e., $\|\Phi^v_{d_i,t}\|$. Then κ is normalized to $\widehat{\kappa}$ to make the sum of all elements equal to 1.

After computing the above matrices and vectors, we can compute the update scores and the coverage scores of the documents in a co-ranking process. So at the c-th iteration, the update and coverage scores of d_i are calculated as:

$$nov(d_i)^{(c)} = \varepsilon_1 \sum_{j \in D_t}^{i \neq j} \widehat{M}^{A}_{i,j} \cdot nov(d_j)^{(c-1)} + \varepsilon_2 \sum_{j \in \{S_s\}} \hat{W}^{AB}_{i,j} \cdot nov(d_j)^{(c-1)} + \frac{(1-\varepsilon_1-\varepsilon_2)}{D+S} \cdot \kappa_{d_i,v} \tag{9}$$

and

$$cov(d_i)^{(c)} = \gamma_1 \sum_{j \in D_t}^{i \neq j} \widehat{M}^{A}_{i,j} \cdot cov(d_j)^{(c-1)} + \gamma_2 \sum_{j \in \{S_s\}} \widehat{M}^{AB}_{i,j} \cdot cov(d_j)^{(c-1)} + \frac{(1-\gamma_1-\gamma_2)}{D+S} \cdot \kappa_{d_i,v}, \tag{10}$$

where we set γ and ε as decay parameters in random walks. Initially, we set $nov(d_i)$ and $cov(d_i)$ as $\frac{1}{D_t}$, respectively. After each iteration c, we normalize $nov(d_i)^{(c)}$ and $cov(d_i)^{(c)}$ and calculate the *saliency* score of each document d_i as follows:

$$sco(d_i)^{(c)} = nov(d_i)^{(c)} + cov(d_i)^{(c)} \tag{11}$$

Following Eq. 9 and 10, for each given viewpoint $v \in \mathcal{V}_C \cup \mathcal{V}_L$, we rank documents in $\mathcal{D}_t$ to a ranking list $\mathcal{R}_v$, thus we apply Algorithm 1 to select documents to generate the viewpoint summary at time t. Eventually, we generate a set of summaries $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_T\}$ as the time-aware summarization result.

Algorithm 1: Time-aware multi-viewpoint summarization at time period t

Input:
Viewpoints $\mathcal{V}_C$ and $\mathcal{V}_L$, ranking list $\{R_v\}_{v \in \mathcal{V}_C \cup \mathcal{V}_L}$,
summaries $\{\mathcal{S}_s\}_{s=1}^{t-1}$, $\mathcal{D}_t$, probability distributions π_t, θ_t, ϕ_t,
probability distributions $\{\pi_s\}_{s=1}^{t-1}$, $\{\theta_s\}_{s=1}^{t-1}$, $\{\phi_s\}_{s=1}^{t-1}$
Output: Multi-viewpoint summary $\mathcal{S}_t$ at t;

```
Ω ← null; T ← predefined threshold; L ← length of summary
while |Ω| < L do
    for each v do
        d_i = top document in R_v;
        R_v = R_v − d_i;
        if max_{d_j∈Ω} sim(d_i, d_j | v, t) < T then
            Ω = Ω + d_i;
    if |Ω| = L then
        S_t = Ω;
        Break;
```

5. EXPERIMENTAL SETUP

In §5.1, we formulate three research questions to guide our experiments; we describe our dataset in §5.2 and specify how data was labeled in §5.3; §5.4 details the parameters used, and §5.5 details our evaluation metrics; the baselines are described in §5.6.

5.1 Research questions

The research questions that guide the remainder of the paper are: (RQ1) How does our viewpoint tweet topic model (VTTM) perform in time-aware viewpoint modeling? Does it help detect time-aware viewpoint drift? (See §6.1.) (RQ2) What is the performance of cross-language viewpoint alignment? Can it help detect common viewpoints from multilingual social text streams? (See §6.2.) (RQ3) How does our end-to-end time-aware multi-viewpoint summarization method (TAMVS) perform? Does it outperform baselines? What is the effect if we only consider *novelty* or *coverage*? (See §6.3.)

5.2 Dataset

In order to assess the performance of our methods, we collect a dataset of microblogs in two languages. We define multilingual queries about 6 well-known topics in 2014 and crawl English and Chinese microblogs via the Twitter streaming API[1] and a Sina Weibo[2] crawler, respectively. Table 2 provides descriptive statistics about the dataset. The tweets and weibos are posted between January, 2014 and August, 2014.

Table 2: Six topics in our dataset. The first column shows the topic name. The second and third column shows the number of English tweets and Chinese weibos per topic respectively. Each item is divided into two parts: the number of documents annotated, and the number of documents for each topic.

Topic	# tweets	# weibos
1. The World Economic Forum	2,000/2,000	1,978/1,978
2. Whaling hunting	566/566	1,072/1,072
3. FIFA Worldcup 2014	1,120/1,963	1,801/1,801
4. Missing MH370	3,124/6,308	4,725/4,725
5. Anti-Chinese in Vietnam 2014	825/2,001	1,095/1,095
6. Sinking of the MV Sewol	403/2,000	1,400/1,881

To evaluate the effectiveness of time-aware viewpoint summarization methods in our dataset, we used a crowdsourcing platform and had workers to label the ground truth in our dataset in their native language (i.e., Chinese or English); §5.3 details the annotations we obtained. In total, 8,308 English tweets and 12,071 Chinese weibos were annotated.

5.3 Crowdsourcing labeling

We obtain our annotations using the CrowdTruth platform [24] and assess the annotations using the CrowdTruth metrics [5].

The *Topic* annotation task gathers relevant tweets for each topic introduced in Table 2, and relevant topic mentions from each given tweet. Based on the answers gathered from the crowd we construct for each topic type a set of relevant tweets and a set of relevant topic mentions. Following the CrowdTruth approach, each tweet is assigned a topic type relevance score and each topic mention a relevance score. The *Sentiment* annotation task captures the sentiment and the intensity (i.e., high, medium, low) of the tweets and their topic mentions. The crowd provides the sentiment and the intensity of each topic mention and the overall sentiment and intensity of the tweet. The *Novelty ranking* task provides a ranking of the tweets based on how much new information they bring in with regard to a given topic. As data preparation, the tweets of a given topic are sorted chronologically and split by day. The crowdsourcing task is a pair-wise comparison of the tweets by following the approach: every tweet of a particular day is compared to all the following tweets, resulting in $\frac{n(n-1)}{2}$ comparison pairs per day, where n is the total number of tweets published on that day. Given the summary of the topic, for each pair of tweets, the crowd indicates which tweet is more salient with regard to the topic. By analyzing these judgments we provide, per day, a ranked list of salient tweets.

Table 3 provides an overview of the annotations gathered. On each task we applied the CrowdTruth metrics [5] in order to identify and remove spam, low-quality workers and their annotations.

[1] https://dev.twitter.com/docs/streaming-apis

[2] Chinese microblogging platform, http://www.weibo.com.

Table 3: Crowdsourcing task results overview.

Task	Topic	Sentiment	Novelty ranking
Units	6,225	5,317	5,211
Jobs	92	77	82
#Total workers	6,337	6,555	5,336
#Unique workers	557	500	341
#Spam workers	1,085	1,334	1,284
#Total judgments	43,575	53,170	78,165
#Spam judgments	7,562	10,519	14,475
Total cost	$1,136	$1,328	$1,444

Only the quality annotations were used as ground truth basis for further experiments. We validate the results by performing manual evaluation of the annotations. We extract a pool of workers, evenly distributed between low and high-quality, and annotate them in the following way: 0 for quality work and 1 for low-quality work. These scores are then used to compute the precision, recall, accuracy and F1-score, in order to confirm the CrowdTruth metrics accuracy. Overall, we obtain high scores for each of the measures (above 0.85) and across tasks, which indicates that the low-quality workers were correctly separated from quality workers.

5.4 Parameters

Following existing topic models [20], for the weighted parameter $\alpha_{v,t}$ and β_t, we set $\alpha_{u,t}$ to $50/V$ and β_t to 0.5. For the hyperparameters χ and δ in VTTM, we set $\chi = \delta = 0.5$. The default number of viewpoints in VTTM is set to 20. To optimize the number of viewpoints, we compare the performance at different values (see below). In time-aware multi-viewpoint summarization we set the parameter $\varepsilon_1 = \varepsilon_2 = 0.4$ in Eq. 9 and $\gamma_1 = \gamma_2 = 0.4$ in Eq. 10; the convergence threshold in co-ranking is set to 0.0001. The length of the summary L is set to 200 words per time period.

5.5 Evaluation metrics

To assess VTTM, we adapt the *purity* and *accuracy* evaluation metrics, which are widely used in topic modeling and clustering experiments [37, 40]. To evaluate the performance of time-aware multi-viewpoint summarization, we measure the quality of summaries by counting overlapping textual units between the generated results and the ground truth results. We adopt the ROUGE evaluation metrics: ROUGE-1 (unigram), ROUGE-2 (bigram) and ROUGE-W (weighted longest common sequence).

Statistical significance of observed differences between the performance of two runs is tested using a two-tailed paired t-test and is denoted using $^{\blacktriangle}$ (or $^{\blacktriangledown}$) for strong significance for $\alpha = 0.01$; or $^{\vartriangle}$ (or $^{\triangledown}$) for weak significance for $\alpha = 0.05$.

5.6 Baselines and comparisons

We list the methods and baselines that we consider in Table 4. We divide our methods into 3 groups according to the phases A, B, and C specified in §4. We write VTTM for the dynamic viewpoint model we proposed in §4.2. In the context of **RQ1**, we write VTTM-S for the stationary viewpoint modeling method. We write CLVA for the LDA-based viewpoint alignment method in phase B. In the context of **RQ2**, we write CLVA-T for the alignment method that applies term frequency in viewpoint similarity calculation, CLVA-E for the alignment method that only checks the consistency of entities. We write TaMVS for the overall process described in §4, which includes dynamic viewpoint modeling, cross-language viewpoint alignment and time-aware viewpoint summarization, and TaMVS-V for the viewpoint summarization method without considering cross-language viewpoint alignment. In the context of **RQ3** we use TaMVSN and TaMVSC to denote variations of TaMVS that only consider *Novelty* and *Coverage*, respectively.

Table 4: Our methods and baselines used for comparison.

Acronym	Gloss	Reference
Dynamic viewpoint modeling		
VTTM	Dynamic viewpoint modeling in (A)	§4.2
VTTM-S	Stationary viewpoint modeling in (A)	§4.2
Cross-language viewpoint alignment		
CLVA	LDA-based strategy in (B)	§4.3
CLVA-T	Term similarity based strategy in (B)	§4.3
CLVA-E	Entity similarity based strategy in (B)	§4.3
Time-aware multi-viewpoint summarization		
TaMVS	Summarization strategy defined in (C)	§4.4
TaMVS-V	TaMVS without phase B	§4.4
TaMVSN	TaMVS only considering *novelty* in (C)	§4.4
TaMVSC	TaMVS only considering *coverage* in (C)	§4.4
Topic models		
Sen-TM	Sentiment LDA based contrastive summarization	[28]
TAM	Topic-aspect model based contrastive summarization	[36]
Summarization		
CoRUS	Co-Ranking update summarization	[44]
IUS	Incremental update summarization	[33]
LexRank	LexRank algorithm for summarization	[13]

No previous work has addressed the same task as we do in this paper. However, some existing work can be considered as baselines in our experiments. To assess the contribution of VTTM in dynamic viewpoint modeling, our baselines include recent work on stationary viewpoint modeling. We use the Topic-aspect model [36, TAM] and the Sentiment-topic model [28, Sen-TM] as baselines for topic models. As baselines for summarization, we use three representative summarization algorithms, i.e., LexRank, IUS and CoRUS, as baselines: (1) the LexRank algorithm [13] ranks sentences via a Markov random walk strategy; (2) the IUS algorithm [33] generates an incremental update summary for given text streams; (3) the CoRUS algorithm [44] generates an update summary using a co-ranking strategy, but without VTTM.

6. RESULTS AND DISCUSSION

We compare VTTM to baselines for viewpoint modeling in social text streams, examine the performance of CLVA for cross-language viewpoint alignment as well as the end-to-end summarization performance of TaMVS.

6.1 Viewpoint modeling

First, Table 5 shows three example viewpoints output by VTTM. Column 1 shows the entities included by each viewpoint, column 2 shows topics attached with the entity in the viewpoint, columns 3, 4, 5 show the probability of positive, neutral and negative sentiment, respectively; column 6 shows the time period of the viewpoint. For a viewpoint about "China-Japan_relations" in Table 5, we find that its topic changes from "#The World Economic Forum" on 2014-01-26 to "#Anti-Chinese in Vietnam" on 2014-06-03.

Next, we address **RQ1** and test whether VTTM is effective for the viewpoint modeling task in social text streams. Table 6 shows the evaluation results for viewpoint modeling in terms of purity and accuracy for English tweets and Chinese weibos. For both languages, we find that VTTM outperforms TAM for all topics in terms of purity and accuracy. VTTM achieves an increase in pu-

Table 5: Task: dynamic viewpoint modeling. RQ1: Example viewpoints produced by VTTM. Column 1 lists the entities corresponding to the viewpoints; Column 2 list the topics in viewpoints, Columns 3, 4 and 5 list the probabilities of positive, neutral and negative labels for each topic, respectively. Column 6 shows the time interval of each viewpoint.

Entity	Topic	Positive	Neutral	Negative	Time interval
Search_for_Malaysia_Airlines_Flight_370	#Missing MH370	0.077	0.422	0.501	2014-03-27
Whaling_in_ Japan	#Whaling hunting	0.015	0.317	0.668	2014-05-05
China-Japan_relations	#The World Economic Forum	0.110	0.166	0.724	2014-01-26
	#Anti-Chinese in Vietnam	0.017	0.621	0.362	2014-06-03

rity over TAM of up to 23.4%, while accuracy increases by up to 21.4%. Compared with Sen-LDA, VTTM offers an increase of up to 12.0%, whereas accuracy increases by up to 12.6%. We look at those unsuccessful results made by VTTM, and find that for 67.2% of those documents the sentiment labeling results are incorrect, whereas for 75.4% of those documents the topic prediction results are incorrect. Another aspect of **RQ1** concerns *viewpoint drift*, i.e., changes of statistical properties. Table 7 contrasts the average performance of VTTM and VTTM-S (the stationary version of VTTM) for all periods in terms of Accuracy. For both languages, VTTM outperforms VTTM-S for each topic. We conclude that VTTM responds better to topic drift than VTTM-S, which neglects the dependency of viewpoints between two adjacent intervals.

6.2 Cross-language viewpoint alignment

To detect the number of common viewpoints between documents in two languages, we evaluate the ROUGE performance of TaMVS with varying numbers of common viewpoints $|\mathcal{V}_C|$. Using the same numbering of topics as in Table 2, Fig. 4 shows the number of shared viewpoints $\mathcal{V}_C$ for our 6 test topics; we find that Weibo users have more common viewpoints with Twitter users on the topics "#Missing MH370" and "#FIFA Worldcup 2014" than on other topics. To test the effectiveness of our cross-language viewpoint alignment strategy in **RQ2**, we examine the performance of CLVA for every topic; see Table 8. CLVA outperforms the other two methods, CLVA-T and CLVA-E, for each topic. CLVA-T outperforms CLVA-E on the cross-language viewpoint alignment task.

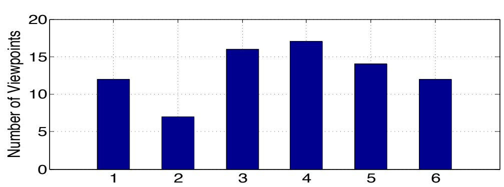


Figure 4: Task: cross-language viewpoint alignment. RQ2: Length of common viewpoints $\mathcal{V}_C$ in 6 topics. The numbers on the x-axis correspond to the topic numbers in Table 2.

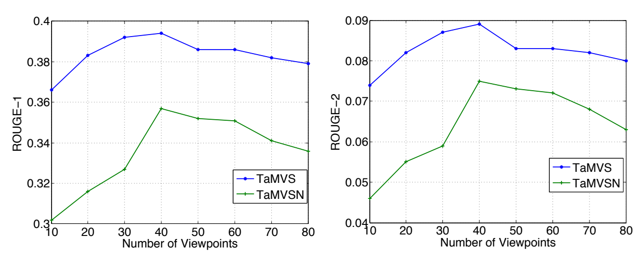
6.3 Overall performance

Tables 9 and 10 show the per topic time-aware multi-viewpoint summarization performance of all methods in terms of the ROUGE metrics. We begin by examining the importance of cross-language viewpoint alignment. Looking at Table 9, we see that TaMVS (columns 2–4) significantly outperforms TaMVS-V in which we leave out the cross-language viewpoint alignment step for each topic, and that it does so for all metrics (columns 5–7). This shows the importance of cross-language viewpoint alignment in multi-viewpoint summarization.

Turning to **RQ3**, to determine the contribution of novelty and coverage, we turn to Table 9, where columns 2–4, 8–10 and 11–13 show the performance of TaMVS, TaMVSN, and TaMVSC, respectively in terms of the ROUGE metrics. Recall that TaMVSN only considers novelty in phase C and that TaMVSC only considers coverage in phase C. We find that TaMVS, which combines novelty and coverage, outperforms both TaMVSN and TaMVSC on all topics. After TaMVS, TaMVSN, which only includes novelty during the summarization process, performs best. Thus, novelty is the most important part of our multi-viewpoint summarization process.

Turning to Table 10, we find that TaMVS outperforms the baselines on all test topics in terms of ROUGE-1, and in several cases significantly so. In terms of ROUGE-2, we see a similar picture: TaMVS outperforms the baselines, and in several cases significantly so. Among the baselines, LexRank gets the worst performance simply because it ignores the dynamic patterns during viewpoint modeling. And CoRUS achieves the second best performance, which indicates the importance of update summarization in our viewpoint summarization. TaMVS achieves 3.2% and 7.5% increases over CoRUS in terms of ROUGE-1 and ROUGE-2, respectively, and 12.1% and 37.1% increases over IUS in terms of ROUGE-1 and ROUGE-2. Compared to Sen-TM, TaMVS achieves a statistical significant improvement of up to 28.1% in terms of ROUGE-1 and 63.4% in terms of ROUGE-2. Interestingly, TaMVS performs better on test topics that have higher scores for dynamic viewpoint modeling (phase A, see Table 6), which underlines the importance of dynamic viewpoint modeling in time-aware multi-viewpoint summarization.

We now analyze the influence of the number of viewpoints. Fig. 5 plots the average ROUGE performance curves for TaMVS and TaMVSN with varying numbers of viewpoints. We that find for both metrics and methods, the performance peaks when the number of viewpoints equals 40, i.e., higher than our default value of 20.



(a) ROUGE-1 (b) ROUGE-2

Figure 5: Task: time-aware multi-viewpoint summarization. RQ3: Performance with different numbers of viewpoints, in terms of ROUGE-1 (a) and ROUGE-2 (b).

7. CONCLUSION AND FUTURE WORK

We have considered the task of time-aware multi-viewpoint summarization of social text streams. We have identified four main challenges: ambiguous entities, viewpoint drift, multi-linguality,

Table 6: Task: dynamic viewpoint modeling. RQ1: Comparison of methods. Purity is abbreviated as pur., Accuracy as acc. We use $^\blacktriangle$ to denote statistically significant improvements of VTTM over the baseline TAM.

	English tweets						Chinese weibos					
	VTTM		TAM		Sen-LDA		VTTM		TAM		Sen-LDA	
Topic	pur.	acc.	pur.	acc.	pur.	acc.	pur.	acc.	pur.	acc.	pur.	acc.
The World Economic Forum	**0.497**$^\blacktriangle$	**0.516**$^\blacktriangle$	0.401	0.415	0.419	0.425	**0.441**$^\blacktriangle$	**0.472**$^\blacktriangle$	0.352	0.371	0.391	0.407
Whaling hunting	**0.454**	**0.463**	0.432	0.435	0.451	0.462	0.493	0.505	0.442	0.458	**0.501**	**0.513**
FIFA Worldcup 2014	**0.472**$^\triangle$	**0.423**$^\triangle$	0.432	0.442	0.445	0.451	**0.541**$^\blacktriangle$	**0.561**$^\blacktriangle$	0.432	0.442	0.483	0.497
Missing MH370	**0.463**$^\blacktriangle$	**0.471**$^\blacktriangle$	0.391	0.403	0.427	0.445	**0.501**$^\blacktriangle$	**0.542**$^\blacktriangle$	0.343	0.352	0.451	0.462
Anti-Chinese in Vietnam	**0.491**$^\blacktriangle$	**0.511**$^\blacktriangle$	0.406	0.415	0.452	0.557	**0.522**$^\blacktriangle$	**0.541**$^\blacktriangle$	0.482	0.495	0.503	0.517
Sinking of the MV Sewol	**0.425**	**0.438**	0.361	0.372	0.407	0.411	**0.625**$^\blacktriangle$	**0.642**$^\blacktriangle$	0.497	0.507	0.559	0.572
Overall	**0.474**$^\blacktriangle$	**0.482**$^\blacktriangle$	0.384	0.397	0.417	0.428	**0.524**$^\blacktriangle$	**0.543**$^\blacktriangle$	0.437	0.452	0.482	0.504

Table 7: Task: dynamic viewpoint modeling. RQ1: Contrasting the performance of VTTM and VTTM-S in the Chinese viewpoint modeling task.

	VTTM		VTTM-S	
Topic	pur.	acc.	pur.	acc.
The World Economic Forum	**0.497**	**0.516**	0.496	0.513
Whaling hunting	**0.454**	**0.463**	0.449	0.459
FIFA Worldcup 2014	**0.472**	**0.423**	0.441	0.459
Missing MH370	**0.463**	**0.471**	0.433	0.448
Anti-Chinese in Vietnam	**0.491**	**0.511**	0.456	0.471
Sinking of the MV Sewol	**0.425**	**0.438**	0.422	0.435
Overall	**0.474**	**0.482**	0.461	0.474

Table 8: Task: cross-language viewpoint alignment. RQ2: Performance of CLVA in cross-language viewpoints alignment task, in terms of Accuracy.

Topic	CLVA	CLVA-T	CLVA-E
The World Economic Forum	**0.754**	0.613	0.591
Whaling hunting	**0.737**	0.671	0.622
FIFA Worldcup 2014	**0.643**	0.588	0.521
Missing MH370	**0.727**	0.611	0.524
Anti-Chinese in Vietnam	**0.787**	0.732	0.655
Sinking of the MV Sewol	**0.854**	0.712	0.659
Overall	**0.711**	0.669	0.615

and the shortness of social text streams. We have proposed a dynamic viewpoint modeling strategy to infer multiple viewpoints in the given multilingual social text steams, in which we jointly model topics, entities and sentiment labels. After cross-language viewpoint alignment, we apply a random walk ranking strategy to extract documents to tackle the time-aware multi-viewpoint summarization problem. We have demonstrated the effectiveness of our proposed method by showing a significant improvement over various baselines tested with a manually annotated dataset. Our viewpoint tweet topic model is helpful for detecting the viewpoint drift phenomenon and summarizing viewpoints over time.

Although we focused mostly on microblogs, our methods are broadly applicable to other settings with opinionated content, such as comment sites or product reviews. Limitations of our work include its ignorance of viewpoint dependencies, viewpoint diversity and, being based on LDA, its predefined number of viewpoints. As to future work, *contrastive* viewpoints in multilingual text streams are worth considering. Also, the transfer of our approach to a non-parametric extension should give new insights and an extrinsic online user evaluation would give deeper insights into the performance of our approach. A novel graphical model that includes dynamic time bins instead of the fixed time granularities, is another direction for future research. Finally, discovering new entities that are not included by Wikipedia will help our approach to explore realtime viewpoints.

Acknowledgments. We thank Kai Wu, Wei Zhang, Hongya Song and Shanshan Huang at Shandong University for collecting the dataset for us. This research was supported by Ahold, Amsterdam Data Science, Blendle, the Bloomberg Research Grant program, the Dutch national program COMMIT, Elsevier, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the ESF Research Network Program ELIAS, the Royal Dutch Academy of Sciences (KNAW) under the Elite Network Shifts project, the Microsoft Research Ph.D. program, the Netherlands eScience Center under project number 027.-012.105, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs 727.011.005, 612.001.116, HOR-11-10, 640.006.013, 612.066.930, CI-14-25, SH-322-15, 652.002.001, 612.001.551, 652.001.003, the Yahoo Faculty Research and Engagement Program, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

8. REFERENCES

[1] A. Ahmed and E. P. Xing. Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. In *UAI*, 2012.

[2] M. Albakour, C. Macdonald, I. Ounis, et al. On sparsity and drift for effective real-time filtering in microblogs. In *CIKM*. ACM, 2013.

[3] J. Allan. Introduction to topic detection and tracking. In *Topic detection and tracking*, pages 1–16. Springer, 2002.

[4] L. AlSumait, D. Barbará, and C. Domeniconi. On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *ICDM*, 2008.

[5] L. Aroyo and C. Welty. The three sides of CrowdTruth. *Journal of Human Computation*, 1:31–34, 2014.

[6] D. M. Blei and J. D. Lafferty. Dynamic topic models. In *ICML*, 2006.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[8] D. M. Blei, T. L. Griffiths, and M. I. Jordan. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57(2), 2010.

[9] D. Chakrabarti and K. Punera. Event summarization using tweets. In *ICWSM*, 2011.

[10] T. Copeck, A. Kazantseva, A. Kennedy, A. Kunadze, D. Inkpen, and S. Szpakowicz. Update summary update. In *TAC*, 2008.

[11] J.-Y. Delort and E. Alfonseca. Dualsum: a topic-model based approach for update summarization. In *EACL*, 2012.

[12] Q. Diao and J. Jiang. Recurrent Chinese restaurant process with a duration-based discount for event identification from Twitter. In *SDM*, 2014.

[13] G. Erkan and D. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.*, 22:457–479, 2004.

Table 9: Task: time-aware multi-viewpoint summarization. RQ2 and RQ3: ROUGE performance of all VTTM-based methods in time-aware viewpoint summarization. ROUGE-1 is abbreviated as R-1, ROUGE-2 as R-2 and ROUGE-W as R-W. Statistically significant differences are with respect to TaMVS-V.

	TaMVS			TaMVS-V			TaMVSN			TaMVSC		
Topic	R-1	R-2	R-W	R-1	R-2	R-W	R-1	R-2	R-W	R-1	R-2	R-W
The World Economic Forum	**0.383**$^{\blacktriangle}$	**0.082**$^{\blacktriangle}$	**0.184**$^{\blacktriangle}$	0.316	0.055	0.131	0.369	0.057	0.136	0.351	0.052	0.145
Whaling hunting	**0.294**$^{\blacktriangle}$	**0.047**$^{\blacktriangle}$	**0.152**$^{\blacktriangle}$	0.197	0.037	0.092	0.221	0.032	0.138	0.221	0.032	0.138
FIFA Worldcup 2014	**0.436**$^{\blacktriangle}$	**0.094**$^{\blacktriangle}$	**0.202**$^{\blacktriangle}$	0.351	0.068	0.142	0.404	0.083	0.194	0.404	0.083	0.194
Missing MH370	**0.425**$^{\blacktriangle}$	**0.087**$^{\vartriangle}$	**0.232**$^{\blacktriangle}$	0.324	0.074	0.132	0.402	0.082	0.170	0.394	0.079	0.162
Anti-Chinese in Vietnam	**0.409**$^{\blacktriangle}$	**0.065**$^{\blacktriangle}$	**0.169**$^{\blacktriangle}$	0.226	0.046	0.127	0.403	0.047	0.139	0.395	0.042	0.129
Sinking of the MV Sewol	**0.373**$^{\blacktriangle}$	**0.064**$^{\blacktriangle}$	**0.171**$^{\blacktriangle}$	0.314	0.051	0.131	0.362	0.055	0.134	0.341	0.046	0.133
Overall	**0.387**$^{\blacktriangle}$	**0.085**$^{\blacktriangle}$	**0.188**$^{\blacktriangle}$	0.318	0.061	0.132	0.377	0.062	0.139	0.359	0.056	0.150

Table 10: Task: time-aware multi-viewpoint summarization. RQ3: Per topic performance of all methods. ROUGE-1 is abbreviated as R-1 and ROUGE-2 as R-2. We use $^{\blacktriangle}$ ($^{\vartriangle}$) to denote strong (weak) statistically significant improvements of TaMVS over CoRUS.

	TaMVS		TAM		Sen-TM		LexRank		IUS		CoRUS	
Topic	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2	R-1	R-2
The World Economic Forum	**0.383**$^{\vartriangle}$	**0.082**$^{\vartriangle}$	0.295	0.047	0.298	0.051	0.321	0.072	0.347	0.066	0.372	0.077
Whaling hunting	**0.294**	**0.047**	0.237	0.038	0.224	0.032	0.221	0.032	0.273	0.042	0.292	0.044
FIFA Worldcup 2014	**0.436**$^{\blacktriangle}$	**0.094**$^{\vartriangle}$	0.347	0.076	0.355	0.081	0.342	0.073	0.378	0.080	0.383	0.089
Missing MH370	**0.425**$^{\blacktriangle}$	**0.087**$^{\blacktriangle}$	0.297	0.049	0.314	0.062	0.341	0.072	0.358	0.071	0.352	0.082
Anti-Chinese in Vietnam	**0.409**$^{\blacktriangle}$	**0.065**$^{\blacktriangle}$	0.222	0.039	0.239	0.042	0.238	0.040	0.269	0.046	0.253	0.052
Sinking of the MV Sewol	**0.373**$^{\vartriangle}$	**0.064**	0.284	0.044	0.295	0.049	0.251	0.040	0.327	0.071	0.369	0.062
Overall	**0.387**$^{\vartriangle}$	**0.085**	0.297	0.047	0.302	0.052	0.309	0.052	0.345	0.062	0.375	0.079

[14] G. Erkan and D. R. Radev. LexPageRank: Prestige in multi-document text summarization. In *EMNLP*, 2004.

[15] Y. Fang, L. Si, N. Somasundaram, et al. Mining contrastive opinions on political texts using cross-perspective topic model. In *WSDM*, 2012.

[16] S. Fisher and B. Roark. Query-focused supervised sentence ranking for update summaries. *TAC*, 2008.

[17] K. Ganesan, C. Zhai, and J. Han. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. In *COLING*, 2010.

[18] K. Ganesan, C. Zhai, and E. Viegas. Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions. In *WWW*. ACM, 2012.

[19] W. Gao, P. Li, and K. Darwish. Joint topic modeling for event summarization across news and social media streams. In *CIKM*. ACM, 2012.

[20] T. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101: 5228–5235, 2004.

[21] T. Hofmann. Probabilistic latent semantic indexing. In *SIGIR*. ACM, 1999.

[22] M. Hu and B. Liu. Mining opinion features in customer reviews. In *AAAI*, 2004.

[23] M. Hu and B. Liu. Opinion extraction and summarization on the web. In *AAAI*, 2006.

[24] O. Inel et al. CrowdTruth: Machine-human computation framework for harnessing disagreement in gathering annotated data. In *ISWC*. Springer, 2014.

[25] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda. Topic tracking model for analyzing consumer purchase behavior. In *IJCAI 2009*, volume 9, pages 1427–1432, 2009.

[26] H. D. Kim and C. Zhai. Generating comparative summaries of contradictory opinions in text. In *CIKM*. ACM, 2009.

[27] H. D. Kim, M. G. Castellanos, M. Hsu, et al. Ranking explanatory sentences for opinion summarization. In *SIGIR*. ACM, 2013.

[28] F. Li et al. Structure-aware review mining and summarization. In *COLING*, 2010.

[29] P. Li et al. Generating aspect-oriented multi-document summarization with event-aspect model. In *EMNLP*, 2011.

[30] B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In *WWW*, 2005.

[31] Y. Lu, C. Zhai, and N. Sundaresan. Rated aspect summarization of short comments. In *WWW*. ACM, 2009.

[32] R. McCreadie, C. Macdonald, I. Ounis, et al. Scalable distributed event detection for Twitter. In *IEEE Int. Conf. on Big Data*, 2013.

[33] R. McCreadie, C. Macdonald, and I. Ounis. Incremental update summarization: Adaptive sentence selection based on prevalence and novelty. In *CIKM*. ACM, 2014.

[34] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW*. ACM, 2007.

[35] E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In *WSDM*. ACM, 2012.

[36] M. Paul and R. Girju. A two-dimensional topic-aspect model for discovering multi-faceted topics. *AAAI*, 2010.

[37] M. J. Paul, C. Zhai, and R. Girju. Summarizing contrastive viewpoints in opinionated text. In *EMNLP*, 2010.

[38] M.-H. Peetz, M. de Rijke, and R. Kaptein. Estimating reputation polarity on microblog posts. *Inf. Proc. & Man.*, 52(2):193–216, 2016.

[39] D. Ramage, S. T. Dumais, and D. J. Liebling. Characterizing microblogs with topic models. In *ICWSM*, 2010.

[40] Z. Ren and M. de Rijke. Summarizing contrastive themes via hierarchical non-parametric processes. In *SIGIR*. ACM, 2015.

[41] Z. Ren et al. Personalized time-aware tweets summarization. In *SIGIR*. ACM, 2013.

[42] L. Shou, Z. Wang, K. Chen, and G. Chen. Sumblr: Continuous summarization of evolving tweet streams. In *SIGIR*. ACM, 2013.

[43] O. van Laere et al. DEESSE: entity-driven exploratory and serendipitous search system. In *CIKM*. ACM, 2014.

[44] X. Wan. Update summarization based on co-ranking with constraints. In *COLING*, 2012.

[45] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *KDD*. ACM, 2011.

[46] D. Wang et al. Comparative document summarization via discriminative sentence selection. *TKDD*, 6(3):12, 2012.

[47] X. Wei, J. Sun, and X. Wang. Dynamic mixture models for multiple time series. In *IJCAI 2007*, pages 2909–2914, 2007.

[48] S.-H. Yang et al. Large-scale high-precision topic modeling on twitter. In *KDD*. ACM, 2014.

[49] K. Zhai and J. Boyd-Graber. Online latent dirichlet allocation with infinite vocabulary. In *ICML*, 2013.