

Topic-Driven Reader Comments Summarization*

Zongyang Ma, Aixin Sun, Quan Yuan, Gao Cong
School of Computer Engineering, Nanyang Technological University, Singapore 639798
{zma4, qyuan1}@e.ntu.edu.sg, {axsun, gaocong}@ntu.edu.sg

ABSTRACT

Readers of a news article often read its comments contributed by other readers. By reading comments, readers obtain not only complementary information about this news article but also the opinions from other readers. However, the existing ranking mechanisms for comments (*e.g.*, by recency or by user rating) fail to offer an overall picture of topics discussed in comments. In this paper, we first propose to study **Topic-driven Reader Comments Summarization** (Torcs) problem. We observe that many news articles from a news stream are related to each other; so are their comments. Hence, news articles and their associated comments provide context information for user commenting. To implicitly capture the context information, we propose two topic models to address the Torcs problem, namely, Master-Slave Topic Model (MSTM) and Extended Master-Slave Topic Model (EXTM). Both models treat a news article as a master document and each of its comments as a slave document. MSTM model constrains that the topics discussed in comments have to be derived from the commenting news article. On the other hand, EXTM model allows generating words of comments using both the topics derived from the commenting news article, and the topics derived from all comments themselves. Both models are used to group comments into topic clusters. We then use two ranking mechanisms Maximal Marginal Relevance (MMR) and Rating & Length (RL) to select a few most representative comments from each comment cluster. To evaluate the two models, we conducted experiments on 1005 Yahoo! News articles with more than one million comments. Our experimental results show that EXTM significantly outperforms MSTM by perplexity. Through a user study, we also confirm that the comment summary generated by EXTM achieves better intra-cluster topic cohesion and inter-cluster topic diversity.

Categories and Subject Descriptors

H.3.1 [INFORMATION STORAGE AND RETRIEVAL]: Content Analysis and Indexing—*Abstracting methods*

*This work was partially supported by MINDEF-NTU-DIRP/2010/03 and MOE AcRF Tier-1 Grant RG13/10, Singapore.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.

Keywords

Comments Summarization, Topic Model, Master-Slave Document

1. INTRODUCTION

The popularity of Web 2.0 platforms and applications prompts users' engagement in expressing and sharing their opinions online by commenting. Not only commenting through social platforms, users are allowed to comment on news articles on most news websites. For instance, Yahoo! News allows users to comment on its news articles, to rate comments, as well as to reply other readers' comments. Through reading others' comments, a reader may get additional information about an event reported in a news article, including the background, the other related events, and opinions of other readers. It has been reported that such commenting mechanism enables news sites to extend their user volume as well as to increase advertising revenue [33].

Due to the large readership, a popular news article may easily accumulate thousands of comments within a short time period. For instance, on average, there are 1059 comments per article among the most-commented articles collected from Yahoo! News in our study. It is time-consuming for users to read all comments of an article and they may feel disincentivized to read unorganized comments. Currently, comments can be ranked in a number of ways to ease users' reading overload. Taking Yahoo! News as an example, comments can be ranked by recency, popularity (*e.g.*, based on user ratings) and the number of replies a comment receiving. None of the mentioned ranking mechanisms offers readers an overview of all topics discussed among commenters. In particular, the user-rating based ranking suffers from the *rich get richer* effect. Consequently comments about one dominating topic may occupy all top-ranked positions while comments on other topics are ranked much lower. Nevertheless, to the best of our knowledge, *there is no service or existing research work on summarizing comments of an article to offer an easy overview of all topics discussed in the comments.*

In this paper, we take the first step to study the problem of **Topic-driven Reader Comments Summarization** (Torcs). We aim to identify the latent topics among all comments of an article and subsequently select a small number of most salient comments to give a concise representation of all comments according to the identified topics. Torcs benefits users by offering an alternative way of quickly ascertaining overall topics of thousands of comments of an article in addition to the existing comments ranking mechanisms. Moreover, Torcs enables a user to selectively read comments on one specific topic chosen by her from all latent topics identified. We argue that Torcs not only benefits Web users, but more importantly offers great value to mobile users. Due to the relatively limited screen size, mobile users do not have the luxury of reading a

large number of comments without scrolling. Grouping comments by latent topics and selecting the most salient comments under each topic bring a more organized way of reading comments through mobile devices.

One key challenge in Torcs is to group comments by their latent topics. Considering a comment is usually a piece of short text¹, at first glance, techniques on short text clustering and classification (e.g., search snippets clustering [29, 31] and classification [9]) can be directly applied. However, these techniques do not consider the *news-comment*, *comment-comment*, and *news-news* relationships. First, as defined in Merriam-Webster dictionary, a comment is “a note explaining, illustrating, or criticizing the meaning of a writing” or “an observation or remark expressing an opinion or attitude”. In news commenting, a comment is always associated with a news article. For this reason, we also call the news article *master document* and its comments *slave documents*. Second, for a given news article, we observe that a subset of its comments may discuss the topics mentioned in the news article, but usually not all of its comments. For instance, a news article talking about tax policy in US may have comments on tax policy, comments on health care reform, or gossip of persons mentioned in the article. That is, comments may extend the discussion with topics related to the news article or even more general topics completely irrelevant to the news article. Third, in a news stream, many news articles are on similar topics (or even the same event). For example, many news articles are about the ongoing US presidential election.

To identify the latent topics in comments with consideration of the relationships between comments and news articles, we propose two topic models, namely, *Master-Slave Topic Model* (MSTM) and *Extended Master-Slave Topic Model* (EXTM). The former assumes that all the topics discussed in comments are derived from the corresponding news articles. The latter is based on our observation that comments may discuss the topics derived from news articles as well as the extended topics not directly derived from news articles. In both models, we assume that each comment has exactly one topic due to the short length of most comments. Such simple topic assignment makes the comment clustering efficient. A cluster is formed by simply grouping comments with the same topic. After clustering process, we select representative comments using two methods, namely, Maximal Marginal Relevance (MMR) and Rating & Length (RL). The former is widely used in document summarization to select comments based on topical relevance while avoiding redundancy; the latter considers the user rating and the length of comment for representative comments selection. Users get more comprehensive understanding of the comments by giving topical and distinctive comments than representing summary merely using keywords. In our evaluation, we observe that the proposed topic-driven approach avoids redundancy problem faced by the word-based clustering approach to some extent. Our main contributions are summarized as follows:

- To the best of our knowledge, we are the first to define and study the problem of *Topic-Driven Reader Comments Summarization* (Torcs), which aims to summarize comments into topic clusters with respect to their associated news articles. Different from existing works on short text clustering, comment clustering is guided by the context information of their associated news articles.
- We present two topic models, MSTM and EXTM, to integrate the generative process of news articles and that of comments in a unified way. Both models are based on the observation that topics of news articles have significant effect

on topics of comments, and a single comment usually is specific to one topic. EXTM further considers topics that are generated from comments themselves which are not strongly related to news articles, simulating the generative process of news-comments document in a more natural manner.

- We conducted experiments on 1005 most-commented Yahoo! news articles with more than one million comments. We first evaluated MSTM and EXTM by perplexity, a standard metric to evaluate topic models, and show that EXTM significantly outperforms MSTM. We then conducted user study to evaluate summaries generated by 6 methods. These 6 methods involve 3 comment clustering techniques (MSTM, EXTM, and CLUTO as a non-topic-model approach) and 2 comment ranking schemes (MMR and RL). Our user study shows that EXTM-MMR yields the best intra-cluster topic cohesion and inter-cluster topic diversity among the summaries generated by all methods.

The rest of the paper is organized as follows. Section 2 surveys the existing studies on comments and recent advances in topic models. In Section 3, we define the Torcs problem, detail the two models proposed, and describe the two comment ranking schemes. The evaluation of MSTM and EXTM using perplexity is reported in Section 4, followed by the user study in Section 5. Finally, Section 6 concludes this paper.

2. RELATED WORK

We begin with a brief overview of the research works conducted on comments. We then review the related topic models, and lastly we discuss the differences between Torcs and short text clustering.

Being the most prevalent type of user-generated content, comments have gained much interest from the research community. Depending on the type of resources/objects being commented, the research works can be broadly categorized into two groups: study of comments on non-textual resources (e.g., photo, video), and study of comments on textual-resources (e.g., news article, blog post, and product description).

Comments on non-textual resources. For non-textual objects like photos and videos, content-based search solely relying on low-level features (e.g., color, texture, and shape) remains challenging [16]. Comments therefore become valuable resource to describe these non-textual resources to improve searching. It is reported that comment-based approach outperforms visual content-based approach for aesthetic photo search [30], and video search accuracy is improved by 15% by incorporating comments [38]. In [34], the authors presented a detailed analysis of YouTube video comments and aimed to reveal the relationship between the content and the ratings of comments.

Most similar to our work is on clustering YouTube comments by using topic model reported in [20]. In their study, for each YouTube video, a topic model is trained on all its comments. The comments are clustered by their topic assignment. Within each cluster, a small set of comments is selected using PageRank-style algorithm on the comment graph. In this graph, a link between two comments exists if the recently published comment contains at least m words from the earlier comment (where m is a predefined number). Our approach is significantly different. We consider not only the topic distribution in comments but also the topic distribution in news articles, and the latter would guide the overall discussion in comments. More specifically, we train our topic models using both the set of news articles and all of their comments to take care of the news-news, news-comment, and comment-comment relationships.

¹More than 60% of comments in our data collection has fewer than 10 words each.

Moreover, we use different approaches to select comments from a cluster, based on the comment’s topical representativeness as well as the user rating.

Comments on textual resources. The study of the comments on product review mainly focuses on sentiment analysis and product feature/aspect identification from comments [11, 18, 37]. The way of summarizing product comments is therefore significantly different from ours. On the other hand, blog posts share large similarity with news articles. Our work is therefore more related to the study of comments on blog posts and news articles. Mishne and Glance conducted a large-scale analysis of blog comments and showed that comments help in improving blog search recall and facilitate better ranking of blog search results [26]. More related to our work, Hu *et al.* proposed to summarize blog post by incorporating information from its associated comments [14]. The main idea is to rank comments by their importance considering three types of relations (*i.e.*, mention, quotation, and topic) among them. The blog post is summarized by extracting the sentences containing the keywords identified from the important comments as well as the content of the blog post. The comments, however, are not summarized.

Comments on news articles have been studied from different aspects, including spam detection [19], discussion structure extraction [32], comment volume prediction [36], comment rating prediction [21], comment-based news recommendation [33] and comment ranking [10]. The collaborative-filtering approach in [33] for personalized news recommendation considers ratings as well as features derived from social networks. For comment ranking, Dalal *et al.* proposed a unified ranking framework for multi-objective comment ranking [10]; the comments can be ranked by ratings, commenter reputation, or recency. To the best of our knowledge, there is no existing work on summarizing comments of news articles.

Topic Model. Latent Dirichlet Allocation (LDA) was first proposed in [5], which provides an unsupervised way to discover latent topics in text collection. The promising performance of LDA inspires various extensions. In the following, we name a few. Correlated Topic Model [4] captures relations of topics based on a more natural assumption that topics are not independent and gives a better fit of the data. Dynamic Topic Model [3] incorporates time feature into LDA, aiming to analyze topics evolution. Author-Topic Model [28] utilizes authorship information to simultaneously model author interest and document topic. Multi-grain Topic Model [35] extracts ratable aspects of objects from online reviews. In [25], the authors formulated Topic Model with network structure regularization to discover topical community. Several topic-sentiment models have also been proposed [18, 22]. This type of models generates topics as well as sentiment using review collection.

Most germane to this work are the correspondence LDA (corrLDA) model [2, 6] and the topic-perspective model [23], both proposed for analyzing social tag annotation (*e.g.*, tags annotated to documents in social bookmark systems like *delicious.com*). The corrLDA model assumes that tag topics are generated due to word topics of the annotated document; hence the model captures strong relations between document topics and its corresponding tag topics. The topic-perspective model, on the other hand, relaxes this rigid assumption. The authors argued that tags could also be generated from user perspective (*e.g.*, tags like *happy*, *delicious*, and so on reflect perspective of users themselves). This model therefore incorporates a switch, which controls tag generation from document topic or user perspective. Each user has a user perspective distribution to describe her preference and interest.

However, both models cannot be applied directly to comments

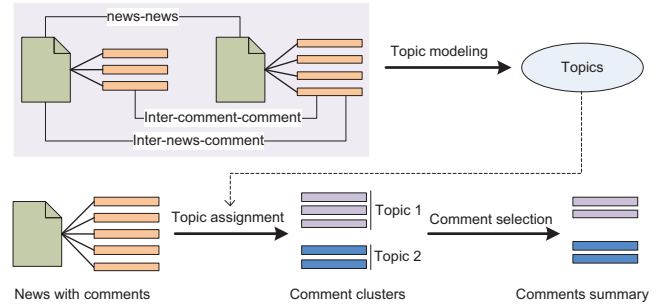


Figure 1: Topic-driven reader comments summarization

because of the significant differences between tags and comments. First, the primary motivations for tagging are organizational and social [24]. Hence, a tag is usually assigned to a set of documents related to the keyword defined by the user (*i.e.*, the tag itself). Naturally, each tag leads to multiple topics derived from its set of annotated documents. A comment, on the other hand, is an expression from a user to one particular article without organizational motivation. We observe that most comments are short and specific to one point (or one topic). Moreover, the social motivation for commenting is much weak compared to tagging evidenced by a large number of comments from anonymous users. Many users prefer to voice out through commenting without releasing their identity.

Our proposed MSTM model is similar to CorrLDA but with the constraint that words from a single comment share the same topic derived from the news article (*i.e.*, the master document). The same constraint is imposed to the proposed EXTM model. However, different from MSTM, EXTM allows a comment to have topic derived from not only the news article but also other comments in the collection by using a switch. The switch in our model is at document-level not at tag-level as in topic-perspective model. Another major difference between EXTM and topic-perspective model is that in EXTM we do not model user information due to the aforementioned reason.

Short Text Clustering. Most studies on short text clustering focus on clustering search result snippets. A detailed survey is given in [8]. One key research issue is the similarity measure between two short texts due to the word sparsity. The state-of-the-art approach is to enrich short text representation by using external knowledge bases like WordNet and Wikipedia [9, 15, 27]. In [9], a Topic Model-based approach is proposed to derive multiple granularity latent topics to enrich short text representation. A graph-based approach for short text clustering is reported in [31]. Each snippet is represented as a graph of topics and the clusters are realized by graph-cut. On the other hand, a transfer learning approach is proposed in [17] to cluster tweets (short text with limited 140 characters in Twitter) by utilizing topically related long texts as the auxiliary data. In these studies, short texts (*e.g.*, search snippets) are independent from each other. In our problem setting, a comment is associated with a master document (*i.e.*, news article). Moreover, the topic distribution among the comments of a news article is largely affected by the topics discussed in the article.

3. TORCS PROBLEM AND MODELS

In this section, we first describe the Torcs problem and then present two generative models that extend LDA. The models integrate both generative process of news article content and generative process of comment content.

3.1 TORCS Problem

Given a news article and its associated comments received from a news stream, the task of *Topic-driven Reader Comment Summarization* is to generate a summary of comments in the form of comment clusters, where (i) comments in each cluster are on one major topic, and (ii) each comment cluster is represented by a few salient comments.

Note that, different from most existing document summarization tasks where a summary is usually a set of sentences extracted or generated from the given document(s), our comment summary is a subset of comments because a comment is usually very short (see Figure 3). Consequently, the two major challenges in addressing the TORCS problem are (i) identifying the topics discussed among comments, and (ii) identifying the most representative comments from each topical comment cluster.

In our problem setting, a comment exists only if it is associated with a news article. Therefore, in our discussion, we also refer to a news article as a *master document* and its associated comments *slave documents*. Further, a master document and its associated slave documents can be considered as “a unified document with star structure²”, shown in Figure 1.

Recall that the first challenge is to identify the topics discussed among the comments of a given news article. The most straightforward approach is to cluster them based on content similarity (e.g., cosine similarity over word features of the comments). However, this simple approach largely ignores the context information of the comments (e.g., related news or comments). Considering all news articles received from a news stream, many news articles are related to each other (e.g., reporting the same event, or events about the same personal and so on); so are their comments. In fact, grouping news articles about the same event or topic is one of the major research issue in topic detection and tracking (TDT) [1]. Between news articles and comments, we observe at least three kinds of relationships: *news-news*, *comment-comment*, and *news-comment* relationship, illustrated in the top left corner of Figure 1.

News-news: Two news articles are related if they report the same event (e.g., USA presidential election); a series of news articles are often published to trace the status of an event. News articles may also refer to recently published articles if reporting similar type of events (e.g., traffic accidents, natural disaster).

Comment-comment: There are two types of comment-comment relationships: *intra-comment-comment* relationship to relate the comments of the same news article (e.g., reply or on similar topics), and *inter-comment-comment* relationship to relate comments of different news articles. That is, readers may extend the discussion in comments to refer to the event(s) reported in multiple news articles including the current article she is commenting on. In other words, such a comment can be applied to any of the news articles reporting the event. For example, a reader may give her comment to support a candidate/party in the election. Such comments are strongly related to each other but not closely related to any one particular news article.

News-comment: Similarly, there are two types of news-comment relationships: *intra-news-comment* relationship to relate the comments of its associated news article, and *inter-news-comment* relationship to relate the comments of one news article to another news article. That is, a user may comment the current news article based on the information reported in another news article.

²Note that, in this paper we do not consider the reply relationship. It is a part of our future work to incorporate reply relationship in our models.

Based on the three types of relationships, we argue that in order to accurately identify the topics discussed among the comments associated with a news article, one has to consider the context information (e.g., the related news articles/comments). In this paper, we propose two topic models to identify the topics from recently published (e.g., within the past two weeks) news and comments. Given a news article, whose comments are to be summarized, we first assign a topic to each of its comments and then group the comments by the assigned topics to form topic clusters. From each topic cluster, the most representative comments are then identified to form the comment summary. These two steps are illustrated in the lower part of Figure 1.

In the following, we present the two topic models, Master-Slave Topic Model (MSTM) and Extended Master-Slave Topic Model (EXTM). The former implicitly models the news-news, intra-news-comment, intra-comment-comment relationships. EXTM extends MSTM by modeling the remaining two types of relationships: inter-news-comment and inter-comment-comment. The graphical representation of the two models are shown in Figure 2 and the notations are summarized in Table 1.

3.2 Master-Slave Topic Model (MSTM)

We aim to discover latent topics in slave documents (comments) with respect to their associated master documents (news articles). In a generative model, the generation of slave document relies on its associated master document, such as the generative process in corrLDA [6].

Similar to corrLDA, our proposed Master-Slave Topic Model assumes that all topics of the slave documents are directly derived from topic distribution of master document. That is, the generative process of the slave documents solely depends on the generative process of master document. However, due to the different vocabularies used in news articles and comments, different words for news articles and comments respectively could be generated for the same topic. We therefore name the same topic *master topic* and *slave topic* when we refer to the topic in the generative process of master and slave documents respectively.

A major difference between MSTM and corrLDA is that MSTM imposes a constraint that each slave document has exactly one slave topic. Recall that most comments are very short. Plotted in Figure 3(c), among one million comments in our dataset, more than 60% of comments each has fewer than 10 words and more than 80% of comments each has fewer than 16 words. Probably because of the length, we observe that most comments are specific to one point. We therefore constrain that each slave document has exactly one slave topic.

The graphical representation of MSTM is shown in Figure 2(a) and the generative process is summarized in Procedure 1. In MSTM, each master document has multiple master topics, and the generative process of the master document is same as in LDA (Lines 4-6 in Procedure 1). MSTM generates the unique topic of each slave document based on uniform word topic distribution of its associated master document, and then generates all words of slave document given the unique slave topic (Lines 7-9).

The exact inference is intractable. We therefore utilize Gibbs Sampling (an approximate approach) to inference the latent variables. Similar to the Collapsed Gibbs Sampling algorithm presented in [12], we construct a Markov chain on latent topics. At each transaction step, a topic of word from master document or a topic of slave document is generated with respect to its conditional probability.

The topic of word from master document is drawn from the following conditional probability, which is the same as LDA.

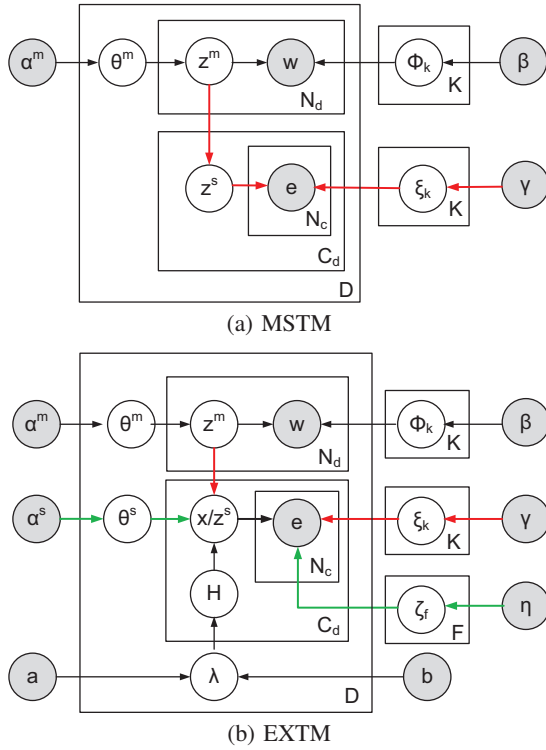


Figure 2: MSTM and EXTM. The red line illustrates slave topics are derived from news articles; the green line illustrates the extended topics are derived from comments.

$$P(z_i^m = k | w_i = v, z_{-i}^m, \alpha^m, \beta) \propto \frac{Q_{kd, \sim i}^{KD} + \alpha^m}{\sum_{k'} Q_{k'd, \sim i}^{KD} + K \alpha^m} \times \frac{Q_{vk, \sim i}^{VK} + \beta}{\sum_{v'} Q_{v'k, \sim i}^{VK} + V \beta}$$

The topic of slave document is drawn from conditional probability:

$$P(z_c^s = k | z_{-c}^m, e_{-c}, \gamma) \propto \frac{Q_{kd}^{KD}}{\sum_{k'} Q_{k'd}^{KD}} \times \frac{\prod_{e_j \in E_c^d, e_j = u_j} \prod_{g=0}^{N_{dc}^e - 1} (Q_{u'jk, \sim c}^{UK} + \gamma + g)}{\prod_{y=0}^{N_{dc}^e - 1} (\sum_{u'} Q_{u'k, \sim c}^{UK} + U \gamma + y)}$$

3.3 Extended Master-Slave Topic Model (EXTM)

Extended Master-Slave Topic Model aims to simulate the generative process of both master document and slave document in a more natural way. MSTM follows a rigid constraint that all topics of slave documents have to be derived from the topic distribution of their master documents. EXTM relaxes this rigid constraint and allows some slave documents to have topics irrelevant to their master documents. We call this type of topic *extended topic*. As discussed earlier, a reader is not constrained by the topic presented in the news article when commenting. She may provide some additional information (e.g., background information about a news article) which is not covered in the news article. Extended topics are introduced in the EXTM model to take care of such topics.

The graphical representation of EXTM is shown in Figure 2(b) and the generative process is summarized in Procedure 2. In EXTM, the generative process of master document is same as that in MSTM (Lines 5-7 in Procedure 2). Similar to MSTM, each slave document has one unique topic. However, the topic can be either derived from topic distribution of its associated master document, or from extended topic distribution of all comments. We use a switch H to decide whether the topic is a slave topic or an extended topic (Line

Table 1: Notations for MSTM and EXTM

D, C	Number of master documents, of slave documents
N_d, N_c	Number of words in master document d , in slave document c
V, U	Vocabulary size of master documents, of slave documents
w, e	Word in master document, in slave document
E_c^d	Set of words in slave document c of master document d
N_{dc}^e	Number of times word e occurs in slave document c of master document d
K	Number of master topics
F	Number of extended topics
z^m, z^s, x	Master topic, slave topic, and extended topic
θ^m, θ^s	Multinomial distribution over master topics, over extended topics
α^m, α^s	Dirichlet prior vector for θ^m , and θ^s
ϕ_k, ξ_k, ζ_f	Multinomial distribution over words of master document for master topics, over words of slave document for slave topics, over words of slave document for extended topics
β, γ, η	Dirichlet prior vector for ϕ_k, ξ_k , and ζ_f
H	Switch to decide whether the topic is a slave topic or an extended topic
λ, a, b	Probability of topic of slave document being slave topic, and beta prior for λ
$Q_{kd, \sim i}^{KD}$	Number of times master topic k is assigned to master document d , excluding the current word w_i
$Q_{vk, \sim i}^{VK}$	Number of times word v of master document is generated from master topic k , excluding the current word w_i
$Q_{uk, \sim c}^{UK}$	Number of times word u of slave document is generated from slave topic k , excluding all words in slave document c
$Q_{uf, \sim c}^{UF}$	Number of times word u of slave document is generated from extended topic f , excluding all words in slave document c
$Q_{H=1, \sim c}^{CD}$	Number of times H is 1 in document d , excluding the current slave document c
$Q_{H=0, \sim c}^{CD}$	Number of times H is 0 in document d , excluding the current slave document c
$Q_{fd, \sim c}^{FD}$	Number of times extended topic f is assigned to slave documents of master document d , excluding the current slave document c

10 and 13). If H equals 1, words of slave document are drawn from the slave topic sampled from uniform word topic distribution of its associated master document (Lines 11 and 12); if H equals 0, words of slave document are drawn from the extended topic sampled from extended topic distribution θ^s (Lines 14 and 15).

Similar to that in MSTM, we utilize Gibbs Sampling to inference the latent variables of EXTM. Because the sampling topic of words from master document is the same as MSTM, we only present sampling topic of slave document. Reported in Table 2, the topic of slave document is drawn from conditional probability $P(H = 1, z_c^s = k | z_{-c}^m, e_{-c}, \gamma, a, b)$ when $H = 1$, and drawn from probability $P(H = 0, x_c = f | x_{-c}, e_{-c}, \alpha^s, \eta, a, b)$ when $H = 0$.

3.4 Salient Comments Selection

Because of the assumption that each comment has exactly one topic, topic clustering becomes straightforward by grouping the comments of the same topic together. The importance of each comment cluster is evaluated by the number of comments contained in the cluster. The next major task is to select a few most salient comments from each cluster for a concise representation of each comment cluster. Considering each comment is analogous to a sentence because of its short length, most sentence ranking/selection techniques studied in document summarization can be directly applied. We evaluate two comment selection techniques, namely, Maximal Marginal Relevance (MMR) and Rating & Length (RL).

Maximal Marginal Relevance. MMR [7] is a commonly used technique in document summarization to select topically relevant sentences while avoiding redundancy. Given a cluster of comments

Table 2: Conditional probability of topic of slave document in EXTM

$$P(H = 1, z_c^s = k | \mathbf{z}_{-c}^m, \mathbf{e}_{-c}, \gamma, a, b) \propto \frac{Q_{H=1, -c}^{CD} + a}{Q_{H=1, -c}^{CD} + Q_{H=0, -c}^{CD} + a + b} \times \frac{Q_{kd}^{KD}}{\sum_{k'} Q_{k'd}^{KD}} \times \frac{\prod_{e_j \in E_c^d, e_j = u_j} \prod_{g=0}^{N_{dc}^j - 1} (Q_{u_j k, -c}^{UK} + \gamma + g)}{\prod_{y=0}^{N_{dc} - 1} (\sum_{u'} Q_{u' k, -c}^{UK} + U\gamma + y)}$$

$$P(H = 0, x_c = f | \mathbf{x}_{-c}, \mathbf{e}_{-c}, \alpha^s, \eta, a, b) \propto \frac{Q_{H=0, -c}^{CD} + b}{Q_{H=1, -c}^{CD} + Q_{H=0, -c}^{CD} + a + b} \times \frac{Q_{fd, -c}^{FD} + \alpha^s}{\sum_{f'} Q_{f'd, -c}^{FD} + F\alpha^s} \times \frac{\prod_{e_j \in E_c^d, e_j = u_j} \prod_{g=0}^{N_{dc}^j - 1} (Q_{u_j f, -c}^{UF} + \eta + g)}{\prod_{y=0}^{N_{dc} - 1} (\sum_{u'} Q_{u' f, -c}^{UF} + U\eta + y)}$$

Procedure 1: Generative process of MSTM

- 1 For each of the D master documents d , draw $\theta_d^m \sim \text{Dirichlet}(\alpha^m)$;
 - 2 For each of the K master topics k , draw $\phi_k \sim \text{Dirichlet}(\beta)$;
 - 3 For each of the K slave topics k , draw $\xi_k \sim \text{Dirichlet}(\gamma)$;
 - 4 **for** each word w_i in master document d **do**
 - 5 Choose a master topic $z_i^m \sim \text{Multinomial}(\theta_d^m)$;
 - 6 Generate a word $w_i \sim \text{Multinomial}(\phi_{z_i^m})$;
 - 7 **for** each slave document c of master document d **do**
 - 8 Choose a slave topic $z_c^s \sim \text{Uniform}(z_{w_1}^m, \dots, z_{w_{N_d}}^m)$;
 - 9 Generate all words in $c \sim \text{Multinomial}(\xi_{z_c^s})$;
-

to be summarized, we compute its centroid feature vector, denoted by c_m , to be the query vector whose relevance is computed for each of the member comments in the cluster using MMR. Let C_s be the set of comments selected so far, the score of a candidate comment c to be selected is computed using the following equation. In our experiments, we use cosine similarity to define $\text{sim}(\cdot, \cdot)$ and set $\delta = 0.8$.

$$\text{Score}(c) = \delta \times \text{sim}(c, c_m) - (1 - \delta) \times \max_{c' \in C_s} (\text{sim}(c, c'))$$

Because of the short length, we do not use $TF \times IDF$ based word weighting scheme to define the comment feature vectors. Instead, words in comments are weighed based on the topic distribution generated from ξ_k or ξ_f (recall that in topic modeling, each topic is a probability distribution over words).

Rating & Length. Instead of solely relying on content of comments, with RL, we utilize the ratings from readers in comment selection. Rating has always been considered important features in comments [10, 13]. In Yahoo! News website, readers evaluate others' comments by clicking "like" or "dislike" button. The more likes or dislikes a comment receives, the more attention it gains from the readers. We sum the likes and dislikes a comment receiving to be its rating score. In addition, we also consider the length of the comments in number of words. It has been reported that comment length as a feature can be utilized to improve prediction accuracy of ranking comments [13]. We normalize the rating scores $\text{Rating}(\cdot)$ and length scores $\text{Length}(\cdot)$ in $[0, 1]$ using min-max normalization, and then rank comments by the linear combination of the two scores, where σ is a factor to balance rating score and length score. In our experiments, we set $\sigma = 0.6$ to give more weights to the rating score.

$$\text{Score}(c) = \sigma \times \text{Rating}(c) + (1 - \sigma) \times \text{Length}(c)$$

4. EXPERIMENTS

To evaluate the proposed solution for Torcs problem, we conducted experiments using the most-commented news articles and

Procedure 2: Generative process of EXTM

- 1 For each of the D master documents d , draw $\theta_d^m \sim \text{Dirichlet}(\alpha^m)$ and draw $\theta_d^s \sim \text{Dirichlet}(\alpha^s)$;
 - 2 For each of the K master topics k , draw $\phi_k \sim \text{Dirichlet}(\beta)$;
 - 3 For each of the K slave topics k , draw $\xi_k \sim \text{Dirichlet}(\gamma)$;
 - 4 For each of the F extended topics f , draw $\zeta_f \sim \text{Dirichlet}(\eta)$;
 - 5 **for** each word w_i in master document d **do**
 - 6 Choose a master topic $z_i^m \sim \text{Multinomial}(\theta_d^m)$;
 - 7 Generate a word $w_i \sim \text{Multinomial}(\phi_{z_i^m})$;
 - 8 **for** each master document d , draw $\lambda_d \sim \text{Beta}(a, b)$ **do**
 - 9 For each slave document c , draw a switch $H \sim \text{Binomial}(\lambda_d)$;
 - 10 **if** $H = 1$ **then**
 - 11 Choose a slave topic $z_c^s \sim \text{Uniform}(z_{w_1}^m, \dots, z_{w_{N_d}}^m)$;
 - 12 Generate all words in $c \sim \text{Multinomial}(\xi_{z_c^s})$;
 - 13 **else** $/* H = 0 */$
 - 14 Choose an extended topic $x_c \sim \text{Multinomial}(\theta_d^s)$;
 - 15 Generate all words in $c \sim \text{Multinomial}(\zeta_{x_c})$;
-

their comments from Yahoo! News. We conducted two sets of evaluations, to evaluate the proposed MSTM and EXTM models and to evaluate the resultant summaries, respectively. In the following, we detail the dataset and report the evaluation of the two topic models using perplexity. The resultant summaries are evaluated through a user study in Section 5.

4.1 Data Set

To obtain news articles with reasonable number of comments, we crawled the top-100 most-commented news articles from Yahoo! News³ on daily basis, from 29 Apr 2012 to 11 May 2012. After duplicate removal, our dataset has 1005 unique news articles. The preprocessing of the news articles and comments includes word stemming, removal of stopwords, removal of words appeared fewer than 3 times in the collection.

The number of comments per article is plotted in Figure 3(a). On average, each news article has 1059 comments with the minimum and maximum number of comments per article being 65 and 18,758 respectively. The total number of comments in our dataset is 1,064,655, slightly above 1 million. Plotted in Figures 3(b) and 3(c), most comments are short with fewer than 10 words. Nevertheless, the vocabulary size of comments is 66,064 while the vocabulary size of news articles is 21,533, about one-third that of comments. That is, comments often extend topics of news articles and bring additional words to the vocabulary.

³<http://news.yahoo.com/all-sections--most-commented/most-popular/1.html>

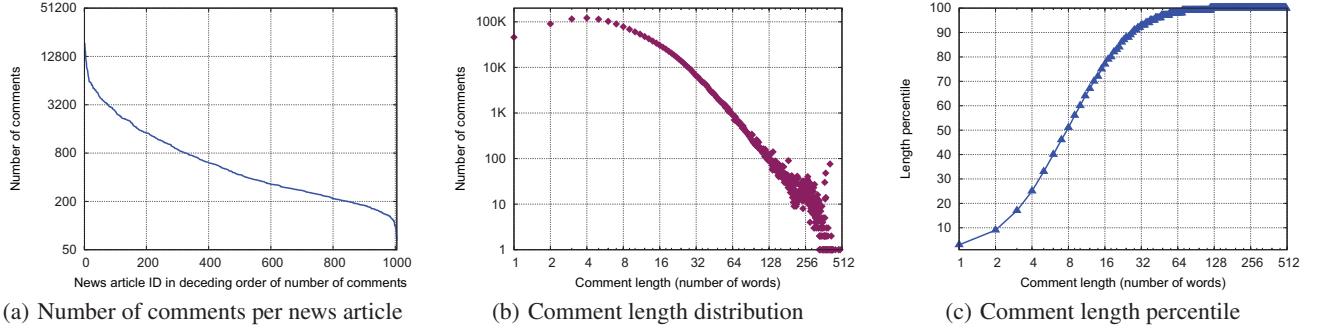


Figure 3: Comments per news article and comment length distribution

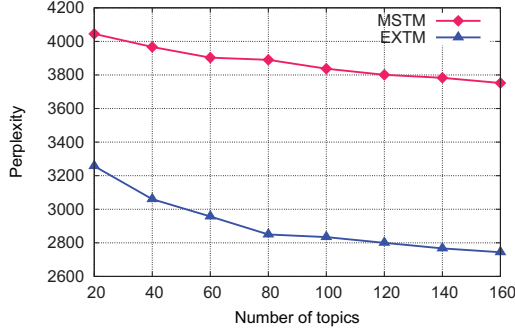


Figure 4: MSTM and EXTM with different number of topics. In MSTM, the number of topics equals the number of slave topics; in EXTM, number of topics is the sum of number of slave topics and extended topics.

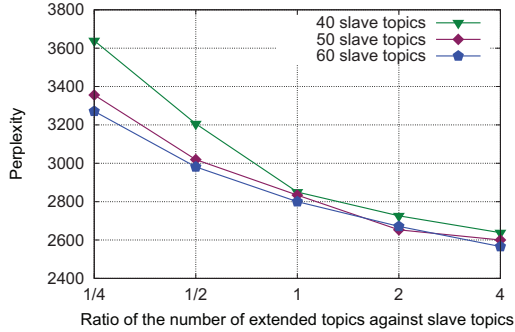


Figure 5: EXTM with different number of extended topics.

4.2 Evaluation

Because both master and slave documents in our dataset are not manually labeled, we adopt a type of density estimation – perplexity [5], which is the standard metric for evaluating topic models. It measures model ability of generalizing unseen data; lower perplexity indicates larger likelihood and better model performance. Perplexity is defined in the equation below where D_{test} indicates the set of test documents (*i.e.*, the held-out set not used in training the model), and $p(E_c^d)$ indicates the probabilities of words in slave document c of master document d .

$$Perplexity(D_{test}) = \exp\left\{-\frac{\sum_{d=1}^{D_{test}} \sum_{c=1}^{C_d} \log p(E_c^d)}{\sum_{d=1}^{D_{test}} \sum_{c=1}^{C_d} N_{dc}}\right\}$$

In many cases, topic model is sensitive to the choices of hyper-parameters. However, in our experiments, through extensive evaluations, we observe that both our MSTM and EXTM models are not very sensitive to the hyper-parameters. In our following evaluation, we fix their settings: $\alpha^m = 0.1$, $\alpha^s = 0.1$, $\beta = 0.01$, $\gamma = 0.01$, $\eta = 0.01$, $a = 0.5$ and $b = 0.5$. We calculate the perplexity of the two models using 10% held-out test set (100 documents) and the results reported are averaged over 3 runs.

In general, the larger the number of topics defined, the better the generative performance for topic model. We therefore compared MSTM and EXTM by setting the same number of the topics for both models. More specifically, the number of slave topics in MSTM was the sum of the number of slave topics and number of extended topics in EXTM. Note that, in EXTM, we also set the number of slave topics same as the number of extended topics. Plotted in Figure 4, EXTM outperforms MSTM for all numbers of topics evaluated from 20 to 160. Observe that the differences in perplexities of the two models grow along the increment of number of topics. This result suggests that in EXTM, allowing words in comments to be generated from not only slave topics but also extended topics enhances the model generalization ability and guides the words of the same topic to cluster together with higher probability.

In the next set of evaluation, we evaluated the effect of setting different number of slave topics and extended topics. We fixed the number of slave topics to be 40, 50 and 60 respectively, then set the number of extended topics to be 1/4 to 4 times of slave topics. The perplexity result is plotted in Figure 5. Observe that by doubling the number of extended topics for a fixed number of slave topics (*e.g.*, 40), the perplexity drops significantly and the slope of dropping is larger when the number of extended topics is small. Differences in perplexities between curves for any two fixed number of slave topics become smaller with increasing number of extended topics. This is because there exists an inherent number of extended topics in the dataset, and setting a larger number does not benefit the model performance much.

4.3 Topic Discovery

We now present example topics discovered by EXTM. The example topics presented are generated by setting both the number of master/slave topics and number of extended topics to be 50. Due to page limit, we randomly select 8 master topics from the 50 topics derived from news articles and list their top words in the order of corresponding probabilities ($p(w|z^m)$) in Table 3 (see the second column). Observe that EXTM discovers comprehensive and distinctive master topics: Topic 10 is about presidential campaign in US; Topic 23 refers to Obama’s speech on the death of Bin Laden;

Table 3: Example topical keywords for master topics and their corresponding slave topics and extended topics.

ID	Master Topic	Slave Topic	Extended Topic
0	police protest people day immigrate illegal officer report building arrest occupy law sheriff court movement	people protest occupy ows job police time movement day america street riot idiot try property	illegal law america country mexico people joe obama border immigrate sheriff enforce government job immigrant
8	job time people month add april report hire times twitter launch suggest company plan look	job america people pay union country rich class money republican worker china business middle corporation	job unemployment obama people rate look month million economy america drop unemployed create count report
9	zimmerman martin death florida prosecutor police jones charged riot website king black people murder march university	people death kill murder martin guy zimmerman life prison police justice trial crime law america	black white people sharpton zimmerman race racist riot media martin trayvon violence jackson hate kill america racial
10	romney obama campaign republican president mitt candidate presidential former politics primary voter gop democratic nominee	obama romney bush america president job republican mitt vote country policy people economy gop tax war debt money	romney obama mccain mitt palin vote president endorse pick republican running sarah gop john
23	afghanistan laden bin pakistan obama attack kill america president war osama release raid house death	obama bin laden bush president kill osama credit seal romney america killing terrorist war republican	obama bin bush laden america president war afghanistan troops kill iraq osama people country romney
25	church smith million god mormon fox company scout international kimmel rupert catholic belief dinner faith religion religious	god jesus sin bible christ people lord love christian homosexual believe obama gods gay life	people god religion muslim christian believe bible islam world jesus church america religious mormon hate
27	cut rate tax billion spend debt vote student percent economy loan increase million budget business	tax pay cut rich job income obama romney rate class people bush money america spend	money tax pay spend people obama government cut america dollar country debt war military billion
48	police found car told time inside officer death die investigate hit injury gun son vehicle	shot fire story bullet gun hole wife kill family shoot warning attack happen die husband	road car people driver driving kid bear happen family accident time stupid sunbathe child drive

Table 4: Least relevant extended topics derived from comments

ID	Extended topic keywords
20	animal wild cheetah people attack pet cat pictures taking tame human stupid try maul play
11	ticket trash threw judge money lottery woman store winner found lady throw winning claim sign
15	family god bless parent little prayer heart avery love life sad story time hope pray
4	kid school parent bully child girl people children little time grade sexual teacher suspend home
6	girl scout boys play women team priest school bishop little people bar rape boy own
47	eat shrimp food kfc people chicken china baby asia chip time tiger fast america fry
24	yahoo story article comment news read people post time report reading articles picture video headline
13	movie bond character avenge love movies people film animal inmate craig james watch prison time

Topic 27 is on tax policy and Topic 48 focuses on car accident. All these master topics are related to hot events during the crawling period. This result suggests that EXTM can mine semantically meaningful and distinctive master topics from news articles.

To illustrate the relationship between the master topics, slave topics, and extended topics. Columns 3 and 4 in Table 3 report the most relevant⁴ slave topic and extended topic for the 8 selected master topics. Comparing master topics and their corresponding slave topics and extended topics, we observe that EXTM model captures strong relationships between the three types of topics. Consider master topics and slave topics first. The table shows that master topics and their corresponding slave topics share a large number of common keywords. For instance, both master and slave topics for Topic 10 are about presidential campaign; that for Topic 23 are about the death of Bin Laden. Recall that the size of slave document vocabulary is three times of the size of master document vocabulary. This result suggests the generative process of slave documents is under control of that of master documents to a large extent. Now we consider the extended topics. Recall that extended

topics can be considered as extensions of master topics from another perspectives. Take Topic 9 as an example, both master and slave topics are closely related to the event itself⁵ while the extended topic captures the discussion on alleged race issues (e.g., black, white, race, racist, and racial). Similarly, the extended topic associated with Topic 8 concentrates more on unemployment when talking about job issue.

To further illustrate that some extended topics are not specific to any particular news article, in Table 4, we list the 8 extended topics which are least relevant to master topics. Observe that most these 8 topics are casual discussions and not specific to some events happened during the crawling period. In particular, Topic 24 is about Yahoo! News in general (e.g., the selection of its news articles and the quality of services); Topic 15 is more on reader sentiment on news stories.

5. USER STUDY

In this section, we report our user study on the evaluation of the summaries generated by the proposed methods. Document sum-

⁴The topic relevance is determined by cosine similarity over the word-feature-vector based on the multinomial distributions of the words for topics.

⁵http://en.wikipedia.org/wiki/Shooting_of_Trayvon_Martin

marization is commonly evaluated against a ground truth summary (e.g., a human generated summary or a set of manually-labeled sentences). In our setting, however, generating such ground-truth labeling becomes extremely difficult because of the large number of comments to each news article (i.e., on average 1059 comments per news article). Moreover, *topic* is not predefined making it even harder to manually label the comments. We therefore choose to evaluate the summaries generated by the different methods for a subset of news articles.

Methods. By applying 3 clustering methods CLUTO, MSTM and EXTM and 2 ranking methods RL and MMR, we evaluated 6 methods in total.

We evaluated CLUTO⁶ as an alternative method without using topic modeling. The bisecting clustering algorithm (with default parameter settings) provided by CLUTO was used to cluster comments of a news article due to its high efficiency and promising performance. The number of clusters was set to 50. Note that bisecting clustering was applied independently on slave documents of each master document. We use CLUTO-MMR and CLUTO-RL to denote the comment summaries selected by MMR and RL respectively, based on the clusters generated by CLUTO. Note that we used $TF \times IDF$ based word weighting scheme to define the comment feature vectors in CLUTO-MMR due to the lack of word topic weights.

For the two topic models MSTM and EXTM, the comments are clustered based on the topics discovered by the corresponding model. The number of topics for MSTM was set to 100. Correspondingly, we set the number of slave topics to be 50 and the number of extended topics to be 50 for EXTM. Similarly, we use postfixes -MMR and -RL to distinguish the MMR and RL comment selection method.

Evaluation Setting and Criteria. We randomly selected 50 articles from the 1005 news articles for user evaluation. To make the manual evaluation feasible, we fixed the number of comments selected in a comment summary to be 15 for any given article. More specifically, given all comments of a news article, a summarization method first selects the top-3 largest clusters (by number of comments) to represent the 3 major topics discovered from the comments. From each cluster, 5 most representative comments are selected to represent that topic. In short, a comment summary for a news article contains exactly 15 comments covering 3 major topics, by any of the 6 methods. To evaluate the 6 methods, a volunteer therefore needs to read $50 \times 15 \times 6 = 4500$ comments in total.

For each comment summary generated by a method, a volunteer gives a rating from 1 (lowest) to 5 (highest) to each of the following three criteria used in our evaluation.

- *Topic cohesion*: To what extent the 5 comments in each comment cluster are topically cohesive.
- *Topic diversity*: To what extent the 3 topic clusters cover 3 different topics.
- *News relatedness*: To what extent the 3 topic clusters are related to the content of the news article.

Evaluation Results. The ratings from three volunteers (undergraduate students in computer science major) are averaged by the method producing the comment summary, reported in Figure 6. In addition to each of the three criteria, we also report the average rating across the three criteria in the last column of the bar chart. We make three observations.

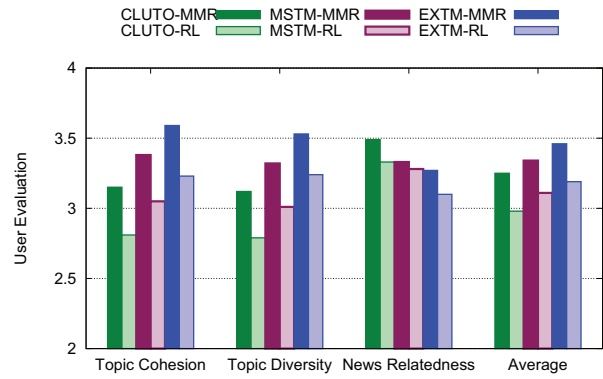


Figure 6: User Evaluation

First, across all three clustering methods, MMR ranking outperforms RL ranking. A possible reason is that RL suffers from rating sparsity and its resultant ranking by length is biased to long (and maybe even redundant) comments, which hurts topic cohesion and topic diversity. Conversely, MMR tries to select comments that are closest to the centroid of a comment cluster but with different word distributions (due to its redundancy control). In particular, EXTM-MMR yields the best topic cohesion rating of 3.59 and the best topic diversity rating of 3.52. It outperforms CLUTO-MMR by 14% on topic cohesion and by 13% on topic diversity, respectively.

Second, we observe that comparing with traditional clustering approach CLUTO, topic model based methods are more likely to generate clusters with better intra-cohesive and inter-diverse topics. However, CLUTO based methods yield the best news relatedness. This is because that CLUTO is independently applied on the comments of every news article. The clustering is therefore not affected by any additional (or contextual) information provided in other news articles/comments. Both MSTM and EXTM based methods capture such contextual information by definition. Nevertheless, such contextual information may not be known to our volunteers due to the random selection of the news articles for evaluation; hence the news relatedness of topic model based methods are rated lower.

Third, between MSTM and EXTM, EXTM-based methods outperform MSTM-based methods on topic cohesion and topic diversity, but are slightly poorer on news relatedness. The latter is probably because EXTM discovers extended topics discussed among comments only, but not in news articles. Considering all evaluation criteria of equally importance (i.e., the average of the three), EXTM-MMR is the best performing method among the 6 methods evaluated.

6. CONCLUSION

In this paper, we first define the topic-driven reader comments summarization problem. The main objective of the research is to present to users a summary of comments covering the major topics discussed among the readers. After analyzing the important relationships between news and comments, we propose two generative models, namely, Master-Slave Topic Model and Extended Master-Slave Topic Model respectively, to implicitly model these relationships. Both models treat news article as master document and comment slave document and assume each slave document has exactly one topic. While MSTM constrains that the topic of a slave document has to be derived from its master document, EXTM allows the topic of a slave document derived from all slave documents. Evaluated using perplexity, in our experiments EXTM significantly outperforms MSTM for grouping comments with the same top-

⁶<http://glaros.dtc.umn.edu/gkhome/views/cluto>

ics. To generate the comment summary, we utilize two comment ranking schemes MMR and RL to select the most representative comments from each comment cluster. Our user study shows that EXTM-MMR outperforms the other 5 methods by topic-cohesion and topic-diversity measures.

Both our proposed models ignore the reply relationship among comments which provides rich structure information about comments. In our future work, we plan to incorporate reply relationship in our models to take advantage of the rich structural information in comments. We also observe that the topics discovered might be affected by some extremely long comments. To improve our model for better handling long comments is also part of our future work.

7. REFERENCES

- [1] J. Allan, editor. *Topic detection and tracking*. Kluwer Academic Publishers, 2002.
- [2] D. M. Blei and M. I. Jordan. Modeling annotated data. In *SIGIR*, pages 127–134, 2003.
- [3] D. M. Blei and J. D. Lafferty. Dynamic topic models. In *ICML*, 2006.
- [4] D. M. Blei and J. D. Lafferty. A correlated topic model of science. *AAS*, 1(1):17–35, 2007.
- [5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.
- [6] M. Bundschuh, S. Yu, V. Tresp, A. Rettinger, M. Dejori, and H.-P. Kriegel. Hierarchical bayesian models for collaborative tagging systems. In *ICDM*, pages 728–733, 2009.
- [7] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336, 1998.
- [8] C. Carpineto, S. Osipiński, G. Romano, and D. Weiss. A survey of web clustering engines. *ACM Comput. Surv.*, 41(3):17:1–17:38, 2009.
- [9] M. Chen, X. Jin, and D. Shen. Short text classification improved by learning multi-granularity topics. In *IJCAI*, pages 1776–1781, 2011.
- [10] O. Dalal, S. H. Sengemedu, and S. Sanyal. Multi-objective ranking of comments on web. In *WWW*, pages 419–428, 2012.
- [11] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *WWW*, pages 519–528, 2003.
- [12] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(suppl. 1):5228–5235, 2004.
- [13] C.-F. Hsu, E. Khabiri, and J. Caverlee. Ranking comments on the social web. In *Proc. Int’l Conf. on Computational Science and Engineering (Vol. 04)*, pages 90–97, 2009.
- [14] M. Hu, A. Sun, and E.-P. Lim. Comments-oriented document summarization: understanding documents with readers’ feedback. In *SIGIR*, pages 291–298, 2008.
- [15] X. Hu, N. Sun, C. Zhang, and T.-S. Chua. Exploiting internal and external semantics for the clustering of short texts using world knowledge. In *CIKM*, pages 919–928. ACM, 2009.
- [16] R. Jain and P. Sinha. Content without context is meaningless. In *ACM Multimedia*, pages 1259–1268, 2010.
- [17] O. Jin, N. N. Liu, K. Zhao, Y. Yu, and Q. Yang. Transferring topical knowledge from auxiliary long texts for short text clustering. In *CIKM*, pages 775–784, 2011.
- [18] Y. Jo and A. H. Oh. Aspect and sentiment unification model for online review analysis. In *WSDM*, pages 815–824, 2011.
- [19] R. Kant, S. H. Sengemedu, and K. S. Kumar. Comment spam detection by sequence mining. In *WSDM*, pages 183–192, 2012.
- [20] E. Khabiri, J. Caverlee, and C.-F. Hsu. Summarizing user-contributed comments. In *ICWSM*, 2011.
- [21] E. Khabiri, C.-F. Hsu, and J. Caverlee. Analyzing and predicting community preference of socially generated metadata: A case study on comments in the digg community. In *ICWSM*, 2009.
- [22] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *CIKM*, pages 375–384, 2009.
- [23] C. Lu, X. Hu, X. Chen, J.-R. Park, T. He, and Z. Li. The topic-perspective model for social tagging systems. In *KDD*, pages 683–692, 2010.
- [24] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *ACM HyperText’06*, pages 31–40, 2006.
- [25] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. In *WWW*, pages 101–110, 2008.
- [26] G. Mishne and N. Glance. Leave a reply: An analysis of weblog comments. In *In Third annual workshop on the Weblogging ecosystem*, 2006.
- [27] X. H. Phan, M. L. Nguyen, and S. Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *WWW*, pages 91–100. ACM, 2008.
- [28] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *UAI*, pages 487–494. AUAI Press, 2004.
- [29] M. Sahami and T. D. Heilman. A web-based kernel function for measuring the similarity of short text snippets. In *WWW*, pages 377–386, 2006.
- [30] J. San Pedro, T. Yeh, and N. Oliver. Leveraging user comments for aesthetic aware image search reranking. In *WWW*, pages 439–448, 2012.
- [31] U. Scaiella, P. Ferragina, A. Marino, and M. Ciaranita. Topical clustering of search results. In *WSDM*, pages 223–232, 2012.
- [32] A. Schuth, M. Marx, and M. de Rijke. Extracting the discussion structure in comments on news-articles. In *WIDM*, pages 97–104, 2007.
- [33] E. Shmueli, A. Kagian, Y. Koren, and R. Lempel. Care to comment?: recommendations for commenting on news stories. In *WWW*, pages 429–438, 2012.
- [34] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. How useful are your comments?: analyzing and predicting youtube comments and comment ratings. In *WWW*, pages 891–900, 2010.
- [35] I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *WWW*, pages 111–120, 2008.
- [36] M. Tsagkias, W. Weerkamp, and M. de Rijke. Predicting the volume of comments on online news stories. In *CIKM*, pages 1765–1768, 2009.
- [37] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis on review text data: a rating regression approach. In *KDD*, pages 783–792, 2010.
- [38] W. G. Yee, A. Yates, S. Liu, and O. Frieder. Are web user comments useful for search. In *Workshop on Large-Scale Distributed Systems for Information Retrieval (LSDS-IR) at SIGIR’09*, pages 63–70, 2009.