ABSTRACT
Timeline Generation aims at summarizing news from different epochs and telling readers how an event evolves. It is a new challenge that combines salience ranking with novelty detection. For long-term public events, the main topic usually includes various aspects across different epochs and each aspect has its own evolving pattern. Existing approaches neglect such hierarchical topic structure involved in the news corpus in timeline generation. In this paper, we develop a novel time-dependent Hierarchical Dirichlet Model (HDM) for timeline generation. Our model can aptly detect different levels of topic information across corpus and such structure is further used for sentence selection. Based on the topic mined from HDM, sentences are selected by considering different aspects such as relevance, coherence and coverage. We develop experimental systems to evaluate 8 long-term events that public concern. Performance comparison between different systems demonstrates the effectiveness of our model in terms of ROUGE metrics.

Categories and Subject Descriptors
I.7.5 [Document Capture]: [Document Analysis]

General Terms
Design, Algorithms, Theory, Experimentation

Keywords
Timeline generation, Dirichlet Process, Topic Model

1. INTRODUCTION
Faced with thousands of news articles, readers usually try to ask the general questions such as the beginning, the evolutionary pattern and the end of an event. General search engines simply return top ranking articles according to query relevance and fail to trace how a specific event goes. Even though the documents can be ranked in a chronological order, readers are tired of reading the overwhelming number of documents in details. Timeline generation gives an ideal solution to this problem by providing readers with a faster and better access to understand news.

Timeline generation for an evolving topic cares for not only the salience, but also the novelty and evolutionary process of the events. In this paper, we focus on extractive approach for timeline generation, which tries to select desired sentences from news corpus. There are two main challenges for long-term timeline generation. First, for extractive methods, we want to select the sentences that can best generalize topics at each time, so how to model information hidden within the news corpus becomes important. As for long-term event, one main topic is usually described from various specific aspects and each aspect has its own evolutionary pattern. For example, the news about “Greek Debt-government Crisis” may involve different sub-events such as “the cause of debt crisis”, “Greek public protests”, “Rescue packages” while each of them contains different stages. Detecting such hierarchical structure is important for later sentence selection. Second, it is very common that the themes of a corpus evolve over time, and thus topics of adjacent time epochs usually exhibit strong correlations. How to incorporate time dependency into the hierarchical structure is also a hard problem.

The problem of timeline generation was firstly proposed by Allan et al. by extracting clusters of noun phases and name entities. Later they built up a system to provide timelines which consist of one sentence per date by considering usefulness and novelty. Chieu et al. built a similar system in unit of sentences with interest and burstiness. However, these methods seldom explored the evolutionary characteristics of news. Recently, Yan et al. extended the graph based sentence ranking algorithm used in traditional multi-document summarization (MDS) to timeline generation by projecting sentences from different times into one plane. They further explored the timeline task from the optimization of a function by considering the combination of different respects such as relevance, coverage, coherence and diversity. Time dependency is considered in Yan et al’s work. However, their approaches treat timeline generation as a sentence ranking or optimization problem and seldom explore the topic information hidden within the corpus or the structure of news information.

Topic detection or mining is not a new task and many algorithms have been proposed in TDT (Topic Detection and Tracking) task or summarization task. Among existing approaches, topic models such as Latent Dirichlet Allocation (LDA) or Hierarchical Dirichlet Pro-
cesses (HDP) have their advantages in capturing latent topics within document collection due to its clear probabilistic interpretations.

As we get back to the two challenges previously discussed, we find that topic models show promising properties to handle them: First, Blei et al. [15] introduced a hLDA model which organizes topics into a tree structure with depth L. Topics located close to each other in the tree tend to share a similar topic-word distribution and the children of a certain node tend to be treated as subtopics talking about minor aspects. Such tree structure can well help model the hierarchical topic structure in the long term event. One major problem hLDA is that LDA based models require users to specify a cluster number at each time epoch for each corpus, which is awkward in real case. In addition, the time-dependent nature of new corpus data cannot be addressed in hLDA model. Second, to deal with time-correlated data, many approaches have been proposed based on topic approaches, aiming to discover the evolving patterns in the corpus as well as the snapshot topics at each time epoch [1, 2, 3, 4, 5, 6, 8, 9, 10].

In this paper, we develop a novel topic approach denoted as time-dependent Hierarchical Dirichlet Model (HDM) for timeline generation. Our HDM model can be viewed as an extended combination of hLDA and multi-level topic-correlated Dirichlet Process (DP) and tries to model the topic distribution in a hierarchical tree structure. Such structure is named as a time-correlated Dirichlet process in both horizontal (how many children does a node have) and vertical (how deep is the tree) direction. We approach HDM by specifying a generative probabilistic model for hierarchical structure and taking a Bayesian perspective on the problem of learning the tree structure automatically from the data. The non-parametric nature of DP means that we do not have to know a priori the number of topics or depth as they will readily accommodate growing data collections. From that structure, we can clearly identify the evolution pattern of different aspects involved in the event by locating the topic position in the tree and further, based on that, we construct a timeline ranking framework by considering the relevance, coherence and coverage of each sentence. We build an experimental system on 8 real long-term events that public concern. The effectiveness of model is verified through ROUGE matrix.

The rest of paper is organized as follows. Section 2 describes related work. Section 3 briefly illustrates Dirichlet Process (DP) and Hierarchical Dirichlet Processes (HDP). The details of HDM is shown in Section 4. We present experimental results in Section 5 and conclude this paper in Section 6.

2. RELATED WORK

Timeline Generation.

The task of timeline generation is firstly proposed by Swan and Allan [28] by using named entities for sentence selection in different time epochs. Since then, researchers tried to generate timeline by considering different respects such as usefulness and novelty [3] or interest and burstiness [10]. Yan et al. [39] extended the graph based sentence ranking algorithm in multi-document summarization (MDS) to timeline generation by projecting sentences from different temporal corpus into one plane. Another work from Yan et al. [39] transforms this task to an optimization problem by considering the relevance, coverage and coherence of the sentences. Existing approaches seldom explore the topic information lied in the corpus or the structure of news information. Li et al. [19] exploited topic model to capture the dynamic evolution of topics in timeline generation.

Topic Models.

Recently, bayesian topic models such as LDA [6] or HDP [30] have shown the power text mining for its clear probabilistic interpretation. In topic model, each document is denoted as a mixture of different topics and each topic is presented as a distribution over words. hLDA [15] extends LDA model to a multi-level tree structure where each node is associated with a topic distribution over words. It can be inferred from a nested Chinese restaurant process. Based on one MCMC strategy. HDP can be specialized as an LDA-based topic model where the number of clusters can be automatically inferred from data. Therefore, HDP is more practical when users have little knowledge about the content to be analyzed.

Learning from time-correlated corpus.

Learning evolutionary topics from a time-correlated corpus aims to preserve the smoothness of clustering results over time, while fitting the data of each epoch [1, 2, 3, 4, 5, 6, 8, 9, 10]. Among existing works, the approaches in [1, 3, 8, 9] utilized time-dependent DP for topic modeling, while others focused on extending LDA to dynamic topic models [5, 6]. In fact, incorporating time dependencies into DP mixture models is a hot topic in the research of Bayesian nonparametric [7, 23, 25, 35, 43]. For instance, Wang et al. [35] focused on detecting the simultaneous bursting of some topics in multiple text streams. Zhang et al. [36] further extended [35] where they adjusted the timestamps of all documents to synchronize multiple streams and then learned a common topic model. Wang et al. [43] introduced an approach based on Hierarchical Dirichlet Process to discover interesting cluster evolutionary pattern from correlated time-varying text corpora.

3. DP AND HDP

Dirichlet Process (DP) can be considered as a distribution over distributions [12]. A DP denoted by \( DP(\alpha, G_0) \) is parameterized by a base measure \( G_0 \) and a concentration parameter. We write \( G \sim DP(\alpha, G_0) \) for a draw of distribution \( G \) from the Dirichlet process. \( G \) itself is a distribution over a given parameter space \( \theta \), therefore we can draw parameters \( \theta_{1:N} \) from \( G \) following a Polya urn distribution [4] also known as a Chinese Restaurant process [30] or through a stick breaking construction [27]. Under the framework of Gibbses sampling, we have:

\[
\theta_{i+1} | \theta_{1:i}, G_0, \alpha \sim \sum_k \frac{m_k}{i-1+\alpha} \delta(\phi_k) + \frac{\alpha}{i-1+\alpha} G_0
\]

where \( \phi_k \) denotes the distinct values among parameter \( \theta \) and \( m_k \) is the number of parameters \( \theta \) having value \( \phi_k \).

As for HDP, we model each document as a DP and each word \( w \) in document \( d \) will be associated with a topic sampled from the random \( G_d \), where \( G_d \sim DP(\alpha, G_0) \). The random measure \( G_d \) represents the document-specific mixing vector over infinite number of topics. In HDP, the document specific measures \( G_d \) are tied together by modeling the
base measure \( G_0 \) itself as a random measure sampled from \( DP(\gamma, H) \).

HDP model can be more easily explained from the Chinese Restaurant Process metaphor, where each document is referred to a restaurant and words are compared to customers. Customers in the restaurant sit around different tables and each table \( b \) is associated with a dish (topic) \( \phi_b \) according to the dish menu. To associate a topic with customer \( w \), we proceed as follows: In restaurant \( d \), the customer can sit on table \( b_{db} \) that has \( n_{db} \) with probability \( \frac{n_{db}}{\sum_b n_{db}} \), and shares the dish \( \phi_{db} \) or picks a new table \( b^{new} \) with probability \( \frac{\alpha}{1 + \alpha} \). If he chooses a new table, he has to order a dish for that table from the global menu. A dish \( \phi_k \) that is served at \( m_k \) across all the restaurants is ordered from the global menu with probability \( \frac{m_k}{\sum_k m_k + \gamma} \), or a new dish \( k^{new} \) that has never been served in any restaurant with probability \( \frac{\gamma}{\sum_k m_k + \gamma} \) and \( \phi^{new} \sim H \). Let \( \theta_{db} \) denotes the dish associated with customer \( w_{di} \), then putting them together we have

\[
\theta_{di} | \theta_{d,i-1}, \alpha, \phi \sim \sum_b \frac{n_{db}}{i + \alpha} \delta_{\phi_{db}} + \frac{\alpha}{i + \alpha} \delta_{\phi^{new}} (2)
\]

\[
\phi^{new}_{db} \sim \sum_k \frac{m_k}{\gamma + \sum_k m_k} \delta_{\phi_k} + \frac{\gamma}{\gamma + \sum_k m_k} H (3)
\]

### 4. HDM FOR TIMELINE GENERATION

#### 4.1 Problem Formulation

Here we firstly give a standard formulation of the task. Given a query, \( Q = \{w_{qi}\}_{i=1}^m \), where \( w_{qi} \) is the word in the query, we get a set of query related documents from the Internet. The corpus is divided into a series of document collections according to the published time as \( C = \{C^t\}_{t=1}^T \), where \( C^t = \{D^t_i\}_{i=1}^{N_t} \), corresponding to the document collection published at time \( t \). \( D^t_i \) denotes document \( i \) at time \( t \) and \( N_t \) denotes the number of documents published at time \( t \). Document \( D^t_i \) is formulated as a collection of words \( D^t_i = \{w_{in}^t\}_{n=1}^{N_t} \). \( V \) denotes the vocabulary size. The output of the algorithm is a series of timelines \( T = \{T^t\}_{t=1}^T \) and \( T^t \subset C^t \).

#### 4.2 HDM

HDM can be viewed as an extension of Hierarchical LDA model [15] where we wish to discover common usage patterns (or topics) given a collection of documents and try to organize these topics into a hierarchy. However, HDM is different from hLDA for its non-parametric nature and capability of modeling time-correlated data.

HDM can be better explained from a Chinese restaurant perspective. There are a series of restaurants in the city and each restaurant is associated with a restaurant-specific dish menu. Such menu is drawn from a global dish menu shared across all restaurants in the whole city. In each restaurant, there are infinite number of tables. When a customer comes into the restaurant, he would choose a table and a dish according to the local restaurant dish menu. An demonstration of restaurant menu is illustrated in Figure 3. From Figure [3] we observe that the Chinese restaurant interpretation for HDM is different from other topic models such as LDA, HDP and sLDA in that the dishes in the menu are organized as a hierarchical tree structure. The intuitive idea for the construction of such menu is that it can be used to represent the hierarchical topic structure in the corpus. For instance, the route of meat \( \rightarrow \) pork \( \rightarrow \) tenderloin \( \rightarrow \) grilled tenderloin can be compared to the hierarchical structure of topic \( \rightarrow \) subtopic \( \rightarrow \) three-level topic \( \rightarrow \) forth-level topic described in Section 1, say Greek crisis \( \rightarrow \) Greek public protests \( \rightarrow \) Indigent Citizens Movement \( \rightarrow \) people get killed. There is a pool of dishes \( H \), from which the everyday city dish menu \( G_t \) is drawn, where \( t \) denotes the index of the day. It is worth noting that the global dish menu changes each day in that residents in the city are fed up with having the same dishes every day. Another thing to mention is residents in the town would preserve part of dishes they already had yesterday on today’s menu since some of them are so delicious and they want to have them again. So \( G_t \), which denotes the city dish menu for day \( t \), is drawn from the combination of base menu \( H \) and yesterday’s menu \( G_{t-1} \) as follows:

\[
G_t \sim DP(\alpha, F(\Delta)G_{t-1} + (1 - F(\Delta))H) (4)
\]

where \( F(\Delta) = \exp(-\Delta/\lambda) \), and it controls the influence of neighboring data. \( \lambda \) is the decay factor of the time-decaying kernel. Such theme has been introduced in a couple of previous works which try to address time dependency in time-correlated data [1][2][25][43].

---

**Figure 1:** Graphical representation of (a) DP and (b) HDP

**Figure 2:** Example of dish menu in Chinese Restaurant Process for HDT.
is a bit critical since he may not feel like having any type of dish (or sub-dishes) appeared on the restaurant menu. If that is the case, the servant would give him the city menu $G_t$, letting him choose and then again base menu $H$ if he is still not satisfied. The problem is, sometimes the customer is so critical that even the global menu $H$ can not satisfy him. In that case, the cook in the restaurant would invent a new kind of meat (if he already decided to have meat) or a new cooking method for tenderloin (if he already decided to have tenderloin) and then add this newly invented dish to the menu list.

Getting back to our data, each sentence is compared to customer in the Chinese Restaurant metaphor document to restaurant and document collection at each time to restaurants at that day. For each sentence $s$, it is linked with a vector of dish (topic) $\theta_1, \theta_2, ..., \theta_s$ to document $H$ is drawn successively from $G_t$ given $\theta_{t-1}$. In Chinese restaurant metaphor, $\theta_1$ can be compared to meat, $\theta_2$ to pork, $\theta_3$ to tenderloin and so on. The plate diagram and generative story for our model are given in Figure 3 and Figure 4.

**Figure 3: Illustration of Chinese Restaurant Metaphor for HDM.**

**Figure 4: Graphical illustration for HDT model.**

### 4.3 Inference

For model inference, we use a straightforward Gibbs sampler based on the Chinese Restaurant Metaphor. Due to the space limit, we just briefly describe this part and skip the details that can be found in Teh et al’s work [30].

#### sample table $B$ for current customer $s$:

$$P(B_s = B | w, B - b_s) \propto \begin{cases} n_B P(s|w, z_B) & \text{if } B \text{ is used} \\ \gamma P(s|w, z_{new}) & \text{if } B \text{ is new} \end{cases} \quad (5)$$

where $n_B$ denotes the number of customers sitting at table $B$ and $z_B$ denotes the dish served at table $B$. $P(s|w, z_B)$ and $P(s|w, z_{new})$ denote the probability of sentence $s$ generated by topic $z_B$ and new topic respectively, which would be described in detail in Appendix A.

#### sample dish $z_B$ for the new table:

If the customer chooses a new table, we have to sample a dish for this table according to the dish list.

$$P(z_{new} = z | w \in s, z) \propto \begin{cases} m_{t'}^w P(s|w, z) & \text{if } z \text{ is used} \\ \alpha P(s|w, z_{new}) & \text{if } z \text{ is new} \end{cases} \quad (6)$$

According to Equ 4, we have $m_t^w = F(\Delta)m_{t-1}^w + (1 - F(\Delta))m_t^w$ where $m_t^w$ denotes the number of tables serving dish $z$ at time $t$.

#### re-sample dish $z_B$ for each table:

Since the process of dish sampling actually changes the component member of tables after each iteration, we need to re-sample dish for each table as follows:

$$P(z_B = z | w, z, B) \propto \begin{cases} m_{t'} P(B|w, z) & \text{if } z \text{ is used} \\ \alpha P(B|w, z_{new}) & \text{if } z \text{ is new} \end{cases} \quad (7)$$

where $P(B|w, z)$ and $P(B|w, z_{new})$ denotes the probability that all sentences around table $B$ are generated by topic $z$ and new topic respectively, which will be described in Appendix A.

### 4.4 Tree-based Sentence Selection

In HDM, each node from the tree is associated with a distribution over vocabularies and each sentence is represented by a path in the tree. We assume that sentences (or words) sharing similar paths should be more similar to each other because of the sharing topics. Let $L$ denote the set of leaf nodes in the tree. The similarity between two words $w_1$ and $w_2$ is obtained by first calculating the Jensen-Shannon
Let \( T_i \) denote a collection of words. \( T_i = \{w|w \in T_i\} \).

\[
q(T|node_i) = \frac{1}{|T|} \sum_{w \in T} p(w|node_i).
\]

And the KL divergence between two collections of words are defined as follows:

\[
KL(T_1||T_2) = \sum_{node_i \in L} q(T_1|node_i) \log \frac{q(T_1|node_i)}{q(T_2|node_i)}
\]

JS divergence is then transformed into a similarity measure by an decreasing logistic function \( \zeta_1(x) = 1/(1 + e^x) \) and increasing logistic function \( \zeta_2(x) = e^x/(1 + e^x) \) to refine the relevance in the range of \((0,1)\). A good timeline should properly consider the following key requirements:\(^{[21]}\):

1. **Focus**: The timeline for epoch \( t \) should be related to the given query. Since the Query distribution is too sparse, we adopt the strategy taken in Yan et al.’s work.\(^{[11]}\). Query expansion is introduced by pseudo-relevance feedback to enlarge \( Q \). We retrieve \( top - \kappa \) snippets, which is denoted as \( Q' \) and use \( Q' \) to approximate \( Q \).

\[
\xi_F(I_t) = \zeta_1[JS(S_t||Q')]
\]

2. **Coherence**: A timeline consists of a series of individual but correlated sentences. News evolves over time and a good component timeline should be coherent with neighboring documents so that the timeline can track the evolutionary pattern of news rather than peaks and bounds.

\[
\xi_C H(I_t) = \zeta_1[JS(S_t||C_{t-1})]
\]

3. **Coverage**: Timeline candidate should keep alignment with source alignment at epoch \( t \):

\[
\xi_C V(I_t) = \zeta_1[JS(S_t||C_t)]
\]

4. **Non-redundancy**: Non-redundancy measures the novelty degree of any of the sentence \( s \) compared with other sentences within \( I_t \).

\[
\xi_{NR}(I_t) = \frac{1}{|I_t|} \sum_{s \in I_t} \zeta_2(JS(s||I_t - s))
\]

(8)

Given the source collection, timeline is scored based on the weighted combination of these four requirements. The score function is illustrated in Eqn.(14).

\[
S(I_t) = w_1 \cdot \xi_F(I_t) + w_2 \cdot \xi_C H(I_t) + w_3 \cdot \xi_C V(I_t) + w_4 \xi_{NR}(I_t)
\]

(9)

\[\sum w_i = 1.\]  Given the dataset, our timeline generation problem is transformed to the optimization problem as follows:

\[
I_t^* = \arg\max S(I_t^*)
\]

(10)

We applied the sentence selection strategy in Yan et al’s work \(^{[11]}\) that literately generate \( I_t^* \) to approximate \( I_t \) by maximizing function \( S \) based on the timeline generation in the last iteration.

---

**5. EXPERIMENTS**

**5.1 Datasets and Experiments Setup**

There is no existing standard evaluation datasets for timeline generation. In this paper, we build 8 datasets talking about the real long-term events that draw public attention and use the golden standards to evaluate the performance of different models. The events include "Iraq War", "Apple Inc", "Financial Crisis", "Greek Debt Crisis", "Arab Spring", "American Presidential Election 2012" and "North Korea’s nuclear crisis". We downloaded 9935 news articles from different resources. The details of datasets are illustrated at Table1 and Table2. Dataset 1 "Iraq War", 2 "Apple Inc", 6 "Afghanistan War" and 8 "North Korea’s Nuclear Crisis" are used as training sets for parameter tuning and the rest are used for testing.

**Preprocessing**: We firstly remove stop-words in documents using a stop-word list of 598 words and the remaining words are stemmed by Porter Stemmer.\(^3\) Besides stemming and stop-word removal, we also extract text snippets with the toolkit provided by Yan et al.\(^{[40]}\). Then we compress the corpus by discarding non-events texts. Timeline for each day is truncated to the same length of 30 words.

**Post-processing**: Since timeline candidates are extracted directly from original news corpus, sentences tend to contain less meaningful information and are sometimes redundant for readers to read. So we apply sentence compression algorithm, making the timeline much conciser. Sentence compression techniques have been well explored in many existing works\(^{[14, 22, 42]}\). We adopt the strategy introduced by \(^{[22]}\) as shown in Figure 7.

**5.2 Evaluation Metrics**

We adopt ROUGE toolkit (version 1.5.5) for performance evaluation. The timeline quality is measured by counting the number of overlapping units, such as N-gram, word sequences and word pairs between candidate timeline CT and the ground-truth timelines GT. Several automatic evaluation methods are implemented in ROUGE and each of them is evaluated.

---

\(^3\) \texttt{http://tartarus.org/~martin/PorterStemmer/}

---

**Table 1: News Sources for timeline generation**

| News Sources | Nation | News Sources | Nation |
|--------------|--------|--------------|--------|
| ABC          | US     | Wanshington Post | US    |
| CNN          | US     | New York Times | US    |
| BBC          | UK     | Xinhua        | China |
## 5.4 Performance Comparison with Baselines

We implement the following algorithms as the baseline systems. For fairness we conduct the same preprocessing and post-processing for all algorithms.

**Random:** The methods that select sentences randomly for timeline generation.

**Centroid:** The method that applies MEAD algorithm [24], which has been widely used in MDS for sentence selection according to centroid value, positional value, and first-sentence overlap.

**GMDS:** The Graph-based MDS sentence selection strategy proposed by Wan et al. [31] that constructs a sentence connectivity graph based on cosine similarity and then selects important sentences based on centrality.

**Chieu:** A timeline system proposed by Chieu et al. [10] by considering interest and burstiness.

**ETTS:** A timeline system proposed by Yan et al. [41] by projecting sentences from different temporal corpus into one plane and developing the graph based ranking algorithm.

![Figure 8: Overall Performance of Different Models.](image)

### Table 3: Comparison with different baselines

| Topics                  | Financial Crisis | Greek Debt Crisis |
|-------------------------|------------------|-------------------|
| systems                 | R-1      | R-2      | R-W      | R-1  | R-2   | R-W   |
| HDT                     | 0.386   | 0.078   | 0.144   | 0.370  | 0.062  | 0.136  |
| Centroid                | 0.297   | 0.039   | 0.081   | 0.279  | 0.031  | 0.087  |
| GMDS                    | 0.286   | 0.044   | 0.088   | 0.288  | 0.033  | 0.081  |
| Chieu                   | 0.335   | 0.056   | 0.115   | 0.331  | 0.048  | 0.119  |
| ETTS                    | 0.340   | 0.054   | 0.120   | 0.327  | 0.044  | 0.120  |
| Random                  | 0.247   | 0.031   | 0.074   | 0.213  | 0.028  | 0.068  |

![http://www.wikipedia.org/](image)

![http://mturk.com](image)
measures to capture the time dependencies. However, since ETTS is a sentence ranking algorithm in nature, its sentence selection strategy may be biased for neglecting the evolutionary pattern of topic information across the corpus. tHDT achieves the best results for its capability in detecting the hierarchical structures in the corpus. HDM outperforms ETTS by 6.2%, 12.9% and 11.7% with regard to the overall performance in ROUGE-1, ROUGE-2 and ROUGE-W respectively.

5.5 Comparison with Other Topic Models

To illustrate the effectiveness of our topic model, we provide six other baseline systems which adopt different aspect modeling techniques.

**tHDP**: HDP is a time-dependent HDP model without considering hierarchical structure of topics. It is a simple version HDM.

**Stand-HDP**: Stand-HDP is standard HDP model that treats different epochs as a series of independent HDPs without considering time dependence.

**Dyn-LDA**: A dynamic LDA\[^5\] where topic-word distribution and popularity are linked across epochs by including Markovian assumption.

**Stand-LDA**: Standard LDA topic model without considering background or temporal information.\[^6\]

5.6 Qualitative Evaluation of Topic Modeling

Finally, we present qualitative evaluations for topic modeling results of HDM. Figure ?? (a) presents first level topics extracted from Greek Debt Crisis dataset and correspondent top words for each topic. intensity corresponds to the number of sentences assigned to the certain topic at a specific time. As we can see, each topic does describe real aspect within the event. According to the top word, we find that Topic 1 talks about general concept for Greek Debt Crisis, Topic 2 talks about response or rescue from international community, Topic 3 discusses protests from Greek public and Topic 4 discusses measures that Greek government take. We also find an interesting phenomena that protests (Topic 3) usually peaks right after domestic measures (Topic 4) or international response (Topic 2), which makes sense in real case where people go on protests for the measure of policy of the government. Figure ?? (b) presents sub-topics (second-level topics) for Topic 3, denoted as Topic 3-1, 3-2 and 3-3. According to top words and peak time, we can find Topic 3-1 presents strike and protest on May 2010 while Topic 3-2 corresponds to the demonstration around May 2011. These two subtopics capture the differences and specifics between the two procedures of protest and demonstration within Greek debt crisis. We can see top words in Topic 3-1 tend to be more violent than words in Topic 3-2, including the information that people got killed in May, 2010. Topic 3-3 tend to describe the general or background perspective in Greek protest. Figure 11 describes presents the hierarchical topic structure for American Election 2012, where the labels are manually given.

\[^5\]http://www.nytimes.com/2010/05/06/world/europe/06greece.html

---

Table 4: Comparison with Other Topic Models

| Systems   | Greek Debt Crisis | Financial Crisis |
|-----------|-------------------|------------------|
|           | R-1 | R-2 | R-W | R-1 | R-2 | R-W |
| HDM       | 0.386 | 0.078 | 0.144 | 0.370 | 0.082 | 0.130 |
| t-HDP     | 0.360 | 0.064 | 0.128 | 0.338 | 0.053 | 0.113 |
| Stand-HDP | 0.348 | 0.055 | 0.114 | 0.323 | 0.050 | 0.108 |
| Dyn-LDA   | 0.358 | 0.060 | 0.120 | 0.330 | 0.048 | 0.111 |
| Stand-LDA | 0.334 | 0.046 | 0.110 | 0.319 | 0.050 | 0.107 |

The overall results are shown in Figure ?? and details are listed in Table 4. As we can see, HDM is better than tHDP and HDP, which verifies HDM’s ability in modeling hierarchical topic structure in long-term news corpus. It is also better than baselines based on LDA and Dyn-LDA. Compared with Table 3 and Figure ?? we find that most topic based models can achieve better results than baselines in
6. CONCLUSION
In this paper, we develop a novel topic model denoted as time-dependent Hierarchical Dirichlet Model (HDM) to explore the hierarchical topic structure for timeline generation. Our model aptly combines Dirichlet Tree with Dirichlet Processes and can automatically learn the structure of trees across corpus. Different levels of Markovian time dependency and background information are considered for tree structure construction. We build an experimental system on 8 real long-term events that public concern. Experimental results illustrate the effectiveness of our proposed model.

7. REFERENCES
[1] A. Ahmed and E. P. Xing. Dynamic non-parametric mixture models and the recurrent chinese restaurant process. Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2007.
[2] A. Ahmed and E. P. Xing. Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. arXiv preprint arXiv:1203.3463, 2012.
[3] J. Allan, R. Gupta, and V. Khandelwal. Temporal summaries of new topics. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pages 10–18. ACM, 2001.
[4] D. Blackwell and J. B. MacQueen. Ferguson distributions via pólya urn schemes. The annals of statistics, pages 353–355, 1973.
[5] D. M. Blei and J. D. Lafferty. Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning, pages 113–120. ACM, 2006.
[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. the Journal of machine Learning research, 3:993–1022, 2003.
[7] F. Caron, M. Davy, and A. Doucet. Generalized polya urn for time-varying dirichlet process mixtures. arXiv preprint arXiv:1206.5254, 2012.
[8] D. Chakrabarti, R. Kumar, and A. Tomkins. Evolutionary clustering. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 554–560. ACM, 2006.
[9] Y. Chi, X. Song, D. Zhou, K. Hino, and B. L. Tseng. Evolutionary spectral clustering by incorporating temporal smoothness. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 153–162. ACM, 2007.
[10] H. L. Chieu and Y. K. Lee. Query based event extraction along a timeline. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pages 425–432. ACM, 2004.
[11] A. Feng and J. Allan. Finding and linking incidents in news. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pages 821–830. ACM, 2007.
[12] T. S. Ferguson. A bayesian analysis of some nonparametric problems. The annals of statistics, pages 209–230, 1973.
[13] G. P. C. Fung, J. X. Yu, H. Liu, and P. S. Yu. Time-dependent event hierarchy construction. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 300–309. ACM, 2007.
[14] D. Gillick and B. Favre. A scalable global model for summarization. In Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing, pages 10–18. Association for Computational Linguistics, 2009.
[15] D. M. B. T. L. Griffiths and M. I. J. B. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. In Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference, volume 16, page 17. MIT Press, 2004.
[16] A. Gruber, Y. Weiss, and M. Rosen-Zvi. Hidden topic markov models. In International Conference on Artificial Intelligence and Statistics, pages 163–170, 2007.
[17] J. Li and C. Cardie. Timeline generation: Tracking individuals on twitter. Proceedings of the 23rd international conference on World wide web, 2014.
[18] J. Li, C. Cardie, and S. Li. Topicspam: a topic-model-based approach for spam detection. In Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics, 2013.
[19] J. Li and S. Li. Evolutionary hierarchical dirichlet process for timeline summarization.
[20] J. Li, M. Ott, and C. Cardie. Identifying manipulated offerings on review portals. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Wash, pages 18–21, 2013.
[21] L. Li, K. Zhou, G.-R. Xue, H. Zha, and Y. Yu. Enhancing diversity, coverage and balance for summarization through structure learning. In Proceedings of the 18th international conference on World wide web, pages 71–80. ACM, 2009.
[22] P. Li, Y. Wang, W. Gao, and J. Jiang. Generating aspect-oriented multi-document summarization with event-aspect model. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 1137–1146. Association for Computational Linguistics, 2011.
[23] I. Pruteanu-Malinici, L. Ren, J. Paisley, E. Wang, and L. Carin. Hierarchical bayesian modeling of topics in time-stamped documents. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(6):996–1011, 2010.
[24] D. Radev, H. Jing, M. Styś, and D. Tam. Centroid-based summarization of multiple documents. Information Processing & Management, 40(6):919–938, 2004.
[25] L. Ren, D. B. Dunson, and L. Carin. The dynamic hierarchical dirichlet process. In Proceedings of the 25th international conference on Machine learning, pages 824–831. ACM, 2008.
[26] K. Salomatin, Y. Yang, and A. Lad. Multi-field correlated topic modeling. In SDM, pages 628–637, 2009.
[27] J. Sethuraman. A constructive definition of dirichlet priors. Technical report, DTIC Document, 1991.
[28] R. Swan and J. Allan. Automatic generation of
overview timelines. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 49–56. ACM, 2000.

[29] Y. W. Teh and G. Haffari. Hierarchical dirichlet trees for information retrieval. 2009.

[30] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[31] X. Wan and J. Yang. Multi-document summarization using cluster-based link analysis. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 299–306. ACM, 2008.

[32] C. Wang, D. Blei, and D. Heckerman. Continuous time dynamic topic models. *arXiv preprint arXiv:1206.3298*, 2012.

[33] J. L. S. L. X. Wang and Y. T. B. Chang. Update summarization using a multi-level hierarchical dirichlet process model.

[34] X. Wang and A. McCallum. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433. ACM, 2006.

[35] X. Wang, C. Zhai, X. Hu, and R. Sproat. Mining correlated bursty topic patterns from coordinated text streams. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 784–793. ACM, 2007.

[36] X. Wang, K. Zhang, X. Jin, and D. Shen. Mining common topics from multiple asynchronous text streams. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 192–201. ACM, 2009.

[37] T. Xu, Z. Zhang, P. S. Yu, and B. Long. Dirichlet process based evolutionary clustering. In *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*, pages 648–657. IEEE, 2008.

[38] T. Xu, Z. Zhang, P. S. Yu, and B. Long. Evolutionary clustering by hierarchical dirichlet process with hidden markov state. In *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*, pages 658–667. IEEE, 2008.

[39] R. Yan, L. Kong, C. Huang, X. Wan, X. Li, and Y. Zhang. Timeline generation through evolutionary trans-temporal summarization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 433–443. Association for Computational Linguistics, 2011.

[40] R. Yan, Y. Li, Y. Zhang, and X. Li. Event recognition from news webpages through latent ingredients extraction. *Information Retrieval Technology*, pages 490–501, 2010.

[41] R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, and Y. Zhang. Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 745–754. ACM, 2011.

[42] D. Zajic, B. J. Dorr, J. Lin, and R. Schwartz. Multi-candidate reduction: Sentence compression as a tool for document summarization tasks. *Information Processing & Management*, 43(6):1549–1570, 2007.

[43] J. Zhang, Y. Song, C. Zhang, and S. Liu. Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1079–1088. ACM, 2010.