Evolutionary Hierarchical Dirichlet Process for Timeline Summarization

Jiwei Li
School of Computer Science
Cornell University
Ithaca, NY, 14853
jl3226@cornell.edu

Sujian Li
Laboratory of Computational Linguistics
Peking University
Bejing, P.R.China, 150001
lisujian@pku.edu.cn

Abstract

Timeline summarization aims at generating concise summaries and giving readers a faster and better access to understand the evolution of news. It is a new challenge which combines salience ranking problem with novelty detection. Previous researches in this field seldom explore the evolutionary pattern of topics such as birth, splitting, merging, developing and death. In this paper, we develop a novel model called Evolutionary Hierarchical Dirichlet Process(EHDP) to capture the topic evolution pattern in timeline summarization. In EHDP, time varying information is formulated as a series of HDPs by considering time-dependent information. Experiments on 6 different datasets which contain 3156 documents demonstrates the good performance of our system with regard to ROUGE scores.

1 Introduction

Faced with thousands of news articles, people usually try to ask the general aspects such as the beginning, the evolutionary pattern and the end. General search engines simply return the top ranking articles according to query relevance and fail to trace how a specific event goes. Timeline summarization, which aims at generating a series of concise summaries for news collection published at different epochs can give readers a faster and better access to understand the evolution of news.

The key of timeline summarization is how to select sentences which can tell readers the evolutionary pattern of topics in the event. It is very common that the themes of a corpus evolve over time, and topics of adjacent epochs usually exhibit strong correlations. Thus, it is important to model topics across different documents and over different time periods to detect how the events evolve.

The task of timelime summarization is firstly proposed by Allan et al.(2001) by extracting clusters of noun phases and name entities. Chieu et al.(2004) 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.(2011) extended the graph based sentence ranking algorithm used in traditional multi-document summarization (MDS) to timeline generation by projecting sentences from different time into one plane. They further explored the timeline task from the optimization of a function considering the combination of different respects such as relevance, coverage, coherence and diversity (Yan et al., 2011b). However, their approaches just treat timeline generation as a sentence ranking or optimization problem and seldom explore the topic information lied in the corpus.

Recently, topic models have been widely used for capturing the dynamics of topics via time. Many dynamic approaches based on LDA model (Blei et al., 2003) or Hierarchical Dirichelt Processes(HDP) (Teh et al., 2006) have been proposed to discover the evolving patterns in the corpus as well as the snapshot clusters at each time epoch (Blei and Lafferty, 2006; Chakrabarti et al., 2006; Wang and McCallum, 2007; Caron et al., 2007; Ren et al., 2008; Ahmed and Xing, 2008; Zhang et al., 2010).

In this paper, we propose EHDP: a evolutionary hierarchical Dirichlet process (HDP) model for timeline summarization. In EHDP, each HDP is built for multiple corpora at each time epoch, and the time dependencies are incorporated into epochs under the Markovian assumptions. Topic popularity and topic-word distribution can be inferred from a Chinese Restaurant Process (CRP). Sentences are selected into timelines by considering different aspects such as topic relevance, coverage and coherence. We built the evaluation sys-

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 556–560,
Sofia, Bulgaria, August 4-9 2013.

tems which contain 6 real datasets and performance of different models is evaluated according to the ROUGE metrics. Experimental results demonstrate the effectiveness of our model .

2 EHDP for Timeline Summarization

2.1 Problem Formulation

Given a general query $Q = \{w_{qi}\}_{i=1}^{i=Q_n}$, we firstly obtain a set of query related documents. We notate different corpus as $C = \{C^t\}_{t=1}^{t=T}$ according to their published time where $C^t = \{D_{ti}\}_{i=1}^{i=N_t}$ denotes the document collection published at epoch t. Document D_i^t is formulated as a collection of sentences $\{s_{ij}^t\}_{j=1}^{j=N_{ti}}$. Each sentence is presented with a series of words $s_{ij}^t = \{w_{ijl}^t\}_{l=1}^{l=N_{ij}^t}$ and associated with a topic θ_{ij}^t. V denotes the vocabulary size. The output of the algorithm is a series of timelines summarization $I = \{I^t\}_{t=1}^{t=T}$ where $I^t \subset C^t$

2.2 EHDP

Our EHDP model is illustrated in Figure 2. Specifically, each corpus C^t is modeled as a HDP. These HDP shares an identical base measure G_0, which serves as an overall bookkeeping of overall measures. We use G_0^t to denote the base measure at each epoch and draw the local measure G_i^t for each document at time t from G_0^t. In EHDP, each sentence is assigned to an aspect θ_{ij}^t with the consideration of words within current sentence.

To consider time dependency information in EHDP, we link all time specific base measures G_0^t with a temporal Dirichlet mixture model as follows:

$$G_0^t \sim DP(\gamma^t, \frac{1}{K}G_0 + \frac{1}{K}\sum_{\delta=0}^{\Delta} F(v,\delta)\cdot G_0^{t-\delta}) \quad (1)$$

where $F(v,\delta) = exp(-\delta/v)$ denotes the exponential kernel function that controls the influence of neighboring corpus. K denotes the normalization factor where $K = 1 + \sum_{\delta=0}^{\Delta} F(v,\delta)$. Δ is the time width and λ is the decay factor. In Chinese Restaurant Process (CRP), each document is referred to a restaurant and sentences are compared to customers. Customers in the restaurant sit around different tables and each table b_{in}^t is associated with a dish (topic) Ψ_{in}^t according to the dish menu. Let m_{tk} denote the number of tables enjoying dish k in all restaurants at epoch t, $m_{tk} = \sum_{i=1}^{N^t}\sum_{n=1}^{N_{ib}^t} 1(\Psi_{in}^t = k)$. We redefine another parameter M_{tk} to incorporate time dependency into EHDP.

for each epoch $t \in [1, T]$
1. draw global measure $G_0^t \sim DP(\alpha, \frac{1}{K}G_0 + \frac{1}{K}\sum_{\delta=0}^{\Delta} F(v,\delta)G_0^{t-\delta})$
2. for each document D_i^t at epoch t,
   2.1 draw local measure $G_i^t \sim DP(\gamma, G_0^t)$
   2.2 for each sentence s_{ij}^t in D_i^t
      draw aspect $\theta_{ij}^t \sim G_i^t$
         for $w \in s_{ij}^t$ draw $w \sim f(w)|\theta_{ij}^t$

Figure 1: Generation process for EHDP

$$M_{tk} = \sum_{\delta=0}^{\Delta} F(v,\delta)\cdot m_{t-\delta,k} \quad (2)$$

Let n_{ib}^t denote the number of sentences sitting around table b, in document i at epoch t. In CRP for EHDP, when a new customer s_{ij}^t comes in, he can sit on the existing table with probability $n_{ib}^t/(n_i^t - 1 + \gamma)$, sharing the dish (topic) Ψ_{ib}^t served at that table or picking a new table with probability $\gamma/(n_i^t - 1 + \gamma)$. The customer has to select a dish from the global dish menu if he chooses a new table. A dish that has already been shared in the global menu would be chosen with probability $M_k^t/(\sum_k M_k^t + \alpha)$ and a new dish with probability $\alpha/(\sum_k M_k^t + \alpha)$.

$$\theta_{ij}^t|\theta_{i1}^t, ..., \theta_{ij-1}^t, \alpha \sim \sum_{\phi_{tb}=\theta_{ij}} \frac{n_{ib}^t}{n_i^t - 1 + \gamma}\delta_{\phi_{jb}} + \frac{\gamma}{n_i^t - 1 + \gamma}\delta_{\phi_{jb}^{new}}$$

$$\phi_{ti}^{new}|\phi, \alpha \sim \sum_k \frac{M_{tk}}{\sum_i M_{ti} + \alpha}\delta_{\phi_k} + \frac{\alpha}{\sum_i M_{ti} + \alpha}G_0 \quad (3)$$

We can see that EHDP degenerates into a series of independent HDPs when $\Delta = 0$ and one global HDP when $\Delta = T$ and $v = \infty$, as discussed in Amred and Xings work (2008).

2.3 Sentence Selection Strategy

The task of timeline summarization aims to produce a summary for each time and the generated summary should meet criteria such as relevance , coverage and coherence (Li et al., 2009). To care for these three criteria, we propose a topic scoring algorithm based on Kullback-Leibler(KL) divergence. We introduce the decreasing logistic function $\zeta(x) = 1/(1 + e^x)$ to map the distance into interval (0,1).

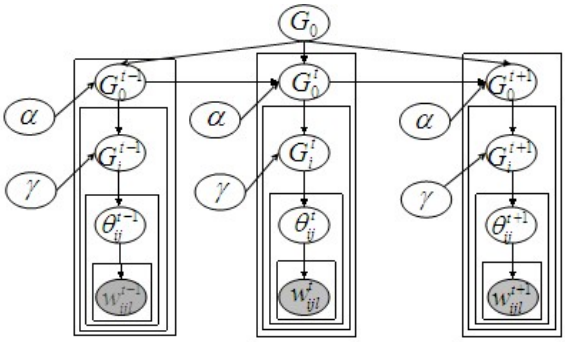

Figure 2: Graphical model of EHDP.

Relevance: the summary should be related with the proposed query Q.

$$F_R(I^t) = \zeta(KL(I^t||Q))$$

Coverage: the summary should highly generalize important topics mentioned in document collection at epoch t.

$$F_{Cv}(I^t) = \zeta(KL(I^t||C^t))$$

Coherence: News evolves over time and a good component summary is coherent with neighboring corpus so that a timeline tracks the gradual evolution trajectory for multiple correlative news.

$$F_{Ch}(I^t) = \frac{\sum_{\delta=-\Delta/2}^{\delta=\Delta/2} F(v,\delta) \cdot \zeta(KL(I^t||C^{t-\delta}))}{\sum_{\delta=-\Delta/2}^{\delta=\Delta/2} F(v,\delta)}$$

Let $Score(I^t)$ denote the score of the summary and it is calculated in Equ.(4).

$$Score(I^t) = \lambda_1 F_R(I^t) + \lambda_2 F_{Cv}(I^t) + \lambda_3 F_{C_h}(I^t) \quad (4)$$

$\sum_i \lambda_i = 1$. Sentences with higher score are selected into timeline. To avoid aspect redundancy, MMR strategy (Goldstein et al., 1999) is adopted in the process of sentence selection.

3 Experiments

3.1 Experiments set-up

We downloaded 3156 news articles from selected sources such as BBC, New York Times and CNN with various time spans and built the evaluation systems which contains 6 real datasets. The news belongs to different categories of Rule of Interpretation (ROI) (Kumaran and Allan, 2004). Detailed statistics are shown in Table 1. Dataset 2(Deepwater Horizon oil spill), 3(Haiti Earthquake) and 5(Hurricane Sandy) are used as training data and the rest are used as test data. Summary at each epoch is truncated to the same length of 50 words.

New Source	Nation	News Source	Nation
BBC	UK	New York Times	US
Guardian	UK	Washington Post	US
CNN	US	Fox News	US
ABC	US	MSNBC	US

Table 1: New sources of datasets

News Subjects (Query)	#docs	#epoch
1.Michael Jackson Death	744	162
2.Deepwater Horizon oil spill	642	127
3.Haiti Earthquake	247	83
4.American Presidential Election	1246	286
5.Hurricane Sandy	317	58
6.Jerry Sandusky Sexual Abuse	320	74

Table 2: Detailed information for datasets

Summaries produced by baseline systems and ours are automatically evaluated through ROUGE evaluation metrics (Lin and Hovy, 2003). For the space limit, we only report three ROUGE ROUGE-2-F and ROUGE-W-F score. Reference timeline in ROUGE evaluation is manually generated by using Amazon Mechanical Turk[1]. Workers were asked to generate reference timeline for news at each epoch in less than 50 words and we collect 790 timelines in total.

3.2 Parameter Tuning

To tune the parameters $\lambda(i = 1, 2, 3)$ and v in our system, we adopt a gradient search strategy. We firstly fix λ_i to 1/3. Then we perform experiments on with setting different values of $v/\#epoch$ in the range from 0.02 to 0.2 at the interval of 0.02. We find that the Rouge score reaches its peak at round 0.1 and drops afterwards in the experiments. Next, we set the value of v is set to $0.1 \cdot \#epoch$ and gradually change the value of λ_1 from 0 to 1 with interval of 0.05, with simultaneously fixing λ_2 and λ_3 to the same value of $(1 - \lambda_1)/2$. The performance gets better as λ_1 increases from 0 to 0.25 and then declines. Then we set the value of λ_1 to 0.25 and change the value of λ_2 from 0 to 0.75 with interval of 0.05. And the value of λ_2 is set to 0.4, and λ_3 is set to 0.35 correspondingly.

3.3 Comparison with other topic models

In this subsection, we compare our model with 4 topic model baselines on the test data. *Stand-HDP(1)*: A topic approach that models different time epochs as a series of independent HDPs without considering time dependency. *Stand-HDP(2)*:

[1]http://mturk.com

	M.J. Death		US Election		S. Sexual Abuse	
System	R2	RW	R2	RW	R2	RW
EHDP	0.089	0.130	0.081	0.154	0.086	0.152
Stand-HDP(1)	0.080	0.127	0.075	0.134	0.072	0.138
Stand-HDP(2)	0.077	0.124	0.072	0.127	0.071	0.131
Dyn-LDA	0.080	0.129	0.073	0.130	0.077	0.134
Stan-LDA	0.072	0.117	0.065	0.122	0.071	0.121

Table 3: Comparison with topic models

	M.J. Death		US Election		S. Sexual Abuse	
System	R2	RW	R2	RW	R2	RW
EHDP	0.089	0.130	0.081	0.154	0.086	0.152
Centroid	0.057	0.101	0.054	0.098	0.060	0.132
Manifold	0.053	0.108	0.060	0.111	0.069	0.128
ETS	0.078	0.120	0.073	0.130	0.075	0.135
Chieu	0.064	0.107	0.064	0.122	0.071	0.131

Table 4: Comparison with other baselines

A global HDP which models the whole time span as a restaurant. The third baseline, *Dynamic-LDA* is based on Blei and Laffery(2007)'s work and *Stan-LDA* is based on standard LDA model. In LDA based models, aspect number is predefined as 80 [2]. Experimental results of different models are shown in Table 2. As we can see, EHDP achieves better results than the two standard HDP baselines where time information is not adequately considered. We also find an interesting result that Stan-HDP performs better than Stan-LDA. This is partly because new aspects can be automatically detected in HDP. As we know, how to determine topic number in the LDA-based models is still an open problem.

3.4 Comparison with other baselines

We implement several baselines used in traditional summarization or timeline summarization for comparison. (1) $Centroid$ applies the MEAD algorithm (Radev et al., 2004) according to the features including centroid value, position and first-sentence overlap. (2) $Manifold$ is a graph based unsupervised method for summarization, and the score of each sentence is got from the propagation through the graph (Wan et al., 2007). (3) ETS is the timeline summarization approach developed by Yan et al., (2011a), which is a graph based approach with optimized global and local biased summarization. (4) $Chieu$ is the timeline system provided by (Chieu and Lee, 2004) utilizing interest and bursty ranking but neglecting trans-temporal news evolution. As we can see from Table 3, $Centroid$ and $Manifold$ get the worst results. This is probably because methods in multi-document summarization only care about sentence selection and neglect the novelty detection task. We can also see that EHDP under our proposed framework outputs existing timeline summarization approaches ETS and chieu. Our approach outputs Yan et al.,(2011a)s model by 6.9% and 9.3% respectively with regard to the average score of ROUGE-2-F and ROUGE-W-F.

[2]In our experiments, the aspect number is set as 50, 80, 100 and 120 respectively and we select the best performed result with the aspect number as 80

4 Conclusion

In this paper we present an evolutionary HDP model for timeline summarization. Our EHDP extends original HDP by incorporating time dependencies and background information. We also develop an effective sentence selection strategy for candidate in the summaries. Experimental results on real multi-time news demonstrate the effectiveness of our topic model.

Oct. 3, 2012
S1: The first debate between President Obama and Mitt Romney, so long anticipated, quickly sunk into an unenlightening recitation of tired talking points and mendacity. S2. Mr. Romney wants to restore the Bush-era tax cut that expires at the end of this year and largely benefits the wealthy
Oct. 11, 2012
S1: The vice presidential debate took place on Thursday, October 11 at Kentucky'sCentre College, and was moderated by Martha Raddatz. S2: The first and only debate between Vice President Joe Biden and Congressman Paul Ryan focused on domestic and foreign policy. The domestic policy segments included questions on health care, abortion
Oct. 16, 2012
S1. President Obama fights back in his second debate with Mitt Romney, banishing some of the doubts he raised in their first showdown. S2: The second debate dealt primarily with domestic affairs and include some segues into foreign policy. including taxes, unemployment, job creation, the national debt, energy and women's rights, both legal and

Table 5: Selected timeline summarization generated by EHDP for American Presidential Election

5 Acknowledgement

This research has been supported by NSFC grants (No.61273278), National Key Technology RD Program (No:2011BAH1B0403), National 863 Program (No.2012AA011101) and National Social Science Foundation (No.12ZD227).

References

Amr Ahmed and Eric Xing. Dynamic non-parametric mixture models and the recurrent chinese restaurant process. 2008. In *SDM*.

James Allan, Rahul Gupta and Vikas Khandelwal. Temporal summaries of new topics. 2001. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*

David Blei, Andrew Ng and Micheal Jordan. 2003. Latent dirichlet allocation. In *Journal of Machine Learning Research*.

David Blei and John Lafferty. Dynamic topic models. 2006. In *Proceedings of the 23rd international conference on Machine learning*.

Francois Carol, Manuel Davy and Arnaud Doucet. Generalized poly urn for time-varying dirichlet process mixtures. 2007. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*.

Deepayan Chakrabarti, Ravi Kumar and Andrew Tomkins. Evolutionary Clustering. In*Proceedings of the 12th ACM SIGKDD international conference Knowledge discoveryand data mining*.

Hai-Leong Chieu and Yoong-Keok 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*

Giridhar Kumaran and James Allan. 2004. Text classification and named entities for new event detection. In *Proceedings of the 27th annual international ACM SIGIR04*.

Liangda Li, Ke Zhou, Gui-Rong Xue, Hongyuan Zha and Yong Yu. Enhancing diversity, coverage and balance for summarization through structure learning. In *Proceedings of the 18th international conference on World wide web*.

Chin-Yew Lin and Eduard Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the Human Language Technology Conference of the NAACL. 2003*.

Dragomar Radev, Hongyan. Jing, and Malgorzata Stys. 2004. Centroid-based summarization of multiple documents. In *Information Processing and Management*.

Lu Ren, David Dunson and Lawrence Carin. The dynamic hierarchical Dirichlet process. 2008. In *Proceedings of the 25th international conference on Machine Learning*.

Xiaojun Wan, Jianwu Yang and Jianguo Xiao. 2007. Manifold-ranking based topic-focused multi-document summarization. In *Proceedings of International Joint Conference on Artificial Intelligence*.

Xuerui Wang and Andrew MaCallum. 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*.

Yee Whye Teh, Michael Jordan, Matthew Beal and David Blei. Hierarchical Dirichlet Processes. In *American Statistical Association.*

Rui Yan, Liang Kong, Congrui Huang, Xiaojun Wan, Xiaoming Li and Yan Zhang. 2011a. 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*.

Rui Yan, Liang Kong, Congrui Huang, Xiaojun Wan, Jahna Otterbacher, Xiaoming Li and Yan Zhang. Timeline Generation Evolutionary Trans-Temporal Summarization. 2011b. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Jianwen Zhang, Yangqiu Song, Changshui Zhang and Shixia Liu. 2010. 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*.