Exploring Temporal Patterns in Emergency Department Triage Notes with Topic Models

Simon Kocbek[1,2]**, Karin Verspoor**[2,3]**, Wray Buntine**[4]
[1]Computer Science & Info Tech, RMIT University, Australia
[2]Dept of Computing and Information Systems, The University of Melbourne, Australia
[3]Health and Biomedical Informatics Centre, The University of Melbourne, Australia
[4]Faculty of IT, Monash University, Australia
`skocbek@gmail.com, karin.verspoor@unimelb.edu.au,`
`wray.buntine@monash.edu`

Abstract

Topic modeling is an unsupervised machine-learning task of discovering topics, the underlying thematic structure in a text corpus. Dynamic topic models are capable of analysing the time evolution of topics. This paper explores the application of dynamic topic models on emergency department triage notes to identify particular types of disease or injury events, and to detect the temporal nature of these events.

1 Introduction

Recording of a patient's presenting complaints forms part of the standard triage procedure at most Australian hospital Emergency Departments (EDs). The complaints are typically recorded as brief notes, which capture the reason the patient has come to the ED. These notes represent the first point of contact of a patient with the hospital, and are a source of timely information about the health of the community served by the hospital. For instance, outbreaks of viruses or increased activity of spiders and snakes can be detected by monitoring ED visits.

The range of reasons for patient visits to the ED is diverse, including both accidents or injuries and disease. Topic modeling of ED triage notes provides a strategy for identifying patterns in this diverse data, i.e., for abstracting over individual patient visits to characterise trends in the health of the community. This abstraction gives a valuable snapshot of the health issues affecting the community.

Given the temporal nature of many health and injury events, including seasonal variation in viral load and day-of-the-week variation in events such as alcohol-related accidents, we expect that temporal patterns can be discerned in this data.

In this work, we explore the application of dynamic topic models on ED triage notes to identify particular types of disease or injury events, and to detect the temporal nature of these events. This analysis provides insight into the changing health needs of the community. Our findings have potential application for public health surveillance applications, where emerging issues of public concern can be detected and an appropriate response can be planned.

2 Related work

We treat each triage note is as one short document. It is known that it is very changeling for topic models to handle very short texts (Zhao et al. 2011) and various forms of tweet pooling on hashtag and/or author can be used to overcome this (Mehrotra et al 2013). For triage notes, however, its not clear what discrete variable could be used to pool on. Therefore we have not used any methods to account for the short documents.

Topic models have been only recently applied to analyse electronic health records data. Initial research suggests that the specific characteristics of the clinical language affect the methods and results of topic modeling techniques (Cohen et al, 2013). Topic modeling of Intensive Care Unit (ICU) progress notes to stratify the risk and mortality prediction for the hospital has been performed (Lehman et al., 2012). In that work, a non-parametric topic modeling method is used to discover topics as shared groups of co-occurring UMLS concepts. Salleb-Aouissi (Salleb-Aouissi

et al., 2011) used topic models to show that infant colic has causes that can be illuminated by analysing a large corpus of paediatric notes.

Different models to discover topics have been used in previous work, mostly extending Latent Dirichlet Allocation (LDA) (Blei et al., 2003). LDA assumes that documents are admixtures of topics, where topics represent distributions over a fixed set of vocabularies (represented with a multinomial). Effectively, each word in a document is assigned to a topic so the word probabilities for the document become an admixture. Topic models are made dynamic by allowing time-evolution of parts of the model. An early model, the dynamic topic model (DTM) (Blei and Lafferty 2006) did this using Gaussians to represent evolution as a chain of means, and a logistic map to project vector data into probabilities. Later models used different tricks with exponential family models to extend the original LDA into the time domain (Ahmed and Xing, 2010; Xuerui and MacCallum, 2006) or non-parametric methods (Chen, Ding and Buntine, 2012).

3 Data and Methods

3.1 Data

The data for this study was obtained from the Royal Melbourne Hospital Emergency Department (ED) where triage notes for 57,984 patients over time period of 12 months (August 2010 – July 2011) were collected. We ignored 1,124 entries since they contained an empty triage note field. The average note length is 118 characters.

The triage notes are written in natural language but contain substantial numbers of abbreviations (e.g., R for right, b/c for because, ped for pedestrian), specialised clinical concepts (e.g., dementia, colonoscopy), and even patient biometric data such as blood pressure or temperature. They are also often not grammatically well-formed and often have spelling errors; they may contain a series of brief descriptive phrases and use of punctuation is inconsistent. As an example, consider the note "*alledge assault kick to head. lac to L eyebrow. ?LOC nil neck pain pupils dialated reactive ETOH tonight*". For the work described in this paper, we did not do any specialised processing of abbreviations or clinical concepts.

3.2 Dynamic topic model

The dynamic topic model we use allows components of the model to change over "epochs," where in our case epochs are month or weekend-weekday periods. Moreover, it is a first order Markov model so the components depend on that of their previous epoch. The topic model is given in graphical form in Figure 1. There are K topics, D_e documents per epoch e and each document has L_d words in it (varying per document d). Each document has topic proportions, a probability vector of $\vec{\theta}_d$. Average topic proportions for the epoch e (a prior for $\vec{\theta}_d$) are given by $\vec{\mu}_e$. The word vector for a topic in an epoch is given by $\vec{\phi}_{ek}$. for topic indexes $k=1,...,K$. The word vector depends on its previous counterpart, so $\vec{\phi}_{ek}$ depends on $\vec{\phi}_{(e-1)k}$. All dependencies of probability vectors on probability vectors are done with the Pitman-Yor process which allows efficient learning to be developed using blocked, collapsed Gibbs sampled data (Buntine and Mishra, 2014). The algorithm is implemented in C with a set of libraries for non-parametric topic models.
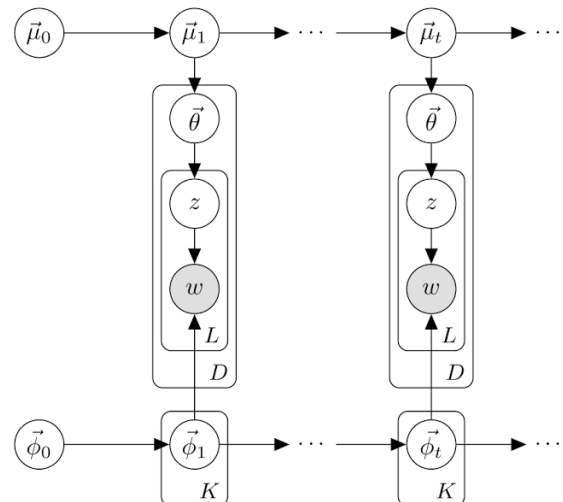

Figure 1: Graphical representation of the dynamic topic model (for epoch *1* and and *t*).

3.3 Running the experiments

For each patient we extracted their ED arrival date and the triage note assigned to the patient. The data was organised into two distinct temporal representations: by month, and by weekdays-weekends. Triage notes were then preprocessed to be in the right format for the topic modeling software. We used the Mallet (McCallum, 2002) stop list to filter out the most common words. The non-parametric dynamic topic model was applied to look for topics. Experiments for 10, 20, 30 and 40 topics were run. All the models were first initialized with 20 major Gibbs cycles of a standard topic model. We then ran the dynamic topic model with 500 (months) or 200 (weekend-weekdays) cycles.

Fewer cycles in the weekend-weekdays model were used due to time constraints.

Manual examination of topics was then performed with the goal of finding coherent and intuitively interpretable topics. Topics were presented with their top words, where we ranked the latter by fraction of their total occurrences in the topic. We also calculated normalized PMI (Han et al. 2014) to measure the coherency of each topic.

To compare topics over time, topic probabilities were calculated. Higher probability for a time period means that the topic is more likely to occur. The top ranked words for each topic were compared between epochs since different words may have different probabilities over time.

4   Results

While we measured coherence using normalised PMI, the raw results were poor, because of the large number of out of vocabulary words in the triage note content. Therefore, for the purposes of the current study we used visual inspection to evaluate topics. For month periods as epochs, 36 out of 40 topics were viewed as coherent; thus we viewed the model to be informative. We display those with interesting time structure here.

Figure 2 illustrates changes in proportions of 9 selected topics over a year. Top representative words for these topics are shown in Table 1. The topics offer a certain degree of coherence and could be interpreted as given in Table 1.

Topic	Problem	Top representative words
T1	Flu	aches, runny, chills, flu-like, fever
T2	Asthma	sentences, speaking, ventolin, talk
T3	Angina	gtn, patch, anginine, spray, aspirin
T4	Arm	foosh, rotation, shortening, rotated
T5	Insect	bite, spider, touch, burn, warm, rabies
T6	VDA	grey, code, packer, street, narcan, heroin
T7	Blood	gb, transfusion, abnormal, wcc
T8	Panic Attack	attack, panic, attacks, anxious
T9	Hernia	inguinal, hernia, testicular, hiatus

Table 1: Identified topics and representative words for the months model.

Figure 2 shows Flu and Angina peaks in Winter months. On the other hand, topics related to Arm, Insect, Drugs/Alcohol, and Hernia injuries and problems show peaks in warmer months. The Asthma topic shows a brief peak in Spring and the Panic Attacks topic increases in Autumn. The Blood topic slowly increases over time.

In Figure 3 we illustrate probabilities of 3 selected topics with interesting time structure for the weekend-weekdays model with 20 topics run. Top words for these topics appear in Table 2.
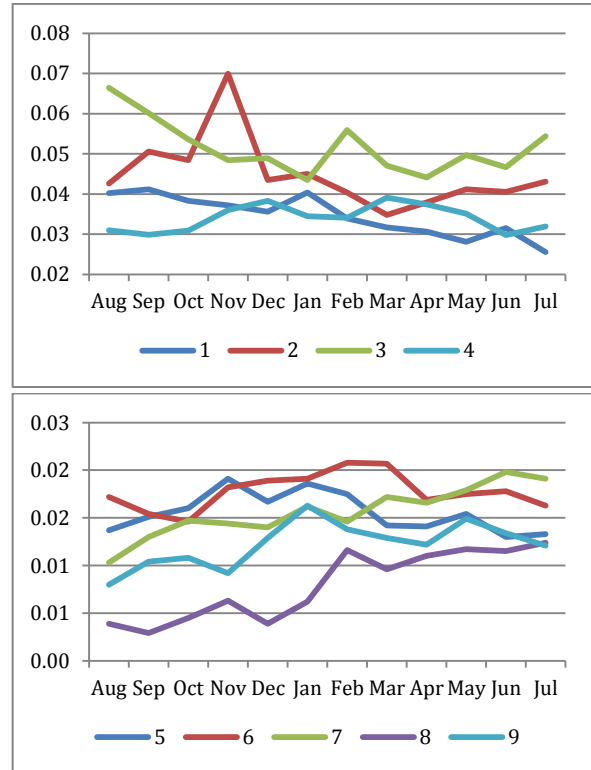

Figure 2: Probabilities of 9 topics over time (the months model). Note the scales of the top and bottom figures differ.
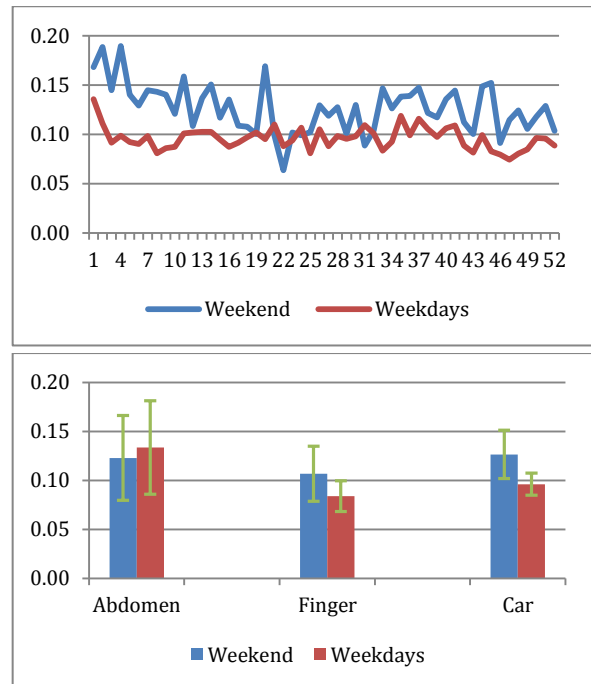

Figure 3: Probabilities of the Car topic (top), and average and standard deviation values for probabilities related to abdomen, finger and car problems (bottom).

Topic	Top representative words
Car	car, loc, driver, hit, speed, head
Finger	finger, cut, vasc, intact, rom, hand
Abdomen	abdo, flank, chronic, lower

Table 2: Identified topics and representative words for the weekend-weekdays model.

Figure 3 (top) shows probabilities of the car accidents topic. Each point on the x axis represents a whole week, where data is divided on weekdays (red) and weekends (blue). The bottom chart shows average and standard deviation values for probabilities for topics related to abdomen, finger and car problems. These topics were selected because their probability values demonstrate a clear division between weekends and weekdays. The Car and Finger topics have higher probabilities on weekends, while the Abdomen topic shows peaks on Weekdays.

5 Discussion

Specific characteristics of the clinical language affect the performance of topic modeling on this data. Triage notes contained considerable numbers of abbreviations, spelling mistakes, clinical concepts and multi-word phrases that the methods do not treat as a unit. Despite such problems, our analysis of ED triage notes with dynamic topic models offers some interesting conclusions.

First, we showed that topic models confirm some expected patterns in the data. For example, probability peaks in the Flu topic correspond to the influenza season in Australia (Winter). This is also the case for the Angina topic where, although the topic has a brief peak in February, we see more angina-related words in colder months and fewer in Spring/Summer. Both influenza and angina are known to be more common in Winter; this is effectively reflected in the topic models.

Several topics with peaks in warmer months also capture expected results. For example, patients seem to have more problems related to insect and animal bites in warmer months. This is expected since people spend more time outdoors, and the insects are more active, in spring and summer. The Arm and Hernia topics also peak between October and May, when people spend more time outdoors doing sports like swimming, rock climbing, volleyball, and similar. The Asthma topic has a brief peak in spring, when pollen-related problems are known to occur.

The results in Figure 2 also lead to some nontrivial conclusions. An interesting topic is VDA related to violence, and drug and alcohol problems. Looking at VDA, we can notice an increase of these issues between January and March. A more detailed analysis will be needed, but these results suggest that broader non-health related (e.g., criminal) surveillance might also be possible using this data and our methods.

Figure 2 also raises some questions. The Blood and Panic Attack topics show a constant and slow increase in probability. With current analysis of only a single year's worth of data, we are not sure about reasons for that.

Results on Figure 3 show interesting patterns when comparing weekdays and weekends. The top chart indicates that car related accidents more likely occur on weekends. Based on the Department of Infrastructure and Transport's report (BITRE, 2011), around half of all fatal crashes in Australia occur during weekends. Considering also non-fatal incidents, we view our results as informative. Peaks of finger and abdomen related problems raise questions about their meaning and further analysis will be needed.

Please note that the weekend-weekdays results should be interpreted with caution. Although the model might discover some patterns, it is not customized for analysing "periodic effects" in data. During the learning, the model tries to track things between epochs that are radically different. A weekend epoch is conditioned on the previous week, but results demonstrate essential difference, which poses challenges to the model. The models need to be adapted to deal better with such expected variation.

6 Conclusion

In this paper we have presented results of applying dynamic topic models to ED triage notes. The results should be viewed as an exploration that is indicative of the potential of the method.

In the future we plan to address several issues in this paper. We plan to address some of the specific characteristics of clinical and medical language with pre-processing techniques such as using MetaMap (Aronson and Lang, 2010) to recognise clinical concepts. We should also add periodicity modelling to the topic model, however, this is a more substantial project.

There is still substantial analysis that should be performed to more deeply explore these initial results, in particular to understand the statistical significance of the results. While the data set is not small, more years of data are required to establish any regular periodic effects. Moreover, we also need to further understand why the topic model worked quite well despite the lack of handling for short texts.

Finally, we plan to design a human evaluation to directly assess topic coherence and modify the PMI analysis to adjust for out-of-vocabulary words.

Acknowledgments

Reference

Ahmed A and Xing EP. 2010. *Timeline: A Dynamic Hierarchical Dirichlet Process Model for Recovering Birth/Death and Evolution of Topics in Text Stream*, In Proc. UAI 2010.

Aronson AR and Lang FM. 2010. *An overview of MetaMap: historical perspective and recent advances.* Journal of the American Medical Informatics Association 17:229-236.

BITRE - Bureau Of Infrastructure, Transport And Regional Economics. Australian Federal Department Of Infrastructure And Transport. 2011. *Road Deaths Australia - 2010 Statistical Summary*, https://www.bitre.gov.au/.

Blei D, Ng A, Jordan M. 2003. *Latent Dirichlet allocation.* Journal of Machine Learning Research 3: 993–1022.

Blei D and Lafferty JD. 2006. *Dynamic topic models.* In Proc. ICML, pp 113–120. ACM.

Buntine W and Mishra S. 2014. *Experiments with non-parametric topic models.* Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.

Chen C, Ding N, Buntine W. 2012. *Dependent Hierarchical Normalized Random Measures for Dynamic Topic Modeling*, ICML 2012.

Cohen R, Elhadad M, Elhadad N. 2013. *Redundancy in electronic health record corpora: Analysis, impact on text mining performance and mitigation strategies.* BMC Bioinformatics 14: 10. doi: 10.1186/1471-2105-14-10.

Lau JH, Newman D, and Baldwin T. 2014. *Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality.* Proceedings of the European Chapter of the Association for Computational Linguistics.

Lehman LW, Saeed M, Long W, Lee J, Mark R. 2012. *Risk stratification of ICU patients using topic models inferred from unstructured progress notes.* In: Proc. AMIA. 505–511.

McCallum AK. 2002. *MALLET: A Machine Learning for Language Toolkit.*, http://mallet.cs.umass.edu.

Mehrotra R, Sanner S, Buntine W, and Xie L. 2013. *Improving lda topic models for microblogs via tweet pooling and automatic labelling.* The 36th Annual ACM SIGIR Conference, 889–892.

Salleb-Aouissi A, Radeva A, Passonneau R, Tomar A, Waltz D, et al. 2011. *Diving into a large corpus of pediatric notes.* In: Proc. ICMLWorkshop on Learning from Unstructured Clinical Text.

Xuerui W, and McCallum A. 2006. *Topics over time: a non-Markov continuous-time model of topical trends.* In Proc. SIGKDD, ACM 2006.

Zhao WX, Jiang J, Weng J, He J, Lim EP, Yan H, Li X. 2011. *Comparing twitter and traditional media using topic models.* In Proceedings of ECIR'11, pages 338–349, Berlin, Heidelberg. Springer-Verlag.