Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220607

Discovering Key Topics in Emergency Medical Dispatch from Free Text Dispatcher Observations

Pablo FERRI[a,1], Carlos SÁEZ[a], Antonio FÉLIX-DE CASTRO[b], Purificación SÁNCHEZ-CUESTA[b] and Juan M GARCÍA-GÓMEZ[a]

[a] *Biomedical Data Science Laboratory (BDSLab), Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València (UPV), Valencia, Spain*

[b] *Conselleria de Sanitat Universal i Salut Pública, Generalitat Valenciana (GVA), Valencia, Spain*

Abstract. The objective of this work was to discover key topics latent in free text dispatcher observations registered during emergency medical calls. We used a total of 1374931 independent retrospective cases from the Valencian emergency medical dispatch service in Spain, from 2014 to 2019. Text fields were preprocessed to reduce vocabulary size and filter noise, removing accent and punctuation marks, along with uninformative and infrequent words. Key topics were inferred from the multinomial probabilities over words conditioned on each topic from a Latent Dirichlet Allocation model, trained following an online mini-batch variational approach. The optimal number of topics was set analyzing the values of a topic coherence measure, based on the normalized pointwise mutual information, across multiple validation K-folds. Our results support the presence of 15 key topics latent in free text dispatcher observations, related with: ambulance request; chest pain and heart attack; respiratory distress; head falls and blows; fever, chills, vomiting and diarrhea; heart failure; syncope; limb injuries; public service body request; thoracic and abdominal pain; stroke and blood pressure abnormalities; pill intake; diabetes; bleeding; consciousness. The discovery of these topics implies the automatic characterization of a huge volume of complex unstructured data containing relevant information linked to emergency medical call incidents. Hence, results from this work could lead to the update of structured emergency triage algorithms to directly include this latent information in the triage process, resulting in a positive impact in patient wellbeing and health services sustainability.

Keywords. Medical emergencies, emergency medical calls, emergency medical dispatch, natural language processing, topic discovery, latent dirichlet allocation.

1. Introduction

Emergency medical dispatch entails the reception and management of demands for medical assistance in an emergency medical services system [1]. It involves emergency medical calls attendance and events triage according to their priority, process generally managed by emergency medical dispatchers. These mediators tend to follow a clinical protocol focused on a small set of structured clinical variables [2].

[1] Corresponding Author, E-mail: pabferb2@upv.es.

In the Valencian Community (Spain), the triage of emergency medical call incidents (EMCI) is currently assisted by an in-house triage protocol, a clinical decision tree based on the collection of structured variables. The dispatcher raises questions to the caller until reaching a final tree node, which has a priority assigned to it, the incident priority.

However, information not covered by the decision tree is also registered during the call in an unstructured manner in free text fields. This information, complementary to that provided by the structured variables, cannot be taken into account automatically by the clinical protocols, and thus, it is left unused.

We have studied in previous works that considering these free text dispatcher observations notably improves EMCI triage. Specifically, we have developed DeepEMC[2], a deep learning model able to automatically deal with structured and unstructured information in real time, providing performance increases of 12.5%, 17.5% and 5.1% in terms of macro F1-score in life-threatening, admissible response delay and emergency system jurisdiction prediction, respect to the current in-house triage protocol of the Valencian emergency medical dispatch service [3].

In addition, prior studies have shown the potential of text mining techniques and, concretely, topic extraction methods, to infer high-level information from huge amounts of unstructured medical data [4], [5].

Given the utmost relevance of free text dispatcher observations in EMCI triage and the availability of methods to explore them from a machine learning perspective, we present in this work an unsupervised analysis of these free text fields, with the aim of 1) discovering and understanding what information dispatchers report during emergency call incidents and 2) exploring how this latent information is distributed across incidents.

2. Methods

A total of 1374931 free text dispatcher observations linked to EMCI of the Health Services Department of the Valencian Community, were compiled in retrospective from 2014 to 2019. Data use was approved by the Institutional Review Board of the GVA and any information that may disclose the identity of the patient was discarded prior to any analysis. Given the data source, our available free text fields were written in Spanish.

A set of preprocessing operations were carried out in order to reduce dimensionality to enhance posterior topic extraction processes. Dispatcher observations were converted to lowercase and then, as text fields were written in Spanish, accents marks were deleted. Punctuation marks were also discarded along with stopwords. Words not appearing at least 50 times in the corpus were dropped, resulting in a vocabulary reduction from 74914 to 4584—discarding 94% of terms— while keeping around 96% of the total word counts in the corpus. Finally, text fields were tokenized.

Data was split using a holdout [6] methodology, with proportions of 80% for training and then 20% for testing. Next, cross-validation [6] splits were conducted over the training set, taking K=4, without allowing repetition.

Topics were inferred from the multinomial probabilities over words conditioned on each topic from a Latent Dirichlet Allocation (LDA) [7] model. We preferred LDA over Latent Semantic Analysis [8] because LDA offers a generative modeling approach, and LDA over Probabilistic Latent Semantic Analysis [9] (PLSA) because the number of parameters estimated in PLSA grows linearly with the number of training documents and generalization to new documents is easier with LDA. LDA is a hierarchical generative Bayesian model, which assumes the existence of K latent topics in a collection of text documents. Next we present the generative process of LDA, to generate a corpus D of M documents each one with N_d words:

1. Choose $N_d \sim Poisson(\xi)$.
2. Choose $\theta_d \sim Dirichlet(\alpha)$.
3. For each of the N_d words w_{di}:
    a. Choose a topic index $z_{di} \sim Multinomial(\theta_d)$, $z_{di} \in \{1, \dots, K\}$.
    b. Choose a word w_{di} from $p(w_{di}|z_{di}, \beta)$, a multinomial probability conditioned on the topic z_{di}.

The LDA model was trained following an online mini-batch variational inference approach [10]. The optimal number of topics was set analyzing the values of a topic coherence measure, where word context vectors were created using the normalized pointwise mutual information [11]. The distance among word context vectors was calculated with the cosine distance, obtaining the final coherence score as the arithmetic mean of all distances, following the procedures described in [12].

We tested different number of topics, specifically 5, 10, 15, 20, 25, 30 and 35. For each combination, we trained four LDA models, one per training K-fold, and calculated the aforementioned topic coherence measure in their respective validation folds. That number of topics offering the best overall performance across the validation K-folds was considered as the optimal number of topics.

Finally, we retrained the model with all the training data using the optimal configuration. For each topic, the most probable words were extracted and studied to infer topic semantics and naming it. After that, we derived the topic distribution in the training and the test corpora, to understand which were the most frequent topics in dispatcher free text fields, as well as to evaluate potential overfitting issues.

3. Results

Figure 1 shows the value of the topic coherence performance metric across the different K-folds, for each number of topics combination:

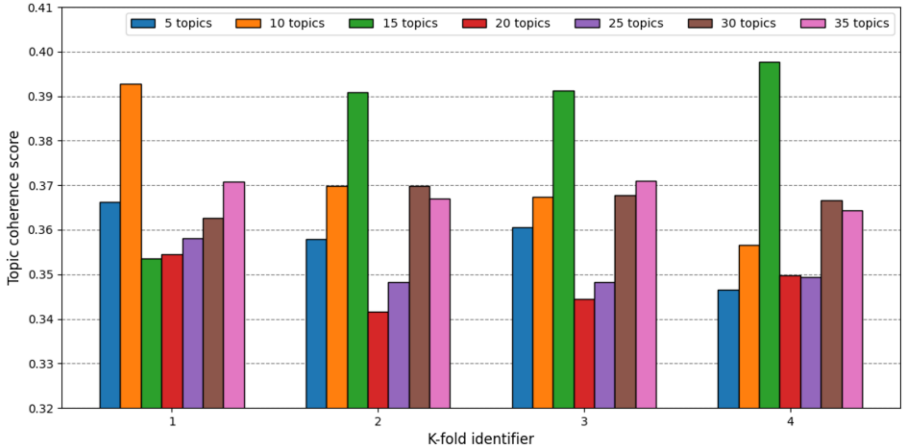

Figure 1. Number of topics selection. Topic coherence across K-folds over training set.

It can be appreciated that the optimal topic coherence value is reached at 15 topics.

Figure 2 displays the 8 words with higher associated probability respect to each topic. Each topic has been named according to the semantics its defining words represent:

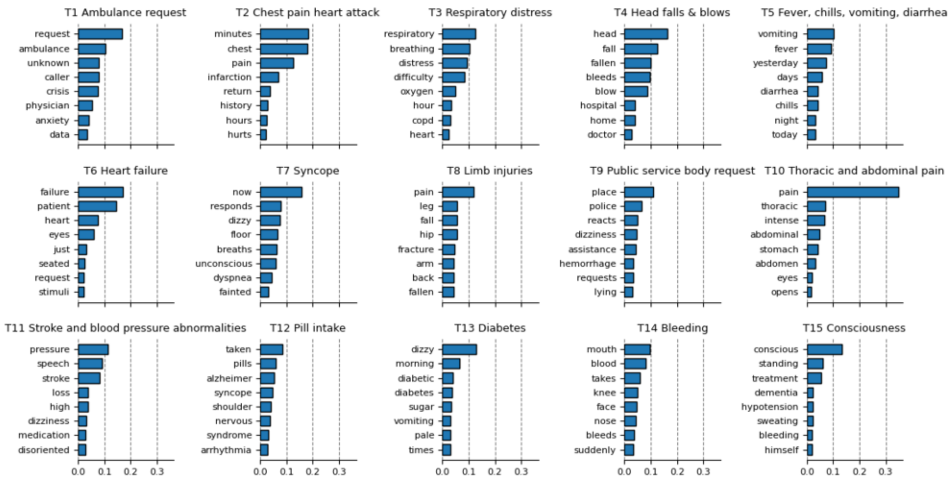

Figure 2. Topics discovered, described by their 8 words with highest probability conditioned on each topic. Word probabilities conditioned on each topic are represented in the x-axis.

It can be observed the presence of 13 clinical topics—T2, T3, T4, T5, T6, T7, T8, T10, T11, T12, T13, T14, T15—along with 2 resource dispatch topics—T1, T9. Likewise, most predominant semantics in the clinical topics are cardiovascular disorders—T2, T6, T11—and injuries T4, T8, T14.

Figure 3 presents the distribution of topics in the training and test corpora, sorted by its frequency of appearance in descending order:

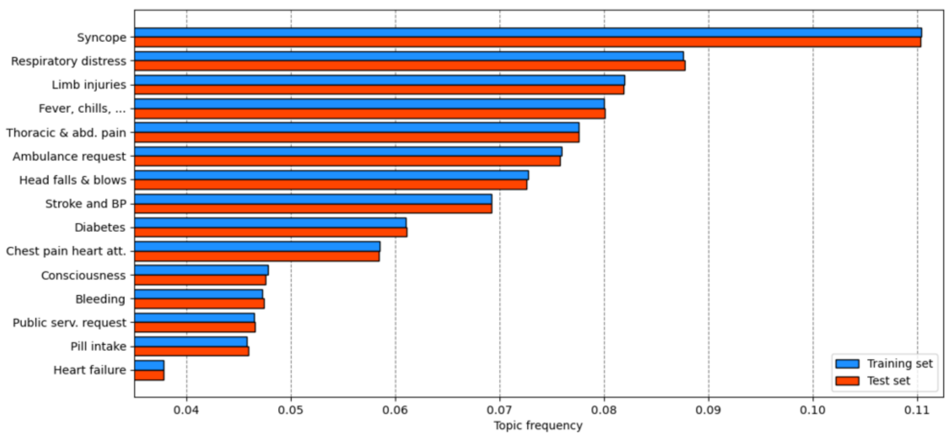

Figure 3. Topics distribution in the train and test corpus, sorted by frequency in descending order.

It can be inferred from this figure that there are not over-represented or under-represented topics. Likewise, there is a strong similarity between training and test topic distributions. Both are good signs indicating that overfitting does not seem to be present.

4. Discussion

The characterization of casuistry latent in complex unstructured data carried out in this work may lead emergency medical professionals to redefine structured decision tree algorithms in order to improve emergency medical dispatch processes.

Although the majority of topics are well-defined and delimited, some topics would require further study to evaluate the presence of topic mixtures and subtopics.

Future work will include studying relations among the topics found and potential clusters bound to the structured variables also registered during the incident. Finally, it is of interest to study why some words appear in different contexts, i.e., topics, despite having similar meanings, such as chest pain and thoracic pain.

5. Conclusions

This work has tackled the discovery of key topics in emergency medical dispatch from free text dispatcher observations. A pipeline comprising word filtering operations, number of topics selection and Latent Dirichlet Allocation model training, has been applied over 1374931 independent retrospective cases from the Valencian emergency medical dispatch service in Spain. Results support the existence of 15 latent topics, whose consideration could lead to the improvement of clinical triage protocols, deriving in turn, in a positive impact in patient wellbeing and health services sustainability.

References

[1] J. J. Clawson and K. B. Dernocoeur, *Principles of emergency medical dispatch*. Salt Lake City, Utah: Priority Press, 2003.
[2] S. J. Stratton, "Triage By Emergency Medical Dispatchers," *Prehosp. Disaster med.*, vol. 7, no. 3, pp. 263–269, Sep. 1992, doi: 10.1017/S1049023X00039601.
[3] P. Ferri *et al.*, "Deep ensemble multitask classification of emergency medical call incidents combining multimodal data improves emergency medical dispatch," *Artificial Intelligence in Medicine*, vol. 117, p. 102088, Jul. 2021, doi: 10.1016/j.artmed.2021.102088.
[4] X. Cheng, Q. Cao, and S. S. Liao, "An overview of literature on COVID-19, MERS and SARS: Using text mining and latent Dirichlet allocation," *Journal of Information Science*, p. 0165551520954674, Aug. 2020, doi: 10.1177/0165551520954674.
[5] J. Pérez, A. Pérez, A. Casillas, and K. Gojenola, "Cardiology record multi-label classification using latent Dirichlet allocation," *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 111–119, Oct. 2018, doi: 10.1016/j.cmpb.2018.07.002.
[6] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, San Francisco, CA, USA, Aug. 1995, pp. 1137–1143.
[7] D. M. Blei, "Latent Dirichlet Allocation," p. 30.
[8] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Am. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, Sep. 1990, doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9.
[9] T. Hofmann, "Probabilistic Latent Semantic Indexing," p. 8.
[10] M. Hoffman, F. R. Bach, and D. M. Blei, "Online Learning for Latent Dirichlet Allocation," p. 9.
[11] M. Röder, A. Both, and A. Hinneburg, "Exploring the Space of Topic Coherence Measures," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Shanghai China, Feb. 2015, pp. 399–408. doi: 10.1145/2684822.2685324.
[12] S. Syed and M. Spruit, "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation," in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct. 2017, pp. 165–174. doi: 10.1109/DSAA.2017.61.