

Semantic Clustering to Augment Qualitative Content Analysis in Exploring Reasons for Emergency Department Transfer Delays

Laura-Maria PELTONEN^{a,1}, Sanna SALANTERÄ^{a,b} and Hans MOEN^c

^a*Department of Nursing Science, University of Turku, Turku, Finland*

^b*Turku University Hospital, Turku, Finland*

^c*Department of Future Technologies, University of Turku, Turku, Finland*

Abstract. The aim of the study was to explore emergency department transfer delays and to assess the potential of using a semantic clustering approach to augment the content analysis of transfer delay data. Data were collected over a period of 5 months from two hospitals. A set of (unique) phrases describing reasons for transfer delays (n=333) were clustered using the k-means with 1) cluster centroids initiated in an unsupervised fashion and 2) a semi-supervised version where the cluster centroids were initiated with keywords. The unsupervised algorithm clustered 77 % and the semi-supervised 86 % of the phrases to suitable clusters. We chose the better performing approach to augment our content analysis. Three main categories for transfer delays were found as a result. These included 1) insufficient staffing resources, 2) transportation and bed issues, and 3) patient and care related reasons. The findings inform the audit of organisational processes, accuracy of staffing and workflow to reduce transfer delays. Future research should explore implications of semantic clustering approaches to other narrative data sets in health service research.

Keywords. Emergency department, health service research, k-means clustering, qualitative content analysis, transfer delay

1. Introduction

A transfer delay can be defined as a situation when a patient is medically ready to be transferred from a unit such as an emergency department (ED) into further care or home, but still occupies a hospital bed. Patient transfer delays are associated with treatment delay and negative patient outcomes, such as increased length of stay, higher hospital mortality [1–2] and higher costs of care [3]. Studies on transfer delays in critical care have reported that organisational issues account for some of the delays [1], with a common reason being insufficient availability of inpatient ward beds [3]. Some transfer delays, such as a deteriorating health condition, cannot be influenced by organisational arrangements, but others could be reduced to improve patient outcomes and reduce costs of care. To date, there is a lack of knowledge about the reasons for patient transfer delays

¹ Corresponding Author, Laura-Maria Peltonen, Department of Nursing Science, 20014 University of Turku, Finland; E-mail: laura-maria.peltonen@utu.fi.

from EDs into further care. Further, there is a need to explore these reasons to reduce them and to find ways to estimate their impact on care provision. Our ultimate goal is to describe the reasons for ED patient transfer delays. We want to test semantic clustering approaches to augment the qualitative content analysis of our narrative data with an inductive and a deductive approach. Although, clustering methods are widely used in other fields (see e.g. [4–5]) there is a lack of literature on the use of these methods on narrative data in health service research. Such methods have the potential to support the analysis of big e.g., hospital wide narrative data sets, which previously have been impossible to analyse as manually processing of large amounts of information is typically difficult and time consuming.

2. Methods

2.1 Data collection

The study had an interrupted time series design. Data were collected in two EDs in Finland for four weeks at the time during five intervals in 2015–2016. Departments were purposefully chosen; one was from the north and the other from the south of the country. Nursing shift leaders manually documented transfer delay reasons. Ethical review was done by the University of Turku Ethics Committee (ID: 13/2015).

2.2 Clustering

We used the word2vec toolkit [6] to train semantic word vectors (embeddings) for each unique word in a large corpus of clinical text consisting of nursing and physician notes from patients admitted to a Finnish university hospital. It consists of 136 million tokens (1.5 million unique tokens). We used NLTK (Natural Language Toolkit for Python) [7] for the pre-processing of our text. Initial testing showed that better scores were achieved when normalizing the words with stemming (Snowball stemmer for Finnish). In addition we performed tokenization, lowercasing and stopword removal. We generated a sentence vector for each documented transfer delay phrase by summing their constituent word vectors, with the addition of first normalizing and multiplying the word vectors by their inverse sentence frequency (c.f. inverse document frequency (IDF)) derived from the training corpus. These vectors were finally used as input into two clustering approaches using the k-means algorithm [8].

- *The unsupervised k-means clustering algorithm with cluster centers initiated in an unsupervised fashion (k-means++).* Here we used the implementation in scikit-learn [9]. We set the expected number of clusters to be generated ($n=8$) based on a consultation with domain experts.
- *K-means clustering with cluster centers initialized with keywords.* As a second clustering approach we tested a semi-supervised clustering approach where we first manually defined keywords and key phrases for each of the eight clusters (one or more per cluster) provided by a domain expert. Next we generated a vector for each cluster centroid by averaging the vector of each keyword (where each keyword vector were generated in the same way as the care delay statements). Finally, we applied k-means with these vectors as cluster centroids.

2.3 Evaluation of automatically generated clusters

We developed a gold standard for the data set by having domain experts manually cluster the transfer delays for our automated cluster evaluation. We used the adjusted Rand index [10] for the automatic evaluation of the generated clusters against the manually made gold standard. This index describes the agreement between two clusters (partitions) as a value or score between -1 and 1 (where 1 equals identical). Although this score can be difficult to interpret directly, it is useful for comparing two or more generated clusters when such a gold standard is available. Finally, we manually assessed the automatically generated clusters. Each item in each cluster was rated on a four-class scale: 1) suits this cluster, 2) suits this or another cluster, 3) suits another cluster but not this, and 4) cannot be analysed due to unclear phrase.

2.4 Synthesis of findings

We used the automatically generated clusters of the better performing clustering approach (i.e. keyword initialized k-means) as a basis for our qualitative content analysis [11-12]. We continued with grouping the developed clusters into higher abstraction levels by merging similar categories by discerning them from those that were dissimilar until we discovered a set of main categories that no longer could be merged. We only focused on the manifest content as the data set consist of free text written by shift leaders.

3. Results

A total of 333 unique reasons were used for the automatic clustering from a set of 600 documented phrases for patient transfer delays. Seven phrases were excluded as they lacked vector representation.

3.1 Performance of the automatic clustering approaches

The automatic evaluation against the gold standard showed an adjusted Rand index score of 0.4508 for the k-means with cluster centers initiated in an unsupervised fashion and 0.5337 when cluster centers were initiated with keywords. A total of 277 out of 333 phrases were assessed to suit the suggested cluster (including classes 1 and 2) when using the keyword initialized clustering based on the manual evaluation, while the respective number for the unsupervised approach was 248 out of 333 (Table 1).

Table 1. Contingency table of the manual evaluation with number of the evaluated phrases per rating class of the unsupervised k-means clustering and keyword initialized k-means clustering results.

Rating class	Unsupervised k-means clustering (n)	Keyword initialized k-means clustering (n)
1) Suits this cluster	154	178
2) Suits this or another cluster	94	99
3) Suits another cluster but not this	73	44
4) Cannot be analysed due to unclear phrase	12	12
1 and 2	248	277

3.2 Reasons for patient transfer delays

Our augmented qualitative content analysis resulted in three main categories of reasons behind transfer delays from EDs into further care. These included 1) insufficient staffing resources, 2) transportation and receiving unit bed issues, and 3) patient and care related issues. The insufficient staffing resources main category included two sub categories that commonly covered examples of waiting for paperwork, such as the electronic health record notes to be completed or a physician's order, as well as waiting for a particular transfer related task to be completed by what seemed to be busy professionals. Individual examples of the busyness of professionals included inability to find time to care for a patient or transfer the patient due to haste. The transportation and bed issues main category also contained two sub categories. A need to wait for permission from the receiving unit to transfer the patient and a lack of space on the receiving unit both within and beyond the hospital were commonly reported issues. The other sub category with frequently reported reasons for waiting regarded transportation means that included examples like waiting for an ambulance or a taxi. The final main category, namely, patient and care related reasons included four sub categories. The first of these focused on the wait of different examination and laboratory values, such as blood samples, x-rays and CT-scans. The second regarded the time for obtaining consultations from specialist, such as neurologists and physiotherapists. The third sub category covered waiting for procedures, such as chest tube drainage or central catheter placement, or follow-up time after a cardioversion. The fourth sub-category included patient related issues, such as a change in a patient's health condition that required attention and re-assessment or a request by the patient for a rest before leaving.

4. Discussion

Content analysis belongs to the most commonly used analysis methods in descriptive qualitative studies in health sciences as it is a feasible method for analysis in many different contexts and settings. But up until now, qualitative content analysis has been limited to the amount of workload possible to be completed by researchers manually. In our study, the clustering algorithms were able to cluster 77–86% of the phrases to suitable clusters. Despite the advantage of the approach where cluster centers are initialized with keywords provided by domain experts, the fully unsupervised approach performed comparable (only 0.08 below in adjusted Rand index). According to our results, it is feasible that both unsupervised and semi-supervised semantic clustering can be used in inductive and deductive qualitative content analyses [12]. An unsupervised approach is particularly useful when prior knowledge and a theoretical framework is lacking and no keywords can be provided. In the future, other clustering approaches could be tested if it is difficult to state a sensible number of clusters and for a more advanced approach to the qualitative content analysis. This includes hierarchical clustering and methods that try to estimate the most sensible number of clusters to be used (see e.g. [13]). It is important to acknowledge that the content analysis process is not free from interpretation [11]. It requires the researcher to be completely familiar with the data, which usually necessitates several iterations of reading the data [12]. A critique against qualitative content analysis states that the method is often used in a simple manner and a deeper data interpretation should be visible [11]. The purpose of using a machine-driven semantic clustering approach is not to shift towards simple or superficial

analyses, but rather to augment the content analysis process when dealing with a large data set that would be difficult to analyse manually. Here, clustering shows potential in developing descriptive categories in the re-contextualisation phase of the analysis process [11]. The findings showed that patient transfer delay reasons in EDs can be classified into insufficient staffing resources, transportation and bed issues, as well as patient and care related reasons. These findings may be used to inform audits and developmental work of organisational processes, staffing adequacy and use, and changes in workflow to reduce transfer delays. This has the potential to improve quality of care and patient outcomes as well as reduce costs of care. It is important to find appropriate ways to analyze, classify and summarize data as ample information is collected in health services every instant. Clustering can be seen as one such approach. More research is warranted to explore the implications of these methods in other narrative health service data sets. Study limitations include the weakness of the manually collected data with missing entries and a possibly skewed representation between data collection months. In conclusion, semantic clustering has the potential to support researches in both inductive and deductive content analysis of big narrative data sets. The results showed organisational process, staffing and workflow -related issues that potentially could be addressed to reduce patient transfer delays from EDs.

Acknowledgements

This research was supported by the Academy of Finland (315376).

References

- [1] Grech C, et al. The delay in transfer between the emergency department and the critical care unit for patients with an acute cardiac event--in hospital factors. *Aust Crit Care*. 2001 Nov;14(4):139-45.
- [2] Chalfin DB, et al. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Crit Care Med*. 2007 Jun;35(6):1477-83.
- [3] Johnson DW, et al. Delay of transfer from the intensive care unit: a prospective observational study of incidence, causes, and financial impact. *Crit Care*. 2013 Jul;17(4):R128.
- [4] Maimon O, Rokach L. Clustering methods. In: Rokach L, Maimon O. editors. *Data mining and knowledge discovery handbook*. Boston: Springer; 2005. p. 321-52.
- [5] Grira N, Crucianu M, Boujemaa N. Unsupervised and semi-supervised clustering: a brief survey. *A review of machine learning techniques for processing multimedia content*. 2004 Jul;1: 9-16.
- [6] Mikolov T, et al. Distributed representations of words and phrases and their compositionality. *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*. Advances in Neural Information Processing Systems 26; 2013 Dec 5-10; Nevada. CA. 2014 p. 3111-9.
- [7] Bird S, Loper E, Klein E. *Natural Language Processing with Python*. O'Reilly Media Inc., Sebastopol, California, USA. 2009.
- [8] Lloyd SP. Least squares quantization in PCM. *Information Theory IEEE Transactions* 1982 28(2):129-37.
- [9] Pedregosa F. et al., Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12: 2825-30.
- [10] Hubert L, Arabie P. Comparing partitions. *J. Classif*. 1985 Dec;2(1):193-218.
- [11] Lindgren BM, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. *Int J Nurs Stud*. 2020 Aug;108:103632.
- [12] Elo S, Kyngäs H. The qualitative content analysis process. *J Adv Nurs* 2008 Apr;62(1):107-15.
- [13] Kriegl HP, Kröger P, Sander J, Zimek A. Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2011 Apr;1(3):231-40.