

Using text mining to explore concepts associated with acute confusion in cardiac patients documentation

Laura-Maria Murtola^{1,2}, Hans Moen³, Lotta Kauhanen¹, Heljä Lundgrén-Laine^{1,2},
Tapio Salakoski⁴ and Sanna Salanterä^{1,2}

¹University of Turku, Department of Nursing Science, Lemminkäisenkatu 1, 20520 Turku, Finland

²Hospital District of Southwest Finland, Kiinamylynkatu 4-8, 20520 Turku, Finland

³Norwegian University of Science and Technology, Department of Computer and Information Science, 7491 Trondheim, Norway

⁴University of Turku, Department of Information Technology, Joukahaisenkatu 3-5 A, 20520 Turku, Finland
lmemur@utu.fi, anloka@utu.fi, hans.moen@idi.ntnu.no, hklula@utu.fi, tapio.salakoski@utu.fi, sansala@utu.fi

1. Introduction

Acute confusion and *delirium* in adult hospitalized patients are severe conditions that lengthen hospital stay, decrease quality of life and increase mortality [1]. Delirium is “an altered state of consciousness accompanied by a change in cognition that develops over a few hours or days and tends to have a fluctuating course” [2]. Acute confusion is considered to be a broader concept than delirium, including anticipative factors to the medical diagnosis [3]. In this research we study acute confusion. Several risk factors may be associated with acute confusion. Elimination or reduction of precipitating factors, such as pain, may decrease acute confusion [3].

Several screening tools based on different criteria are used to detect acute confusion and delirium, such as the Confusion Assessment Method (CAM), the Delirium Rating Scale-Revised-98 (DRS-R98) and the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [4]. But still, the condition remains under diagnosed as up to 70% of cases may pass unnoticed [5]. It is a clinical challenge to recognize patients that suffer from acute confusion, which is also reflected in the documentation related to patient care. Therefore, new methods are needed to aid clinicians in identifying these patients.

Using natural language processing (NLP) and text mining is a potential approach to getting a better understanding of the features of acute confusion. However, Boolean searching through querying with concepts as they are described in dictionaries and medical literature is likely to provide poor results in terms of retrieval coverage. The reason for this would be the specialized languages used in care notes [6], heavily influenced by factors such as the context of care, profession and ward type. A common way of overcoming this problem is to do manual annotation of the text with labels denoting concepts of interest. The next step would then be to train machine-learning algorithms to recognize these annotated labels in unseen text [7]. However, manual annotation is very costly in terms of time and human resources.

Methods for building statistical language models from text, such as random indexing [8], could be utilized to improve coverage and flexibility when querying the data. By analyzing large amounts of unlabeled text, such methods may be used to find near synonyms and concepts corresponding to the manually annotated information and terms found in established vocabularies (c.f. named entities), and possibly identify novel properties and traits related to the condition. In this study, we will thus use a combination of manual annotations, named entities and co-occurrence information, and also investigate the use of topic models derived from the corpus using named entities as seed words.

2. The Aim of the Study and Research Questions

The aim of this study is to identify reported concepts associated with acute confusion in cardiac patient's electronic patient documents. These can be used to train machines to help clinicians to identify acute confusion in the future. The study focuses on the following questions:

- What has been written in patient documents about acute confusion of cardiac patients?
- What common concepts are associated with acute confusion in the documentation?
- What is the connection between the documentation of acute confusion and the theoretical criteria for the diagnosis of acute confusion?

3. Methods

Cardiac patients are at increased risk of postoperative acute confusion [9], therefore we will explore cardiac patients documents as a use case. The data consists of 23,528 cardiac patients' electronic health records that were admitted to one Finnish university hospital between 2005 and 2009 with any type of heart problem. We will use named entities from medical vocabularies (Metatesaurus, FinMeSH, ICD-10 and the thesaurus to search and to index publications of nursing science named Hoidokki), domain experts and query expansion (random indexing). With these terms we will query the text to search for other frequently co-occurring words and concepts. We also plan to look at words co-occurring to these (named entities, annotated words, corpus-level co-occurring words) on sentence level, and possibly also on document level. In this way we aim to extract and explore contextual similar words that are associated with acute confusion to detect undiagnosed patients with symptoms of acute confusion in the data set. Domain experts will then manually evaluate the findings based on defined criteria of acute confusion using a validated instrument (CAM). Cases of acute confusion discovered will be compared with diagnosed cases, which function as baseline.

4. Results

The expected primary outcomes are concepts associated with acute confusion and the prevalence of these concepts in cardiac patients documents. Secondary outcomes may provide information about associations between acute confusion and other still unknown factors.

5. Conclusions

This study aims to identify reported concepts associated with acute confusion in cardiac patients' electronic patient documents to improve the identification of the condition to develop care. Text mining may be one solution to the challenges in the identification of acute confusion and also, other similar situations in the clinical setting. The primary outcomes can be used to develop an automatized search instrument for the detection of acute confusion, to hasten treatment initiation and to improve recovery. Secondary outcomes may provide information to be used in the development of care practices to prevent acute confusion. Appropriate identification of acute confusion and quickly initiated treatment would improve recovery and quality of life, as well as decrease length of hospital stay and reduce costs of care.

References

1. Van Rompaey, B., Schuurmans, M.J., Shortridge-Baggett, L.M., Truijien, S., Elseviers, M., Bossaert, L.: Long term outcome after delirium in the intensive care unit. *J Clin Nurs* **18**(23) (2009) 3349–3357
2. American Psychiatric Association: Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC (2000)
3. Sendelbach, S., Guthrie, P.F.: Evidence-Based Guideline. Acute Confusion/Delirium Identification, Assessment, Treatment, and Prevention. *J Gerontol Nurs* **35**(11) (2009) 11–18
4. Ryan, D.J., O'Regan, N.A., Caoimh, R.Ó., Clare, J., O'Connor, M., Leonard, M., McFarland, J., Tighe, S., O'Sullivan, K., Trzepacz, P.T., Meagher, D., Timmons, S.: Delirium in an adult acute hospital population: predictors, prevalence and detection. *BMJ Open* **3**(1) (2013)
5. Collins, N., Blanchard, MR., Tookman, A., Sampson, EL.: Detection of delirium in the acute hospital. *Age Ageing*. **39**(1) (2010) 131–135
6. Friedman, C., Kra, P., Rzhetsky, A.: Two biomedical sublanguages: a description based on the theories of Zellig Harris. *J Biomed Inform* **35**(4) (2002) 222–235
7. Velupillai, S.: Automatic Classification of Factuality Levels -- A Case Study on Swedish Diagnoses and the Impact of Local Context. In Proc. The Fourth International Symposium on Languages in Biology and Medicine -- LBM 2011, Singapore (2011)
8. Kanerva, P, Kristoferson, J, and Holst, A.: Random indexing of text samples for latent semantic analysis. In Proceedings of the 22nd Annual Conference of the Cognitive Science Society, page 1036, Philadelphia, Pennsylvania. Erlbaum (2000)
9. Koster S, Hensens AG, Schuurmans MJ, van der Palen J.: Risk factors of delirium after cardiac surgery: a systematic review. *Eur J Cardiovasc Nurs* **10**(4) (2011) 197–204