Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 47 (2018) 173–185

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Applying matrix factorization techniques to compare experts' categorization process during case formulation task performed by concept maps

Action editor: Takashi Hashimoto

Piotr Kaczmarek [a,*], Anna Słysz [b]

[a] Poznań University of Technology, Institute of Control and Information Engineering, Poznań, Poland
[b] Adam Mickiewicz University, Institute of Psychology, Poznań, Poland

Received 29 November 2016; received in revised form 16 August 2017; accepted 4 October 2017
Available online 13 October 2017

Abstract

The aim of this paper was to present a method to enable the analysis of the process of categorization of patients' testimonials and the comparison of individual categories created by professionals. A complex diagnostic task (case conceptualization) was employed to study the categorization function in professional thinking. Two groups of psychotherapists (30 people in each group) served as subjects of the research. The main objective of the study was to find an appropriate representation of concept maps enabling a comparison of both the categories and the structures between experts. In the comparison process, only the information about the premises justifying each given category was taken into account and represented by a concept-testimonials matrix. Three different elements weighting schemes and matrix factorization-based unsupervised clustering methods were analyzed in the context of consistency and ability to establish main semantic groups of concepts common to the majority of experts. Moreover, special attention was paid to determining the number of main semantic classes. The study showed that even the used representation was similar to the task of documents indexing there was some discrepancy. The highest accuracy in generating main semantic groups was achieved using the PCA and K-Means (*nKM*) (the average false positive rate in clusters was 32%). This method outperformed Tempered PLSA (the average false positive rate per cluster was 52%). It was demonstrated that in analyzed task the *nKM* method allowed comparing the similarity of concepts even when they were created by various experts using different conceptual apparatus.
© 2017 Elsevier B.V. All rights reserved.

Keywords: Categorization; Case conceptualization; Concept map; Concept maps comparison; Matrix factorization; K-means; PCA; PLSA

1. Introduction

Categorization is the natural way of recognizing things and experiences. Two essential dimensions are used in psychological analyses to characterize notional categories: the scope of a category (broad vs. narrow) and its complexity (small vs. large number of cognitive perspectives taken into consideration in the thinking process) (Massaro & Ferguson, 1993). The applications of notional categories include ordering the thoughts of experts in various fields. One of the fields where psychotherapists use cognitive categorization is case conceptualization. The case conceptualization competence is one of the technical diagnostic skills. It consists of reaching an adequate assessment of a given person's problem and forming hypotheses concerning the causes and psychological, interpersonal or behavioral

* Corresponding author.
E-mail addresses: piotr.kaczmarek@put.poznan.pl (P. Kaczmarek), aslysz@amu.edu.pl (A. Słysz).

https://doi.org/10.1016/j.cogsys.2017.10.001
1389-0417/© 2017 Elsevier B.V. All rights reserved.

determinants of the diagnosed patient's problems; its elements include classification of information (categorization) and creation of a coherent enough image of the patient/-client (Eells and Eells, 1997). The analysis of diagnosing accuracy usually considers final stage of a diagnostic process namely correctness of nosological diagnosis. Whereas there are only a few studies related to the correctness of functional diagnosis enabling to discover and assess components of the diagnostic process (Bruchmuller, Margraf, & Schneider, 2012; Mitchell, Vaze, & Rao, 2009; O'neill et al., 2008). The aim of this paper is to present a method to enable the analysis of elementary cognitive processes, in particular, the process of categorization of patients' testimonials and the comparison of individual categories created by professionals.

In the present work, we used concept maps to study case conceptualizations performed by experts (psychotherapists). Concept maps are a tool for representing knowledge in a given field (Novak, 2010). They take into account mutual relations between ideas. This kind of maps is used in the field of education to evaluate the knowledge of students or to activate students in the learning process (Safar, Jafer, & Alqadiri, 2014). In the majority of applications, concept maps represent a well-defined set of keywords and terms and relations between them (Novak, 2010). In this case, the connections explain a given unknown concept based on its relationships with other terms. A relationship is usually defined by a linking word or a phrase. Such a representation allows evaluating the knowledge of subjects and detecting erroneous understanding of the concepts. Concept maps have also been used to study the structure of expert knowledge for case conceptualization (Mayfield, Kardash, & Kivlighan, 1999). In this case, concept maps describe a problem, and concept represents a state, premise or other factors of different kind referring to the problem. Thus a relation between two terms does not strictly explain their meaning or semantic relationship but mostly describes the role and importance to the problem (the case) being an underlying factor. So far, comparison of maps has been performed quantitatively by calculating factors referring to their topology or their building process (building time, building strategy, etc.) (Mayfield et al., 1999; Williams, 1998). A more extensive comparison and validation of concept maps describing case conceptualization require two fundamental problems to be solved. The first involves comparing categories (concepts). The second demands a method for analyzing a map structure where the categories defined by experts are not strictly comparable or might differ in their semantic fields (i.e., depending on a map, some categories might be very specific, while other are general, or some concepts might overlap with others). In the present paper, the first issue is addressed. It is the first stage of the development of a method for automatic analysis and comparison of concept maps. The natural method of categories comparison is a semantic analysis of category names, yet in this case, such an analysis needs to be performed by another expert, and its results are difficult to verify. Moreover, it can be expected that experts belonging to different therapeutic schools would use different conceptual apparatuses to refer to the same semantic fields, which may necessitate the inclusion of therapeutic schools in the semantic analysis as well, making it even more arbitrary.

The unsupervised clustering requires a model which let to represent categorized items in unified space. In application to documents clustering and also in computer vision usually, the bag-of-words models are used. These models describe each item by a vector of terms frequency. The terms frequency can be used directly or weighted to reflect the importance of given term for the document (e.g. using tf-idf method) (Robertson, 2004). In general, the bag-of-words model (BoWM) neglects relationship other than co-occurrence of the features. In some modifications for text classification, N-grams can be included to overcome this limitation. Similarly, in recommender systems where the collaborative filtering is used the model of users and products is described by a set of recommendations (Su & Khoshgoftaar, 2009). In all mentioned above applications, the model is described by sparse, high-dimensional $N \times M$ observations matrix where N is number of items (i.e. articles, products, users, etc.) and M is number of features. As the features space usually is sparse and non-orthogonal, direct comparison of items or application of similarity-based clustering methods (such as K-means) are not feasible. Therefore matrix factorization techniques can be used to reduce space dimensionality and sparseness, to orthogonalize observations matrix, or to infer latent variables. There are several matrix factorization techniques used to documents indexing, namely singular values decomposition (SVD), probabilistic latent semantic analysis (PLSA), non-negative matrix factorization (NMF), Latent Dirichlet Allocation (LDA) or Principal Component Analysis (PCA) together with K-means (KM) clustering algorithms. All of them let to discover latent variables which can be interpreted as documents topics. It is not clear which of these methods is the best, and it seems that the performance might be dependent on application and dataset properties. These methods can be characterized concerning their coherence, classification accuracy in reduced space and overfitting tendency. Coherence reflects an ability to discover latent variables which agree with semantic categories created by human judges. In this context, lower perplexity is presented by NMF, PLSA, and LDA than by SVD (Hofmann, 2001; Kakkonen, Myller, & Sutinen, 2006; Stevens, Kegelmeyer, Andrzejewski, & Buttler, 2012). Although the PCA is not able to discover coherent latent variables, it may facilitate dimensionality reduction of initial data and thus in co-operation with other matrix factorization methods improve low-rank approximation (Vozalis & Margaritis, 2007). It was also shown that in most cases direct clustering using latent variables generated by NMF and PLSA works better than clustering using K-means (Ding, Li, & Peng, 2008) or indexing based on KNN method (Kakkonen et al., 2006). A very

Download English Version:

https://daneshyari.com/en/article/6853815

Download Persian Version:

https://daneshyari.com/article/6853815

Daneshyari.com