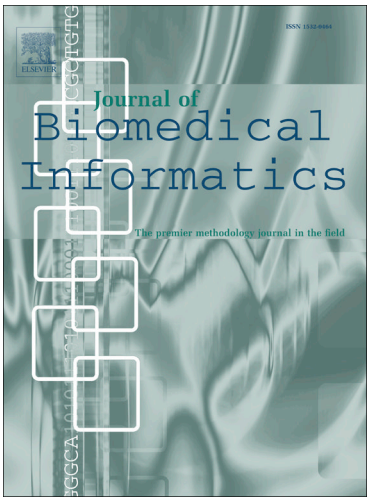

Automatic Recognition of Symptom Severity From Psychiatric Evaluation Records

Travis R. Goodwin, Ramon Maldonado, Sanda M. Harabagiu

PII: S1532-0464(17)30119-3
DOI: http://dx.doi.org/10.1016/j.jbi.2017.05.020
Reference: YJBIN 2792

To appear in: *Journal of Biomedical Informatics*

Accepted Date: 26 May 2017

Please cite this article as: Goodwin, T.R., Maldonado, R., Harabagiu, S.M., Automatic Recognition of Symptom Severity From Psychiatric Evaluation Records, *Journal of Biomedical Informatics* (2017), doi: http://dx.doi.org/10.1016/j.jbi.2017.05.020

Automatic Recognition of Symptom Severity From Psychiatric Evaluation Records

Travis R. Goodwin*, Ramon Maldonado, Sanda M. Harabagiu

Human Language Technology Research Institute, Department of Computer Science, The University of Texas at Dallas, Richardson, TX, USA

Abstract

This paper presents a novel method for automatically recognizing symptom severity by using natural language processing of psychiatric evaluation records to extract features that are processed by machine learning techniques to assign a severity score to each record evaluated in the 2016 RDoC for Psychiatry Challenge from CEGS/N-GRID. The natural language processing techniques focused on (a) discerning the discourse information expressed in questions and answers; (b) identifying medical concepts that relate to mental disorders; and (c) accounting for the role of negation. The machine learning techniques rely on the assumptions that (1) the severity of a patient's positive valence symptoms exists on a latent continuous spectrum; and (2) all the patient's answers and narratives documented in the psychological evaluation records are informed by the patient's latent severity score along this spectrum. These assumptions motivated our two-step machine learning framework for automatically recognizing psychological symptom severity. In the first step, the latent continuous severity score is inferred from each record; in the second step, the severity score is mapped to one of the four discrete severity levels used in the CEGS/N-GRID challenge. We evaluated three methods for inferring the latent severity score associated with each record: (i) pointwise ridge regression; (ii) pairwise comparison-based classification; and (iii) a hybrid approach combining pointwise regression and the pairwise classifier. The second step was implemented using a tree of cascading support vector machine (SVM) classifiers. While the official evaluation results indicate that all three methods are promising, the hybrid approach not only outperformed the pairwise and pointwise methods, but also produced the second highest performance of all submissions to the CEGS/N-GRID challenge with a normalized MAE score of 84.093% (where higher numbers indicate better performance). These evaluation results enabled us to observe that, for this task, considering pairwise information can produce more accurate severity scores than pointwise regression – an approach widely used in other systems for assigning severity scores. Moreover, our analysis indicates that using a cascading SVM tree outperforms traditional SVM classification methods for the purpose of determining discrete severity levels.

1. Introduction

In 2008, the National Institute of Mental Health (NIMH) included an aim to "[d]evelop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures" in its new strategic plan. The framework that implements this aim was named the Research Domain Criteria project, or RDoC. RDoC fostered research that integrates multiple forms of information to better understand all the dimensions of functioning underlying the full range of patient behavior, from normal to abnormal. As reported in Cuthbert [1], an NIMH workgroup was convened in early 2009 to devise an approach for RDoC. The workgroup determined five major domains of functioning: (1) negative va-

*Corresponding author

Email addresses: travis@utdallas.edu (Travis R. Goodwin), ramon@utdallas.edu (Ramon Maldonado), sanda@utdallas.edu (Sanda M. Harabagiu)

URL: www.hlt.utdallas.edu/~travis (Travis R. Goodwin), www.hlt.utdallas.edu/~ramon (Ramon Maldonado), www.hlt.utdallas.edu/~sanda (Sanda M. Harabagiu)