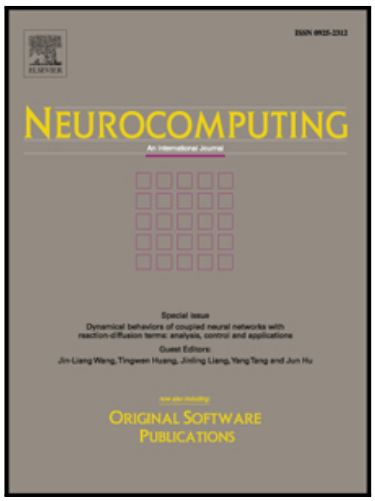
Accepted Manuscript



Discriminant document embeddings with an extreme learning machine for classifying clinical narratives

Paula Lauren, Guangzhi Qu, Feng Zhang, Amaury Lendasse

PII: S0925-2312(17)31415-7
DOI: 10.1016/j.neucom.2017.01.117
Reference: NEUCOM 18805

To appear in: *Neurocomputing*

Received date: 8 July 2016
Revised date: 31 December 2016
Accepted date: 3 January 2017

Please cite this article as: Paula Lauren, Guangzhi Qu, Feng Zhang, Amaury Lendasse, Discriminant document embeddings with an extreme learning machine for classifying clinical narratives, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.01.117

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Discriminant document embeddings with an extreme learning machine for classifying clinical narratives

Paula Lauren[a,], Guangzhi Qu[a], Feng Zang[b], Amaury Lendasse[c,d]

[a] *Computer Science and Engineering Department, Oakland University, Rochester, USA*
[b] *China University of Geosciences, Wuhan, China*
[c] *Department of Mechanical and Industrial Engineering The University of Iowa, Iowa City, USA*
[d] *Department of Business Management and Analytics, Arcada University of Applied Sciences, Helsinki, Finland*

Abstract

The unstructured nature of clinical narratives makes them complex for automatically extracting information. Feature learning is an important precursor to document classification, a sub-discipline of natural language processing (NLP). In NLP, word and document embeddings are an effective approach for generating word and document representations (vectors) in a low-dimensional space. This paper uses skip-gram and paragraph vectors-distributed bag of words (PV-DBOW) with multiple discriminant analysis (MDA) to arrive at discriminant document embeddings. A kernel-based extreme learning machine (ELM) is used to map the clinical texts to the medical code. Experimental results on clinical texts indicate overall improvement especially for the minority classes.

Keywords: Document classification, Feature learning, Word embeddings, Document embeddings, Skip-gram, PV-DBOW, Multiple discriminant analysis, Extreme learning machines, Clinical narratives

1. Introduction

Clinical narratives contain the clinical encounter as observed by the healthcare professional with a patient. The data from clinical narratives enable qual-

*Corresponding author
Email address: gqu@oakland.edu (Guangzhi Qu)

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/6864735

Download Persian Version:

https://daneshyari.com/article/6864735

Daneshyari.com