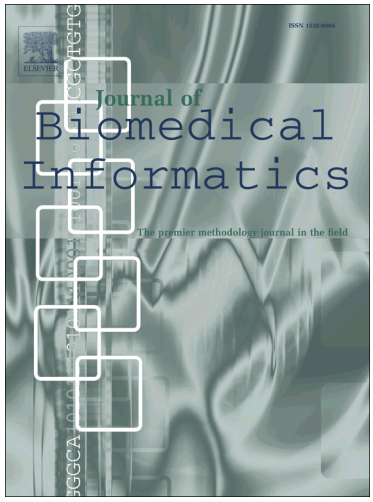

Stress Modelling and Prediction in Presence of Scarce Data

Alban Maxhuni, Pablo Hernandez-Leal, L. Enrique Sucar, Venet Osmani, Eduardo F. Morales, Oscar Mayora

PII: S1532-0464(16)30109-5
DOI: http://dx.doi.org/10.1016/j.jbi.2016.08.023
Reference: YJBIN 2631

To appear in: *Journal of Biomedical Informatics*

Received Date: 21 February 2016
Revised Date: 18 August 2016
Accepted Date: 26 August 2016

Please cite this article as: Maxhuni, A., Hernandez-Leal, P., Enrique Sucar, L., Osmani, V., Morales, E.F., Mayora, O., Stress Modelling and Prediction in Presence of Scarce Data, *Journal of Biomedical Informatics* (2016), doi: http://dx.doi.org/10.1016/j.jbi.2016.08.023

Stress Modelling and Prediction in Presence of Scarce Data

Alban Maxhuni[a,*], Pablo Hernandez-Leal[c,d], L. Enrique Sucar[c], Venet Osmani[b], Eduardo F. Morales[c], Oscar Mayora[b]

[a] *DISI, University of Trento*
Via Sommarive 9, Povo - Trento, Italy
[b] *CREATE-NET*
Via alla Cascata 56/D Povo - Trento, Italy
[c] *Instituto Nacional de Astrofísica, Óptica y Electrónica*
Luis Enrique Erro #1, Sta. María Tonantzintla, Puebla, México
[d] *Centrum Wiskunde & Informatica*
Science Park 123, Amsterdam, The Netherlands

Abstract

Objective: Stress at work is a significant occupational health concern. Recent studies have used various sensing modalities to model stress behaviour based on non-obtrusive data obtained from smartphones. However, when the data for a subject is scarce it becomes a challenge to obtain a good model.
Methods: We propose an approach based on a combination of techniques: semi-supervised learning, ensemble methods and transfer learning to build a model of a subject with scarce data. Our approach is based on the comparison of decision trees to select the closest subject for knowledge transfer.
Results: We present a real-life, unconstrained study carried out with 30 employees within two organisations. The results show that using information (instances or model) from *similar* subjects can improve the accuracy of the subjects with scarce data. However, using transfer learning from dissimilar subjects can have a detrimental effect on the accuracy. Our proposed ensemble approach increased the accuracy by $\approx 10\%$ to 71.58% compared to not using any transfer learning technique.
Conclusions: In contrast to high precision but highly obtrusive sensors, using smartphone sensors for measuring daily behaviours allowed us to quantify behaviour changes, relevant to occupational stress. Furthermore, we have shown that use of transfer learning to select data from close models is a useful approach to improve accuracy in presence of scarce data.

Keywords: Stress modelling, Transfer learning, Semi-supervised learning, Ensemble methods

1. Introduction

Stress is a physiological response to mental, emotional, or other physical challenges that humans confront in their real-life activities, including in their working environments. Continuous exposure to

*Corresponding author
Email addresses: alban.maxhuni@disi.unitn.it (Alban Maxhuni), Pablo.Hernandez@cwi.nl (Pablo Hernandez-Leal), esucar@inaoep.mx (L. Enrique Sucar), venet.osmani@create-net.org (Venet Osmani), emorales@inaoep.mx (Eduardo F. Morales), omayora@create-net.org (Oscar Mayora)