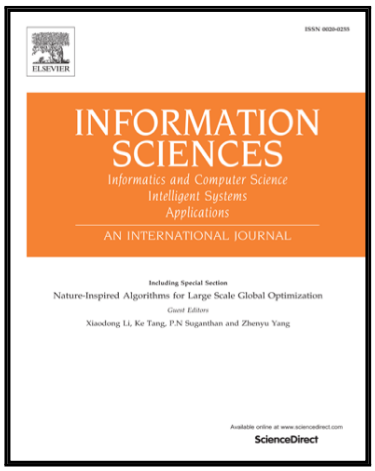

Positive unlabeled learning for building recommender systems in a parliamentary setting

Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete, Luis Redondo-Expósito

PII: S0020-0255(17)31167-2
DOI: 10.1016/j.ins.2017.12.046
Reference: INS 13340

To appear in: *Information Sciences*

Received date: 4 May 2017
Revised date: 22 September 2017
Accepted date: 25 December 2017

Please cite this article as: Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete, Luis Redondo-Expósito, Positive unlabeled learning for building recommender systems in a parliamentary setting, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.12.046

Positive unlabeled learning for building recommender systems in a parliamentary setting

Luis M. de Campos[a,*], Juan M. Fernández-Luna[a], Juan F. Huete[a], Luis Redondo-Expósito[a]

[a] *Departamento de Ciencias de la Computación e Inteligencia Artificial, ETSI Informática y de Telecomunicación, CITIC-UGR, Universidad de Granada, 18071 Granada, Spain*

Abstract

Our goal is to learn about the political interests and preferences of Members of Parliament (MPs) by mining their parliamentary activity in order to develop a recommendation/filtering system to determine how relevant documents should be distributed among MPs. We propose the use of positive unlabeled learning to tackle this problem since we only have information about relevant documents (the interventions of each MP in debates) but not about irrelevant documents and so it is not possible to use standard binary classifiers which have been trained with positive and negative examples. Additionally, we have also developed a new positive unlabeled learning algorithm that compares favourably with: a) a baseline approach which assumes that every intervention by any other MP is irrelevant; b) another well-known positive unlabeled learning method; and c) an approach based on information retrieval methods that matches documents and legislators' representations. The experiments have been conducted with data from the regional Spanish Andalusian Parliament.

Keywords: positive unlabeled learning, content-based recommender systems, parliamentary documents, k-means, support vector machines

1. Introduction

In the information society we live in today, enterprises, institutions and individuals can easily access vast amounts of information. In many cases, users do not need to actively search for the information they need but play a more passive role and are constantly bombarded with advertising, news, e-mails, etc. The problem is then to separate the wheat from the chaff so as to determine what is interesting, important or useful and what is not. This task is hard

*Corresponding author

Email addresses: lci@decsai.ugr.es (Luis M. de Campos), jmfluna@decsai.ugr.es (Juan M. Fernández-Luna), jhg@decsai.ugr.es (Juan F. Huete), luisre@decsai.ugr.es (Luis Redondo-Expósito)