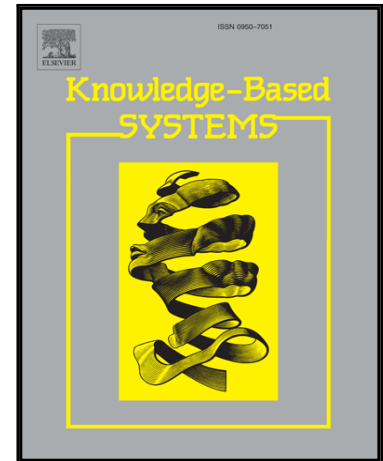


Accepted Manuscript

Cross-ratio uninorms as an effective aggregation mechanism in Sentiment Analysis

Orestes Appel, Francisco Chiclana, Jenny Carter, Hamido Fujita

PII: S0950-7051(17)30112-0
DOI: [10.1016/j.knosys.2017.02.028](https://doi.org/10.1016/j.knosys.2017.02.028)
Reference: KNOSYS 3842



To appear in: *Knowledge-Based Systems*

Received date: 9 December 2016
Revised date: 24 February 2017
Accepted date: 25 February 2017

Please cite this article as: Orestes Appel, Francisco Chiclana, Jenny Carter, Hamido Fujita, Cross-ratio uninorms as an effective aggregation mechanism in Sentiment Analysis, *Knowledge-Based Systems* (2017), doi: [10.1016/j.knosys.2017.02.028](https://doi.org/10.1016/j.knosys.2017.02.028)

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.

Cross-ratio uninorms as an effective aggregation mechanism in Sentiment Analysis

Orestes Appel^{a,b}, Francisco Chiclana^a, Jenny Carter^a, Hamido Fujita^c

^a*Centre for Computational Intelligence, De Montfort University, Leicester, UK*

^b*Bissett School of Business, Mount Royal University, Calgary, Canada*

^c*Iwate Prefectural University, Takizawa, Iwate, Japan*

Abstract

There are situations in which lexicon-based methods for Sentiment Analysis (SA) are not able to generate a classification output for specific instances of a given dataset. Most often, the reason for this situation is the absence of specific terms in the sentiment lexicon required in the classification effort. In such cases, there were only two possible paths to follow: (1) add terms to the lexicon (off-line process) by human intervention to guarantee no noise is introduced into the lexicon, which prevents the classification system to provide an immediate answer; or (2) use the services of a word-frequency dictionary (on-line process), which is computationally costly to build. This paper investigates an alternative approach to compensate for the lack of ability of a lexicon-based method to produce a classification output. The method is based on the combination of the classification outputs of non lexicon-based tools. Specifically, firstly the outcome values of applying two or more non-lexicon classification methods are obtained. Secondly, these non-lexicon outcomes are fused using a uninorm based approach, which has been proved to have desirable compensation properties as required in the SA context, to generate the classification output the lexicon based approach is unable to achieve. Experimental results based on the execution of two well-known supervised machine learning algorithms, namely Naïve Bayes and Maximum Entropy, and the application of a cross-ratio uninorm operator are presented. Performance indices associated to options (1) and (2) above are compared against the results obtained using the proposed approach for two different datasets. Additionally, the performance of the proposed cross-ratio uninorm operator based approach is also compared when the aggregation operator used is the arithmetic mean instead. It is shown that the combination of non lexicon-based classification methods with specific uninorm operators improves the classification performance of lexicon-based methods, and it enables the offering of an alternative solution to the SA classification problem when needed. The proposed aggregation method could be used as well as a replacement of ensemble averaging techniques commonly applied when combining the results of several machine learning classifiers' outputs.

Keywords: Cross-ratio Uninorms, Semantic Orientation Aggregation, Hybrid Sentiment Analysis, Supervised Machine Learning, Naïve Bayes, Maximum Entropy

1. Introduction

Nowadays, understanding the opinions being conveyed by a variety of sources is an important task. The rates of volume and speed at which these waves of information are delivered to us

Email addresses: orestes.appel@my365.dmu.ac.uk (Orestes Appel), oappel@mtroyal.ca (Orestes Appel), chiclana@dmu.ac.uk (Francisco Chiclana), jennyc@dmu.ac.uk (Jenny Carter), issam@iwate-pu.ac.jp (Hamido Fujita)

Download English Version:

<https://daneshyari.com/en/article/4946182>

Download Persian Version:

<https://daneshyari.com/article/4946182>

[Daneshyari.com](https://daneshyari.com)