

Accepted Manuscript

Semi-Supervised Learning for Big Social Data Analysis

Amir Hussain, Erik Cambria

PII: S0925-2312(17)31636-3
DOI: [10.1016/j.neucom.2017.10.010](https://doi.org/10.1016/j.neucom.2017.10.010)
Reference: NEUCOM 18990

To appear in: *Neurocomputing*

Received date: 1 May 2017
Revised date: 6 October 2017
Accepted date: 6 October 2017

Please cite this article as: Amir Hussain, Erik Cambria, Semi-Supervised Learning for Big Social Data Analysis, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2017.10.010](https://doi.org/10.1016/j.neucom.2017.10.010)



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.

Semi-Supervised Learning for Big Social Data Analysis

Amir Hussain^a, Erik Cambria^{b,c}

^a*Department of Computing Science and Mathematics, University of Stirling, UK*

^b*School of Computer Science and Engineering, Nanyang Technological University, Singapore*

^c*Corresponding author*

Abstract

In an era of social media and connectivity, web users are becoming increasingly enthusiastic about interacting, sharing, and working together through online collaborative media. More recently, this collective intelligence has spread to many different areas, with a growing impact on everyday life, such as in education, health, commerce and tourism, leading to an exponential growth in the size of the social Web. However, the distillation of knowledge from such unstructured Big data is, an extremely challenging task. Consequently, the semantic and multimodal contents of the Web in this present day are, whilst being well suited for human use, still barely accessible to machines. In this work, we explore the potential of a novel semi-supervised learning model based on the combined use of random projection scaling as part of a vector space model, and support vector machines to perform reasoning on a knowledge base. The latter is developed by merging a graph representation of commonsense with a linguistic resource for the lexical representation of affect. Comparative simulation results show a significant improvement in tasks such as emotion recognition and polarity detection, and pave the way for development of future semi-supervised learning approaches to big social data analytics.

1. Introduction

With the advent of social networks, web communities, blogs, Wikipedia, and other forms of online collaborative media, the way people express their opinions and sentiments has radically changed in recent years [1]. These new tools have facilitated the creation of original content, ideas, and opinions, connecting millions of people through the World Wide Web, in a financially and labour-effective manner. This has made a huge source of information and opinions easily available by the mere click of a mouse.

As a result, the distillation of knowledge from this huge amount of unstructured information comes into vital play for marketers looking to create and shape brand and product identities. The practical purpose this encapsulates has led to the emerging field of big social data analysis, which deals with information retrieval and knowledge discovery from natural language and social networks using graph mining and natural language processing (NLP) techniques to distill knowledge and opinions from the huge amount of information on the World Wide Web. Sentic computing [2] tackles these crucial issues by exploiting affective commonsense reasoning, modeled upon the intrinsically human capacity to interpret cognitive and affective information associated with natural language, so as to infer new knowledge and make decisions in connection with one's social and emotional values, censors, and ideals. In other words, we can say that commonsense computing techniques are applied to narrow the semantic gap between word-level natural language data and the concept-level opinions conveyed by these.

In the past, graph mining techniques and multi-dimensionality reduction techniques [3] were employed on a knowledge base obtained by merging ConceptNet [4], a directed graph representation of commonsense knowledge, with WordNet-Affect (WNA) [5], a linguistic resource for the lexical representation of affect. Our research fits within the sentic computing framework and aims to exploit machine learning for developing a cognitive model for emotion recognition in natural language text. Unlike purely syntactical techniques, concept-based approaches can

Email addresses: ahu@cs.stir.ac.uk (Amir Hussain), cambria@ntu.edu.sg (Erik Cambria)

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات