Chapter 12
Learning Algorithms of Sentiment Analysis:
A Comparative Approach to Improve Data Goodness

Suania Acampa
University of Naples Federico II, Italy

Ciro Clemente De Falco
University of Naples Federico II, Italy

Domenico Trezza
University of Naples Federico II, Italy

ABSTRACT

The uncritical application of automatic analysis techniques can be insidious. For this reason, the scientific community is very interested in the supervised approach. Can this be enough? This chapter aims to these issues by comparing three machine learning approaches to measuring the sentiment. The case study is the analysis of the sentiment expressed by the Italians on Twitter during the first post-lockdown day. To start the supervised model, it has been necessary to build a stratified sample of tweets by daily and classifying them manually. The model to be test provides for further analysis at the end of the process useful for comparing the three models: index will be built on the tweets processed with the aim of detecting the goodness of the results produced. The comparison of the three algorithms helps the authors to understand not only which is the best approach for the Italian language but tries to understand which strategy is to verify the quality of the data obtained.

The work is the result of a joint work of the three authors, however the paragraph "Big corpora and Digital Methods: a critical approach to improve data goodness" and "Sentiment Analysis and Main Text Classification Algorithms" are by Suania Acampa; the paragraph "Supervised Learning Algorithms used in the analysis "and" Index Scores and Tweet Characteristics "are by Ciro Clemente De Falco; the

paragraph "Data and Methods to compare some Supervised Algorithms" and the paragraph "Analysis. Path, Models and Comparisons" are by Domenico Trezza.

INTRODUCTION

Big Corpora and Digital Methods: A Critical Approach to Improve Data Goodness

The ubiquity of digital technologies and the popularity of opinion-rich platforms such as social media and review sites generates a large and rapid amount of user-generated data encoded in natural language daily. Reviews, tweets, likes, links, shares, texts, posts, tags etc.; these are only part of the billions of digital traces that we leave on the web every day, through which it is possible to accurately trace the tastes, opinions, and attitudes of everyone. Big corpora represent a profitable empirical basis for all those who investigate social phenomena on the net. The production and increasing availability of data offers new possible forms of knowledge of social complexity that social researchers cannot ignore. The data revolution is considered as *"the sum of the disruptive social and technological changes that are transforming the routine of construction, management and analysis of data consolidated within the various scientific disciplines"* (Amaturo, Aragona, 2017, p.1). The new digital technologies and big data allow social research to move from the construction of empirical bases through interrogation to the construction of empirical bases through survey. Big data allows us to measure complex phenomena in detail in real time, thanks to the evolution of IT tools and techniques such as artificial intelligence, machine learning, and natural language processing. This promotes interdisciplinarity between different scientific areas and provides social researchers solid empirical bases for experimenting and integrating new and traditional approaches to social research. These technologies push the social sciences into a scenario in which *"web-mediated research [...] is already transforming the way researchers practice traditional research methods transposed to the web"* (Amaturo and Punziano, 2016, 35, 36).

To be able to describe and analyse this wealth of information, social scientists have also begun to use computational analytical methods to assemble, filter and interpret user generated data encoded in natural language. Text mining is part of this context, a branch of data mining that allows you to analyse vast textual corpora in different languages by extracting high quality information with very limited manual intervention. Natural language processing (NLP) is the area of machine learning dedicated to the meaning of the written word.

A very profitable branch of natural language processing is sentiment analysis: it consists in the extraction and analysis of the opinions that users express on the web towards products, services, topics or characters. With language processing and text analysis, sentiment analysis identifies subjective information in sources. The main objective is to determine the general polarity of a text (whether it is a review or a comment) and classify it into three categories: positive, negative or neutral. Sentiment analysis techniques are divided according to the type of approach used: lexicon based or machine learning approach. The machine learning approach treats sentiment classification as a question of general text classification. This approach to classification is divided between unsupervised and supervised learning models. In supervised models it is necessary to arrange a training set labelled with the indication of the polarity of the feeling (negative, positive, neutral) that the algorithm will use to predict the polarity of other textual content contained in the test set. The machine learning approach has the advantage of not

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/learning-algorithms-of-sentiment-analysis/287459

Related Content

Melbourne's Advanced Rail Transportation: Innovative Systems and Their Future Perspective
Koorosh Gharehbaghi, Ken Farnesand Matt Myers (2020). *International Journal of Strategic Engineering (pp. 24-36).*
www.irma-international.org/article/melbournes-advanced-rail-transportation/255140

Supporting Equity Through Mathematical Writing
Christie Sullivan Martinand Drew Polly (2024). *Illuminating and Advancing the Path for Mathematical Writing Research (pp. 174-192).*
www.irma-international.org/chapter/supporting-equity-through-mathematical-writing/334150

An Overview of Disaster and Emergency Management Systems Models
Dilshad Sarwar (2018). *International Journal of Strategic Engineering (pp. 24-37).*
www.irma-international.org/article/an-overview-of-disaster-and-emergency-management-systems-models/196602

An Overview of Disaster and Emergency Management Systems Models
Dilshad Sarwar (2018). *International Journal of Strategic Engineering (pp. 24-37).*
www.irma-international.org/article/an-overview-of-disaster-and-emergency-management-systems-models/196602

Community Empowerment Initiatives of Faith-Based NGOs: A Case Study in Bangladesh
Shofiqur Rahman Chowdhuryand Haris Abd Wahab (2019). *Social Research Methodology and New Techniques in Analysis, Interpretation, and Writing (pp. 75-105).*
www.irma-international.org/chapter/community-empowerment-initiatives-of-faith-based-ngos/220332