

Implementation of Text Mining in Socio-Economic Research

Konstantin Malysenko, Humanitarian and Pedagogical Academy, Crimean Federal University Named After V.I. Vernadsky, Russia*

Vadim Malysenko, Humanitarian and Pedagogical Academy, Crimean Federal University Named After V.I. Vernadsky, Russia

Marina Anashkina, Humanitarian and Pedagogical Academy, Crimean Federal University Named After V.I. Vernadsky, Russia

Dmitry Anashkin, Crimean Federal University Named After V.I. Vernadsky, Russia

ABSTRACT

This work aims to analyze insights from social networks for identification of population satisfaction with pay level in Russia using the text mining approach. For this, a sentiment analysis framework was developed, which integrates Twitter mining tools and a sentiment index. Sentiments were extracted using Twitter mining and then recoded and substituted into the sentiment formula. The results of sentiment analysis indicate low satisfaction with levels of pay among Russians. Twitter was chosen as the object of research, as one of the most active and independent networks in Russia. It is possible that some of the tweets belong to authors who are not living in Russia at the moment, but their number is not significant and their interest in this issue, in the authors' opinion, only enhances the relevance of the problem under study.

KEYWORDS

Big Data, Content, Data Mining, Mood Index, Sentiment Analysis

INTRODUCTION

In modern conditions, the mood of the population forms the basis of many spheres of society. The influence of this indicator can be traced in the financial and economic spheres.

For example, in their work "Early warning indicators? The effect of consumer and investor sentiments on the restaurant industry," Yost et al. (2020) examine the impact of cyclical fluctuations in the consumer confidence index and the volatility index as early warning indicators of changes in the financial performance of restaurant business entities. Wu et al. (2019) focus on the relationship of consumer spending with changes in the confidence index. The results show that planning consumption growth by considering the mood of consumers and businesses leads to many economic benefits.

The mood (satisfaction) of the population is a set of citizens' opinions regarding the studied object. In this study on population satisfaction, we consider the average assessment of respondents' views obtained by processing messages posted on a social network regarding wage levels (Juhro & Iyke, 2020; Peláez et al., 2020; Yan et al., 2019).

DOI: 10.4018/IJBDCN.341263

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

In world practice, special leading indicators (confidence indices) are actively used for these purposes, based solely on data obtained through a mass survey of citizens. Various economic actors use these indicators to develop effective development policies, but the state apparatus can also use them to assess social tension and economic confidence.

The development of modern technologies opens up a wide range of opportunities for analyzing various spheres of society, most of which have moved partially or entirely to the Internet environment. The massive volume of users generates an increase in information flows, resulting in the possibility of processing and extracting information, which is impossible to obtain during the application of classical methods of text processing and data analysis, becoming of enormous importance. Social networks are a vast information base regarding the population's opinion on various issues today.

OBJECTIVES OF THE STUDY

The effectiveness of the above methods for determining consumer sentiment at the present stage of society's development is declining. Traditional surveys and questionnaires are gradually moving to the Internet. Modern users often try to avoid this form of opinion assessment, which results from the population's distrust due to the growth of Internet fraud. The same applies to traditional contact methods for questionnaires. At the same time, conducting research based on statistical data may not always reflect the actual state of the issue under study due to the influence of numerous factors. Among these are the inability to fully cover subjects' economic activities, their underestimation of the results of their activities, and the growth of the shadow economy. In addition, a significant factor involves the retrospective nature of statistical data. According to some indicators, collecting and processing information takes a long time—up to six months. Thus, a study based on statistical data allows us to obtain a result reflecting consumer sentiment only over past periods. At the same time, to develop effective management decisions in various spheres of society's economic and social life, it is necessary to have a current idea of this indicator.

Notably, in Russian practice, the use of big data processing programs to determine the population's opinion and assess the degree of satisfaction with the economic situation and confidence in state policy in this area has been poorly studied. However, such studies are actively conducted all over the world. Many foreign scientists have studied the possibilities, features, and effectiveness of the application of big data.

Thus, at the moment, no methods exist for using text mining technologies in Russia to determine the population's satisfaction. The possibilities of adapting technologies to work with the Russian text have not been studied, which does not allow for the development of their application in the socio-economic sphere of the country.

The purpose of this work is to assess the population of the Russian Federation's satisfaction with their incomes based on the analysis of social network data using text mining tools. Studying this issue will allow us to get an idea of the features of the use of text mining technologies in the processing of texts from Russian-language sources, as well as to develop a methodology for building on this basis the wage satisfaction index as one of the components of the composite index of economic sentiment.

LITERATURE REVIEW

To solve the problems of analyzing information on the Internet and extracting users' opinions, researchers have begun actively developing big data technologies; in particular, they are investigating and expanding the possibilities of their application for conducting socio-economic research. These technologies comprise a set of various methods, techniques, and tools, of which data mining methods are of the greatest interest for the study of socio-economic phenomena (Alarifi et al., 2020; El Alaoui et al., 2019; Kannan & Kothamasu, 2020; Odendaal et al., 2020; Riahla et al., 2021; Song & Shin, 2019).

19 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/article/implementation-of-text-mining-in-socio-economic-research/341263

Related Content

Architecture Design Approaches and Issues in Cross Layer Systems

Andrea F. Cattoni, Troels B. Sørensen and Preben E. Mogensen (2012). *Using Cross-Layer Techniques for Communication Systems* (pp. 29-52).

www.irma-international.org/chapter/architecture-design-approaches-issues-cross/65664

Design of a Temperature-Constrained Governor to Save Energy in an Open Multi-Frequency Green Router

Alfio Lombardo, Diego Reforgiato, Vincenzo Riccobene and Giovanni Schembra (2013). *International Journal of Business Data Communications and Networking* (pp. 16-34).

www.irma-international.org/article/design-of-a-temperature-constrained-governor-to-save-energy-in-an-open-multi-frequency-green-router/95721

Game Theory for Collaboration in Future Networks

José André Moura, Rui Neto Marinheiro and João Carlos Silva (2014). *Convergence of Broadband, Broadcast, and Cellular Network Technologies* (pp. 94-123).

www.irma-international.org/chapter/game-theory-for-collaboration-in-future-networks/108090

Wireless Proxy: Distributed System to Mitigate the Effects of User Mobility Over Streaming Services on IEEE 802.11 Wireless LANs

Manuel Vilas, Xabiel G. Pañeda, David Melendi, Roberto García and Víctor García (2007). *International Journal of Business Data Communications and Networking* (pp. 1-18).

www.irma-international.org/article/wireless-proxy-distributed-system-mitigate/1441

Orphan Node Connected Management in Multi-Hop Clustering-Based Routing Protocols for Wireless Sensor Networks

Wassim Jerbi, Abderrahmen Guerazi and Hafedh Trabelsi (2019). *International Journal of Interdisciplinary Telecommunications and Networking* (pp. 1-16).

www.irma-international.org/article/orphan-node-connected-management-in-multi-hop-clustering-based-routing-protocols-for-wireless-sensor-networks/235466