OPINION WG2 – Coding error analysis

Johannes B. Gruber

2023-12-04

Table of contents

This document contains abstracts where agreement on whether they are relevant was especially low. We use this to see what is still unclear and where our codebook needs improvement.

1 Relevant 1

No (or unclear) ● ● Yes

ID: 14110

Agreement Score: 0.14

Proposed preprocessing methods for manipulate text of tweet

Social media provides plentiful information to study people's ideas and **opinions** about events in this world, such as political, economic issues, disasters and others. The tweets that published must study and classify. In a classification system for text mining, an important step is the preprocessing phase. It is often underestimated. This paper uses Twitter data as a case study. The aim of the paper is to propose methods for preprocessing for each text of tweet. In this paper the propose in preprocessing steps that include proposed manipulate Hashtag state method that used to extract another important word to assist in other processing processes and proposed enhancement stemming algorithm to return the entered word to its source without Affix, which are both Suffix and Prefix extensions and final results of the preprocessing process are presented based on number of words after preprocessing. © 2019 Mattingley Publishing. All rights reserved.

No (or unclear) ● Yes

ID: 1949

Agreement Score: 0.14

Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model

With an increase in the number of active users on OSNs (Online Social Networks), the propagation of fake news became obvious. OSNs provide a platform for users to interact with others by expressing their **opinions**, resharing content into different networks, etc. In addition to these, interactions with posts are also collected, termed as social engagement patterns. By taking these social engagement patterns (by analyzing infectious disease spread analogy), SENAD(Social Engagement-based News Authenticity Detection) model is proposed, which detects the authenticity of news articles shared on Twitter based on the authenticity and bias of the users who are engaging with these articles. The proposed SENAD model incorporates the novel idea of authenticity score and factors in user social engagement centric measures such as Following-followers ratio, account age, bias, etc. The proposed model significantly improves fake news and fake account detection, as highlighted by classification accuracy of 93.7%. Images embedded with textual data catch more attention than textual messages and play a vital role in quickly propagating fake news. Images published have distinctive features which need special attention for detecting whether it is real or fake. Images get altered or misused to spread fake news. The framework Credibility Neural Network (CredNN) is proposed to assess the credibility of images on OSNs, by utilizing the spatial properties of CNNs to look for physical alterations in an image as well as analyze if the image reflects a negative sentiment since fake images often exhibit either one or both characteristics. The proposed hybrid idea of combining ELA and Sentiment analysis plays a prominent role in detecting fake images with an accuracy of around 76%.

ID: 1129

Agreement Score: 0.29

Machine Learning Framework for Detecting Offensive Swahili Messages in Social Networks with Apache Spark Implementation

Languages morphological context varies by community. The linguistic analysis became more complex due to grammatical variations, cultural, traditional, slang, misspellings, and language variance. Many studies in sentimental analysis have focused on natural language processing and peoples **opinions**. Text language processing takes time, requires lots of storage space, and a fast computer to work in distributed networks. Many developers choose Hadoop and Map Reduce to process Big Data. This study developed a methodology that employs Apache Spark as a text classification processing engine since it is faster in cluster computing systems. African libraries and packages for language lemmatization and stemming are still lacking. The proposed approach was utilized to detect offensive Swahili texts in social networks. Swahili is the third most widely spoken language in Africa. Four different machine learning techniques were tested as benchmarks, with the multinomial logistic model proving to be the most effective. The evaluation measures show that the proposed machine learning framework is versatile and suitable for usage in centralized and distributed systems.

ID: 2901

Agreement Score: 0.29

Topic model for online communities' interests prediction

This paper introduces the investigation of users interests by topic modeling. This work proposes a methodology for end-to-end unsupervised users topics investigation by textual data. Using additive regularization of topic models with large topic counts, tokens frequency filtering, lemmatization and communities information, one found interpretable interests. This paper shows differences of extracted topics from users posts and communities ones, from public user posts and hidden information from subscriptions. These topics differ significantly. This paper proposes a topics evaluation method by testing model prediction on the semi-automated labeled dataset. To get a more stable topic model one suggests using an averaged topic distribution based on users subscriptions. For cases with a large number of topics, one proposes an auto labeling approach solving as a classification problem with a constant number of classes. (C) 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC -ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/) Peer-review under responsibility of the scientific committee of the 8th International Young Scientist Conference on Computational Science.

ID: 6626

Agreement Score: 0.29

Lasso-based variable selection methods in text regression: the case of short texts

Communication through websites is often characterised by short texts, made of few words, such as image captions or tweets. This paper explores the class of supervised learning methods for the analysis of short texts, as an alternative to unsupervised methods, widely employed to infer topics from structured texts. The aim is to assess the effectiveness of text data in social sciences, when they are used as explanatory variables in regression models. To this purpose, we compare different variable selection procedures when text regression models are fitted to real, short, text data. We discuss the results obtained by several variants of lasso, screening-based methods and randomisation-based models, such as sure independence screening and stability selection, in terms of number and importance of selected variables, assessed through goodness-of-fit measures, inclusion frequency and model class reliance. Latent Dirichlet allocation results are also considered as a term of comparison. Our perspective is primarily empirical and our starting point is the analysis of two real case studies, though bootstrap replications of each dataset are considered. The first case study aims at explaining price variations based on the information contained in the description of items on sale on e-commerce platforms. The second regards open questions in surveys on satisfaction ratings. The case studies are different in nature and representative of different kinds of short texts, as, in one case, a concise descriptive text is considered, whereas, in the other case, the text expresses an **opinion**. © 2023, The Author(s).

No (or unclear) ● ● Yes

ID: 4780

Agreement Score: 0.33

Unveiling: An Unexpected Mid-campaign Court Ruling's Consequences and the Limits of Following the Leader

Strong evidence exists that major campaign-relevant events can have substantial impacts on vote intentions. We know less about how information about such events diffuses and why only some events become salient. We posit that voters often become aware of such exogenous events via a media mechanism. As the salience of the policy issue in the media increases, we argue that, under certain conditions, the media primes the voters to defect from their party and its leader. We investigate these processes by studying an unexpected court ruling during the 2015 Canadian federal election campaign. Based on difference-in-differences and text-as-data approaches, we find that an exogenous court ruling related to immigrant integration led to between a 5 and 11 percentage point decline in the leading party's support. Beyond modeling how campaign-relevant events become salient through the media, we provide evidence about circumstances where leaders should not expect party loyalty to override crystallized **opinions**.

No (or unclear) ● Yes ●

ID: 1222

Agreement Score: 0.43

A Neural Network-Inspired Approach for Improved and True Movie Recommendations

In the last decade, sentiment analysis, **opinion** mining, and subjectivity of microblogs in social media have attracted a great deal of attention of researchers. Movie recommendation systems are the tools, which provide valuable services to the users. The data available online are growing gradually because the online activities of users or viewers are increasing day by day. Because of this, big data, analytics, and computational issues have raised. Therefore, we have to improve recommendations services upon the traditional one to make the recommendation system significant and efficient. This article presents the solution for these issues by producing the significant and efficient recommendation services using multivariates (ratings, votes, Twitter likes, and reviews) of movies from multiple external resources which are fetched by the web bot and managed by the Apache Hadoop framework in a distributed manner. Reviews are analyzed by a deep semantic analyzer based on the recurrent neural network (RNN/LSTM attention) with user movie attention (UMA) to produce the emotion. The proposed recommender evaluates multivariates and produces a more significant movie recommendation list according to the taste of the user on a mobile app in an efficient way.

No (or unclear) ● Yes ●

ID: 12428

Agreement Score: 0.47

Research on the Method of Identifying Opinion Leaders Based on Online Word-of-Mouth

Opinion leaders are attracting increasing attention on practitioners and academics. Opinion leaders' online Word-of-Mouth (WOM) plays a guiding and decisive role in reducing risks and uncertainty faced by users in online shopping. It is of great significance of businesses and enterprises to effectively identify **opinion** leaders. This study proposes an integrated method by looking at not only essential indicators of reviewers but also the review characteristics. The RFM model is used to evaluate the activity of reviewers. Four variables L (text length), T (period time), P (with or without a picture) and S (sentiment intensity) are derived to measure review helpfulness from review text. And two effective networks are built using the Artificial Neural Network (ANN). This study utilizes a real-life data set from Dianping.com for analysis and designs three different experiments to verify the identification effect. The results show that this method can scientifically and effectively identify the **opinion** leaders and analyze the influence of **opinion** leaders. © 2020, Springer Nature Switzerland AG.

2 Relevant 2

ID: 6626

Agreement Score: -0.29

Lasso-based variable selection methods in text regression: the case of short texts

Communication through websites is often characterised by short texts, made of few words, such as image captions or tweets. This paper explores the class of supervised learning methods for the analysis of short texts, as an alternative to unsupervised methods, widely employed to infer topics from structured texts. The aim is to assess the effectiveness of text data in social sciences, when they are used as explanatory variables in regression models. To this purpose, we compare different variable selection procedures when text regression models are fitted to real, short, text data. We discuss the results obtained by several variants of lasso, screening-based methods and randomisation-based models, such as sure independence screening and stability selection, in terms of number and importance of selected variables, assessed through goodness-of-fit measures, inclusion frequency and model class reliance. Latent Dirichlet allocation results are also considered as a term of comparison. Our perspective is primarily empirical and our starting point is the analysis of two real case studies, though bootstrap replications of each dataset are considered. The first case study aims at explaining price variations based on the information contained in the description of items on sale on e-commerce platforms. The second regards open questions in surveys on satisfaction ratings. The case studies are different in nature and representative of different kinds of short texts, as, in one case, a concise descriptive text is considered, whereas, in the other case, the text expresses an **opinion**. © 2023, The Author(s).

ID: 12428

Agreement Score: -0.2

Research on the Method of Identifying Opinion Leaders Based on Online Word-of-Mouth

Opinion leaders are attracting increasing attention on practitioners and academics. Opinion leaders' online Word-of-Mouth (WOM) plays a guiding and decisive role in reducing risks and uncertainty faced by users in online shopping. It is of great significance of businesses and enterprises to effectively identify **opinion** leaders. This study proposes an integrated method by looking at not only essential indicators of reviewers but also the review characteristics. The RFM model is used to evaluate the activity of reviewers. Four variables L (text length), T (period time), P (with or without a picture) and S (sentiment intensity) are derived to measure review helpfulness from review text. And two effective networks are built using the Artificial Neural Network (ANN). This study utilizes a real-life data set from Dianping.com for analysis and designs three different experiments to verify the identification effect. The results show that this method can scientifically and effectively identify the **opinion** leaders and analyze the influence of **opinion** leaders. © 2020, Springer Nature Switzerland AG.

Empirical study ● IRRELEVANT ● Methodological paper ● Overview

● ● ● ● ● ● ● ● ● ● ● ● ● ●

ID: 12641

Agreement Score: -0.14

Computing sentiment analysis through aspect-based fuzzy aggregations

The expression of a general **opinion** on a product or service may sometimes be broken down into different sub-**opinions** on the different aspects which characterize such a product or service. Sometimes, the general **opinion** about a product does not have to be the average of the other sub-**opinions**, but other user-dependent factors can make the aggregation of the sub-**opinions** more complex. This work studies some use cases with real **opinions**, in which fuzzy aggregation operators can help represent certain user behaviors, and hence, explain the overall assessment of an **opinion** with respect to partial **opinions** of aspects. Copyright © 2019, the Authors. Published by Atlantis Press. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

ID: 6633

Agreement Score: -0.14

Dissemination Management for Official Statistics Using Artificial Intelligence-Based Media Monitoring

Online media has been essential to our everyday lives as it spreads crucial information to the masses. Unlike traditional media, online news and social media let people freely communicate in two ways, usually by using reactions and comments, which creates public **opinion** and engagement on news or topics. Public **opinion** and engagement are beneficial in creating awareness of important news or issues. However, it will also lead to misuse, misinterpretation, and spreading of false information. For that reason, organisations often use media monitoring as a tool to gain insights about products or services. Nowadays, media intelligence, an artificial intelligence-based media monitoring, has been widely used across organisations to transform massive information from various online media platforms into useful insights. Department of Statistics Malaysia (DOSM), a national statistics provider, uses media intelligence to prevent misuse and misinterpretation of official statistics. It allows DOSM to identify the degree of happiness and centrality among the public on official statistics. This paper presents the application of using Public Maturity Assessment on Official Statistics (PMAOS), a system curated by Intelligence Media Analysis (IMA) software. PMAOS uses Natural Language Processing (NLP), syntactic and semantic techniques to understand the structure of a text, and SentiWordNet Lexicon as a base for sentiment classification. Regular monitoring of official statistics by DOSM in social media and online news found that the degree of centrality at the right source for official statistics is high. The evidence shows no or least occurrence of misuse or misinterpretation of official statistics as it is centralised across government agencies. © 2023, The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd.

Empirical study IRRELEVANT Methodological paper Overview

ID: 12679

Agreement Score: -0.14

Analysis of Patients' Reviews and Enhancing Services in Hospitals

Currently, when it comes to hospitals, a large amount of patients visit daily. If the feedbacks or reviews or **opinions** are not analyzed properly, it is bound to affect the business and reputation of hospitals. So it is necessary to analyze the reviews that the patients have about hospitals. We have proposed an approach for mining of feedbacks of patients which will be useful to find the common features in a hospital which can be +ve or −ve. We successfully generate the summary of the reviews and also classify the reviews with more than 80% accuracy with all the classifiers. 'Majority vote' which is the approach proposed by us proved to be more accurate than other classifiers. © 2020, Springer Nature Singapore Pte Ltd.

Empirical study ● IRRELEVANT ● Methodological paper

ID: 1949

Agreement Score: -0.14

Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model

With an increase in the number of active users on OSNs (Online Social Networks), the propagation of fake news became obvious. OSNs provide a platform for users to interact with others by expressing their **opinions**, resharing content into different networks, etc. In addition to these, interactions with posts are also collected, termed as social engagement patterns. By taking these social engagement patterns (by analyzing infectious disease spread analogy), SENAD(Social Engagement-based News Authenticity Detection) model is proposed, which detects the authenticity of news articles shared on Twitter based on the authenticity and bias of the users who are engaging with these articles. The proposed SENAD model incorporates the novel idea of authenticity score and factors in user social engagement centric measures such as Following-followers ratio, account age, bias, etc. The proposed model significantly improves fake news and fake account detection, as highlighted by classification accuracy of 93.7%. Images embedded with textual data catch more attention than textual messages and play a vital role in quickly propagating fake news. Images published have distinctive features which need special attention for detecting whether it is real or fake. Images get altered or misused to spread fake news. The framework Credibility Neural Network (CredNN) is proposed to assess the credibility of images on OSNs, by utilizing the spatial properties of CNNs to look for physical alterations in an image as well as analyze if the image reflects a negative sentiment since fake images often exhibit either one or both characteristics. The proposed hybrid idea of combining ELA and Sentiment analysis plays a prominent role in detecting fake images with an accuracy of around 76%.

ID: 4638

Agreement Score: -0.14

VAUT: a visual analytics system of spatiotemporal urban topics in reviews

Online review platforms offer customers the opportunities to express valuable feedback and personal views from various aspects on restaurants, products, works of art, or other items. A majority of previous studies on these user-generated reviews are devoted to controversy, bias, and **opinion** analysis. However, little work has been done to study urban characteristics via topic analysis from the city level in reviews. In this paper, we propose a visual analytics system, to visually explore spatiotemporal urban topics for cultural trend discovery, location mining, and decision making. Specifying a topic by users is supported due to the difference between review text and traditional text, such as news and books, and the diversity of topics and users. Sentiment analysis and statistical analysis are adopted to characterize the temporal trend and sentiment and topic geographical distributions of the user-specific topic. Our system allows the user to interactively explore the time-evolving frequency trend and characteristic geographical distributions of a topic in reviews. We evaluate the effectiveness and usefulness of our system using three case studies in different domains.

Empirical study ● IRRELEVANT ● Methodological paper ● Overview

ID: 1222

Agreement Score: -0.14

A Neural Network-Inspired Approach for Improved and True Movie Recommendations

In the last decade, sentiment analysis, **opinion** mining, and subjectivity of microblogs in social media have attracted a great deal of attention of researchers. Movie recommendation systems are the tools, which provide valuable services to the users. The data available online are growing gradually because the online activities of users or viewers are increasing day by day. Because of this, big data, analytics, and computational issues have raised. Therefore, we have to improve recommendations services upon the traditional one to make the recommendation system significant and efficient. This article presents the solution for these issues by producing the significant and efficient recommendation services using multivariates (ratings, votes, Twitter likes, and reviews) of movies from multiple external resources which are fetched by the web bot and managed by the Apache Hadoop framework in a distributed manner. Reviews are analyzed by a deep semantic analyzer based on the recurrent neural network (RNN/LSTM attention) with user movie attention (UMA) to produce the emotion. The proposed recommender evaluates multivariates and produces a more significant movie recommendation list according to the taste of the user on a mobile app in an efficient way.

ID: 1301

Agreement Score: 0

Concentration Areas of Sentiment Lexica in the Word Embedding Space

Sentiment lexicons play an important role in **opinion** mining systems and cognitive linguistics. Previous work aimed mostly at creating sentiment lexicons, but not thorough research into their fundamental properties. In this paper the arrangement of sentiment lexica in the multidimensional space of distributed word representations is studied. A hypothesis on the existence of sentiment lexica concentration areas is introduced and it is tested on the basis of the joint analysis of the distribution of sentiment words and general lexica. The results of the test allow to confirm the proposed hypothesis and discover the words which more than 80% of the sentiment lexica is concentrated around.

Empirical study ● IRRELEVANT ● Methodological paper

ID: 11874

Agreement Score: 0

Multi-task learning using variational auto-encoder for sentiment classification

With the rapid growth of the big data, many approaches in the representation of text for sentiment classification have been successfully proposed in natural language processing. However, these approaches remedy this problem based on single-task supervised objectives learning and do not consider their relative of multiple tasks. Based on these defects, in this work, we consider these tasks are relative, and use weight-shared parameters for learning the representation of text in neural network model, we introduce and study a multi-task approach with variational auto-encoder generative model (MTVAE) by jointly learning them. Experimental results on six subsets of Amazon review data show that the proposed approach can effectively improve the sentiment classification accuracy by other relative tasks. © 2018

ID: 1129

Agreement Score: 0

Machine Learning Framework for Detecting Offensive Swahili Messages in Social Networks with Apache Spark Implementation

Languages morphological context varies by community. The linguistic analysis became more complex due to grammatical variations, cultural, traditional, slang, misspellings, and language variance. Many studies in sentimental analysis have focused on natural language processing and peoples **opinions**. Text language processing takes time, requires lots of storage space, and a fast computer to work in distributed networks. Many developers choose Hadoop and Map Reduce to process Big Data. This study developed a methodology that employs Apache Spark as a text classification processing engine since it is faster in cluster computing systems. African libraries and packages for language lemmatization and stemming are still lacking. The proposed approach was utilized to detect offensive Swahili texts in social networks. Swahili is the third most widely spoken language in Africa. Four different machine learning techniques were tested as benchmarks, with the multinomial logistic model proving to be the most effective. The evaluation measures show that the proposed machine learning framework is versatile and suitable for usage in centralized and distributed systems.

ID: 878

Agreement Score: 0

Tweet Classification with Convolutional Neural Network

Tweet classification task is used to determine **opinion** of tweets. Such **opinion** is useful for making new strategy and taking right decision as per situation. Due to high speed and high availability of internet, large numbers of people are involving in social media to share their **opinion** towards any happening event like sport. It needs to analyze the behavior of people whether they are happy or unhappy towards the event. We consider convolutional neural network model, one of the deep learning approach for tweet classification. We used word embedding technique like word vector for text representation. We used APNews corpus as word embedding technique to give word vector called pre-trained word vector. On top of pre-trained word vector, we apply convolutional neural network to know the polarity of tweet. Here, we map each word of tweet to already pre-trained word vector of APNews corpus. We fetched the tweets from social website and performed preprocess for training and testing purpose. During training phase, we found 87.25% accuracy. During testing phase, we found 79% accuracy of proposed CNN model.

ID: 8694

Agreement Score: 0

Application of Machine Learning Techniques for Information Extraction from Airline Tweet Data

Sentimental data analysis is a way of determining the sentiments shared through text. Today, the social media is a powerful platform to share the sentiments which can be in the form of reviews, **opinions**, etc. Twitter is such a social media platform where we can share our **opinions**, thoughts, or any reviews about any product, events, or services. For commercial service providers, it is important to identify the **attitudes** in the sentiments to improve their services. For this purpose, where these tweets need to be analyzed, the sentimental data analysis has made this task easier. This paper throws light on two different classifiers used to build two different models to classify the airline sentiments into positive, negative, and neutral classes The process starts with data collection followed by the pre-processing the data to remove those words which are irrelevant to identify the tone of the sentiments. Pre-processing includes data segmentation, tokenization, stop word removal, lemmatization, and stemming. The next step is to apply classifiers to classify the tweets. We used two classifiers: Support Vector Machine (SVM) and Artificial Neural Network (ANN). Finally, using the confusion matrix obtained, accuracy, specificity, and sensitivity are calculated. These measures are used to compare the performance of the classifiers which proves the efficiency of the SVM-based classier over the ANN-based classifiers in classifying the tweets. © 2022, The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd.

Empirical study ● IRRELEVANT ● Methodological paper

ID: 2581

Agreement Score: 0

New Filtering Scheme Based on Term Weighting to Improve Object Based Opinion Mining on Tourism Product Reviews

Reviews are an essential thing in tourism industry. Opinion mining used for processing a massive amount of review data, so it can be more useful for the industry. The utilization of filtering can improve the feature extraction result from object based on **opinion** mining and can improve **opinion** classification result generally. However, there is no proven method yet to develop filter data automatically. This work applies several term weighting methods such as TF-IDF mTFIDF and BM25 to develop filter data automatically. The result from this research is the best term weighting method for developing filter data, that can improve the feature extraction and **opinion** mining relatively. TFIDF become the best term weighting method applied for filter data combined with the most frequent objects, The accuracy is 37.98%, the precision is 50.69%, the recall is 44,28%, and F-measure 47.27% for hotel data. Meanwhile, for restaurant data, the accuracy is 37.98%, precision is 50.69%, recall is 44.28%, and F-measure 47.27%. (C) 2019 The Authors. Published by Elsevier B.V.

ID: 8952

Agreement Score: 0.06

Multimodal Social Media Sentiment Analysis Based on Cross-Modal Hierarchical Attention Fusion

With the diversification of data forms on social media, more and more multimodal information mixed with image and text replaces the traditional single text description. Compared with single-modal data, multimodal data can more fully express people's **opinions**, and it also contains richer emotional information. Therefore, the task of multimodal sentiment analysis is gradually becoming a hot topic of current research. However, images and texts do not always complement each other in expressing emotional polarity in real social media. Moreover, the contribution of different modal information to the overall emotional polarity is also different. To solve these problems, a multimodal sentiment analysis method (CMHAF) that integrates topic information is proposed. The method first extracts topical information that highly summarizes the comment content from social media texts. Secondly, the current outstanding pre-training models are used to obtain emotional features of various modalities. And then, we propose cross-modal global fusion and cross-modal high-level semantic fusion methods to combine the features of different levels. At last, we conduct extensive experiments on a real Chinese multimodal dataset. The experimental results show that compared with the baseline method, the proposed method has significant improvement in multiple indicators, and can effectively classify the sentiment of multimodal social media reviews. © 2022, Springer Nature Switzerland AG.

ID: 12735

Agreement Score: 0.14

Opinion mining from online product review and online product's rating

nowadays, if we observe then we can conclude that no any person wants to go outside of his walls to purchase anything. Then he tries to make things work by sitting inside his house. So he takes the support of e commerce sites and tries to achieve his target. So when we are talking about the e commerce sites then of course, feedbacks and suggestions are involved about the products and its related features. So we can see feedback portal to each and every site. Every customer that purchased the same product before shares his experience with another one with the help of such feedback portals. This result into massive amount of data is collected. This throws business analyst or industry experts into difficulties to unable to process or analyze the reviews. Sometimes the data is found difficult to analyze because it requires lot of time to read it and remove the exact orientation of product of which user shares his valuable feedback. So in this paper a more easy and prescribed method based on sentiment analyses rather than semantics is been proposed due to which there will be no any need of using import files or extensions which are required if we start analyzing the feedback semantically. For achieving these results we used **opinion** mining technique which is a part of the Natural Language Processing Domain. We had used the technique of sentiment analysis which helps us in analyzing polarities as positive, negative or sometime as neutral and performing the calculations by consuming low time. © 2020 IJSTR.

Empirical study ● IRRELEVANT ● Methodological paper ● Overview

● ● ● ● ● ● ● ● ● ● ● ● ● ●

ID: 14110

Agreement Score: 0.14

Proposed preprocessing methods for manipulate text of tweet

Social media provides plentiful information to study people's ideas and **opinions** about events in this world, such as political, economic issues, disasters and others. The tweets that published must study and classify. In a classification system for text mining, an important step is the preprocessing phase. It is often underestimated. This paper uses Twitter data as a case study. The aim of the paper is to propose methods for preprocessing for each text of tweet. In this paper the propose in preprocessing steps that include proposed manipulate Hashtag state method that used to extract another important word to assist in other processing processes and proposed enhancement stemming algorithm to return the entered word to its source without Affix, which are both Suffix and Prefix extensions and final results of the preprocessing process are presented based on number of words after preprocessing. © 2019 Mattingley Publishing. All rights reserved.

ID: 1531

Agreement Score: 0.14

Multi-aspect Sentiment Analysis Using Domain Ontologies

Various aspects or characteristic features of an entity come into interplay to create an underlying fabric upon which sentiments blossom. In multi aspect Sentiment Analysis (SA), potentially related aspects of an entity under review are discussed in a single piece of text such as an online review. In this work, we use domain ontologies for enabling multi-aspect Sentiment Analysis. Since, domain ontologies contain the entire domain knowledge, they assist in enhanced aspect identification and detection of the latent or hidden aspects in a review document. We illustrate our approach by developing a system named Ontology driven Multi Aspect Sentiment Analysis (OMASA) system. We provide hotel reviews as input to this system and identify the panorama of explicitly expressed and latent aspects in a review using hotel domain ontology. After detecting the aspects, we link them with the corresponding **opinions** to gauge the sentiment pertaining to the aspects extracted. OMASA first computes sentiment scores for every aspect of the hotel. It then evaluates the overall sentiment score. On comparing with the baseline, the experimental results of OMASA show a marked improvement in the aspect level evaluation metrics Delta(2)(aspect) and rho(aspect) after detecting the hidden aspects. This shows that OMASA has the potential to identify the latent aspects in text thereby improving the quality of SA.

ID: 2859

Agreement Score: 0.14

Research on Tibetan hot words, sensitive words tracking and public opinion classification

The rapid development of Tibetan information technology provides rich resources for Tibetan information processing technology. The construction of Tibetan corpus is the field of Tibetan information processing of basic work. In this paper, we design the system of Tibetan network data collection and web pages preprocessing. It can timely and efficiently access to web resources, and provide a basis for further analysis of Tibetan data. It can establish the Tibetan related corpus, enrich the Tibetan digital resources. It can also alleviate the status of Tibetan corpus data sparse and lack of resources and bring the convenient condition for Tibetan information processing. The hot words reflect the hot spot of Tibetan people's attention in a certain period of time. Firstly, the paper proposes the method for reducing the space dimension of Tibetan news text. It can effectively reduce the complexity of subsequent processing. Secondly, term weighting method is proposed based on improved TFIDF for Tibetan text information extraction. It utilizes the idea that the words of different locations are given different weights to extract the hot words. On sensitive words discovery and classification of public **opinion**, sensitive thesaurus are collected artificially. Through the sensitive thesaurus comparison, the sensitive words are extracted. Classification of public **opinion** words is based on the proposed classification formula and the public **opinion** thesaurus. It will classify one Tibetan text to one public **opinion** class. In this paper, the software is developed to automatically collect Tibetan web pages from the network, preprocess the web pages, extract the text features and hot words, discover the sensitive words and classify the Tibetan text to one public **opinion** class. The experiment shows that the Tibetan hot words extraction is effective and Tibetan classification results of public **opinion** are significant.

ID: 10001

Agreement Score: 0.29

Applying deep learning techniques for sentiment analysis to assess sustainable transport

Users voluntarily generate large amounts of textual content by expressing their **opinions**, in social media and specialized portals, on every possible issue, including transport and sustainability. In this work we have leveraged such User Generated Content to obtain a high accuracy sentiment analysis model which automatically analyses the negative and positive **opinions** expressed in the transport domain. In order to develop such model, we have semiautomatically generated an annotated corpus of **opinions** about transport, which has then been used to fine-tune a large pretrained language model based on recent deep learning techniques. Our empirical results demonstrate the robustness of our approach, which can be applied to automatically process massive amounts of **opinions** about transport. We believe that our method can help to complement data from official statistics and traditional surveys about transport sustainability. Finally, apart from the model and annotated dataset, we also provide a transport classification score with respect to the sustainability of the transport types found in the use case dataset. © 2021 by the authors. Licensee MDPI, Basel, Switzerland.

ID: 5076

Agreement Score: 0.29

Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models

Background: Timely understanding of public perceptions allows public health agencies to provide up-to-date responses to health crises such as infectious diseases outbreaks. Social media such as Twitter provide an unprecedented way for the prompt assessment of the large-scale public response. Objective: The aims of this study were to develop a scheme for a comprehensive public perception analysis of a measles outbreak based on Twitter data and demonstrate the superiority of the convolutional neural network (CNN) models (compared with conventional machine learning methods) on measles outbreak-related tweets classification tasks with a relatively small and highly unbalanced gold standard training set. Methods: We first designed a comprehensive scheme for the analysis of public perception of measles based on tweets, including 3 dimensions: discussion themes, emotions expressed, and **attitude** toward vaccination. All 1,154,156 tweets containing the word measles posted between December 1, 2014, and April 30, 2015, were purchased and downloaded from DiscoverText. com. Two expert annotators curated a gold standard of 1151 tweets (approximately 0.1% of all tweets) based on the 3-dimensional scheme. Next, a tweet classification system based on the CNN framework was developed. We compared the performance of the CNN models to those of 4 conventional machine learning models and another neural network model. We also compared the impact of different word embeddings configurations for the CNN models: (1) Stanford GloVe embedding trained on billions of tweets in the general domain, (2) measles-specific embedding trained on our 1 million measles related tweets, and (3) a combination of the 2 embeddings. Results: Cohen kappa intercoder reliability values for the annotation were: 0.78, 0.72, and 0.80 on the 3 dimensions, respectively. Class distributions within the gold standard were highly unbalanced for all dimensions. The CNN models performed better on all classification tasks than k-nearest neighbors, naive Bayes, support vector machines, or random forest. Detailed comparison between support vector machines and the CNN models showed that the major contributor to the overall superiority of the CNN models is the improvement on recall, especially for classes with low occurrence. The CNN model with the 2 embedding combination led to better performance on discussion themes and emotions expressed (microaveraging F1 scores of 0.7811 and 0.8592, respectively), while the CNN model with Stanford embedding achieved best performance on **attitude** toward vaccination (microaveraging F1 score of 0.8642). Conclusions: The proposed scheme can successfully classify the public's **opinions** and emotions in multiple dimensions, which would facilitate the timely understanding of public perceptions during the

outbreak of an infectious disease. Compared with conventional machine learning methods, our CNN models showed superiority on measles-related tweet classification tasks with a relatively small and highly unbalanced gold standard. With the success of these tasks, our proposed scheme and CNN-based tweets classification system is expected to be useful for the analysis of tweets about other infectious diseases such as influenza and Ebola.

ID: 15205

Agreement Score: 0.29

Predicting political tendency of posts on Facebook

Facebook is the most popular social networking website. Every post on Facebook actually can imply the user s emotion or **opinion**. In this paper, we present our analysis on posts associated with left- and right-wing politics in the United States of America. Our dataset contains posts several related Facebook fan pages. We analyze sentiment of posts for the prediction of left- or right-wing politics. We build sentiment features for the prediction and evaluate prediction performance. The results show that F1-score can be as high as 0.95 when TF-IDF is used with a decision tree. Posts generally involve emotional words. We use the lexical databases for sentiment analysis. Our experiment results show that the sentiment analysis is sensitive to some classification algorithms. © 2018 Association for Computing Machinery.

ID: 4349

Agreement Score: 0.29

Sentiment and Emotion-Aware Multi-Modal Complaint Identification

The expression of displeasure on a consumer's behalf towards an organization, product, or event is denoted via the speech act known as complaint. Customers typically post reviews on retail websites and various social media platforms about the products or services they purchase, and the reviews may include complaints about the products or services. Automatic detection of consumers' complaints about items or services they buy can be critical for organizations and online merchants since they can use this insight to meet the customers' requirements, including handling and addressing the complaints. Previous studies on Complaint Identification (CI) are limited to text. Images posted with the reviews can provide cues to identify complaints better, thus emphasizing the importance of incorporating multi-modal inputs into the process. Furthermore, the customer's emotional state significantly impacts the complaint expression since emotions generally influence any speech act. As a result, the impact of emotion and sentiment on automatic complaint identification must also be investigated. One of the major contributions of this work is the creation of a new dataset- Complaint, Emotion, and Sentiment Annotated Multi-modal Amazon Reviews Dataset (CESAMARD), a collection of **opinionated** texts (reviews) and images of the products posted on the website of the retail giant Amazon. We present an attentionbased multi-modal, adversarial multi-task deep neural network model for complaint detection to demonstrate the utility of the multi-modal dataset. Experimental results indicate that the multi-modality and multi-tasking complaint identification outperforms uni-modal and single-task variants.

ID: 12821

Agreement Score: 0.29

Multi-rule based ensemble feature selection model for sarcasm type detection in Twitter

Sentimental analysis aims at inferring how people express their **opinion** over any piece of text or topic of interest. This article deals with detection of an implicit form of the sentiment, referred to as sarcasm. Sarcasm conveys the opposite of what people try to convey in order to criticize or ridicule in a humorous way. It plays a vital role in social networks since most of the tweets or posts contain sarcastic nuances. Existing approaches towards the study of sarcasm deals only with the detection of sarcasm. In this paper, in addition to detecting sarcasm from text, an approach has been proposed to identify the type of sarcasm. The main motivation behind determining the types of sarcasm is to identify the level of hurt or the true intent behind the sarcastic text. The proposed work aims to improve upon the existing approaches by incorporating a new perspective which classifies the sarcasm based on the level of harshness employed. The major application of the proposed work would be relating the emotional state of a person to the type of sarcasm exhibited by him/her which could provide major insights about the emotional behavior of a person. An ensemble-based feature selection method has been proposed for identifying the optimal set of features needed to detect sarcasm from tweets. This optimal set of features was employed to detect whether the tweet is sarcastic or not. After detecting sarcastic sentences, a multi-rule based approach has been proposed to determine the type of sarcasm. As an initial attempt, sarcasm has been classified into four types, namely, polite sarcasm, rude sarcasm, raging sarcasm, and deadpan sarcasm. The performance and efficiency of the proposed approach has been experimentally analyzed, and change in mood of a person for each sarcastic type has been modelled. The overall accuracy of the proposed ensemble feature selection algorithm for sarcasm detection is around 92.7%, and the proposed multi-rule approach for sarcastic type identification achieves an accuracy of 95.98%, 96.20%, 99.79%, and 86.61% for polite, rude, raging, and deadpan types of sarcasm, respectively. © 2020 Karthik Sundararajan and Anandhakumar Palanisamy.

ID: 4054

Agreement Score: 0.29

Decompositional Argument Mining: A General Purpose Approach for Argument Graph Construction

This work presents an approach decomposing propositions into four functional components and identify the patterns linking those components to determine argument structure. The entities addressed by a proposition are target concepts and the features selected to make a point about the target concepts are aspects. A line of reasoning is followed by providing evidence for the points made about the target concepts via aspects. Opinions on target concepts and **opinions** on aspects are used to support or attack the ideas expressed by target concepts and aspects. The relations between aspects, target concepts, **opinions** on target concepts and aspects are used to infer the argument relations. Propositions are connected iteratively to form a graph structure. The approach is generic in that it is not tuned for a specific corpus and evaluated on three different corpora from the literature: AAEC, AMT, US2016G1tv and achieved an F score of 0.79, 0.77 and 0.64, respectively.

ID: 14854

Agreement Score: 0.29

Sentiment analysis of movie reviews using heterogeneous features

Human disposition has always influenced by others suggestion and reviews. People are always eager to know other's reviews for their profit but, every website contains a very large amount of review text, the average human reader will have trouble in identifying relevant sites, extracting and abstracting the reviews so they cannot reach to the right decision in less time that is why automated sentiment analysis systems are required. In the proposed approach, heterogeneous features such as machine learning based and Lexicon based features and supervised learning algorithms like Naive Bayes (NB) and Linear Support Vector Machine (LSVM) used to build the system model. From implementation and observation, conclude that using proposed heterogeneous features and hybrid approach can get an accurate sentiment analysis system compared to other baseline system. In future for big data, we can use these heterogeneous features for bulding advance and more accurate models using Deep Learning (DL) algorithms. © 2018 IEEE.

Empirical study ● Methodological paper ● something else

ID: 13358

Agreement Score: 0.29

Opinion mining for travel route recommendation using Social Media Networks (Twitter)

Most of the organizations use text analytics to uncover purposeful information from an unstructured text as a result of considering the linguistic communication process techniques area unit extremely difficult. They typically cause several issues because of the inconsistency in syntax and linguistics. Sentiment analysis based on the **opinion** of the users. On twitter, many people post about their experience on the traffic routes. This project discusses the prediction of text mining analysis. On that post collecting from the data set and we find out which path is the best path for the travellers and waiting for commuters. In this project we discuss the traffic mining tweets using the keywords predicting the positive and negative comment on the Twitter. Experimentation involves discussion and comparison of ensemble classifiers over tagged tweets. Finally, it will be finding the best accuracy. © Blue Eyes Intelligence Engineering & Sciences Publication.

Empirical study ● IRRELEVANT ● Methodological paper ●

ID: 2901

Agreement Score: 0.29

Topic model for online communities' interests prediction

This paper introduces the investigation of users interests by topic modeling. This work proposes a methodology for end-to-end unsupervised users topics investigation by textual data. Using additive regularization of topic models with large topic counts, tokens frequency filtering, lemmatization and communities information, one found interpretable interests. This paper shows differences of extracted topics from users posts and communities ones, from public user posts and hidden information from subscriptions. These topics differ significantly. This paper proposes a topics evaluation method by testing model prediction on the semi-automated labeled dataset. To get a more stable topic model one suggests using an averaged topic distribution based on users subscriptions. For cases with a large number of topics, one proposes an auto labeling approach solving as a classification problem with a constant number of classes. (C) 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC -ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/) Peer-review under responsibility of the scientific committee of the 8th International Young Scientist Conference on Computational Science.

● ● ● ● ● ● ● ● ● ● ● ● ●

ID: 15584

Agreement Score: 0.29

NewsAnalyticalToolkit: An online natural language processing platform to analyze news

In today's increasingly divided political climate there is a need for a tool that can compare news articles and organizations so that a user can receive a wider range of views and philosophies. NewsAnalyticalToolkit allows a user to compare news sites and their political articles by coverage, mood, sentiment, and objectivity. The user can sort through the news by topic, which was determined using Natural Language Processing (NLP) and Latent Dirichlet Allocation (LDA). LDA is a probabilistic method used to discover latent topics within a series of documents and cluster them accordingly. Each news article can be considered a mix of multiple topics and LDA assigns a set of topics to each with a probability of it pertaining to that topic. For each topic, a user can then discover the coverage, mood, sentiment and objectivity expressed by each author and site. The mood was determined using IBM Watsons ToneAnalyzerV3, which uses linguistic analysis to detect emotional, social and language tones in written text. The analyzer is based on the theory of psycholinguistics, a field of research that explores the relationship between linguistic behavior and psychological theories. The sentiment and objectivity scores were determined using SentiWordNet, which is a lexical database that groups English words into sets of synonyms and assigns sentiment scores to them. The features were combined to plot an interactive graph of how **opinionated** versus how analytical an article is, so that the user can click through them to get a better understanding of the topic in question. © 2018 SPIE. Downloading of the abstract is permitted for personal use only.

ID: 7931

Agreement Score: 0.29

Data Extraction Approach using Natural Language Processing for Sentiment Analysis

The branch of research known as "sentiment analysis and **opinion** mining" focuses on extracting meaning from the written words of others by studying their thoughts, feelings, judgments, and **attitudes**. Natural language processing is one of the most active study fields in data mining, web mining, and text mining. There are numerous uses of sentiment analysis, such as assessing the impact of events on social networks and gauging public **opinion** on products and services. In the same way as blogs, forums, microblogs, and social networks such as Twitter and Facebook have grown in popularity, sentiment analysis is becoming increasingly important. It is possible to measure sentiments that are captured in digital form using supervised machine learning and lexical-based techniques. © 2022 IEEE

Empirical study ● IRRELEVANT ● Methodological paper

ID: 12321

Agreement Score: 0.29

Language independent sequence labelling for opinion target extraction

In this paper we present a language independent system to model Opinion Target Extraction (OTE) as a sequence labelling task. The system consists of a combination of clustering features implemented on top of a simple set of shallow local features. Experiments on the well known Aspect Based Sentiment Analysis (ABSA) benchmarks show that our approach is very competitive across languages, obtaining, at the time of writing, best results for six languages in seven different datasets. Furthermore, the results provide further insights into the behaviour of clustering features for sequence labeling tasks. Finally, we also show that these results can be outperformed by recent advances in contextual word embeddings and the transformer architecture. The system and models generated in this work are available for public use and to facilitate reproducibility of results. © 2020 Inst. Sci. inf., Univ. Defence in Belgrade. All rights reserved.

ID: 4780

Agreement Score: 0.33

Unveiling: An Unexpected Mid-campaign Court Ruling's Consequences and the Limits of Following the Leader

Strong evidence exists that major campaign-relevant events can have substantial impacts on vote intentions. We know less about how information about such events diffuses and why only some events become salient. We posit that voters often become aware of such exogenous events via a media mechanism. As the salience of the policy issue in the media increases, we argue that, under certain conditions, the media primes the voters to defect from their party and its leader. We investigate these processes by studying an unexpected court ruling during the 2015 Canadian federal election campaign. Based on difference-in-differences and text-as-data approaches, we find that an exogenous court ruling related to immigrant integration led to between a 5 and 11 percentage point decline in the leading party's support. Beyond modeling how campaign-relevant events become salient through the media, we provide evidence about circumstances where leaders should not expect party loyalty to override crystallized **opinions**.

ID: 15098

Agreement Score: 0.43

Customizable twitter bots for improved QoS in product marketing

In recent times, users have started spending a significant amount of their time on social networking platforms (Twitter, Facebook, Orkut), sharing a plethora of personal information and **opinions** on different aspects of life. Internet users around the world make use of this powerful tool of communication a.k.a microblogging, using Twitter, a web application that provides dual benefits of microblogging and social networking. In this paper, we propose an application that uses the open structure of twitter to perform 2 main tasks: (1) analyze data for sentiment analysis and **opinion** mining of a particular product and classify this data into positive, neutral and negative reviews (2) use this data to create automated programs, known as bots, which are useful for generating benign tweets. We use this double-edged sword to provide companies with the opportunity to make use of these automated bots to generate the right tags to advertise products. This enhanced marketing opportunity provides Quality of Service in terms of cost, time andenergy. © 2017 IEEE.

ID: 10162

Agreement Score: 0.43

CoSaD-Code-Mixed Sentiments Analysis for Dravidian Languages

Analyzing sentiments or **opinions** in code-mixed languages is gaining importance due to increase in the use of social media and online platforms especially during the Covid-19 pandemic. In a multilingual society like India, code-mixing and script mixing is quite common as people especially the younger generation are quite familiar in using more than one language. In view of this, the current paper describes the models submitted by our team MUCIC for the shared task in'Sentiments Analysis (SA) for Dravidian Languages in Code-Mixed Text'. The objective of this shared task is to develop and evaluate models for code-mixed datasets in three Dravidian languages, namely: Kannada, Malayalam, and Tamil mixed with English language resulting in Kannada-English (Ka-En), Malayalam-English (Ma-En), and Tamil-English (Ta-En) language pairs. N-grams of char, char sequences, and syllables features are transformed into feature vectors and are used to train three Machine Learning (ML) classifiers with majority voting. The predictions on the Test set obtained average weighted F1-scores of 0.628, 0.726, and 0.619 securing 2nd, 4th, and 5th ranks for Ka-En, Ma-En, and Ta-En language pairs respectively. © 2021 Copyright for this paper by its authors.

ID: 9334

Agreement Score: 0.43

Sentiment Analysis of Snapchat Application's Reviews

Sentiment analysis is a process of extracting **opinion** and subjectivity knowledge from user generated text content without the need to monitor the reviews manually. It can help to obtain an overview on performance of a product or subject based on the reviews from users. The aim of this study is to classify the Snapchat application's reviews into different polarities which are positive, neutral or negative. Next, the most frequent words are identified. Furthermore, Multinomial Naïve Bayes and Random Forest classification algorithm are used to predict the user's rating. The performances of the classification models are evaluated using accuracy, precision, recall and F1-score. The results showed majority of the Snapchat users had a positive experience with the total of 6037 positive reviews. Based on the performance measures, Multinomial Naïve Bayes classification algorithm performed slightly better than the Random Forest classification algorithm in predicting the rating of Snapchat application. Overall, both of the classification algorithms have average performance in predicting user's rating for Snapchat application. © 2021 IEEE.

Empirical study ● Methodological paper ●

ID: 6029

Agreement Score: 0.43

Deep-ABSA: A Multichannel Deep Learning Framework for Aspect-Based Bangla Sentiment Analysis

Nowadays people express their **opinions** on social media. Also provides product reviews on eCommerce websites and responds to various news as comments. It is necessary to know the polarity and aspect of various posts and comments for business, education and security. Aspect-based sentiment analysis (ABSA) predicts text category and polarity. In this paper, we proposed a deep learning framework Deep-ABSA for aspect-based sentiment analysis from Bangla texts. Our proposed framework is a multi-channel architecture. We implemented the word embedding, Bi-LSTM with the attention mechanism for one channel. And for another channel, we adopted the character-embedded convolutional neural network. Finally, we concatenated both channels for adjoining the features of each channel. We obtained an adequate performance from our proposed framework for aspect term analysis from Bangla sentences. © 2023 IEEE.

ID: 15018

Agreement Score: 0.43

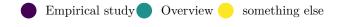**Sentiment analysis using machine learning for business intelligence**

This paper suggests use of sentiment analysis classification as an effective method for examining textual data coming from variety of resources on internet. Sentiment analysis is a method of data mining that evaluates textual data consuming machine learning techniques. Due to tremendous expanse of **opinions** of users, their reviews, feedbacks and suggestions available over the web resources, it is so much indispensable to discover, analyze and consolidate their views for enhanced decision making. Sentiment analysis presents an effective and efficient **opinion** of consumers in real time which can greatly affect the decision making process for business domain. We have seen an increment in level of activity during last ten year period and emphases on exploratory research approaches. We noticed that several procedures are inattentive from the pond of Business Intelligence research. We also recognized potential zones that requisite additional exploration. © 2017 IEEE.

ID: 7674

Agreement Score: 0.43

A New Approach to Analysis of Public Sentiment on Padma Bridge in Bangla Text

Opinion mining or emotion analysis of people is generally called sentiment analysis. Sentiment analysis is necessary for getting customer or user notions about any products or services. This research paper investigated and recorded the public sentiment on Padma Bridge. As a result, the Government of Bangladesh can make decisions easily on future mega-construction projects based on the recorded results. Padma bridge is a mega construction event for Bangladesh due to the big budget and banned World Bank loans. Bangladeshi people express their feelings, suggestions, **opinions**, and thoughts about the Padma Bridge project on Facebook, YouTube and other social media. The main focus of this paper is sentiment analysis of people's reactions to Padma bridge is based on the Bangla comment dataset. We have collected more than 15K data which has two types of sentiment: Positive and Negative. Then we used three machine learning models (SVM, RF, LSVC) and one deep learning model (LSTM) for sentiment analysis. In our proposed system, we used an innovative voting method that can count and compare the sentiments produced by the mentioned ML and DL models. Finally, our model makes decision-based on the maximum voting results. This paper concludes that the voting method technique improves the accuracy by around 7.5% compared to every single ML and DL model. © 2022 IEEE.

ID: 3479

Agreement Score: 0.43

Sentiment analysis for the tweets that contain the word earthquake

The social networks represent modern means of conveying information and states. The accurate access to the messages and states in the shortest period of time might be very useful in case this contributes to reducing the risk of producing destructive events. In case of earthquakes, it is highly important to be aware of the people's state in order to rapidly provide help. There are multiple information methods used to identify the sentiment and the human **opinion**, but the results should be vigilantly managed. The latter are related to the real existing circumstances that might be determined by the environmental factors. The present study aims to identify the people's states when they are posting a short message on Twitter, immediately after a natural event, such as an earthquake. Knowing the **opinion** of those people who were in a calamity earthquake situation, and also knowing the sentiment analysis, a contribution to the quantification of the reaction regarding these events that they were exposed to might be supplied.

ID: 5416

Agreement Score: 0.43

Ensemble Stacking Model for Sentiment Analysis of Emirati and Arabic Dialects
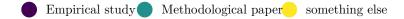
Sentiment analysis is the process of examining people's **opinions** and emotions towards goods, services, organizations, individuals, and other things, through the use of textual data. It involves categorizing text as positive, negative, or neutral to quantify people's beliefs. Social media platforms have become an important source of sentiment analysis data due to their widespread use for sharing **opinions** and information. As the number of social media users continues to grow, the amount of data generated for sentiment analysis also increases. Previous research on sentiment analysis for the Arabic language has mostly focused on Modern Standard Arabic and various dialects such as Egyptian, Saudi, Algerian, Jordanian, Tunisian, and Levantine. However, there has been no research on the utilization of deep-learning approaches for sentiment analysis of the Emirati dialect, which is an informal form of the Arabic language spoken in the United Arab Emirates. It's important to note that each country in the Arab world has its dialect, and some dialects may even have several sub-dialects. The primary aim of this research is to create a highly advanced deep-learning model that can effectively perform sentiment analysis on the Emirati dialect. To achieve this objective, the authors have proposed and utilized seven different deep-learning models for sentiment analysis of the Emirati dialect. Then, an ensemble stacking model was introduced to combine the best-performing deep learning models used in this study. The ensemble stacking deep learning model consisted of deep learning models with a meta-learner layer of classifiers. The first model combined the two best-performing deep learning models, the second combined the four best-performing models, and the final model combined all seven trained deep learning models in this research. The proposed ensemble stacking deep learning model was assessed on four datasets, three versions of the ESAAD Emirati Sentiment Analysis Annotated Dataset, two versions of Twitter-based Benchmark Arabic Sentiment Analysis Dataset (ASAD), an Arabic Company Reviews dataset, and an English dataset known as a Preprocessed Sentiment Analysis Dataset PSAD. The results of the experiments demonstrated that the proposed ensemble stacking model presented an outstanding performance in terms of accuracy and achieved an accuracy of 95.54% for the ESAAD dataset, 96.71% for the ASAD benchmark dataset, 96.65% for the Arabic Company Reviews Dataset, and 98.53% for the Preprocessed Sentiment Analysis Dataset PSAD. © 2023 The Author(s)

ID: 15727

Agreement Score: 0.47

Cross-domain sentiment classification via a bifurcated-LSTM

Sentiment classification plays a vital role in current online commercial transactions because it is critical to understand users' **opinions** and feedbacks in businesses or products. Cross-domain sentiment classification can adopt a well-trained classifier from one source domain to other target domains, which reduces the time and efforts of training new classifiers in these domains. Existing cross-domain sentiment classification methods require data or other information in target domains in order to train their models. However, collecting and processing new corpora require very heavy workload. Besides, the data in target domains may be private and not always available for training. To address these issues, motivated by multi-task learning, we design a Bifurcated-LSTM which takes advantages of attention-based LSTM classifiers along with augmented dataset and orthogonal constraints. This Bifurcated-LSTM can extract domain-invariant sentiment features from the source domain to perform sentiment analysis in different target domains. We conduct extensive experiments on seven classic types of product reviews, and results show that our system leads to significant performance improvement. © Springer International Publishing AG, part of Springer Nature 2018.

ID: 1017

Agreement Score: 0.47

Classification of Code-Mixed Bilingual Phonetic Text Using Sentiment Analysis

The rapid growth of internet facilities has increased the comments, posts, blogs, feedback, etc., on a large scale on social networking sites. These social media data are available in an unstructured form, which includes images, text, and videos. The processing of these data is difficult, but some sentiment analysis, information retrieval, and recommender systems are used to process these unstructured data. To extract the **opinion** and sentiment of internet users from their written social media text, a sentiment analysis system is required to develop, which can work on both monolingual and bilingual phonetic text. Therefore, a sentiment analysis (SA) system is developed, which performs well on different domain datasets. The system performance is tested on four different datasets and achieved better accuracy of 3% on social media datasets, 1.5% on movie reviews, 1.35% on Amazon product reviews, and 4.56% on large Amazon product reviews than the state-of-art techniques. Also, the stemmer (StemVerb) for verbs of the English language is proposed, which improves the SA system's performance.

3 Relevant 3

ID: 3479

Agreement Score: -0.29

Sentiment analysis for the tweets that contain the word earthquake

The social networks represent modern means of conveying information and states. The accurate access to the messages and states in the shortest period of time might be very useful in case this contributes to reducing the risk of producing destructive events. In case of earthquakes, it is highly important to be aware of the people's state in order to rapidly provide help. There are multiple information methods used to identify the sentiment and the human **opinion**, but the results should be vigilantly managed. The latter are related to the real existing circumstances that might be determined by the environmental factors. The present study aims to identify the people's states when they are posting a short message on Twitter, immediately after a natural event, such as an earthquake. Knowing the **opinion** of those people who were in a calamity earthquake situation, and also knowing the sentiment analysis, a contribution to the quantification of the reaction regarding these events that they were exposed to might be supplied.

ID: 12641

Agreement Score: -0.14

Computing sentiment analysis through aspect-based fuzzy aggregations

The expression of a general **opinion** on a product or service may sometimes be broken down into different sub-**opinions** on the different aspects which characterize such a product or service. Sometimes, the general **opinion** about a product does not have to be the average of the other sub-**opinions**, but other user-dependent factors can make the aggregation of the sub-**opinions** more complex. This work studies some use cases with real **opinions**, in which fuzzy aggregation operators can help represent certain user behaviors, and hence, explain the overall assessment of an **opinion** with respect to partial **opinions** of aspects. Copyright © 2019, the Authors. Published by Atlantis Press. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

ID: 1301

Agreement Score: -0.14

Concentration Areas of Sentiment Lexica in the Word Embedding Space

Sentiment lexicons play an important role in **opinion** mining systems and cognitive linguistics. Previous work aimed mostly at creating sentiment lexicons, but not thorough research into their fundamental properties. In this paper the arrangement of sentiment lexica in the multidimensional space of distributed word representations is studied. A hypothesis on the existence of sentiment lexica concentration areas is introduced and it is tested on the basis of the joint analysis of the distribution of sentiment words and general lexica. The results of the test allow to confirm the proposed hypothesis and discover the words which more than 80% of the sentiment lexica is concentrated around.

ID: 6626

Agreement Score: 0.14

Lasso-based variable selection methods in text regression: the case of short texts

Communication through websites is often characterised by short texts, made of few words, such as image captions or tweets. This paper explores the class of supervised learning methods for the analysis of short texts, as an alternative to unsupervised methods, widely employed to infer topics from structured texts. The aim is to assess the effectiveness of text data in social sciences, when they are used as explanatory variables in regression models. To this purpose, we compare different variable selection procedures when text regression models are fitted to real, short, text data. We discuss the results obtained by several variants of lasso, screening-based methods and randomisation-based models, such as sure independence screening and stability selection, in terms of number and importance of selected variables, assessed through goodness-of-fit measures, inclusion frequency and model class reliance. Latent Dirichlet allocation results are also considered as a term of comparison. Our perspective is primarily empirical and our starting point is the analysis of two real case studies, though bootstrap replications of each dataset are considered. The first case study aims at explaining price variations based on the information contained in the description of items on sale on e-commerce platforms. The second regards open questions in surveys on satisfaction ratings. The case studies are different in nature and representative of different kinds of short texts, as, in one case, a concise descriptive text is considered, whereas, in the other case, the text expresses an **opinion**. © 2023, The Author(s).

ID: 1222

Agreement Score: 0.14

A Neural Network-Inspired Approach for Improved and True Movie Recommendations

In the last decade, sentiment analysis, **opinion** mining, and subjectivity of microblogs in social media have attracted a great deal of attention of researchers. Movie recommendation systems are the tools, which provide valuable services to the users. The data available online are growing gradually because the online activities of users or viewers are increasing day by day. Because of this, big data, analytics, and computational issues have raised. Therefore, we have to improve recommendations services upon the traditional one to make the recommendation system significant and efficient. This article presents the solution for these issues by producing the significant and efficient recommendation services using multivariates (ratings, votes, Twitter likes, and reviews) of movies from multiple external resources which are fetched by the web bot and managed by the Apache Hadoop framework in a distributed manner. Reviews are analyzed by a deep semantic analyzer based on the recurrent neural network (RNN/LSTM attention) with user movie attention (UMA) to produce the emotion. The proposed recommender evaluates multivariates and produces a more significant movie recommendation list according to the taste of the user on a mobile app in an efficient way.

ID: 1949

Agreement Score: 0.14

Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model

With an increase in the number of active users on OSNs (Online Social Networks), the propagation of fake news became obvious. OSNs provide a platform for users to interact with others by expressing their **opinions**, resharing content into different networks, etc. In addition to these, interactions with posts are also collected, termed as social engagement patterns. By taking these social engagement patterns (by analyzing infectious disease spread analogy), SENAD(Social Engagement-based News Authenticity Detection) model is proposed, which detects the authenticity of news articles shared on Twitter based on the authenticity and bias of the users who are engaging with these articles. The proposed SENAD model incorporates the novel idea of authenticity score and factors in user social engagement centric measures such as Following-followers ratio, account age, bias, etc. The proposed model significantly improves fake news and fake account detection, as highlighted by classification accuracy of 93.7%. Images embedded with textual data catch more attention than textual messages and play a vital role in quickly propagating fake news. Images published have distinctive features which need special attention for detecting whether it is real or fake. Images get altered or misused to spread fake news. The framework Credibility Neural Network (CredNN) is proposed to assess the credibility of images on OSNs, by utilizing the spatial properties of CNNs to look for physical alterations in an image as well as analyze if the image reflects a negative sentiment since fake images often exhibit either one or both characteristics. The proposed hybrid idea of combining ELA and Sentiment analysis plays a prominent role in detecting fake images with an accuracy of around 76%.

IRRELEVANT ● No (or unclear) ● Yes ●

● ● ● ● ● ● ● ● ● ● ● ● ● ●

ID: 12428

Agreement Score: 0.2

Research on the Method of Identifying Opinion Leaders Based on Online Word-of-Mouth

Opinion leaders are attracting increasing attention on practitioners and academics. Opinion leaders' online Word-of-Mouth (WOM) plays a guiding and decisive role in reducing risks and uncertainty faced by users in online shopping. It is of great significance of businesses and enterprises to effectively identify **opinion** leaders. This study proposes an integrated method by looking at not only essential indicators of reviewers but also the review characteristics. The RFM model is used to evaluate the activity of reviewers. Four variables L (text length), T (period time), P (with or without a picture) and S (sentiment intensity) are derived to measure review helpfulness from review text. And two effective networks are built using the Artificial Neural Network (ANN). This study utilizes a real-life data set from Dianping.com for analysis and designs three different experiments to verify the identification effect. The results show that this method can scientifically and effectively identify the **opinion** leaders and analyze the influence of **opinion** leaders. © 2020, Springer Nature Switzerland AG.

ID: 1129

Agreement Score: 0.29

Machine Learning Framework for Detecting Offensive Swahili Messages in Social Networks with Apache Spark Implementation

Languages morphological context varies by community. The linguistic analysis became more complex due to grammatical variations, cultural, traditional, slang, misspellings, and language variance. Many studies in sentimental analysis have focused on natural language processing and peoples **opinions**. Text language processing takes time, requires lots of storage space, and a fast computer to work in distributed networks. Many developers choose Hadoop and Map Reduce to process Big Data. This study developed a methodology that employs Apache Spark as a text classification processing engine since it is faster in cluster computing systems. African libraries and packages for language lemmatization and stemming are still lacking. The proposed approach was utilized to detect offensive Swahili texts in social networks. Swahili is the third most widely spoken language in Africa. Four different machine learning techniques were tested as benchmarks, with the multinomial logistic model proving to be the most effective. The evaluation measures show that the proposed machine learning framework is versatile and suitable for usage in centralized and distributed systems.

ID: 6633

Agreement Score: 0.29

Dissemination Management for Official Statistics Using Artificial Intelligence-Based Media Monitoring

Online media has been essential to our everyday lives as it spreads crucial information to the masses. Unlike traditional media, online news and social media let people freely communicate in two ways, usually by using reactions and comments, which creates public **opinion** and engagement on news or topics. Public **opinion** and engagement are beneficial in creating awareness of important news or issues. However, it will also lead to misuse, misinterpretation, and spreading of false information. For that reason, organisations often use media monitoring as a tool to gain insights about products or services. Nowadays, media intelligence, an artificial intelligence-based media monitoring, has been widely used across organisations to transform massive information from various online media platforms into useful insights. Department of Statistics Malaysia (DOSM), a national statistics provider, uses media intelligence to prevent misuse and misinterpretation of official statistics. It allows DOSM to identify the degree of happiness and centrality among the public on official statistics. This paper presents the application of using Public Maturity Assessment on Official Statistics (PMAOS), a system curated by Intelligence Media Analysis (IMA) software. PMAOS uses Natural Language Processing (NLP), syntactic and semantic techniques to understand the structure of a text, and SentiWordNet Lexicon as a base for sentiment classification. Regular monitoring of official statistics by DOSM in social media and online news found that the degree of centrality at the right source for official statistics is high. The evidence shows no or least occurrence of misuse or misinterpretation of official statistics as it is centralised across government agencies. © 2023, The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd.

ID: 4054

Agreement Score: 0.29

Decompositional Argument Mining: A General Purpose Approach for Argument Graph Construction

This work presents an approach decomposing propositions into four functional components and identify the patterns linking those components to determine argument structure. The entities addressed by a proposition are target concepts and the features selected to make a point about the target concepts are aspects. A line of reasoning is followed by providing evidence for the points made about the target concepts via aspects. Opinions on target concepts and **opinions** on aspects are used to support or attack the ideas expressed by target concepts and aspects. The relations between aspects, target concepts, **opinions** on target concepts and aspects are used to infer the argument relations. Propositions are connected iteratively to form a graph structure. The approach is generic in that it is not tuned for a specific corpus and evaluated on three different corpora from the literature: AAEC, AMT, US2016G1tv and achieved an F score of 0.79, 0.77 and 0.64, respectively.

ID: 2901

Agreement Score: 0.29

Topic model for online communities' interests prediction

This paper introduces the investigation of users interests by topic modeling. This work proposes a methodology for end-to-end unsupervised users topics investigation by textual data. Using additive regularization of topic models with large topic counts, tokens frequency filtering, lemmatization and communities information, one found interpretable interests. This paper shows differences of extracted topics from users posts and communities ones, from public user posts and hidden information from subscriptions. These topics differ significantly. This paper proposes a topics evaluation method by testing model prediction on the semi-automated labeled dataset. To get a more stable topic model one suggests using an averaged topic distribution based on users subscriptions. For cases with a large number of topics, one proposes an auto labeling approach solving as a classification problem with a constant number of classes. (C) 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC -ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/) Peer-review under responsibility of the scientific committee of the 8th International Young Scientist Conference on Computational Science.

ID: 2581

Agreement Score: 0.29

New Filtering Scheme Based on Term Weighting to Improve Object Based Opinion Mining on Tourism Product Reviews

Reviews are an essential thing in tourism industry. Opinion mining used for processing a massive amount of review data, so it can be more useful for the industry. The utilization of filtering can improve the feature extraction result from object based on **opinion** mining and can improve **opinion** classification result generally. However, there is no proven method yet to develop filter data automatically. This work applies several term weighting methods such as TF-IDF mTFIDF and BM25 to develop filter data automatically. The result from this research is the best term weighting method for developing filter data, that can improve the feature extraction and **opinion** mining relatively. TFIDF become the best term weighting method applied for filter data combined with the most frequent objects, The accuracy is 37.98%, the precision is 50.69%, the recall is 44,28%, and F-measure 47.27% for hotel data. Meanwhile, for restaurant data, the accuracy is 37.98%, precision is 50.69%, recall is 44.28%, and F-measure 47.27%. (C) 2019 The Authors. Published by Elsevier B.V.

● ● ● ● ● ● ● ● ● ● ● ● ● ●

ID: 14110

Agreement Score: 0.43

Proposed preprocessing methods for manipulate text of tweet

Social media provides plentiful information to study people's ideas and **opinions** about events in this world, such as political, economic issues, disasters and others. The tweets that published must study and classify. In a classification system for text mining, an important step is the preprocessing phase. It is often underestimated. This paper uses Twitter data as a case study. The aim of the paper is to propose methods for preprocessing for each text of tweet. In this paper the propose in preprocessing steps that include proposed manipulate Hashtag state method that used to extract another important word to assist in other processing processes and proposed enhancement stemming algorithm to return the entered word to its source without Affix, which are both Suffix and Prefix extensions and final results of the preprocessing process are presented based on number of words after preprocessing. © 2019 Mattingley Publishing. All rights reserved.

IRRELEVANT ● No (or unclear) ● Yes ●

ID: 13358

Agreement Score: 0.43

Opinion mining for travel route recommendation using Social Media Networks (Twitter)

Most of the organizations use text analytics to uncover purposeful information from an unstructured text as a result of considering the linguistic communication process techniques area unit extremely difficult. They typically cause several issues because of the inconsistency in syntax and linguistics. Sentiment analysis based on the **opinion** of the users. On twitter, many people post about their experience on the traffic routes. This project discusses the prediction of text mining analysis. On that post collecting from the data set and we find out which path is the best path for the travellers and waiting for commuters. In this project we discuss the traffic mining tweets using the keywords predicting the positive and negative comment on the Twitter. Experimentation involves discussion and comparison of ensemble classifiers over tagged tweets. Finally, it will be finding the best accuracy. © Blue Eyes Intelligence Engineering & Sciences Publication.

ID: 12679

Agreement Score: 0.43

Analysis of Patients' Reviews and Enhancing Services in Hospitals

Currently, when it comes to hospitals, a large amount of patients visit daily. If the feedbacks or reviews or **opinions** are not analyzed properly, it is bound to affect the business and reputation of hospitals. So it is necessary to analyze the reviews that the patients have about hospitals. We have proposed an approach for mining of feedbacks of patients which will be useful to find the common features in a hospital which can be +ve or −ve. We successfully generate the summary of the reviews and also classify the reviews with more than 80% accuracy with all the classifiers. 'Majority vote' which is the approach proposed by us proved to be more accurate than other classifiers. © 2020, Springer Nature Singapore Pte Ltd.

ID: 4638

Agreement Score: 0.43

VAUT: a visual analytics system of spatiotemporal urban topics in reviews

Online review platforms offer customers the opportunities to express valuable feedback and personal views from various aspects on restaurants, products, works of art, or other items. A majority of previous studies on these user-generated reviews are devoted to controversy, bias, and **opinion** analysis. However, little work has been done to study urban characteristics via topic analysis from the city level in reviews. In this paper, we propose a visual analytics system, to visually explore spatiotemporal urban topics for cultural trend discovery, location mining, and decision making. Specifying a topic by users is supported due to the difference between review text and traditional text, such as news and books, and the diversity of topics and users. Sentiment analysis and statistical analysis are adopted to characterize the temporal trend and sentiment and topic geographical distributions of the user-specific topic. Our system allows the user to interactively explore the time-evolving frequency trend and characteristic geographical distributions of a topic in reviews. We evaluate the effectiveness and usefulness of our system using three case studies in different domains.

●●●●●●●●●●●●●●●●●

ID: 4780

Agreement Score: 0.47

Unveiling: An Unexpected Mid-campaign Court Ruling's Consequences and the Limits of Following the Leader

Strong evidence exists that major campaign-relevant events can have substantial impacts on vote intentions. We know less about how information about such events diffuses and why only some events become salient. We posit that voters often become aware of such exogenous events via a media mechanism. As the salience of the policy issue in the media increases, we argue that, under certain conditions, the media primes the voters to defect from their party and its leader. We investigate these processes by studying an unexpected court ruling during the 2015 Canadian federal election campaign. Based on difference-in-differences and text-as-data approaches, we find that an exogenous court ruling related to immigrant integration led to between a 5 and 11 percentage point decline in the leading party's support. Beyond modeling how campaign-relevant events become salient through the media, we provide evidence about circumstances where leaders should not expect party loyalty to override crystallized **opinions**.