A Survey on Hybrid Framework for Sentiment Analysis using Genetic Algorithm based Feature Reduction

Bhagyaashree Bhyar1, Tejaswini Patil2, Sonali Thosar3

1,2,3Assistant Professor, Dept. of Computer Engg., Dr. D.Y. Patil Inst. of Engineering & Technology, Pune, India

Abstract: Sentiment analysis has become an important opinion mining technique due to the rapid development of Internet technologies and social media. Recent analysis work has represented the effectiveness of different sentiment classification techniques starting from easy rule-based and lexicon-based approaches to a lot of complicated machine learning algorithms. Whereas lexicon-based approaches have suffered from the lack of dictionaries and labeled knowledge, machine learning approaches have fallen short in terms of accuracy. To resolve the quantify ability issue that arises because the feature-set grows, a completely unique genetic rule (GA)-based feature reduction technique is projected. By using this hybrid approach, we have a tendency to are ready to scale back the feature-set size by up to forty-two while not compromising the accuracy. In future, the prediction of the applicability of our proposed work in various areas including security and surveillance, law-and-order, and public administration.

Keywords: Genetic rule, Sentiment analysis, classification technique.

1. Introduction

In today’s era Internet and associated internet technologies have dramatically changed the means our society works. Social networks such as Facebook and Twitter became commonplace for exchanging concepts, sharing data, promoting business and trade, running political and philosophic campaigns, and promoting product and services. Social media is mostly studied from completely different views i.e., collection business intelligence for product and services promotion, watching malicious activities for sleuthing and mitigating cyber-threats, and sentiment analysis for analyzing people's feedback and reviews. Sentiment analysis, often referred as opinion mining, is that the extraction, identification, or characterization of the sentiment from text victimization Natural Language process (NLP), statistics, or machine learning (ML) ways. The field of sentiment analysis has been wide studied by researchers throughout the last few years. Sentiment analysis is generally mentioned within the context of product reviews like; is that this product review positive or negative? Are customers satisfied or dissatisfied? what is more, it additionally helps to answer the Business Intelligence connected queries like; Why aren't shoppers shopping for our product? but, cross-domain insights and applications of sentiment analysis are scarce. The samples of such applications include analysis of user opinion on the politics, sociology, and the science of society. The proposes a hybrid approach to sentiment analysis which employs progressive milliliter algorithms and lexical databases to mechanically analyze archives of on-line documents (e.g., reviews, chats, and social media data). The propose a unique Genetic formula (GA) based mostly answer to feature reduction drawback by developing a bespoke fitness perform. The fitness perform utilizes SentiWord-Net lexicon to calculate the polarity distinction between a class label and possible feature vector (potential solution).

To the most effective of our information, we have a tendency to are the first to use such a hybrid approach with GA based mostly optimized feature selection. This organic process approach for optimum feature selection ends up in multiplied accuracy and higher quantifies ability.

The bespoke fitness perform shows up to forty-two reduced feature-set with none compromise on overall accuracy. Furthermore, so as to demonstrate the feasibleness of the proposed feature reduction formula, to have a tendency to conjointly perform detailed comparison with different feature reduction algorithms including PCA and LSA which ends in system having up to fifteen.4% multiplied accuracy over PCA and up to 40.2% multiplied accuracy over LSA. PCA may be a spatiality reduction procedure that simplifies the complexity in high-dimensional knowledge by reducing an out sized set of variables to a small set that also retains data and trends gift in data. It comes a collection of points onto a smaller dimensional affine topological space of ``best t'’. LSA may be a methodology employed in NLP that discovers a knowledge illustration that includes a lower dimension than the initial linguistics area by analyzing relationships between documents and its terms. It decreases the dimension employing a mathematic technique known as singular value decomposition (SVD).

2. Related Work

In this segment we talk about the conspicuous related research being completed in the zone of opinion investigation
and content mining. Our correlation criteria depends on the two
to examined previously; combination of conclusion
examination approaches in a bound together manner and a
cross-disciplinary application zone. The intrigued to perceive
how client's supposition and his/her social conduct can be
useful in dissecting the current geopolitical circumstance and
uprising.

Medhat et al. [18] presented a comprehensive summary of
the recently planned algorithms, enhancements, and
applications in the space of sentiment analysis. They
additionally mentioned the connected fields to sentiment
analysis e.g., transfer learning, feeling detection, and building
resources. They tried to offer a full image of the sentiment
analysis techniques and connected fields with transient details.

Khan et al. [19] proposed a rule-based domain-independent
method which classifies subjective and objective sentences
from reviews and blog comments. SentiWordNet is used to
calculate the score and to determine the polarity. They showed
their proposed method is effective and it outperforms ML-
based methods with an accuracy of 76.8% at the feedback level
and 86.6% at the sentence level. Our proposed approach is
aligned with these studies as we are also focusing on ML and
lexicon-based methods. However, we are employing GA based
optimized feature selection for training ML algorithms.

Agarwal et al. [20] inspected notion investigation on Twitter
information. They presented POS-specific earlier extremity
highlights furthermore, investigated the utilization of a tree bit
to deter the requirement for dreary element building. Their new
highlights furthermore, the tree piece performed nearly at the
equivalent level and both beat the cutting edge standard
strategies.

Kouloumpis et al. [21] examined the utility of etymological
highlights for distinguishing the assessment of Twitter
messages. They assessed the convenience of the current lexical
assets just as the imaginative language utilized in
microblogging.

[4] introduced a language-autonomous model for slant
examination for short content structures e.g., informal
organizations statuses. They utilized Twitter datasets to
demonstrate glad and miserable opinions and demonstrated that
their framework performed 10% superior to Naive Bayes (NB)
model. These three papers are utilizing feeling investigation on
short-content information i.e., SMS, tweets and so forth. So
also, Pontiki et al. [22] portrayed the angle based supposition
examination. They identified the parts of given target elements
and the assumption communicated for every angle. They
utilized physically commented on surveys of eateries and
workstations as a dataset.

Njolstad et al. [23] proposed, defined, what's more, assessed
four distinctive element classifications formed of 26 article
highlights for notion examination. They utilized diverse ML
techniques to prepare conclusion classifier of Norweign
financial web news articles. They accomplished classification

References


management, technology, and information dimensions," Commun. Assoc.


