

Gender Classification for Intent analysis in Business Intelligence

Praveen Gujjar J¹ | Dr. T Manjunatha¹

¹ SRS College, Chitradurga

² Visvesvaraya Technological University B D T College of Engineering, Davangere-577 004, Karnataka, India,

To Cite this Article

Praveen Gujjar J and Dr. T Manjunatha, "Gender Classification for Intent analysis in Business Intelligence", *International Journal for Modern Trends in Science and Technology*, Vol. 06, Issue 04, April 2020, pp.:100-103.

Article Info

Received on 08-March-2020, Revised on 28-March-2020, Accepted on 31-March-2020, Published on 07-April-2020.

ABSTRACT

Data is the important component in the business intelligence. Internet is the one where it constitutes the huge amount of the data in the form of customer reviews, customer feedback etc., and Gender classification is one of the crucial step for identifying intent of the user. This paper proposed a method for classifying gender and identifying the accuracy of the classification using decision tree, random forest, logistic regression, Support Vector Classifier. Classification purpose Scikit learn is used in this paper. Scikit-learn is a simple and efficient tool for data mining and data analysis. It is built on NumPy, SciPy, and matplotlib. The paper includes advantages of the proposed technique and concludes with the challenges for the marketers when using this technique in their market survey.

KEYWORDS: Intent Analysis, Decision Tree, Random forest, Logistic regression, Support Vector classifier

Copyright © 2014-2020 International Journal for Modern Trends in Science and Technology
All rights reserved.

I. INTRODUCTION

Sentimental analysis (SA) is to determine the tendency or attitude of a communicator through the contextual polarity of their writing or speaking. The Internet has changed the manner in which individuals express their perspectives, by the introduction of web 3.0 user's can have conversation about the product or service with manufacturer or service provider. It is for the most part done through blog entries, online discussions, item survey sites, internet based life, and so on. Users expressing their views, emotions and sentiment through social network sites like Google Plus, Twitter, Facebook etc. Social network is producing a huge volume of notion rich information as tweets, reviews, comments, discussion, blog entries, and so forth. Social media network gives a chance to businesses by giving a

stage to interface with their target customer for advertisement. A user on the most of the cases depends widely on other user's generated content for decision making about the product available in the online. Because of the huge content generated by the users in a daily basis it becomes a difficult job for the ordinary users to analyze the content. Hence there is a great demand to automate the users review. In this context SA has a very vital role to play. Sentiment analysis enlightens the customer whether the data concerning the product is attractive or not before they purchased it. Advertisers and firms utilize this data to comprehend about their product or administrations so that it very well may be offered according to the user's prerequisites. Hence in the context of analysis one can use textual Information retrieval techniques. It mainly focuses on

searching, preparing or analyzing the genuine information present. But, there is some other textual information which may express subjective features. This information's are primarily focused on sentiments, attitudes, opinions, emotions and appraisals, which may form the centre of SA. Intent analysis enhanced version of sentimental analysis. Sentiment analysis tells the statement or emotion is positive, negative or neutral but intent analysis tells intention behind the statement or emotion. This paper deals about gender classification based on the decision tree classifier, random forest classifier, logistic regression and support vector classifier.

II. INTENT ANALYSIS

Sentimental analysis is a term that incorporates numerous undertakings, for example, conclusion extraction, assessment characterization, subjectivity arrangement, rundown of suppositions or spam detection.

In the preliminary stage intent analysis seems like sentimental analysis only. But it is the advanced version of sentimental analysis. For the instance consider the statement "I will buy a new smart phone", here the statement is positive and user may have purchase the new smart phone. Consider the other statement "I am saving like a crazy, flagship cell phone here I come", here there is no word like purchase or buy even though user intention is purchase of best smart phone. An intent analysis tool would tag the second sentence as follows:

intendee = "I"

intention = "buy"

intended object = "Flagship"

The intent analysis is based on grammar-parsing technology, and a portion of natural language processing.

In the recent past wide number of a great deal of work has been done in the field of sentimental analysis by number of scientists and scholars. In its beginning period it was expected binary classification to classify either positive and negative. In the following section let's understand different technique used for sentimental analysis. Pak and Paroubek (2010) proposed techniques to classify the twitter tweets as neutral, positive or negative. They were used twitter API to collect the tweets and for the analyses used Naive based algorithms.

Badenhorst et.al (2012) Proposed that artificial intelligence, machine learning capabilities enhance

the growth in the abilities of machines to say intelligent.

Carlos et.al (2012) analyses the importance of intent for marketing sales and services. With the help of available user comments automatically discovered the intent of the user.

Po-Wei Liang et.al. (2013), Proposed a solution for SA with reference to twitter data. These data are collected using the Twitter API.

Liu (2013), analyzed that there are various challenges and issues with respect to opinion mining in proposing emotions of the people from the data available in the twitter. Further they quoted that varying languages makes opinion mining as challenging task.

Funk and Bontcheva (2008) used NLP techniques as input to machine learning instrument to conduct opinion analysis for business intelligence applications.

Akulick er.al (2017) analyses the intent detection with the help of text mining this research relies upon n-grams, support vector machine for identifying the intent of the customer.

Piryani et.al (2017) analyses the citation pattern of the paper with the help of sentimental analysis.

Praveen Gujjar et. al. (2018) said that the purpose of a business organization is to make profit. The profitability analysis is done to throw light on the current operating performance and efficiency of business firms.

III. PROPOSED METHOD

For the gender classification training data is used person height, weight and shoe size. The proposed classification method includes the following:

1. Decision tree classifier
2. Random forest classifier
3. Logistic regression
4. Support vector

Aforesaid proposal has been implemented in Python Language.

Decision classifier uses the three key terms

1. Impurity
2. Entropy
3. Information gain

Pseudocode for the Decision tree classifier is as shown

```
dtc_clf = dtc_clf.fit(X,Y)
```

```
dtc_prediction = dtc_clf.predict(test_data)
```

Random Forest classifier is assembling of algorithm. It covers more than one algorithm of the same or different classifying objects. It creates a set of decision trees from a randomly selected

subset of the training set. It then aggregates the votes from different decision trees to decide the final class of the test object.

Pseudocode for the Random Forest Classifier is as shown

```
rfc_clf.fit(X,Y)
rfc_prediction = rfc_clf.predict(test_data)
```

Logistic regression is a method for predicting binary classes. The outcome variable is dichotomous in nature. It computes the probability of an event occurrence.

Pseudocode for the Logistic Regression Classifier is as shown

```
l_clf.fit(X,Y)
l_prediction = l_clf.predict(test_data)
```

A Support Vector Machine is a classifier by a separating hyperplane. In two dimensional space, this hyperplane is a line dividing a plane into two parts wherein each class lay in either side.

Pseudocode for the Support Vector Classifier is as shown

```
s_clf.fit(X,Y)
s_prediction = s_clf.predict(test_data)
```

IV. RESULTS AND DISCUSSION

For the results researcher has taken person height, weight and Shoe size. As shown in the example below

```
test_data = [[154, 75, 38], [190, 70, 43],
,[181,65,40]]
test_labels = ['female','male','male']
```

Decision Tree classifier predicts the test data as shown below

Output is ['female' 'male' 'female']

Random Forest Classifier predicts the test data as shown below

Output is ['female' 'male' 'female']

Logistic Regression predicts the test data as shown below

Output is ['male' 'female' 'female']

Support Vector Classifier predicts the test data as shown below

Output is ['female' 'female' 'female']

Accuracy calculation for the classification of the gender and predicting the new person gender problem reveals that Random Forest is the best classifier for this problem. For the test purpose researcher used four basic classification algorithms those are Decision tree classifier, Random forest classifier, and logistic regression and support vector classifier. Among these classifier Random forest classifier is the best classifier for the aforesaid problem.

V. ADVANTAGES

There are plenty of advantages in intent analysis. Few of them are listed below

1. The proposed technique highly applicable in business intelligence.
2. The proposed technique may be used in the placing the Ad's based on user interest.
3. The proposed technique may help decision makers to have product and service benchmarking.

VI. CHALLENGES

Even though there are a quite possible number of advantages in aforesaid analysis, it involves certain challenges to be addressed. This section describes few issues in the following lines.

1. In the Intent analysis there is an issue for emojis.
2. For a biased statements intent analysis may result in wrong intentions.
3. Wide usage of the different languages, intent analysis may difficult to analyze the intentions.

VII. CONCLUSION AND FUTURE SCOPE

A practical, computationally cost-effective gender classification is proposed for the intent analysis in the field of business intelligence the technique used in this paper highly applicable in business intelligence. The proposed technique may help decision makers to set product and service benchmarking. In this paper four classifier is used those are Decision tree classifier, Random forest classifier, Logistic regression and support vector classifier. In this classifier Random forest classifier is considered as the best classifier for classifying the gender. This paper also focused on the challenges faced in the intent analysis with reference to Text message. In the future scope use of the supervised learning and unsupervised learning can be used to address the challenges faced in the intent analysis particularly for emojis

REFERENCES

- [1] A.Pak and P. Paroubek. (2010), "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, pp.1320-1326
- [2] Badenhorst, B. K. and Fitzgerald, W. (2012), "The application of text and sentiment analysis in market research", social sciences and business intelligence in South Africa. Southern African Marketing Research Association (SAMRA), available: <http://www.samra.co.za/wp-content/uploads/2013/05/Baden>

[horstFitzgerald Text-and-sentiment-analysis-Expert-QA-Materials.pdf](#)

- [3] B. Liu, (2012), "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, volume 5, Issue. 1, pp. 1-167.
- [4] Cohan Suja y Carlos and Madhulika Yalamanch (2012), "Intention Analysis for Sales, Marketing and Customer Service", Proceedings of COLING, pp.33-40
- [5] Funk, A., and Bontcheva, K. (2008), "Opinion Analysis for Business Intelligence Applications" Proceedings of the first international workshop on Ontology-supported business intelligence 2008 Article No.3, available: <https://gate.ac.uk/sale/iswc08/obimusing/musing-om-iswc08.pdf>.
- [6] Po-Wei Liang, Bi-Ru Dai, (2013), "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3 - 6, 2013, pp 91-96, ISBN: 978-1-494673-6068-5, <http://doi.ieeecomputersociety.org/10.1109/MDM.2013>.
- [7] Praveen Gujjar and T Manjunatha, (2018), "Profitability Analysis of Indian Information Technology Companies using DuPont Model", Asian Journal of Management, 2019, Volume 9, Issue 3, pp 1105-1108.
- [8] R. Piryani, D. Madhavi, and V. K. Singh, (2017), "Analytical mapping of opinion mining and sentiment analysis research during 2000-2015," Information Processing & Management, vol. 53, no. 1, pp. 122-150.
- [9] Samantha Akulick and El Sayed Mahmoud (2017), "Intent Detection through Text Mining and Analysis", Future Technologies Conference (FTC), pp. 493-496.