Issue Ticketing System Based on Machine Learning

SABITHA JETTY,
Assistant Professor,
Department of Computer Science and Engineering,

PADALA PAVAN KUMAR,
Assistant Professor,
Department of Electronics and Communications Engineering,

Siddhartha Institute of Technology and Sciences,
Narapally, Hyderabad, Telangana – 500 088.

Abstract

Ticketing systems have gained significance and constitute a critical feature as businesses strive for superior customer experience and satisfaction in terms of business viability. Various software businesses have the development of very effective software solutions for problem tracking, several sub-processes and duties inside ticketing systems are still carried out manually. These were done by hand. Tasks, especially in large companies, are bottlenecks. They lead to a decrease in productivity and an increase in reaction time. Machine learning advances may be linked with existing issue tracking and ticketing systems on the market in a unique way to provide maximum operational efficiency in large-scale enterprises' Client Service and Support (CSS) departments that deal with customer reports. This study provides a holistic approach to customer service by addressing three apparently unrelated bottlenecks in the ticketing system: spam detection, ticket assignment, and sentiment analysis.

1. Introduction

Because of the competitive market, each organization's Customer Support and Service (CSS) Department must meet and exceed consumers' expectations. As a result, the CSS department is experiencing major modifications in order to achieve efficient business procedures. As a result, it's critical for businesses to identify slow-running processes, conduct root-cause analyses, and then enhance their performance.

In practise, however, issue resolution systems are unproductive and prone to errors. As consumers, we frequently see pitiful circumstances in which we must wait extended periods of time for a response or resolution to a reported malfunctioning product. Manual work is the primary cause of support service delays.

The major goal of this article is to identify and improve ticketing system bottlenecks. This may be accomplished by using low-cost artificial intelligent components with simple artificial intelligence components. Furthermore, the goal of this project is to create a system that unites the three primary automation components that share the same implementation logic. The result would be an intelligent ticketing system that, in the best-case scenario, might be adopted by whole enterprises.

The result would be an intelligent ticketing system that, in the best-case scenario, might be adopted by whole enterprises. Organizations may, however, incorporate each of these components individually and independently to handle each of the difficulties separately. Although machine learning techniques have been used to solve a variety of document classification problems, their application to automatic spam detection in a ticketing system is still a work in progress.

The main contribution is the implementation of a unique approach, a conservative unanimity method that combines the output of numerous spam filtering classifiers as an aggregation strategy. This method has the lowest rate of false positives. As a whole, handling spam filtering, ticket assignment, and sentiment analysis as a package is unique since no other research article has taken all three factors into account.

2. Literature survey

The ever-increasing problem of unsolicited emails has become a major issue for both businesses and people. A main dataset from a software firm in Liechtenstein was utilised and examined to explore the problem of unwanted emails in the context of ticketing systems. The business has developed its own ticketing system. Spam emails are a problem for them.

Spam emails are sent to the support email address, which is set up to receive client reports and concerns. Incoming emails to this address are automatically passed to the ticketing system, which generates tickets and incidents automatically. Apart from genuine emails, numerous spam emails are also handled; as a result, spam emails produce spam tickets by default.

During the examination of the dataset, important information concerning another type of email was discovered. The software development firm is not only dealing with spam emails, but it has also identified the issue of getting unwanted emails. Unimportant automatic notification emails are sent from various client server email addresses on a regular basis, and these undesired emails produce unwanted tickets. Unwanted emails use the same amount of time as spam emails in terms of time usage.

They use supervised machine learning to solve this problem. The major goal is to develop a system that can accurately filter undesirable emails while also reducing the number of false positives. Furthermore, a majority voting system technique has been used, which takes into account the majority of classifiers' decisions before concluding whether or not an email is undesirable.

Traditional ticketing systems include surveys as standard features for gauging consumer satisfaction. Following the resolution of a problem, the system prompts clients to rate the service by completing a brief customer satisfaction survey. This comprises of a few service-related inquiries delivery. Despite the fact that many people have used this strategy. Customers do not accept it, as it is not commonly acknowledged by enterprises. It takes a long time to complete a survey.

A lot of stuff is created while filing a ticket (reporting a bug, requesting information, asking for help, etc.). To recognise customer attitudes, views, and emotions stated in tickets, an intelligent technique is required. The number of pleased consumers can be converted into monetary gains.

3. Methodology

This section describes the implementation aspect of our system, including the dataset, tools, and machine learning approaches that we employ. To assess the performance of the algorithms, we use primary data. The dataset was retrieved from a software development company's ticketing system. It contains 18,917 entries, all of which are emails created from customer reports, from which tickets are automatically generated.

This technique identified German as the most commonly used language on around 72 percent of the tickets, and English on approximately 28 percent of the tickets. The data corpus we used in our experiment ranges from March 9, 2009 to March 28, 2017. The dataset has 14 columns, three of which we recognised as being important to the classification challenges.

Table 1. Dataset details

No.of. columns	14
Total tickets	18917
Total words in subject	95015
Total words in description	1753667
Average words per subject	516
Average words per description	953

- **Data Preprocessing, Feature Extraction and Training**

Because the only information available from the ticket is text, we begin by cleaning it, removing punctuation and stop words (both "english" and "german"). We use a bag-of-words technique, which means that individual words match to characteristics gathered via the use of a word tokenizer. The term-frequency, inverse document frequency tf-idf statistic is used for weighting characteristics, i.e. words. We train a set of classifiers for our tasks based on the retrieved characteristics from each ticket.

We utilise the pre-processed dataset given above to train and evaluate the performance of a range of machine learning classifiers, i.e. we divide the dataset into training and test sets. We adhere to current best practises by dividing the dataset into 90 percent of the data for training and 10 percent of the data for testing.

- **Spam detection**

Because only 3,880 of the tickets are tagged as spam, we employ a balanced dataset of 3,880 ham tickets, totalling 7,760 tickets in our scenario. Following that, we divided the data into 90 percent training and 10% testing sets, with 6,984 and 776 tickets, respectively. During the training phase, we do 10-fold cross-validation, where the accuracy is calculated by taking the mean of the 10 runs and also calculating the standard deviation.

Furthermore, sentiment analysis is the computer-assisted identification of textual views, sentiments, and subjectivity. Sentiment analysis, unlike spam detection and ticket triage, is a very young and active research subject. Before the year 2000, there was virtually little study on sentiment analysis and opinion mining, and sentiment analysis only became popular around 2004.

The sentiment analysis problem has been used to a variety of areas, first for assessing online product evaluations and then primarily for analysing text on social media, including election analysis, stock market prediction, and medical. There have also been a number of firms that offer online opinion mining services.

4. Result and discussion

During the assessment or, more accurately, testing phase, 10% of the tickets were used, resulting in 776 tickets (388 ham + 388 spam). We extract characteristics from the text provided on the subject and the ticket description for each of the 776 tickets, and each method predicts the class.

We tested different combinations for the MV approach, and the greatest result was obtained by considering the majority vote of the top three performers: SVM + Nave Bayes + Logistic Regression. We also evaluate the weighted majority voting approach (WMV), which assigns weights to each algorithm depending on their cross-validation results.

To reduce the number of false positives even further, we use unanimity with confidence threshold as a next technique [25], which takes into account each algorithm's likelihood of prediction. Every categorised assignment is confident thanks to the scikit-learn toolbox. We use this parameter to set a confidence level for the algorithms in the unanimity strategy's probability estimate.

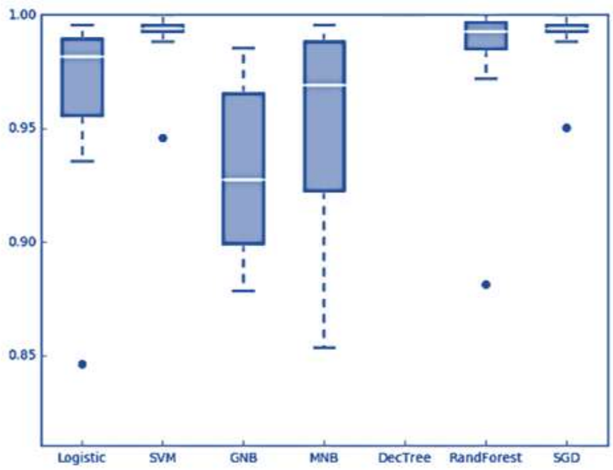

Fig.1 The performance of the classifiers is depicted in a boxplot.

However, while weighing the risks, the latter approach would take more time because a support staff member would have to manually sift these tickets. When the confidence level is increased to 0.95, the precision is 100 percent, which means there are no false positives (no ham ticket may be labelled as spam). However, this permits almost 33% of spam tickets to be labelled as ham, implying that the ticketing system receives more spam tickets.

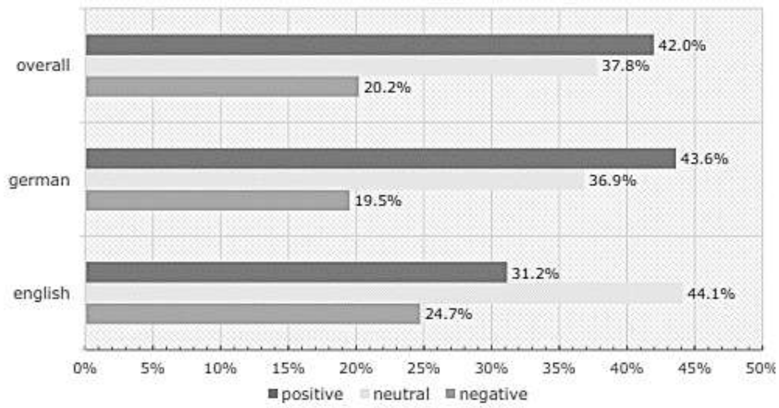

Fig.2 Sentiment analysis

5. Conclusion

The general goal of this research was to develop a system that could provide a completely automated ticketing system that addressed concerns in three areas: spam filtering, automatic ticket assignment, and sentiment analysis. We created a prototype application for a real-world scenario using data from a ticketing system given by a software development business, with the primary goal of identifying spam tickets and, as a secondary step, assigning legitimate tickets to the appropriate department within the organisation.

To solve the issue of false positives in automatic spam detection, we use a hybrid technique that combines various models. In return for enabling more spam tickets to be recognised as real tickets, our unanimity with confidence threshold hybrid model tightly filtered the false positives. We were unable to train our own classifier tailored to our dataset due to a lack of ground truth annotation in the dataset for the sentiment analysis part, so this part remains the first major future work, especially given its importance to the company, as it will enable the company to automatically find and prioritise tickets with unsatisfactory content (negative sentiment).

References

1. K. Blackwell, Achieving high customer satisfaction. ”Course Technology”, 2015.
2. A.-W. Scheer, F. Abolhassan, W. Jost, and M. Kirchmer, Business process automation. Springer, 2004.
3. C. Rygielski, J.-C. Wang, and D. C. Yen, “Data mining techniques for customer relationship management,” Technology in society, vol. 24, no. 4, pp. 483–502, 2002.
4. L. F. Cranor and B. A. LaMacchia, “Spam!” Communications of the ACM, vol. 41, no. 8, pp. 74–83, Aug. 1998. [Online]. Available: http://doi.acm.org/10.1145/280324.280336
5. Fiaz, N. Devi, and S. Aarthi, “Bug tracking and reporting system,” arXiv preprint arXiv: 1309.1232, 2013.
6. T. S. Guzella and W. M. Caminhas, “A review of machine learning approaches to spam filtering,” Expert Systems with Applications, pp. 10 206–10 222, 2009.
7. S. Nazirova, “Survey on spam filtering techniques,” Communications and Network, vol. 20, no. 3, 2011. [Online]. Available: http://dx.doi.org/10.4236/cn.2011.33019
8. J. Smola and B. Scholkopf, “A tutorial on support vector ¨ regression,” Statistics and computing, vol. 14, no. 3, pp. 199– 222, 2004.
9. V. Dedik and B. Rossi, “Automated bug triaging in an industrial context,” in 2016 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Aug 2016, pp. 363–367.
10. D. Cubranic, “Automatic bug triage using text categorization,” in In Proceedings of the Sixteenth International Conference on Software Engineering and Knowledge Engineering, 2004, pp. 92–97.