Abstract

Spam is one of the serious problems faced by internet community globally. Spam Detection is a critical issue in business world. In this paper an intelligent three stage model is presented to perform the spam inclusive outlier identification. The SMS textual dataset is taken as input and than its filtration is done. After that this textual information is converted to the statistical information using fuzzy and assign the weights to dataset. The decision tree algorithm is than applied on this fuzzy weighed dataset to classify the dataset. This algorithm is defined to separate the spam and non spam data values. A comparison of existing Bayesian and proposed Fuzzy based decision tree approach is done. The results shows that the recognition rate is improved using the proposed approach. The work is implemented in weka integrated java environment.

References

- Nikita Spirin, Jiawei Han, Survey on Web Spam Detection: Principles and Algorithms, SIGKDD Explorations Volume 13, Issue 2.
- N. S. Kumar, D. P. Rana, R. G. Mehta, Detecting E-mail Spam Using Spam Word Associations, IJETAE Volume 2, Issue 4, April 2012.
- Zhen Yang, Xiangfei Nie, Weiran Xu, and Jun Guo, An Approach to Spam Detection by Naive Bayes Ensemble Based on Decision Induction, ISDA’06, 2006.
- Matthias Reif, Markus Goldstein, Armin Stahl, Anomaly Detection by Combining Decision Trees and Parametric Densities, 2008 IEEE.
- Kelton Costa; Patricia Ribeiro; Atair Camargo; Victor Rossi; Henrique Martins; Miguel Neves; Ricardo Fabris, Comparison of the Techniques Decision Tree and MLP for Data Mining in Spams Detection To Computer Networks, 2013 IEEE.
- N N R Ranga Suri, An Algorithm for Mining Outliers in Categorical Data through Ranking; 978-1-4673-5116-4/12@2012 IEEE
- Ms. K. Mouthami, Sentiment Analysis and Classification Based On Textual Reviews;
- Farkhund Iqbal, Mining Criminal Networks from Chat Log; 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology 978-0-7695-4880-7/12 © 2012 IEEE.

Index Terms

Computer Science
Fuzzy Systems
Keywords
Spam Detection  Outlier detection  Data mining  fuzzy logic  decision tree (DT)  weka