

# Automated Text Document Classification Using Predictive Network

B. Y. Baravkar

Assistant Professor, Department of Computer Engineering, DGOI, FOE, Maharashtra, India

**Abstract:** *Every day the mass of information available, merely finding the relevant information is not the only task of automatic text classification systems. Instead the automatic text classification systems are supposed to retrieve the relevant information as well as organize according to its degree of relevancy with the given query. The main problem in organizing is to classify which documents are relevant and which are irrelevant. The Automated text classification consists of automatically organizing clustered data. We propose an automatic method of text classification using machine learning based on the disambiguation of the meaning of the word we use the word net word embedding algorithm to eliminate the ambiguity of words so that each word is replaced by its meaning in context. The closest ancestors of the senses of all the undamaged words in a given document are selected as classes for the specified document.*

**Keywords:** Descriptive clustering, feature selection, logistic regression, model selection, machine learning.

## 1. Introduction

The mass of data obtained to us increases. This data would be irrelevant if our capacity to productively get to did not increment too. For most extreme advantage, there is need of devices that permit look, sort, list, store and investigate the accessible information. One of the hopeful region is the is the automatic text classification. Envision ourselves within the sight of impressive number of texts, which are all the more effectively available on the off chance that they are composed into classes as per their topic. Obviously one could request that human read the text and arrange them physically. This assignment is hard if done on hundreds, even a huge number of texts. Thus, it appears to be important to have a computerized application, so here automatic text categorization is presented. Growing the number of Data Mining presentations include the complex structured type of Data Require the use of sensitive language. In this more application cannot solve by using Traditional Data Mining Algorithm

### 1.1 Existing system

Unfortunately Existing approach, especially by using Logic Programming methods, often suffer is not low scalability. When allocating with difficult Data Base schemas but also insufficient predictive performance usage noisy value in real-life applications. Though, leveling approaches be wont to need significant time effort for the data conversion, outcome in behind the compressed representations of the standardized Database, produce are extremely large data with huge Number of extra attribute value frequent NULL values. In this system result is problems have prevented a diverse application of multi Relational Mining, and

challenge to the Data Mining community. This article introduces a Descriptive clustering methodology where flattening is required to bridge the gap between propositional learning algorithms and relational.

### 1.2 Disadvantages of Existing System:

- 1) Not worked on automated text classification
- 2) Everything done manually
- 3) Time consuming process
- 4) No repository Available
- 5) Domain wise analysis is time consuming

## 2. Proposed System

In proposed methodology, data analysis technique, such as clustering it can be classify subset of information example with mutual characteristics. Operator can explore the information by examining Users can explore the data by examining selected example in each group instead of relatively examining the instances of the complete data set. This allows users to focus efficiently on large relevant subsets Data sets, in particular for document collections. In particular, the descriptive grouping consists of automatic combination set of the parallel instance in clusters and automatically generates a description or a synthesis that can be interpreted by man for each group. The description of each cluster allows a user determine the relevance of the group without having to examine its content For text documents, a description suitable for each group can be a multi-word tag, an removed name. Quality is the grouping is important, so that it is aligned with the idea of likeness of the user, but it is similarly imperative to deliver a user with a brief useful instant that correctly replicates the constant of the cluster.

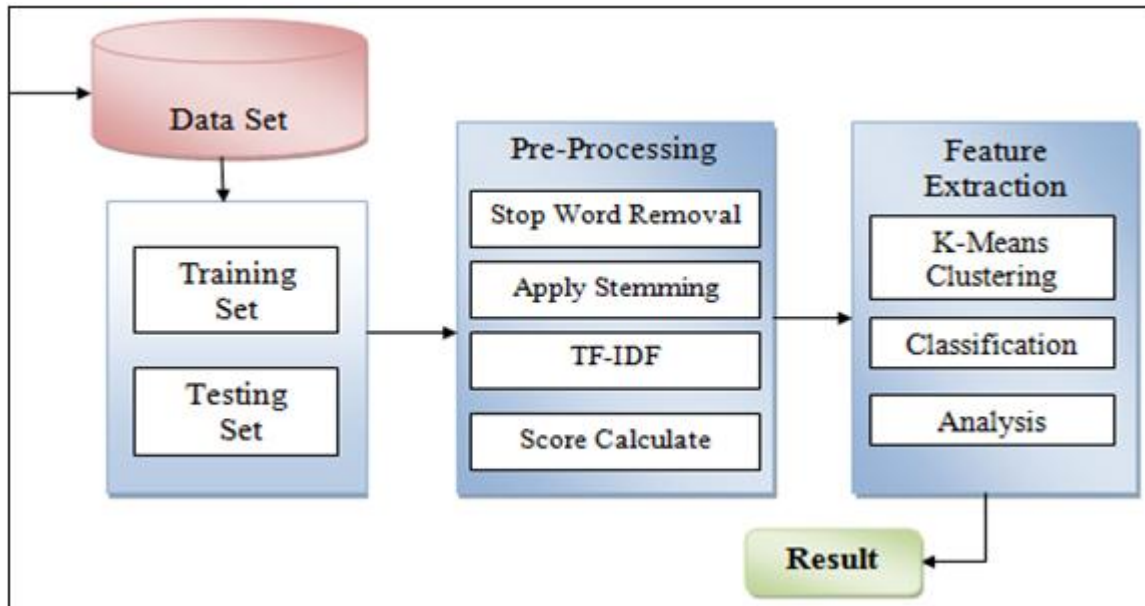


Figure 1: System Architecture

### 3. Result

Classification Using ACO-GA	
Classification	Show Analysis
Paper ID	Classified Domain
A Highly Accurate Prediction Algorithm for Unknown Web ...	cloud computing
A Holistic Approach to Distributed Dimensionality Reducti...	cloud computing
A Hybrid Cloud Approach for Secure Authorized Deduplica...	cloud computing
A Methodology for Direct and Indirect Discrimination Prev...	cloud computing
A Personalized Mobile Search Engine_9	cloud computing
A QoS-Oriented Distributed Routing Protocol for Hybrid W...	wireless sensor network
A Scalable Two-Phase Top-Down Specialization Approach f...	cloud computing
A System for Denial-of-Service Attack Detection Based on ...	cloud computing
Accuracy-Constrained Privacy-Preserving Access Control M...	data mining
An Empirical Performance Evaluation of Relational Keywor...	database security
An Error-Minimizing Framework for Localizing Jammers i...	wireless sensor network
An Incentive Framework_9	mobile computing
Anonymizing Collections of Tree-Structured Data_1	data mining

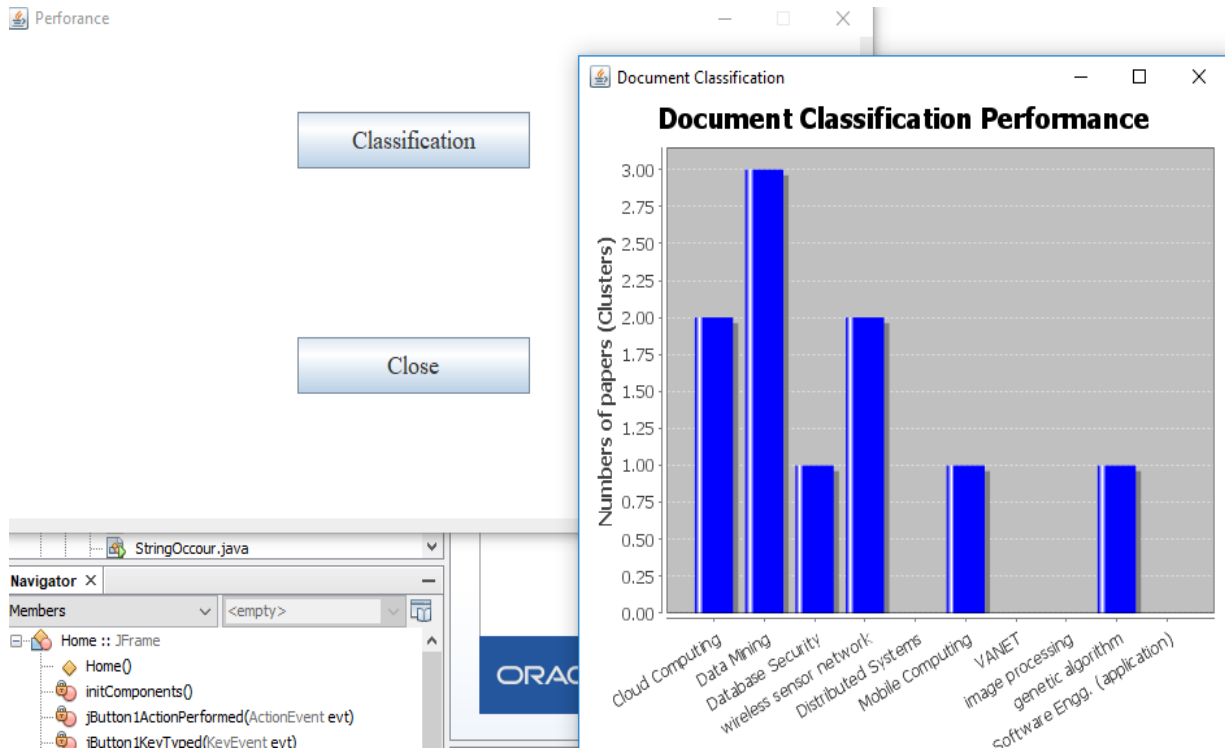


Figure 2: Classification

#### 4. Conclusions

Proposed descriptive Clustering as two coupled predictions activity choose a grouping that is predictive of features and prediction of the cluster assignment of a subset of features. Use predictive performance as a goal criterion, descriptive clustering parameters the number of clusters and the number of functions per Clusters: they are chosen from the model selection. With the result solution, each group is described by a minimum subset of features necessary to predict if an instance belongs to the cluster our hypothesis is that even a user will be able to predict membership in the group of documents using the descriptive features selected by the algorithm. Given Some relevant requirements, a user can quickly identify clusters that probably contain relevant documents.

#### References

- [1] Chi Chen, Member, IEEE, Xiaojie Zhu, Student Member, IEEE, Peisong Shen, Student Member, IEEE, Jiankun Hu, Member, IEEE, Song Guo, Senior Member, IEEE, Zahir Tari, Senior Member, IEEE, and Albert Y. Zomaya, Fellow, IEEE, An Efficient Privacy Preserving Ranked Keyword Search Method(2016)
- [2] J.-T. Chien, Hierarchical theme and topic modeling, IEEE Trans. Neural Netw. Learn. Syst., vol. 27, no. 3, pp. 565578, 2016.
- [3] Bernardini, C. Carpineto, and M. DAmico, Full-subtopic retrieval with keyphrase based search results clustering, in IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intelligent Agent Technol., 2009, pp. 206213.
- [4] Q. Mei, X. Shen, and C. Zhai, Automatic labeling of multinomial topic models, in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2007, pp. 490499.
5. R. Xu and D. Wunsch, Survey of clustering

algorithms, IEEE Trans. Neural Netw., vol. 16, no. 3, pp. 645678, 2005.

#### Author Profile



Assistant Professor, DGOI, FOE, Swami-Chincholi.