Vol.5 Issue 5, Pg.: 53-60
May 2017

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

FEATURE EXTRACTION BASED TEXT MINING WITH NEURAL NETWORK FOR FAULT DETECTION OF RAILWAY SYSTEMS

Akshay Gulhane

Nutan Maharashtra Institute of Engineering and Technology,
Talegaon Dabhade, Pune, India.
akshaysgulhane@gmail.com

Abstract: - In recent years, huge amount of text data is recorded in the forms of repair verbatim in railway maintenance sectors. Efficient text mining of such maintenance data plays an important role in detecting anomalies and improving fault detection efficiency. However, unstructured verbatim, high dimensional data, and imbalanced fault class distribution pose challenges for feature selections and fault detection. The work proposes two level feature extraction based text mining that combines features extracted at both syntax and semantic levels with the aim to improve the fault classification performance. Firstly back propagation algorithm is used to train the data. Then perform an improved X^2 statistics based feature selection at the syntax level to overcome the learning difficulty caused by an imbalanced data set. Then reselect the common features according to both relevance and Hellinger distance. This can be categorized as feature selection at the syntax level. Then, perform a prior LDA based feature selection at the semantic level to reduce the data set into a low dimensional topic space. Then, combine fault features derived from both syntax and semantic levels via serial fusion. Finally, Support Vector Machine (SVM) is used for classification. The proposed method uses fault features at different two levels and enhances the accuracy of fault detection for all fault classes.

Keywords*:* Fault detection, feature extraction, neural networks, railway systems, text mining.

1. Introduction

Text mining is a knowledge-intensive task and is gaining more and more attention in several industrial fields, such as, aerospace, automotive, railway, power, medical, biomedicine, manufacturing, sales and marketing sectors.[1],[2] In a railway field, advanced information technologies, such as sensor networks, RDIF techniques, wireless communication, and Internet cloud, are used to observe the health of the railway systems. In the event of malfunctioning, the diagnostic trouble symptoms are generated and transmitted to the observing center database by wired/wireless communications. After every detection episode a repair verbatim is recorded, which consists of a textual description of the mixture of fault sign (i.e., fault terms), e.g., “Speed Distance Unit (SDU) relevant faults,” a fault sign associated with a specific part, e.g., “SDU,” failure modes (i.e., fault classes), and finally corrective actions, e.g., “replaced SDU,” taken to fix its faults. In railway field, millions of such repair verbatims are generated every year. From repair verbatim data, text mining techniques can be used to establish the associations between fault terms and fault classes such that these associations can be used to improve the accuracy of fault detection.

However, the task of automatic discovery of knowledge from the repair verbatim is a non-trivial exercise most importantly due to the following reasons: High-dimension data, imbalanced fault class distribution, unsupervised text mining models.

This work proposes a two level feature extraction based text mining with neural network for fault detection to meet the aforementioned challenges by automatically analysing the repair verbatims. The main idea is to extract fault features at both the syntax and semantic levels respectively and then combine them to achieve the desired outcome. Considering the fact that the extracted features at each level gives a different importance to a particular aspect of feature spaces and has its deficiencies, the proposed feature combination of two levels may improve the accuracy of fault detection for all fault classes, mainly minority ones. The back propagation algorithm in neural network is used to trained the data or predict data for the efficient results.

At the syntax level, propose an improved ICHI statistics to manage with the feature selection of imbalanced data set. Firstly, overcome the negative effect of imbalanced data set by adjusting the feature weight of minority and majority classes. This makes minority classes comparatively far away from the majority ones. Then, proposed work considers the Hellinger distance [3] as a decision criterion for feature selection, which is shown to be imbalance insensitive. The proposed ICHI can be regarded as feature selections at the syntax level because it mainly uses the document word matrix.

At the semantic level, propose an LDA [4] with prior knowledge (PLDA) to perform the feature extraction. By representing documents in topics rather than word space, this work is able to provide more feature extraction at the semantic level to recompense those extracted at the syntax level. The integration of prior knowledge with the basic LDA is based on the fact that LDA, as an unsupervised model, cannot deal with such issues as selecting topic counts and reducing the adverse effect of common words, which may not produce topics that conform to a user's existing knowledge. Prior knowledge helps us guide topic mining in basic LDA.

Finally, fuse the extracted features derived from the syntax level with the semantic one by serial fusion to improve Support Vector Machine (SVM) based fault detection for all fault classes, especially minority ones.

2. Review of Literature

Implicit feature detection, also known as implicit feature identification, is a crucial aspect of feature specific opinion mining but earlier works have frequently ignored it. Based on the explicit sentences, several Support Vector Machine classifiers can be established to do this job. However, they believe it is possible to do better by using a constrained topic model as an alternative of traditional attribute selection methods. This paper proposes an approach for implicit feature detection based on Support Vector Machine (SVM) and Topic Model(TM). The Topic Model, which included into constraints based on the pre-defined product feature, is established to extract the training attributes for SVM. In the end, several SVM classifiers are constructed to train the selected attributes and utilized to detect the implicit features.[5]

Traditional classification algorithms can be partial in their performance on highly unbalanced datasets. A popular flow of work for countering the problem of class imbalance has been application of a sundry of sampling strategies. In this work, they have focus on designing modifications to Support Vector Machine to appropriately tackle the problem of class imbalance. They incorporate different "rebalance" heuristics in SVM modelling including cost sensitive learning, over sampling, and under sampling. In specific, the SVM variations considered in this work, the novel Granular Support Vector Machines - Repetitive Under sampling algorithm (GSVMRU) is the best in terms of both effectiveness and efficiency. GSVM-RU is effective as it can reduce the negative effect of information loss while maximizing the positive effect of data cleaning in the under sampling process.[6]

Feature selection is a vital pre-processing step for text classification job used to solve the curse of dimensionality problem. Most existing metrics (such as information gain) only access features individually but completely ignore the redundancy between them. This can reduce the overall discriminative power because one features predictive power is weakened by others. On the other side, though all advanced order algorithms (such as Minimum Redundancy Maximum Relevance) take redundancy into account, the high computational complexity renders them inappropriate in the text domain. This work proposes a metric called global information gain (GIG) which can avoid redundancy logically. An efficient feature selection technique called maximizing global information gain (MGIG) is also given. Moreover, MGIG runs significantly faster than the traditional advanced order algorithms, which makes it a proper choice for feature selection in text field.[7]

An amount of feature selection metrics had been explored in text category, among which information gain; chi-square, correlation coefficient and odds ratios are considered most effective. Correlation coefficient and odds ratios are one sided metrics while information gain and chi-square are two sided. Feature selection using one-sided metrics selects the features most indicative of association only, while feature selection using two-sided metrics implicitly joins the features most indicative of membership (such as positive features) and non-membership (such as negative features) by ignoring the symbols of features. The previously never consider the negative features, which are fairly valuable, while the latter cannot make sure the optimal fusion of the two kinds of features especially on imbalanced data. In this paper, they investigated the usefulness of explicit control of that fusion within a proposed feature selection framework. [8]

There are many techniques available to data mining practitioners, including Artificial Neural Networks (ANN), Regression, and Decision Trees. Many practitioners are wary of Neural Networks due to their black box nature, even though they have proven themselves in many conditions. This paper is an overview of artificial neural networks in data mining. The use of neural networks in data mining is a promising field of research particularly given the ready availability of large mass of data sets and the reported capability of neural networks to detect and incorporate relationships between a large numbers of variables. Neural networks are becoming very popular with data mining practitioners, especially in medical research, finance and marketing.[9]

A bi-level feature extraction based text mining that combines features extracted at together syntax and semantic levels with the aim to improve the fault classification performance. They firstly perform an improved X2 statistics based feature selection at the syntax level to conquer the learning difficulty caused by an imbalanced data set. Then they reselect the common features according to both relevance and Hellinger distance. This can be categorized as feature selection at the syntax level. Secondly they perform a prior latent Dirichlet allocation based feature selection at the semantic level to decrease the data set into a low dimensional topic space. Finally, they combine fault features derived from both syntax and semantic levels via serial fusion. [10]

3. The Proposed Work

Extraction based text mining has been a focus of research for several years. This work proposes a two level feature extraction based text mining with neural networks for fault detection to meet the challenges by automatically examining the repair verbatim. It is a non-trivial exercise most importantly due to the following reasons: high-dimension data, unsupervised text mining models, imbalanced fault class distribution. Our main idea is to extract fault features at syntax level, semantic level. Neural network train the data. Then this work combines the fault features derived from syntax level, semantic. Finally proposed work classifies them by using SVM to get desired result. The below fig. 1 block diagram shows the proposed work procedure.

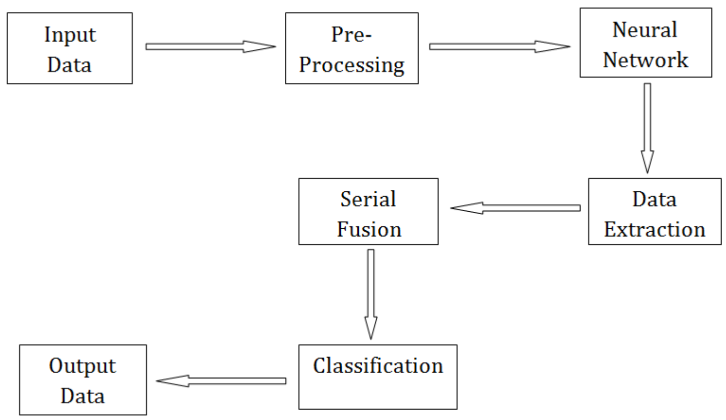

Fig. 1: Proposed work block diagram

Pre-processing: The document annotation helps to filter out the information that is not related for analysis and it provides a specific background for the reliable understanding of the data. Initially, the pre-processing steps the sentence boundary detection, are utilized to part a repair verbatim into partitioned sentences, the stop words are erased to the non-expressive terms, and the lexical matching identifies the right significance of abbreviations. Pre-processing includes normalization, cleaning, transformation, feature selection, etc.

Data Extraction: Data extraction is an initial step for unstructured text analysing. Simplification of text is the work of data extraction. The main work is to recognize phrases and discovers the relation between them. It is suitable for the bulky size of text. It extracts structured information from unstructured information. For data extraction, proposed work have used relationship based prior knowledge extraction and Radial basis function neural network.

Classification: Classification is to find the main theme of document by adding Meta and analysing document. The count of words and from that count decides the topic of the document which was done by the classification technique. Multi SVM is used for classification.

Vol.5 Issue 5, Pg.: 53-60
May 2017

4. Neural Network

Neural networks are appropriate in data-rich environments and are normally used for extracting embedded knowledge in the form of rules, clustering, quantitative evaluation of these rules, classification and regression, self-organization, dimensionality reduction and feature evaluation. In more practical terms neural networks are non-linear statistical data modelling tools. They can be used to model complex relationships between outputs and inputs or to find patterns in data. Neural networks are programmed or trained to recognize, store and associatively retrieve patterns or database entries, to solve combinatorial optimization problems. The main contribution of Neural Networks towards Data Mining stems from rule extraction and evaluation, incremental learning from clustering and dimensionality reduction. In this proposed work used back propagation algorithm neural network.[9]

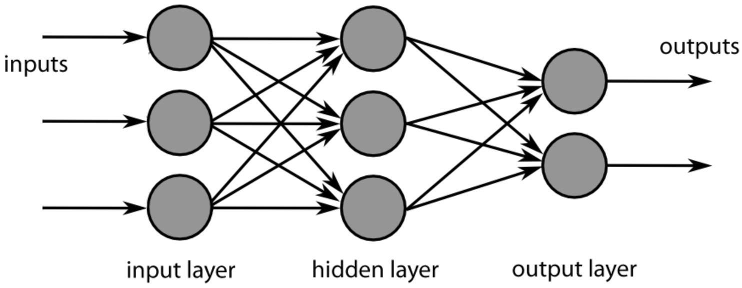

Fig. 2: Neural network layers

4.1 The Back Propagation Algorithm

Back propagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. The back propagation algorithm is used in layered feed forward ANNs. This means that the artificial neurons are organized in layers, and send their signals “forward", and then the errors are propagated backwards.

Algorithm 1: The Back Propagation Algorithm.

```
Initialize the weights in the network (often randomly)
repeat
   for each example e in the training set do
         O = neural-net-output(network, e);
         forward pass
         T = teacher output for e
         Calculate error (T - O) at the output units
         Compute delta-w_i for all weights from hidden layer to output layer;
         backward pass
         Compute delta-w_i for all weights from input layer to hidden layer;
         backward pass continued
         Update the weights in the network
   end
until all examples classified correctly or stopping criterion satisfied
return (network)
```

The back propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. The idea of the back propagation algorithm is to reduce this error, until the ANN learns the training data.

The summary of the technique: 1. Present a training sample to the neural network. 2. Compare the network's output to the desired output from that sample. Calculate the error in each output neuron. 3. For each neuron, calculate what the output should have been, and a scaling factor, how much lower or higher the output must be adjusted to match the desired output. This is the local error. 4. Adjust the weights of each neuron to lower the local error. 5. Repeat the steps above on the neurons at the previous level.

Vol.5 Issue 5, Pg.: 53-60
May 2017

5. ICHI-Based Feature Selection at Syntax Level

The basic idea of the proposed ICHI is to make a minority class far away from the majority one by adjusting weights of fault terms. To facilitate understanding, we firstly define some notations. T_m is the set of fault terms of minority fault classes, T_M the set of fault terms of majority fault classes and T_c, the intersection of T_m and T_M, the common feature set. Let symbol / denote the set difference, T_m=T_c and T_M=T_c are related with minority and majority classes only, respectively, thereby called them as exclusive fault term sets.

5.1 X^2 Statistics and Hellinger Distance

X^2 Statistics can be used to calculate the lack of independence between a term t and a category c_i and can be compared to the ICHI distribution with one degree of freedom to judge extremeness [11]. It is defined as:

$$X^2(t, c_i) = \frac{N[P(t,c_i)P(\bar{t},\bar{c_i}) - P(t,\bar{c_i})P(\bar{t},c_i)]^2}{P(t)P(\bar{t})P(c_i)P(\bar{c_i})} \qquad (1)$$

where N is the total number of documents. (t, c_i) represents the presence of term t and its membership in class c_i, (t, c_i) presence of t but not its membership in c_i, (t, c_i) absence of t but its membership in c_i, and (t, c_i) absence of t and its non-membership in c_i. P(.,.) means the chance of presence/ absence of term t and its membership/non membership in class c_i.

Hellinger distance is a measure of distributional divergence [3]. Given two discrete probability distributions P = $\{p_1, p_2, \ldots, p_n\}$ and Q = $\{q_1, pq_2, \ldots, q_n\}$, their Hellinger distance is defined as:

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\textstyle\sum_i^k (\sqrt{p_1} - \sqrt{q_1})} \qquad (2)$$

The Hellinger distance is a metric satisfying triangle inequality as per definition. $\sqrt{2}$ in the definition is used for ensuring the H(P,Q) ≤ 1 for all probability distributions.

5.2 ICHI Based Feature Selection at Syntax Level

The important steps of ICHI based feature selection are summarized in Algorithm 2. When a fault maintenance document D and a fault term dictionary **Ω** are provided, word set W (i.e., fault term set) is extracted by word segmentation.

Algorithm 2: ICHI Based feature selection at the syntax level.

Data: Maintenance dataset D, fault term dictionary **Ω**, Fault Class Set C
Result: Feature set f_a
begin
W ← Word Set by word segmentation in D according to **Ω**
M ← Word Document Matrix from W and C
for w_i ϵ W and c_j ϵ C **do**
  R(i,j) ← the correlation between fault term w_i and fault class c_j by Eq. (1)
end
$\bar{R}$ ← normalization of R by Eq. (3)
F(i) ← Fault feature (i.e., feature terms) set of fault class i(0≤i≤n) with $\bar{R}$ larger than given threshold
for c_i, c_j ϵ C **do**
  $\bar{F}$(i, j) ← Common fault feature set of c_i and c_j by intersect F(i) and F(j)
end
$\bar{F}$ ← Common fault feature set by union all $\bar{F}$
for c_i ϵ C do
  $\bar{F}$(i) ← Exclusive Feature Set for c_i by excluding $\bar{F}$ from F(i)
  $\bar{F}_w$(i) ← Weight of $\bar{F}$(i) inversing the probability of c_i (i.e.,1/P(c_i))
end
for w_k ϵ F **do**
  H(w_k) ← Hellinger distance of two fault class distribution on terms w_k by Eq. (2)
  $\bar{F}$(i, j) ← Common feature set of c_i and c_j by selecting the hight k features according to Hellinger distance H.

Vol.5 Issue 5, Pg.: 53-60
May 2017

18 **end**
19 $\bar{F}'$ ← Common Feature Set by union all $\bar{F}$(i, j)
20 F_a ← Selected Feature Set formed as $[(\bar{F},\ \overline{F_w}),\ \bar{F}']$
end

According to W and fault classes C, a word document matrix M can be generated (lines 1-2). Then compute correlations R between feature terms and fault classes by X^2 statistics (lines 3-4). In order to compare the correlation between different fault terms and different classes, we normalize them as follows (line 5):

$$\bar{R}(w_i, c_j) = \frac{R(i,j)^2}{\sum_{i=1:m} R(i,j) \times \sum_{j=1:n} R(i,j)} \qquad (3)$$

Where n denotes the number of fault terms contained in W, m denotes the number of fault classes in C. In Eq. (3), the correlation of feature term w_i and fault class c_j depends on the correlations between term w_i and all other fault classes besidesc_j. To the common fault term set F, we need to evaluate the distributive discrimination of each feature on fault classes by calculating its Hellinger distance with these fault classes using Eq. (2) (line 16). Thus, proposed work complete the feature selection of fault term features and get such feature space F_a as [(exclusive feature sets, weights), common feature set] (line 20).

6. PLDA-Based Feature Selection at Semantic Level

In this section, firstly give an introduction to Latent Dirichlet Allocation (LDA) then introduce the extraction of relationship based on prior knowledge. At final we present the proposed PLDA that integrates prior knowledge into LDA to realize the feature selection at the semantic level.

6.1 LDA

Given D documents expressed over W unique words and T topics, LDA outputs the document topic distribution and topic word distribution, both of which can be obtained with Gibbs Sampling[4],[12]. Its key step is the topic updating for each word in each document according to the below formula,

$$P(z_i = j | z\_i, w, \alpha, \beta) \propto \left(\frac{n_{-i,j}^{w_i} + \beta}{n_{-i,j}^{(\cdot)} + W\beta}\right)\left(\frac{n_{-i,j}^{d_i} + \alpha}{n_{-i,j}^{(d_i)} + T\alpha}\right) \qquad (4)$$

where z_i = j is the ith word in a document assigned to topic j, z_i all the topic assignments other than the ith word, i.e., the current one. w = $\{w_1, w_2, \dots, w_n\}$ is denotes a corpus, where each w_i belongs to some document. α and β are hyper parameters for the document topic and topic word Dirichlet distributions, respectively.

6.2 Extraction of Relationship Based knowledge

To better understanding the extraction of prior knowledge, proposed work give three types of relationship between fault terms and fault classes, such as Strong Relationship (SR), Weak Relationship (WR) and Complex Relationship (CR).

Algorithm 3: Relationship-based prior knowledge extraction.

Data: Maintenance dataset D, fault term dictionary Ω, Fault Class Set C, Topic Sets Z
Result: Correlation Γ (w_i, z_k) between term w_i ϵ Z
begin
$\bar{R}$ ← normalization of R by Alg. 1
Ξ ← k clusters (here, set k=8) by k-means analysis of $\bar{R}$ and rank the k cluster in descending order
Θ ← correlation degree of Ξ except for the highest two and lowest two clusters, in which Θ_i (i=1,2,3,4) is Corresponding to ith cluster center of the remaining k-4=4 clusters in ascending order
for w_i ϵ W and c_i ϵ C **do**
  if $\bar{R}(w_i, c_j)$ is the highest or lowest two ranks in Ξ **then**
    $\bar{R}(w_i, c_j)$ is assigned SR or WR between w_i and c_j and the values is set as a positive number (>1) and zero respectively.
  else

Vol.5 Issue 5, Pg.: 53-60
May 2017

7 $\bar{R}(w_i, c_j)$ is considered as CR and is set as a positive value if Θ_i is larger than the threshold t, otherwise, set as a negative value; and the value can be calculated by $\| \Theta_i\text{-t}\|/\text{t}$
8 **end**
9 **end**
10 Each fault classes $c_i \in$ C is pre assigned with two corresponding topics z_{2*1} , z_{2*i+1}, (1≤ i ≤|C|)
11 Γ(w_i, z_k) ← initialize correlation between term w_i and topic $z_k \in$ Z with zeros
12 **for** $w_i \in$ W and $z_k \in$ Z **do**
13 **if** $z_k \in c_j$ **then**
14 Γ(w_i, z_k) is assigned with the value of $\bar{R}(w_i, c_j)$
15 **end**
16 **end**
end

The main steps of prior knowledge extraction are summarized into Algorithm 3. Similar to Algorithm 1 the normalized correlations (R) are evaluated by Line 1. Then R is clustered into 8 clusters Ξ through the K-means clustering method (Line 2). Correlation degree (Θ) between fault terms and fault classes, such as SR, WR and CR, is then assigned to each pairwise term and fault class (Lines 4-8). In this work, each fault class is pre-assigned with two corresponding topics. Then the correlation (Γ) between terms and topics can be obtained (lines 13-15).

6.3 Incorporating Prior Knowledge into LDA

The main idea of incorporate prior knowledge into LDA is to improve the topic updating probabilities by using prior information. That means, in a topic updating process in (Eq. 4), then multiply an additional indicator function δ(w_i, z_j), which denotes a hard constraint of SR and WR from terms to topics. For CR, influence of fault term w_i and fault classes on topic word distribution should be all taken into account. Our basic idea is to determine the association between w_i and C_{zj} , where C_{zj} is the set of fault classes to which topic z_j attached. If they have the higher relevance than a pre given threshold, Γ(w_i, z_j) should be assigned a positive number. Otherwise, if w_i and C_{zj} are lower than the threshold, Γ(w_i, z_j) is set as a negative number. Therefore, (Eq. 4) is revised as follows:

$$P(z_i = j|z\_i, w, \alpha, \beta) \propto \left(\frac{(1+F_{wi,zj})n_{-i,j}^{w_i}+\beta}{\sum_{w\prime}^{W}(1+F_{w\prime,z_j})n_{-i,j}^{(w\prime)}+W\beta}\right)\left(\frac{n_{-i,j}^{(d_i)}+\alpha}{n_{-i,j}^{(d_i)}+T\alpha}\right) \qquad (5)$$

Where $F_{wi,j}$ corresponds to (w_i, z_j) in Algorithm 3 and reflects the correlation of fault term w_i with topic z_j. Then (Eq. 5) is used to change the sampling process for fault data set with CR relationship.

7. Serial Fault Feature Fusion and Classification

The fault feature extracted at the syntax level is fused with those at the semantic level. To facilitate understanding, they denote the processed fault feature from the syntax level as F_a = (a_1, $a_2, a_3, \dots, a_M$) and the one from semantic level $F_b = (b_1, b_2, b_3, \dots, b_N)$, where M and N are the dimension at syntax and semantic levels respectively. Here we adopt a serial fusion method [9], to form a combined featureF_Y.

Classification is to find the main theme of document by adding Meta and analysing document. The count of words and from that count decides the topic of the document which was done by the classification technique. Proposed work used a Multi SVM for classification.

8. Conclusion

Text mining of repair verbatims for fault detection of railway systems poses a big challenge due to unstructured verbatims, high-dimension data, and imbalanced fault classes. In this paper, to improve the fault detection performance, proposed work proposes a two level feature extraction based text mining method with neural network. Firstly trained the data by using back propagation algorithm. Then adjust the exclusive feature weights of various fault classes based on X^2 statistics and their distributions. Next reselect the common features according to both relevance and Hellinger distance. This can be categorized as feature selection at the syntax level. Then, proposed work extracts semantic features by using a prior LDA model to make up for the limitation of fault terms derived from the syntax level. Next, this work combines fault term sets derived from the syntax level and semantic level by serial fusion. Finally, proposed works classify them by using SVM to get desired result.

Efficient feature fusion methods play an important role in feature extraction. Therefore, such powerful methods as parallel feature fusion should be further researched to improve the proposed methods performance. Other merging learning methods should also be explored for better imbalanced classification.

REFERENCES

[1] L. Huang and Y. L. Murphy, "Text mining with application to engineering diagnostics," in Proc. 19th Int. Conf. IEA/AIE, Annecy, France, 2006, pp. 1309-1317.

[2] D. G. Rajpathak, "An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain," Comput. Ind., vol. 64, no. 5, pp. 565-580, Jun. 2013.

[3] D. A. Cieslak and N. V. Chawla, "Learning decision trees for unbalanced data," in Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases-Part I. Berlin, Germany: Springer-Verlag, 2008, pp. 241-256.

[4] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993-1022, Jan. 2003.

[5] W. Wang, H. Xu, and X. Huang, "Implicit feature detection via a constrained topic model and SVM," in Proc. Conf. Empirical Methods Natural Lang. Process., Seattle, WA, USA, 2013, pp. 903-907.

[6] Y. Tang, Y. Zhang, and N. V. Chawla, "SVMs modelling for highly imbalanced classification," IEEE Trans. Syst., Man Cybern., B, vol. 39, no. 1, pp. 281-288, Feb. 2009.

[7] C. Shang, M. Li, S. Feng, Q. Jiang, and J. Pan, "Feature selection via maximizing global information gain for text classification,- Knowl.-Based Syst., vol. 54, pp. 298-309, Dec. 2013.

[8] Z. Zheng, X.Wu, and R. Srihari, "Feature selection for text categorization on imbalanced data," ACM SIGKDD Explorations Newslett.-Spec. Issue Learn. Imbalanced Datasets, vol. 6, no. 1, pp. 80-89, Jun. 2004.

[9] Yashpal Singh, Alok Singh Chauhan, "Neural networks in data mining," JATIT Syst., 2009.

[10] F. Wang, T. Xu, T. Tang, M. Zhou, H. Wang, "Bilevel feature extraction based text mining for fault diagnosis of railway systems," IEEE Trans. Syst., Jan. 2016.

[11] Y. Yang and J. Pedersen, "A comparative study on feature selection in text categorization," in Proc. 14th Int. Conf. Mach. Learn., Nashville, TN, USA, 1997, pp. 412-420.

[12] T. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. United States Amer., vol. 101, no. 1, pp. 5228-5235, 2004.

AUTHOR PROFILE

Akshay Gulhane is a Master of Engineering student at the Department of Computer Engineering, Nutan Institute of Engineering and Technology, Talegaon Dabhade, in Savitribai Phule Pune University, Pune. He received the Bachelor of Engineering degree in Computer Science and Engineering from Sant Gadge Baba Amravati University in 2014.