Optional SVM for Fault Diagnosis of Blast Furnace with Imbalanced Data

Limei LIU, Anna WANG, Mo SHA, Xiyan SUN and Yunlu LI
College of Information Science and Engineering, Northeastern University, Liaoning, 110819 P.R. China. E-mail: liulimei_2006@163.com

(Received on March 28, 2011; accepted on May 9, 2011)

Fault diagnosis for blast furnace is actually a multi-class classification problem because the blast furnace may appear usually many kinds of abnormal states. Moreover, those abnormal states should be monitored and diagnosed timely and what can help workers take effective measures. Support vector machine (SVM) is state-of-the-art for many classification problems currently. But many classification tasks involve imbalanced training examples in practice. Imbalanced dataset learning is an important practical issue in machine learning, especially in support vector machine (SVM). Fault diagnosis for blast furnace is such an imbalanced data problem. A novel algorithm named optional support vector machine is proposed to solve this imbalanced data classification by pruning training sets and adding the unlabeled data and applying edited nearest neighbor (ENN) rules. Firstly, training sets of majority class are pruned in order to reduce the training time. Secondly, the algorithm selects some useful unlabelled training data and adds them to the training sets. Those samples are used to replenish the lack of training samples so that the training sets are representative. However, they may contain some noisy examples. Finally, the edited nearest neighbor rule is removed the noisy examples. The algorithm adds the unlabelled (testing) samples to balance the number of samples between the minority class and the majority one. The real-time producing data of blast furnace are used to running experiment. In order to more accurately diagnose which kinds fault happened, a binary tree multi-class classification method is adopted based on blast furnace characteristics. Simulation results show that the proposed algorithm is feasible and effective.

KEY WORDS: support vector machine (SVM); pruning training set; active learning; imbalanced data classification; edited nearest neighbor (ENN).

1. Introduction

Fault diagnosis for blast furnace is actually a multi-class classification problem because the blast furnace may appear usually many kinds of abnormal states, such as heating, over-development of brim gas flow, hanging, cooling, chimney, lack of brim gas flow, low stock line, slip and etc other common disorders and abnormalities. Moreover, those abnormal states should be monitored and diagnosed timely so that workers may take effective measures and serious accidents will be avoided and many economic losses will be to a minimum. Fault diagnosis for blast furnace is an imbalanced data classification problem that there are a lot of normal samples and a few abnormal samples in reality. Moreover, in many other real-world classification problems, a classifier has to deal with imbalanced training examples. Learning for imbalanced class problems is encountered in a large number of practical applications of machine learning. Various classification methods are used in the study of imbalanced data classification problems. They include decision tree methods, neural networks, linear discriminant analysis, C5.0, multi-layer perceptron (MLP) and support vector machine and so on. But some classifiers require more examples to learn, such as linear discriminant analysis, decision trees, and neural networks. And neural networks are not applicable for the problem with relative high number of features and fewer examples. Among the other classifiers: multi-layer perceptron (MLP) and support vector machine (SVM), it was found that compared to C5.0, MLP was less sensitive and SVM was not sensitive at all. Japkowicz and Stephen (2002) compared the performance of three methods for dealing with class imbalances: over-sampling, under-sampling, and cost-modifying (the weighting methods), with C5.0 as the classifier. The cost-modifying method was found to be most effective among all. Sun et al. (2009) make a comparative study that is the effectiveness of the imbalanced classification strategies. The support vector machine (SVM) is currently one of the state-of-the-art classification techniques and it exhibits good predictive performance due to its ability to model nonlinearities. The class imbalance problem could hinder the performance of standard machine learning methods so that standard SVM is not ideal to it. Li et al. (2008) introduced SVM with uneven margins (SVMUM) that it modified the standard SVM.

It is very important for data imbalance to design new and more adaptive classifiers than before. And these new classifiers should not only study the data distribution but also...
consider different the number of the samples. If data distribution of training samples is different to one of testing samples, there is a lower performance. Therefore, a perfect training process is important to deal with it.

The rest of the paper is structured as follows. Section 2 introduces support vector machines. Section 3 focuses on active learning and imbalanced data classification. Section 4 presents an algorithm of optional SVM for imbalanced data classification. Section 5 describes ironmaking process of blast furnace and some faults of blast furnace and discusses simulation experiments to evaluate the algorithms on the datasets for blast furnace faults, with a particular emphasis on measuring the usefulness of the proposed algorithms. A fixed binary tree model based on one-against-all method for the blast furnace faults will approach the multi-class classification problem. In the end, Section 6 summarizes our findings and presents some discussions.

2. Support Vector Machine

The SVM is suitable for processing a binary classification problem. Given a training set of \( \{ (x_i, y_i) \mid i = 1, \ldots, l \} \) with input data \( x_i \in \mathbb{R}^d \) and corresponding binary class labels \( y_i \in \{-1,1\} \) and testing data \( x_j, y_j \in \mathbb{R}^d \). The SVM is tried to find the hyperplane \( \langle w, b \rangle \) and the labels of testing data \( y_j \). The nonlinear function \( \phi(\cdot) \) maps the input space to a high (possibly infinite)-dimensional feature space. In this feature space, a hyperplane may be constructed that is \( \langle w, b \rangle \) and its margin (or the distance) from the unlabelled examples and having the margin (or the distance) from the unlabelled examples and testing data.

The Lagrange multipliers satisfy . For a new point \( x \), the value \( f(x) \) is determined by evaluating which side of the hyperplane it falls on in feature space. The goal is first to find the set of binary variables by using the combinatorial optimal approach. The variable \( \alpha \) is the solution of the following function.

\[
\begin{align*}
  \max W(\alpha) &= \frac{1}{2} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\
  s.t. & \sum_{i=1}^{l} \alpha_i y_i = 0, \\
  & 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, l
\end{align*}
\]

Where kernel function satisfies equation

\[
k(x_i, x_j) = \phi(x_i)^T \phi(x_j)
\]

3. Active Learning and Imbalanced Data Classification

3.1. Active Learning with SVM for Unlabeled Data

SVM active learning is an SVM-specific algorithm, which uses the margin (or the distance) from the unlabelled example to the classification hyperplane as a measure of the example’s importance for learning. Active learning is a framework that attempts to reduce the cost of annotating training material for statistical learning methods. Active learning entails the control of the learning algorithm over the input data on which it learns, and more specifically, focusing on the problem areas. Due to the complexity of the annotation task, there is a challenge of obtaining sufficient training data. In addition, training samples are not representative. A good idea is to use active learning which is usually implemented as a module in the learning system. Active learning can select an unlabelled example based on its properties as the current model. Then the selected example is labeled and put in the training set.

3.2. Evaluation Criteria of Imbalanced Data Classification

Literature has a detailed introduction of some evaluation criteria, such as under the ROC curve, geometric mean of accuracies, F1-measure, class weighted accuracy, sensitivity, specificity, precision, and overall accuracy, which are used to evaluate the classifier performance on imbalanced data. In this study, precision and F1-measure will be described as following.

In brief, precision and recall are calculated by

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

and

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

Where \( TP \) is the number of positive sample true classified and \( FP \) is the number of positive sample false classified and \( FN \) is the number of positive sample false classified. In order to facilitate those instructions, a confusion matrix is shown in Table 1.

F-measure criterion takes into account both precision and recall that is made by the following formula (5).

\[
F - \text{measure} = \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 (\text{Recall} + \text{Precision})}
\]

When \( \beta = 1 \) is satisfied that it means precision and recall being equally important. Therefore, the F-measure with \( \beta = 1 \) is also known in other literatures. It is given by Eq. (6).

\[
F1 - \text{measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{(\text{Recall} + \text{Precision})}
\]

The above evaluation criteria are used to evaluate the effectiveness of a data proposed method.

| Table 1. Confusion matrix of two-class classification problem. |
|---------------------------------------------------------------|
| Model predicted | Positive | Negative |
|-----------------|----------|----------|
| Actual class    |          |          |
| Positive        | TP       | FN       |
| Negative        | FP       | TN       |

© 2011 ISIJ
3.3. Training Set and Generalization Performance of SVM Classifier

If a training set is representative of the whole data set, in general, a SVM classifier with a larger margin on the training set would have a better generalization performance. Figure 1 shows an ideal SVM classifier about a binary classification problem with a two-dimensional data. However, if the training set is unrepresentative, a maximal margin classifier learnt from an unrepresentative training set may have poor generalization performance, as can be seen in Fig. 2. This figure also describes a binary-class classification on imbalanced data sets that it is different between ideal classifier and actual classifier. The difference is mainly caused by which few positive examples can not correctly reflect the class distribution of change in the imbalanced training data.

In fact, unfortunately, many imbalanced classification problems have only a small number of positive training examples what leads to an SVM classifier with poor generalization capability. And redundancy negative training examples waste much more training time. Therefore, actual classifier must be modified further by the corresponding adjustment.

4. Algorithm of Optional SVM for Imbalanced Data Classification

4.1. Edited Nearest Neighbor Rule

Edited nearest neighbor (ENN) was originally proposed by Wilson (1972). It works by removing noisy examples from the original set. An example is deleted if it is incorrectly classified by its k-nearest neighbors (k = 3, in general). A new method may apply ENN to remove each majority/minority class example from the data set what does not have at least two of its three nearest neighbors from the same class.

4.2. Pruning Training Sets

Imbalanced data conclude superabundant majority samples that are referred to as negative class. Some of majority samples are farther than the others to hyperplane. They are less influence of classification and may be gradually pruned. The training observations corresponding to nonzero Lagrange multipliers \( \alpha \) are called support vectors and they are located close to the decision boundary. The distribution of the support vectors will follow the data distribution. More support vectors will be found in dense input areas and less in more sparse ones. Therefore, selecting some ratio majority samples firstly run experiment and next pruned the others.

4.3. Algorithm of Optional SVM

Herein a new method is described for classifying imbalanced data. According to the usual practice, minority class is positive and majority class is negative one. They \( n_+ \) and \( n_- \) denote the number of minority class and majority class respectively and are satisfied to the inequality \( n_- < n_+ \). The two variables \( n_+ \) and \( n_- \) do not need to calculate or optimize because they are certain number with respect to a provided training set.

The proposed algorithm proceeds as follows:

1. Randomly choose \( n_+ \) majority samples and minority ones respectively as the initial training set \( S_0 \). That is the same number of samples between minority class and majority class. Train the SVM classifies on \( S_0 \) with those training samples.

2. Calculate the distances mean of the minority class and the majority class, and denote \( f_1(x) \) and \( f_2(x) \).

3. Select some ratio \( R_1 \) negative training samples and run an experiment and prune the negative training samples that their distances to \( w, b > 0 \) are \( f_1(x) > f_2(x) \).

4. Select some ratio \( R_2 \) of unlabelled (testing) documents and apply the SVM classifiers to select the positive class samples with their distances to \( w, b > 0 \) that they satisfy to the inequality \( 0 < f_1(x) \leq f_2(x) \). Label them and add them to the training set. Meanwhile, selecting the negative class samples with their distances to \( w, b > 0 \) that they meet to the inequality \( 0 < f_1(x) \leq f_2(x) \). Label them and add them to the training set, too.

5. Apply ENN to remove each noisy example from the data set.

6. Retrain the SVM classifiers with the changed training set.

7. Repeat steps (2) and steps (3) and steps (4) for a predefined number of loops or until the system obtain a predefined performance level.

This algorithm is optional SVM (OSVM) and it is suit-
able which samples of minority class are not representative or samples of majority class are not representative or much more special.

In view of the different data characteristics, OSVM algorithm may run each section separately, as shown in Fig. 3.

5. Simulation Experiments

5.1. Introduction of Ironmaking Process and Some Faults of Blast Furnace

The blast furnace (BF) is an important process unit in the metal industry. Blast furnace melting is reductive into pig iron ore that is continuous production process. In general, a typically furnace is 20–30 meters high and has a diameter of 10 meters. Bigger blast furnace capacity is about 2000–3000 m³, blast furnace is top-down divided into five parts that are furnace throat, furnace stack, furnace waist, furnace abdomen, furnace hearth. Blast furnace melting for raw material consists mainly of iron ore, fuel (coke) and flux (limestone) of three parts. Blast furnace production loaded raw material from top and oxygen are injected through tuyeres in the lower part of the furnace. Some blast furnaces use the other auxiliary fuel, such as pulverized coal (PC), heavy oil, natural gas etc. At high temperatures coke carbon with the oxygen in air blast generated carbon monoxide, which rise in the oven to remove the oxygen in the process of iron ore, and get iron. The molten pig iron and slag collect on the bottom or hearth of the furnace in turn and are tapped more or less regularly through the taphole via runners to torpedo cars or ladles, which are sent downstream for steel processing. Blast furnace gas is expelled from the top of blast furnace and it is the fuel of hot air stove, heating furnace, coke oven after dedusting.

Some parameters value can reflect whether blast furnace state is normal. If some parameters change greatly, abnormal phenomenon of blast furnace would appear. At this time, blast furnace needs an accurate diagnosis and workers need some guidance advice. Therefore, some special parameters may be selected as the criterions of fault diagnosis to detect whether the blast furnace system runs in a normal state. Blast furnace abnormality include heating, over-development of brim gas flow, hanging, cooling, chimney, lack of brim gas flow, low stock line, slip and etc other common disorders and abnormalities. Certainly, the fault of low stock line may be recognized by watching the location of the gauge rod.

5.2. Description of Datasets

According to the complexity of the BF process, 14 measured variables are selected from all measured variables. They are blast volume (m³/min), blast pressure (kPa), top pressure (kPa), differential pressure (kPa), permeability, the top temperature (includes four-point temperature), crossing temperature (including center and edge) (°C), feeding rate (batch/hour), [Si], physical heat (°C) and position of material (m). Blast volume and top pressure are set basic parameters. In the real production process of blast furnace, the thermocouple temperature sensors and pressure sensors can obtain the temperature and pressure, and calculate the number of material per hour on the approval that was expected to material speed. The ratio of air volume and differential pressure is permeability.

The higher temperature the blast furnace has, the more silicon content will be. If silicon content is above standard, heating will happen. If silicon content is below normal, cooling will happen. Right now, if permeability of the blast furnace is bad, hanging will be appeared. At this time, the blast furnace must be reduction pressure. However, if pressure drop suddenly, it would result in blast furnace slip.

In experiment, there are 2600 labeled samples and 1400 unlabelled samples that originate from the real-time production data of blast furnace. Where normal samples is 2000 and fault samples is 600 of labeled samples set.

Because of the existence of a lot of normal data, experimental result becomes normal state of high precision. So selected estimate criterion is F1-measure to avoid this problem.

5.3. Parameters Analysis of OSVM

As discussed in section 4, four parameters of OSVM algorithm impact the performance of the proposed new algorithm. When putting this algorithm in practice one need to consider several important questions how the parameters values are chosen and what is the optimal parameters value. Our experiments evaluate different values of the four parameters. They are R1, R2, C and σ. The first parameter R1 is the ratio of pruned samples that it impacts the training time and performance. Using higher ratio of pruned samples would lead to a lower accurate of learning. Of course, if the ratio of pruned samples is smaller, the training time would become longer. The second one is the added ratio of unlabelled samples, namely R2. The value of parameter R2 is
too small or too large that performance of OSVM is not all satisfactory. The variables $C$, $\sigma$ are respectively the relative parameters of SVM. SVM will have a good performance when variables $C$, $\sigma$ they must be optimal. Generally, $R_1$ and $R_2$ are all random number in intervals $[0, 1]$. Here two variables $R_1$ and $R_2$ may be optimized by some usual method, such as genetic algorithm, particle swarm optimization, etc. This study applies recursive methods that 0.1 is starting point and each increase 0.1. The variable $C$ is limited in $[10^{-1}, 10^0, ..., 10^1]$ and $\sigma$ is limited in $[2^0, 2^1, ..., 2^5]$ respectively. Variation of parameters will be introduced in the following section.

### 5.4. Simulation Experiment and Results Analysis

There are three experiments what are experiment with pruned negative samples, experiment with added unlabelled samples and experiment with standard SVM and optional SVM. All the experiments were carried out in turn in the following. All results reported below are the average from five runs.

Table 2 shows the class code of the blast furnace state and the samples size of training samples and testing samples in the experiments. Normal sample sizes are more than the samples sizes of abnormal state. In order to experimental need, these data are normalized before all the experiment.

Negative class of training samples is randomly selected from the original generated sets according to a pruned ratio. As can be seen from Table 3, training time and testing time has all changed. The smaller pruned ratio $R_1$ is, the shorter total time is. But accurate of testing must not be ideal when whole time is the smallest. Therefore, total time and performance of algorithm must be considered at the same time. With its tiny variation, whole time has not changed much. A part of the meaningful results are displayed in Table 3.

Table 4 compares different added ratio of unlabelled samples that produce different testing results. When the added ratio of unlabelled samples is 0.8, F1-measure is much better. Considering the characteristics of the fault diagnosis of industrial processes, we choose to use a multi-class classification strategy that is based on two-class classification. One-against-all (OAA) scheme is adopted for this study. Because four kinds fault exist certain relevance, a fixed binary tree model can be established, as shown in Fig. 4. Therefore, in order to make a fair comparison between standard SVM and optional SVM algorithm, the same binary tree model is selected.

Table 5 reports the fault diagnosis results between standard SVM and optional SVM. The proposed optional SVM is superior to the standard SVM in F1-measure. The OSVM algorithm has flexibility and it may be suitable for different classification problems.

### 6. Conclusions

A novel algorithm is proposed to solve some imbalanced data classification problems by enhancing the standard SVM. The training sets are relatively more perfect than original one what makes it overall more appropriate fault diagnosis for blast furnace. Moreover, edited nearest neighbor rule is removed the noisy examples and the proposed algorithm is pruning the training set to spend less training time. Simulation results demonstrate clearly that optional SVM
outperforms the standard SVM model and it has more accurate recognition ability.

Based on the finding, future research realizes online fault diagnosis for blast furnace gradually.

Acknowledgment
The authors would like to thank the reviewers and associate editor for their insightful and very useful comments. All of their advices render the paper more solid. This research is supported by National Natural Science Foundation of China (Grant No.61050006).

REFERENCES
1) N. Japkowicz and S. Stephen: Intelli. Data Analysis, 6 (2002), 429.
2) T. W. Liao: Expert Syst. Appl., 35 (2008), 1041.
3) X. W. Chen, B. Gerlach and D. Casasent: Proc. of Int. Joint Conf. on Neural Networks, Montreal, Canada, 3 (2005), 1883.
4) A. X. Sun, E. P. Lim and Y. Liu: Decis. Support Syst., 48 (2009), 191.
5) Y. Y. Li, K. Bontcheva and H. Cunningham: Nat. Lang. Eng., 15 (2008), 241.
6) B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson: Neural Comput., 13 (2001), 1443.
7) G. Schohn and D. Cohn: Proc. of the 17th Int. Conf. on Machine Learning (ICML-00), Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, (2000), 839.
8) C. Campbell, N. Cristianini and A. Smola: Proc. of the 17th Int. Conf. on Machine Learning, ICML-00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, (2000).
9) D. Martens, B. Baesens and T. V. Gestel: IEEE Trans. Knowl. Data Eng., 21 (2009), 178.
10) C. Y. Yang, J. S. Yang and J. J. Wang: Neurocomputing, 73 (2009), 397.
11) D. L Wilson: IEEE Trans. Syst. Man Cybern., 2 (1972), 408.
12) H. Saxen and R. Ostermark: Eur. J. Oper. Res., 89 (1996), 34.
13) H. X. Tian and A. N. Wang: ISIJ Int., 50 (2010), 738.