International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

TO AVOID SELLING LAPSED ITEMS USING ENTROPY SAMPLING METHOD

SATHYA PRIEYA[1], G.AKALYA [2], I.APSARA [3], P.ASHWINI [4]

[1]Associate Professor, Dept. of Computer Science Engineering, Panimalar Engineering College, Tamilnadu, India
[2,34]Student, Dept. of Computer Science Engineering, Panimalar Engineering College, Tamilnadu, India

---

Abstract - Here in existing framework now days in shops they are keeping up old items and lapsed items if any one utilized those items in a few circumstances will be harmed. What's more, a portion of the shop people are changing that all dates or more the cover and making it like a unique items in the wake of terminating time they are changing that all spreads everything. Fundamentally these issues are occurring in healing facility drug additionally there specialists are giving distinctive sorts of medication for various sickness. At whatever point they will understand that therapeutic shop they will give for specific sickness diverse prescription. Here to overcome each one of those issue first client need to keep up every one of the items with id. Presently after login the businessperson account they need to transfer every one of the insights regarding items and they need to keep up make item and terminate date all they need to keep up in the wake of transferring all that these all data will goes to administrator group (carefulness group) now administrator group will deal with that all data and they can investigate and they will give all the data about the item lapsing date if the item will lapse they will send a notice to retailer before 15days of item will terminate. At that point businessperson will make offer for that specific id items then just it won't be squander capable that items. It will demonstrate the fabricate date and terminate date in the event that it was phony it won't demonstrate any outcome .If like that any client discover like that they can send a mail. To administrator they can make a move on that specific shop.

Key Words: Entropy, Sampling method, Imbalanced learning, GIR algorithm.

1. INTRODUCTION

Imbalanced learning has pulled in a lot of premiums in the examination network. The vast majority of the outstanding information mining and AI procedures are proposed to take care of grouping issues concerning sensibly adjusted class circulations. In any case, this supposition isn't in every case valid for a slanted class circulation issue existing in some true informational collections, in which a few classes (the greater parts) are over-spoken to by an enormous number of examples however some others (the minorities) are underrepresented by just a couple. The answers for the class imbalance issue utilizing customary learning methods predisposition the prevailing classes bringing about poor characterization execution. For amazingly multi-class imbalanced information set, imbalanced order execution might be given by conventional classifiers with an almost 100 percent precision for the larger parts and with near 0 percent precision for the minorities. Henceforth, the class-irregularity issue is considered as a noteworthy obstruction to the achievement of exact classifiers. So as to defeat this disadvantage, we present another metric, named entropy-based lopsidedness degree. It has been realized that data entropy can mirror the positive data substance of a given informational collection. Therefore we measure the data substance of each class and acquire the distinctions among them, i.e., EID. So as to limit EID to adjust the informational index in data content, an entropy-based half and half examining methodology is proposed, joining both entropy-based oversampling and entropy-based under-sampling techniques.

This paper takes such an approach in identifying the effect of duplicates on the performance of graph mining. Based on that observation, it proposes a number of heuristics to reduce the number of duplicates generated to significantly improve the performance of these algorithms. Further, we establish their correctness as well as their performance analysis for a number of graph characteristics. Based on these analysis, we show that it is possible to choose the best heuristic whether we have additional information about the graphs or not.

1.1. PROBLEM DEFINITION:

This paper takes such an approach in identifying the effect of duplicates on the performance of graph mining. Based on that observation, it proposes a number of heuristics to reduce the number of duplicates generated to significantly improve the performance of these algorithms. Further, we establish their correctness as well as their performance analysis for a number of graph characteristics. Based on these analysis, we show that it is possible to choose the best heuristic whether we have additional information about the graphs or not.

1.2. LITERATURE SURVEY

Classification is a popular technique used to predict group membership for data samples in datasets. A multi-class or multinomial classification is the problem of classifying instances into more than two classes. With the

emerging technology, the complexity of multi-class data has also increased thereby leading to class imbalance problem. With an imbalanced dataset, a machine learning algorithm cannot make an accurate prediction. Therefore, in this paper Hellinger distance based oversampling method has been proposed. It is useful in balancing the datasets so that minority class can be identified with high accuracy without affecting accuracy of majority class. New synthetic data is generated using this method to achieve balance ratio. Testing has been done on five benchmark datasets using two standard classifiers KNN and C4.5. The evaluation matrix on precision, recall and measure are drawn for two standard classification algorithms. It is observed that Hollinger distance reduces risk of overlapping and skewness of data. Obtained results show increase of 20% in classification accuracy compared to classification of imbalance multi-class dataset.

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM:

In existing framework, the examining techniques have demonstrated their in-sufficiency, for example, causing the issues of over-age and over-lapping by oversampling procedures or the unreasonable loss of huge data by under-examining systems.

2.2. DRAWBACKS IN EXISTING SYSTEM:

The disadvantage is that the execution time is more as wasted in producing candidates every time, it also needs more search space and computational cost is too high.

2.3. PROPOSED SYSTEM:

This paper introduces three sampling based approach, each significantly improving the overall mining cost by reducing the number of duplicates generated. These alternatives provide flexibility to choose the right technique based on graph properties. The entropy of a substance is genuine physical amount and is a positive capacity of the condition of the body like weight, temperature, volume of inward vitality. Entropy is a proportion of the turmoil or irregularity in the framework. This proposed system is combines all of three entropy methods using the GIR algorithm. The GIR stands for Generalized Imbalance Ratio. To verify the effectiveness of the proposed EOS, EUS, and EHS methods, we carry on extensive experiments on two 2D data sets and 12 real-world data sets.

IR is the traditional overall imbalanced measure and EID is our proposed imbalanced degree. For each data set, we perform 5-fold cross validation where the original data set is randomly divided into 5 folds. Each fold is used for testing once while the remaining 4 folds are trained. In each fold, all classification methods are trained 10 times and the results are averaged over 10 runs in order to eliminate the randomness.

2.3.1 ADVANTAGES IN PROPOSED SYSTEM:

The entropy of a substance is genuine physical amount and is a positive capacity of the condition of the body like weight, temperature, volume of inward vitality. Entropy is a proportion of the turmoil or irregularity in the framework.

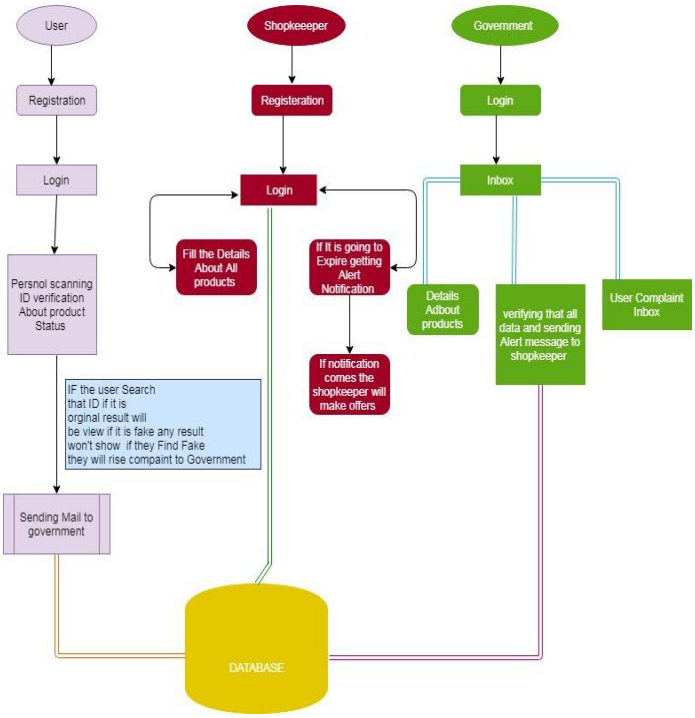

Fig.1

3. EXPERIMENTAL STUDY:

For a given multi-class imbalanced dataset, the first priority is to determine imbalance degree between the multi majorities and the multi-minorities. Most sampling approaches

use imbalance-ratio (IR) as the metric of class imbalance because of its simplicity. However, it is not an informative measure for multi-class problems. On one hand, it just describes class imbalance based on the largest class and the smallest class without considering other classes. On the other hand, the multi-class imbalance may still exist even with a balance in size. As stated in previous methods, the number of representative (effective) minority instances, rather than that of overall minority instances, decides the classification accuracy for minority classes. Therefore, IR is inappropriate to be considered as the measure of class imbalance. In this section, we propose a novel metric to measure the class imbalance, termed entropy based imbalance degree (EID), instead of imbalance-ratio. In this case, we first measure the importance of instances and classes [32, 33], and then present three entropy-based sampling approaches: entropy-based oversampling (EOS)

approach, entropy-based undersampling (EUS) approach, and entropy-based hybrid sampling (EHS) approach.

EID: In information theory, entropy is defined to measure the expected average amount of information contained within a data set. It is generally used as the metric of information content.

EOS: Oversampling technique is effective for imbalanced learning, which is devoted to balance skewed data distribution by generating new minority instances. As for mentioned above, a large number of synthetic sample methods have been proposed (e.g. SMOTE and AdaSyn). We First compute instance-wise statistics (x) for all instances in a data set and class-wise statistics for overall classes using (x). Then instance-wise difference statistics $(_r j j#r) for all instances using _ and class-wise difference statistics for all classes by $(_r j j#r). By definitions, it can be known that instances and classes with less information and lower $(_r j j#r) and are easier to learn, i.e., they are majority instances and classes.

EUS: Unlike oversampling technique, under sampling technique attempts to remove a subset of majority instances to form a balanced data set. Since a great deal of useful information may be lost, and the training for classifiers is hard on the subset of data with these under-representative information, it is necessary to implement detection and recognition of easy-to-learn instances, remove them, and retain difficult to- learn instances.

EHS: Hybrid sampling techniques combine the oversampling and undersampling techniques, adding minority instances and removing majority instances simultaneously in order to eliminate overfitting and prevent the loss of too much information effectively.

4. REQUIREMENT ANALYSIS AND SPECIFICATION

4. 1.HARDWARE ENVIRONMENT

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the system does and not how it should be implemented.

PROCESSOR : PENTIUM IV 2.6 GHz, Intel Core 2 Duo.
RAM : 4 GB DD RAM
MONITOR : 15" COLOR
HARD DISK : 40 GB

4.2. SOFTWARE ENVIRONMENT

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

FRONT END : J2EE (JSP, SERVLETS) JAVASCRIPT
BACK END : MY SQL 5.5
OPERATING SYSTEM : Windows 07
IDE : Eclipse

5. MODULE DESIGN SPECIFICATION

5.1. USER INTERFACE DESIGN

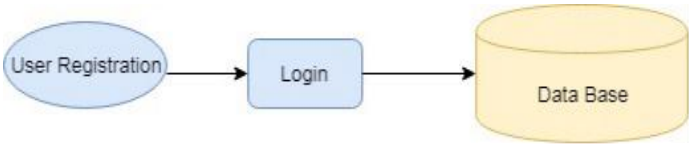

Fig.2

This is the first module of our project. The important role for the user is to move login window to user window. This module has created for the security purpose. In this login page we have to enter login user id and password. It will check username and password is match or not (valid user id and valid password). If we enter any invalid username or password we can't enter into login window to user window it will shows error message. So we are preventing from unauthorized user entering into the login window to user window. It will provide a good security for our project. So server contain user id and password server also check the authentication of the user. It well improves the security and preventing from unauthorized user enters into the network. In our project we are using JSP for creating design. Here we validate the login user and server authentication.

5.2. SHOPKEEPER UPLOADING DETAILS ABOUT PRODUCTS

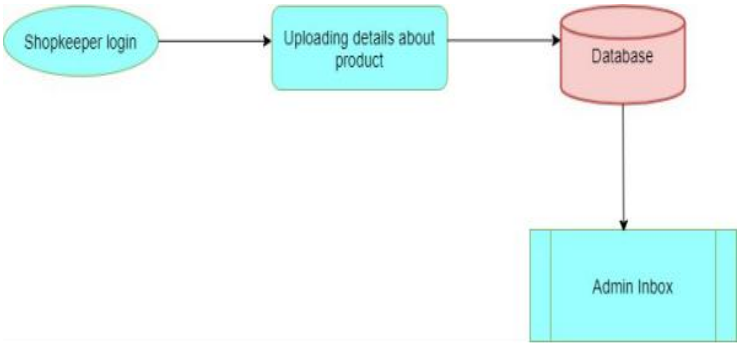

Fig.3

Here user have to check to all the products once .weather all products have the expire date and manufacture

date is available or not if not available don't use that product to get in to shop. After getting that products shopkeeper have to fill all the product details and it will stores in shopkeeper database and admin data base.

5.3.ADMIN VIEW AND MAINTAIN THE PRODUCT STATUS

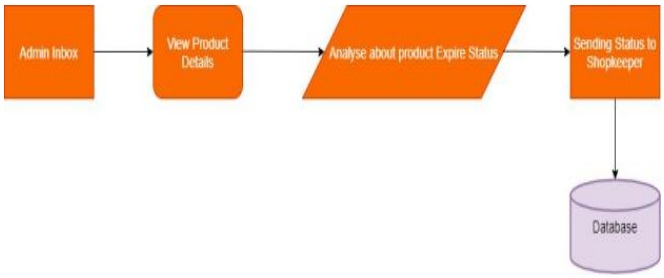

Fig.4

Here admin will calculate that details all those details about product expire date and inform to shopkeeper.

5.4. ADMIN INBOX

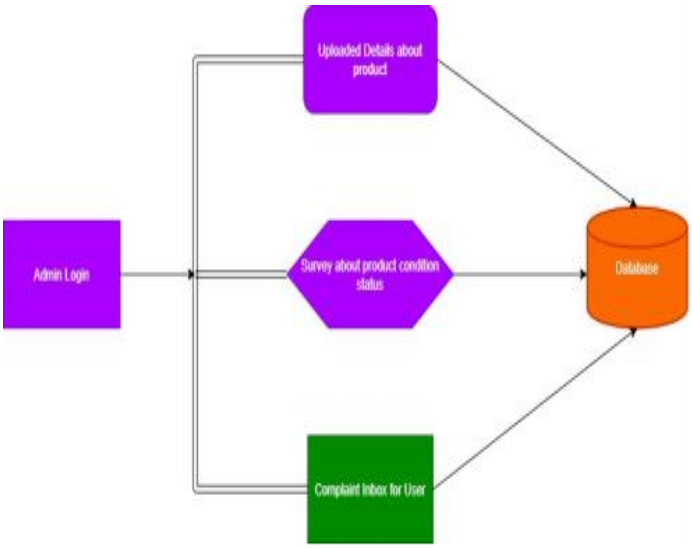

Fig.5

Here the shopkeeper whatever they will that products that all will stores in admin data base. By using that admin data they will calculate that all and provide one analysis and give to shopkeeper before 20 days when the product is going to expire.

5.5ADMIN COMPLAINT INBOX

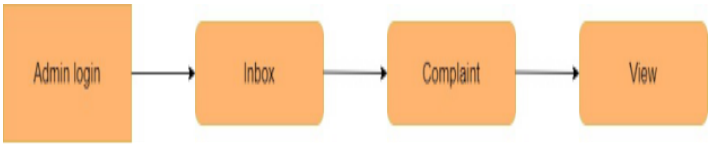

Fig.6

Here customer first they have to be register after login if they want to check that particular product weather that product is in good condition or not if he have any drought they can enter that id if that id have shown any result then that product is original if not show it will be fake . Even if it original if the product was expired they can rise a complaint and it will send to admin. That compliant will stores in admin inbox.

5.6. SHOPKEEPER PRODUCT STATUS INBOX

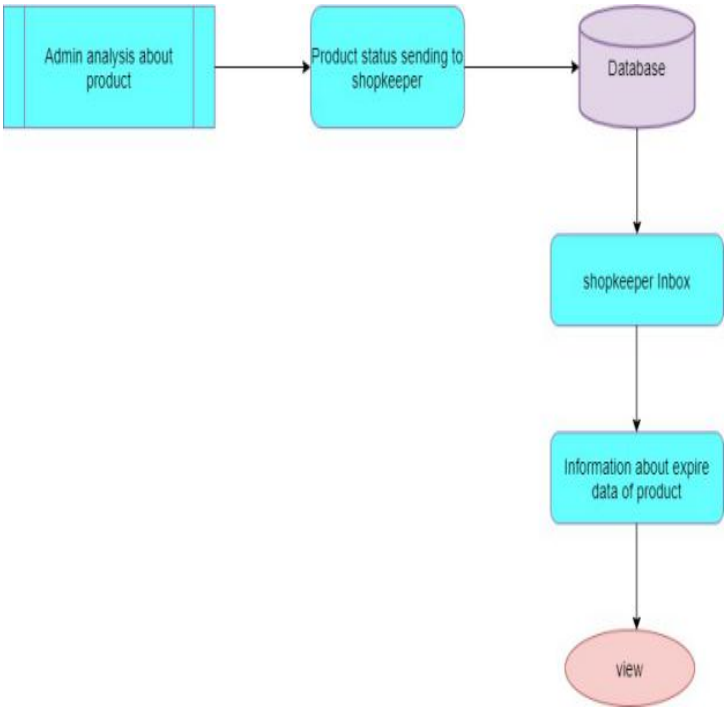

Fig7

If any user send that complaint to admin they will send a warning notification to shop owner .then shopkeeper can see that warning notification in inbox page and another use is shopkeeper upload all the product details that will stores in admin database .if the product is going to expire they will send that alert notification to shopkeeper inbox.

5.7 CUSTOMER VERIFICATION

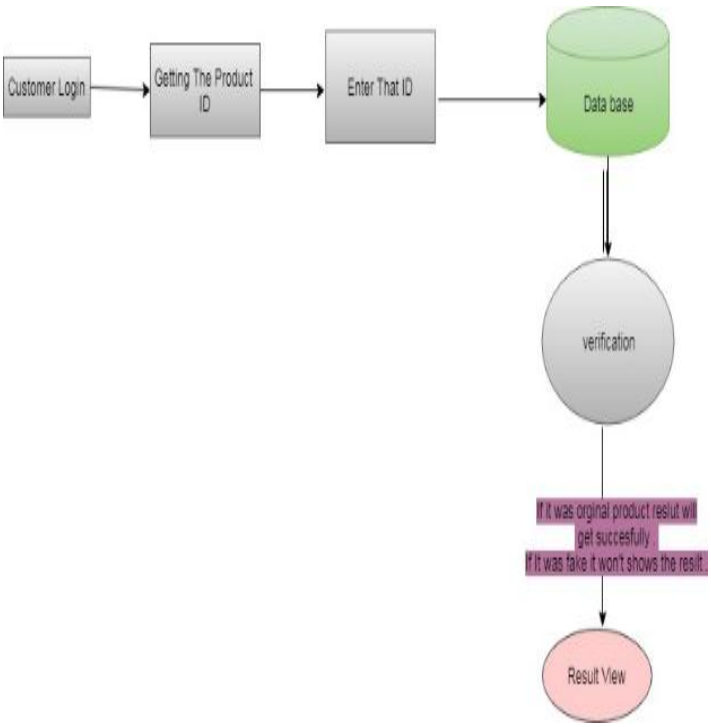

Fig.8

First user have to be register in that account .after login that account if user want to search about any product they can search by using of product Id.

5.8. SENDING COMPLIANT TO ADMIN

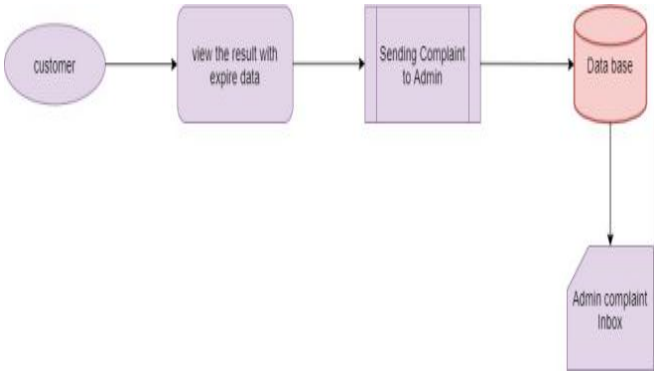

Fig.9

If user fined any wrong product or any expired product means they can directly write a mail and send to admin.

6. REPORT:

6.1. EXPERIMENTAL OUTPUT SCREENSHOTS

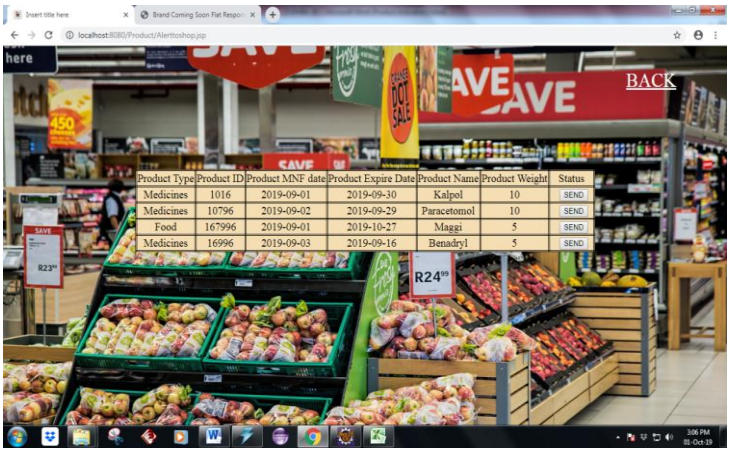

Fig.10

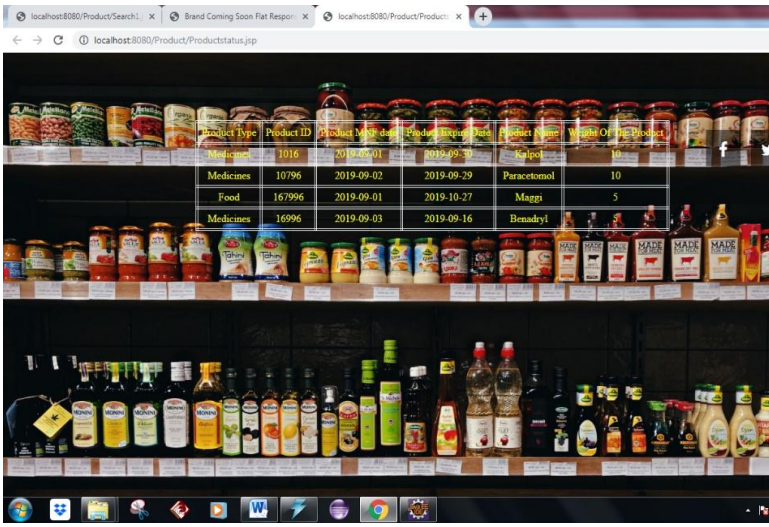

Fig.11

7. CONCLUSION

In this paper, we present three new entropy-based learning approaches, for multi-class unevenness learning issues. For a given imbalanced informational index, the proposed techniques utilize new entropy-based unevenness degrees to gauge the class irregularity as opposed to utilizing conventional unevenness proportion. EOS depends on the data substance of the biggest dominant part class. EOS oversamples different classes until their data substance accomplish the biggest one. EHS depends on the normal data substance of the considerable number of classes, and oversamples the minority classes just as under samples the greater part classes as indicated by EID. The viability of our proposed three techniques is exhibited by the unrivaled learning execution both on manufactured and real-world informational collections. Moreover, since entropy-based

half and half examining can all the more likely safeguard information structure than entropy-based oversampling and entropy-based under-sampling by creating less new minority tests just as expelling less greater part tests to adjust informational indexes, it has more predominance than entropy-based oversampling and entropy-based under-sampling.

8. FUTURE ENHANCEMENTS

In the future, we might want to investigate the hypothetical properties of our proposed lopsidedness measure and broaden it just as our three imbalanced learning techniques for other grouping issues, for example, picture arrangement what's more, move realizing.

9. REFERENCES:

[1] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on knowledge and data engineering, vol. 21, no. 9, pp. 1263–1284, 2009.
[2] Z. Wan, H. He, and B. Tang, "A generative model for sparse hyperparameter determination," IEEE Transactions on Big Data, vol. 4, no. 1, pp. 2–10, March 2018.
[3] C.-T. Lin, T.-Y. Hsieh, Y.-T. Liu, Y.-Y. Lin, C.-N. Fang, Y.-K. Wang, G. Yen, N. R. Pal, and C.-H. Chuang, "Minority oversampling in kernel adaptive subspaces for class imbalanced datasets," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 5, pp. 950– 962, 2018.
[4] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalanced data classification," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 1806–1819, 2017.
[5] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time ev charging scheduling based on deep reinforcement learning," IEEE Transactions on Smart Grid, pp. 1–1, 2018.
[6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority oversampling technique," Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.
[7] T. Zhu, Y. Lin, and Y. Liu, "Synthetic minority oversampling technique for multiclass imbalance problems," Pattern Recognition, vol. 72, pp. 327–340, 2017.
[8] K. E. Bennin, J. Keung, P. Phannachitta, A. Monden, and S. Mensah, "MAHAKIL: Diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction," IEEE Transactions on Software Engineering, 2017.
[9] Z. Wan and H. He, "Answernet: Learning to answer questions," IEEE Transactions on Big Data, pp. 1–1, 2018.
[10] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 2009, pp. 475–482.
[11] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in IEEE International Joint Conference on Neural Networks, 2008, pp. 1322–1328.
[12] S. Chen, H. He, and E. A. Garcia, "RAMOBoost: ranked minority oversampling in boosting," IEEE Transactions on Neural Networks, vol. 21, no. 10, pp. 1624–1642, 2010.
[13] S. Barua, M. M. Islam, X. Yao, and K. Murase, "MWMOTE–majority weighted minority oversampling technique for imbalanced data set learning," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 2, pp. 405–425, 2014.
[14] H. Han, W.-Y. Wang, and B.-H. Mao, "BorderlineSMOTE: a new over-sampling method in imbalanced data sets learning," in International Conference on Intelligent Computing, 2005, pp. 878–887.
[15] X. Yang, Q. Kuang, W. Zhang, and G. Zhang, "Amdo: an over-sampling technique for multi-class imbalanced problems," IEEE Transactions on Knowledge and Data Engineering, vol. PP, no. 99, pp. 1–1, 2017.
[16] L. Li, H. He, J. Li, and W. Li, "Edos: Entropy differencebased oversampling approach for imbalanced learning," in 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018, pp. 1–8.