Triple-O for SHL Recognition Challenge: An Ensemble Framework for Multi-class Imbalance and Training-testing Distribution Inconsistency by OvO Binarization with Confidence Weight of One-class Classification

Jinhua Su[1], Yuanyuan Zhang[2]

[1]Renmin University of China, Beijing, China.
[2]Beijing Baixingkefu Network Technology Co., Ltd., Beijing, China.

Background

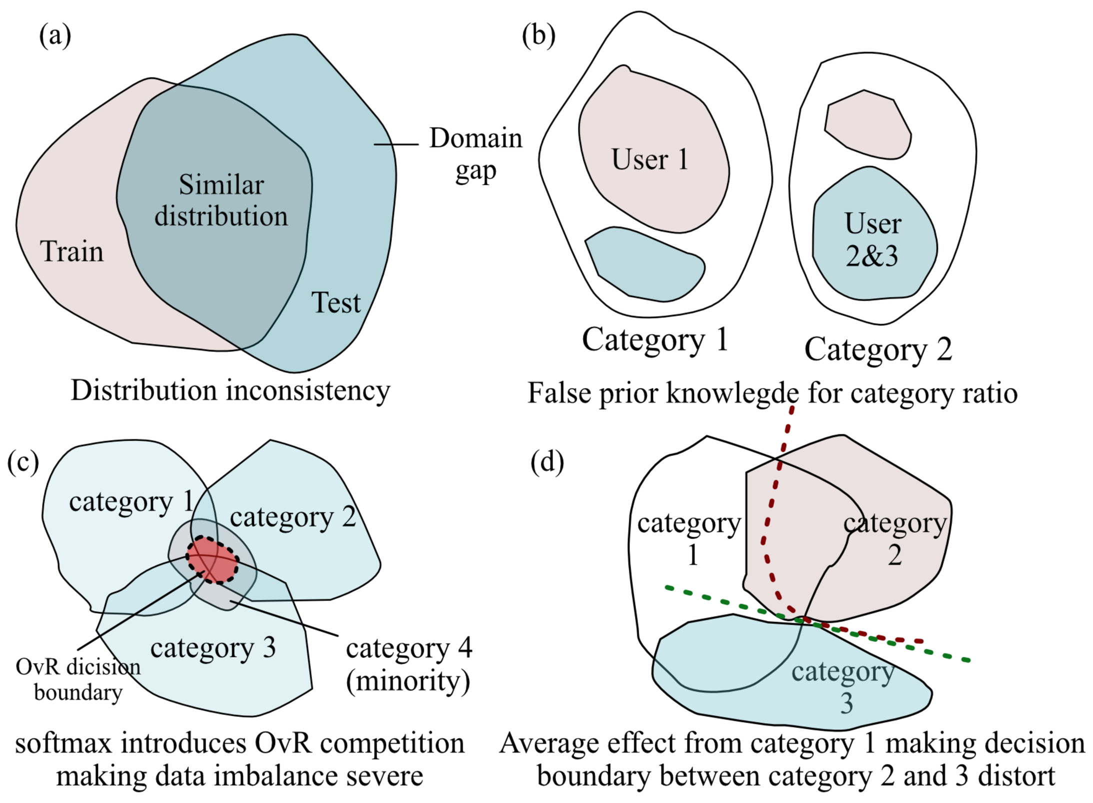
SHL Recognition Challenge provides phone sensor data for recognizing eight modes of locomotion and transportation(activities). Most motion sensor data is user-dependent, so the model's generalization ability requires various data at the user level. The SHL dataset is collected only by three users, and category imbalance and distribution gap between training data and testing data can be significant obstacles for this challenge.

- There is no apparent advantage of the deep and heavy model compared to some naive but robust machine learning methods.
- With re-sampling or other data imbalance methods, models still give the worst prediction for the minor category.
- Overfitting is common; most feeling-good models get a lower ranking. The explicitly underfit model won first place.
- Traning only on the training dataset makes many participants doubt life. Finetuning on the validation dataset works.

SHL challenge also notices that problem and provides user-independent sensor data in 2021[1].

Challenge

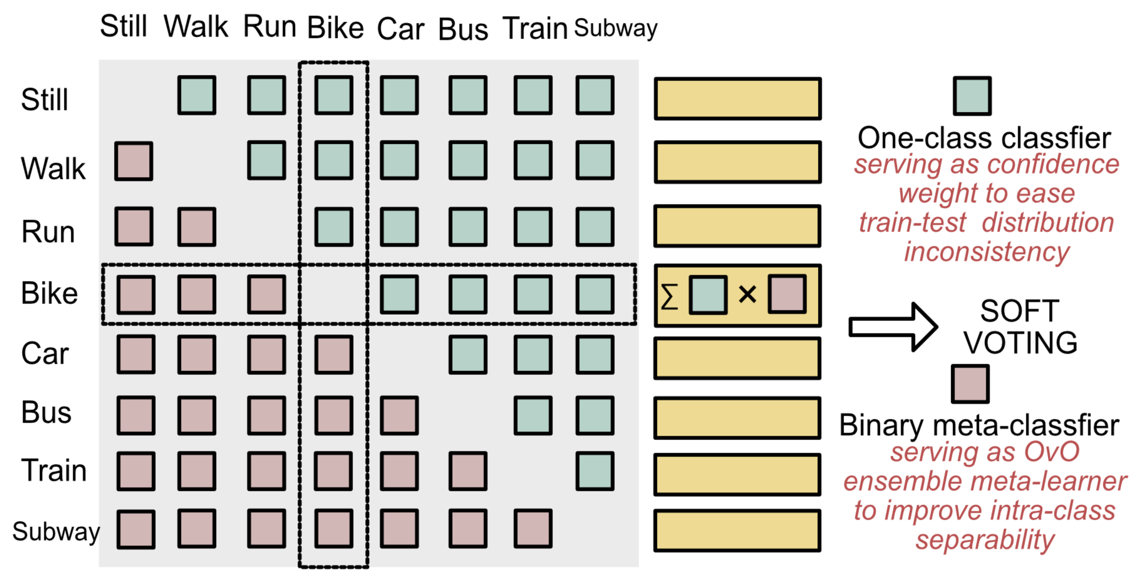
Category imbalance and distribution inconsistency are still the obstacles for the SHL-2021 Challenge.



The distribution inconsistency in user-level and category-level arise a domain adaptation problem for challenge participants. The concept map is shown in the above figure. Different user has different phone-using manners at the same activity. Furthermore, different locomotion and transportation activities depend on the real-world transportation conditions and user commuting habits, which indicates that the category ratio of one user may vary during weekdays and weekends. Multi-modal distribution in different levels makes recognition challenging. Facing the same problem, face recognition use centre loss and marge loss to realize domain adaptation. Lack of user category labels, it is hard to apply centre loss for intra-class compactness directly. However, inter-class separability can be integrated for the activities recognition task. This paper is aimed to alleviate the above problems drawing lessons from the research of face recognition

Solution

This paper puts forward a new ensemble framework called triple-O, using OvO binarization and one-class classification. The results show that the OvO binarization ensemble gets better results on the hard-to-distinguish class than re-sampling and re-weighting.

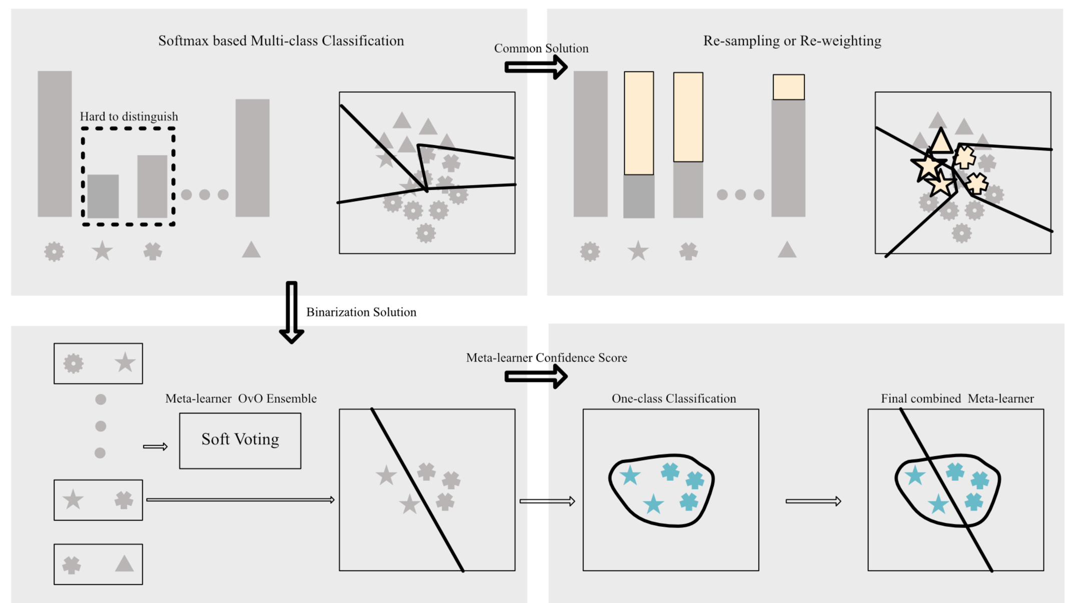


Splitting the multi-class classification task into several OvO(one v.s. one) binary-class classification gives a solution to low intra-class separability. Indeed, [2] has done the experiments on the relatively simple task for this proposal. The results show OvO methods did outperform the OvR methods, especially for the imbalanced situation.

To apply the most straightforward One-class Classification, we calculate the centre of binary data collections. Furthermore, select 90% quantile of the distance between data and centre as the radius r of the sphere. The above OvO voting procedure is to calculate the sum of probability of each binary meta-learner. We estimate 28 spheres(decision boundary for our one-class classifier) for the 28 meta-learner of the OvO ensemble and use the following confidence equation(shown in 1) to get a weighted probability for final prediction.

$$confidence = exp^{-\frac{(d-r)^2}{r^2}}, d > r \qquad (1)$$

Triple-O applies the OvO method to tackle the multi-class imbalance problem caused by the softmax method. Experiments show its efficiency compared to other data imbalance solutions, like re-sampling and re-weighting. What is more, another challenge of this dataset is the training and testing distribution inconsistency. To tackle this problem, we use a one-class classification method to detect the outlier from the training distribution. The predicting confidence can be used as the weight of meta-learner for soft voting.



This ensemble uses the cascade framework, which is creative and needs to be further researched. The OvO ensemble plus One-class classification weighting pipeline is named Triple-O. Triple-O can be plug-and-play and pave the way for exploring complex model structures. To our best knowledge, this method has not been further explored. The most relevant work we can find is [3].

Results

We split the validation dataset into 80% validation-training part and 20% validation-testing part and directly train the model on the validation-training and evaluated it on the validation-testing part. This paper focuses on solving the data imbalance problem for the multi-class classification problem. The results of applying different learning strategies on the meta-learner are shown in the following table. Directly training on the validation-training part still performed relatively worse in the prediction of Run and Bus. We are applying the re-sampling and re-weighting methods whose results show an improvement on the Run and Bus. The F1 for other activities also drops, which is reasonable that the previous give a fairly good prediction for the majority. The mean of all activities F1 score(equals to macro-F1) do improve in a small amount. Using the binarization ensemble will enhance the total parameters; the final results appear to be overfitting. By looking up the training evaluation, 96.2 of macro-F1 indicates the overfitting is not severe.

F1(%)	1	2	3	4	5	6	7	8
Direct	93.5	91.8	67.4	89.4	87.3	71.2	92.7	91.7
Re-Sample	90.5	87.4	74.3	89.4	87.9	75.6	92.3	91.5
Re-Weight	89.2	88.4	73.9	90.1	86.8	77.2	91.8	92.3
OvO	97.6	92.8	84.7	99.9	94.1	85.2	99.3	98.3
Tirple-O	87.6	88.8	81.7	92.3	90.2	88.2	85.3	87.3
OvO+Train	96.4	91.5	80.3	99.9	92.7	81.1	98.3	97.4

Evaluation on different learning strategies on validation dataset Notes: RF is short for random forest.

The results show that every type of activity becomes worse and worse. It is hard to claim the failure of One-class classification for limited data and lack of hyperparameter tuning. If the radius is too small, one-class classification will limit the model generalization ability. On the other hand, whether the sphere is a suitable boundary for data has not been explored.

	1	2	3	4	5	6	7	8
1	4114	88	0	6	0	0	56	0
2	29	4572	0	0	21	0	28	0
3	0	100	204	0	0	0	0	0
4	0	0	0	2082	0	0	0	0
5	92	30	0	0	3516	266	0	145
6	38	486	0	0	0	1784	40	0
7	0	0	0	0	0	0	3487	0
8	0	70	0	0	0	0	0	3999

Finally, add the training-training dataset for the OvO ensemble pretraining to make the model more robust. The confusion matrix of this OvO ensemble evaluated on the validation-testing is shown in the above table.

Conclusions

For SHL Recognition Challenge, the binarization strategy works on the setting of distribution inconsistency and class imbalance. The OvO ensemble method explicitly outperforms in contrast to the re-sampling and re-weight methods. To further tackle distribution inconsistency, we try the One-class classification to re-weight the contribution of each meta-learner to the final prediction.

References

[1] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Paula Lago, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. Summary of the sussex-huawei locomotion-transportation recognition challenge 2021. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021.

[2] Mikel Galar, Alberto Fernández, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8):1761–1776, 2011.

[3] Dong Chen, Xiao-Jun Wang, Changjun Zhou, and Bin Wang. The distance-based balancing ensemble method for data with a high imbalance ratio. *IEEE Access*, 7:68940–68956, 2019.

Contact information:
Jinhua Su
chasesu@ruc.edu.cn
Github: JinhuaSu