USING STACKED GENERALIZATION FOR ANOMALY DETECTION

Miguel Oliveira Sandim

Dissertation developed under the supervision of *Prof. Carlos Soares*
and co-supervision of *Prof. Bernhard Pfahringer*

September 19, 2017

---

1. Motivation

Data Mining has become an important field in the modern world, given the large number of possible applications in many different domains such as marketing, medical research, computer vision, social network analysis, intrusion detection and fraud detection [1].

Anomaly Detection is a very specific but significant topic in this field, given the high number of domains in which it can be applied [2]. In fact, the problem that motivates this field is a very common one and can be easily translated into this question: given a certain amount of data, is it possible to detect observations that deviate from the normal behavior of the data? This question can arise, e.g. in areas such as credit card fraud detection or machine condition monitoring.

The literature regarding Anomaly Detection techniques is very extensive and diverse, with a wide range of techniques that can have different outputs (either an *anomaly score* that indicates how much of a data instance in a dataset is an *anomaly*, or a label – *anomalous* or *normal*), as well as different assumptions (e.g. density based techniques have different underlying assumptions than clustering based techniques). This heterogeneity within Anomaly Detection techniques may cause different techniques to behave differently on the same dataset, which makes the task of choosing the right technique(s) for a specific domain very difficult and data-dependent.

2. Goals

This dissertation intends to address this issue, by using several Anomaly Detection techniques at the same time and then combining their outputs into a single one. This is the idea behind Ensemble Learning methods, which work by generating a group of models (which is designated by *ensemble*) and then combining their predictions into one. Ensemble Learning has proven to improve performance in machine learning applications such as classification, regression, time-series analysis and recommender systems [3]. More specifically this dissertation explores a Stacked Generalization method, which consists in using an extra model that *learns* the best way of combining the group of models.

Therefore this thesis intends to answer the following main research question:

- Can a Stacked Generalization method improve the performance of Anomaly Detection techniques, more specifically the performance of the best technique for a given dataset?

3. Description of the Dissertation

This research work was divided in two different research studies. The first study focused on the performance and diversity of the Anomaly Detection techniques selected and had the following goals:

- Study the performance and diversity of different types of Anomaly Detection techniques on several well-known datasets;
- Assess if this experimental setup contains *accurate* and *diverse* models.

The second one focused on the application of Stacked Generalization to the techniques selected and had the following goals:

- Determine if combining several Anomaly Detection techniques with a model improves the performance of each of the Anomaly Detection techniques used in this study;
- If so, determine how much the performance is improved.

This research work included several state of the art Anomaly Detection techniques: Classification and Regression Trees (CART), Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), Multilayer Perceptron (MLP), One-Class SVM, k-means, Density-based Spatial Clustering of Applications with Noise (DBSCAN) and Local Outlier Factor (LOF).

Several datasets were used in order to access the performance of the Anomaly Detection techniques and Ensemble Learning methods. These datasets were previously used in the Anomaly Detection literature and gathered by Campos et al. [4].

4. Conclusions

Then main conclusions of this dissertation can be briefly summarized as follows:

- Most of the Anomaly Detection techniques used in this study are *accurate* and *diverse* in

the datasets used, therefore having the necessary conditions for the Stacking method overperforming the best technique in each dataset;

- The application of the Stacking method guaranteed higher F1 values than the best Anomaly Detection technique on more than half of the datasets used;
- There is no clear indication whether including Anomaly Detection techniques from different learning modes guarantees higher F1 values. In the datasets where this was true, the best combination was including techniques from all the learning modes available;
- There is not a meta-classifier that clearly outperformed the others in terms of F1 on the datasets, so choosing the appropriate one seems to be very dependent on the dataset;
- Replacing the meta-classifier with the Majority Voting method improved the F1 value on even more datasets, with also a higher mean improvement on the F1. In this case, ensembles with tree-based Anomaly Detection techniques only (CART and Random Forest) were the ones with higher F1 values on most datasets.

References

[1] Charu C Aggarwal. *Data Mining: The Textbook*. Springer Publishing Company, Incorporated, 2015. ISBN: 978-3-319-14142-8.

[2] Rupali Kandhari et al. "Anomaly detection". In: *ACM Computing Surveys* 41.3 (2009), pp. 1–6. ISSN: 03600300. DOI: 10.1145/1541880.1541882.

[3] Charu C Aggarwal. *Outlier Analysis*. Springer Publishing Company, Incorporated, 2017. ISBN: 1461463955, 9781461463955.

[4] Guilherme O. Campos et al. "On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study". In: *Data Mining and Knowledge Discovery* 30.4 (July 2016), pp. 891–927. ISSN: 1573756X. DOI: 10.1007/s10618-015-0444-8. URL: http://link.springer.com/10.1007/s10618-015-0444-8.