Understanding Where Your Classifier Does (Not) Work

Wouter Duivesteijn(✉) and Julia Thaele

Technische Universität Dortmund, Dortmund, Germany
{wouter.duivesteijn,julia.thaele}@tu-dortmund.de

Abstract. FACT, the First G-APD Cherenkov Telescope, detects air showers induced by high-energetic cosmic particles. It is desirable to classify a shower as being induced by a gamma ray or a background particle. Generally, it is nontrivial to get any feedback on the real-life training task, but we can attempt to understand how our classifier works by investigating its performance on Monte Carlo simulated data. To this end, in this paper we present the SCaPE (Soft Classifier Performance Evaluation) model class for Exceptional Model Mining, which is a Local Pattern Mining framework devoted to highlighting unusual interplay between multiple targets. The SCaPE model class highlights subspaces of the search space where the classifier performs particularly well or poorly. These subspaces arrive in terms of conditions on attributes of the data, hence they come in a language a human understands, which should help us understand where our classifier does (not) work.

1 Introduction

The FACT telescope [1,2] is an Imaging Air Cherenkov Telescope, designed to detect light emitted by secondary particles, generated by high-energetic cosmic particles interacting with the atmosphere of the Earth. For astrophysical reasons, it is important to classify the light as resulting from the atmosphere being hit by a gamma ray or a proton; the latter occur much more frequently, but the former are the more interesting in gamma astronomy. Currently, one of the used classifiers is a random forest, whose performance needs our detailed attention.

The problem with training a classifier on real astrophysical data is that there is no clear feedback. Based on the observed light, we could deduce whether the inducing particle is a gamma ray or a proton. Then, we can look in the direction from which the particle originated, and strive to find an astrophysical source generating gamma rays. But even if we find such a source, there is no certain way of telling what kind of particle induced the original observation. Effectively, we are dealing with a feedbackless learning task, and it is typically hard to finetune a classifier without feedback.

This Nectar Track submission presents the paper [4]. A significantly longer version of that paper appeared as a technical report [5].

A. Bifet et al. (Eds.): ECML PKDD 2015, Part III, LNAI 9286, pp. 250–253, 2015.
DOI: 10.1007/978-3-319-23461-8_24

To study our learning performance, we turn to Monte Carlo data. We simulate particle interactions with the atmosphere, as well as reflections of the resulting Cherenkov light with telescope mirrors on the one hand and the FACT camera electronics on the other hand. This gives us a dataset of camera images that is equivalent in form to a dataset we would get from real astrophysical observations, except that we also know the true label of our classification task. By training our random forest on this dataset, we obtain the soft classifier probabilities for each record. Through studying the interaction between the binary ground truth that we already knew and the soft classifier probabilities we learned from the data, we can understand where our classifier performs exceptionally.

We study this interaction with Exceptional Model Mining (EMM) [3,7]: a Local Pattern Mining framework, seeking coherent subsets of the dataset where multiple targets interact in an unusual way. We present the SCaPE (Soft Classifier Performance Evaluation) model class for EMM, seeking subgroups for which a soft classifier represents a ground truth exceptionally well or poorly. This provides us with insight where our classifier does (not) work.

2 Related Work

Previous work exists on discovering subgroups displaying unusual interaction between multiple targets, for instance in the previously developed model classes for EMM: correlation, regression, Bayesian network, and classification (cf. [3,7]). The last of these model classes is particularly related to the SCaPE model class, with some major differences. Most notably, the classification model class *investigates* classifier *behavior* in the *absence* of a ground truth, whereas the SCaPE model class *evaluates* classifier *performance* in the *presence* of a ground truth. Hence, the two model classes are different means to achieve different ends.

Automated guidance to improve a classifier has been studied in the data mining subfield of meta-learning: how can knowledge about learning be put to use to improve the performance of a learning algorithm? In almost all of the existing meta-learning work, the focus is on letting the machine learn how the machine can perform better. By contrast, the SCaPE model class for EMM focuses on providing *understanding* to the domain expert where his/her classifier works well or fails. As such, SCaPE provides progress on the path sketched by Vanschoren and Blockeel [10, Section 5]: "We hope to advance toward a meta-learning approach that can explain not only *when*, but also *why* an algorithm works or fails [...]". Vilalta and Drissi [11, Section 4.3.1] do devote a subsubsection to "Finding regions in the feature space [...]", but this is in the context of algorithm selection.

A very recent first inroad towards peeking into the classifier black box is the method by Henelius et al. [6], who strive to find groups of attributes whose interactions affect the predictive performance of a given classifier. This is more akin to the classification model class for EMM. While Henelius et al. study hard classifiers, the SCaPE model class is designed for soft classifiers.

3 The SCaPE Model Class for EMM

Exceptional Model Mining (EMM) [3,7] is a framework within *Pattern Mining* [8]: the broad subfield of data mining where only a part of the data is described at a time, ignoring the coherence of the remainder. EMM is a supervised variant of Pattern Mining, typically invoked in a multi-target setting: there are several attributes $t_1, \ldots, t_m$ that are singled out as the *targets* of EMM. The goal of EMM is to find subgroups of the datasets where these targets display an unusual *interaction*. This interaction is captured by the definition of a *model class*, and subgroups are deemed interesting when their model is exceptional, which is captured by the definition of a *quality measure*.

In the SCaPE model class for EMM, we assume a dataset Ω, which is a bag of N records of the form $x = (a_1, \ldots, a_k, b, r)$. We call $\{a_1, \ldots, a_k\}$ the *descriptive attributes*, or *descriptors*, whose domain is unrestricted. The remaining two attributes, b and r, are the *targets*. The first, b, is the *binary target*; we will denote its values by 0 and 1. The second, r, is the *real-valued target*, taking values in $\mathbb{R}$. The goal of the SCaPE model class is to find subgroups for which the soft classifier outputs, as captured by r, represent the ground truth, as captured by b. In [4] and [5], we define a quality measure to assess this quality in a subgroup. Conceptually, the real-valued target r imposes a total order on the records of the dataset. The quality measure considers the ranking of the values of the binary target b under this order, and computes an average ranking loss [9]. This average ranking loss is computed for the entire dataset, and for each subgroup under consideration; the quality of a subgroup is compared to the overall quality in the dataset at hand. Subgroups with a higher-than-usual average ranking loss highlight areas of poor classifier performance, and subgroups with a lower-than-usual average ranking loss highlight areas of good classifier performance.

4 Experimental Results

In [4], we have presented subgroups found on the Monte-Carlo simulated FACT data, along with astrophysical interpretations. Additionally, in [5], we have presented results on nine UCI datasets. These results showcase what the SCaPE model class can unearth in your dataset, and describe problematic areas of the search space for these well-known datasets, which forms an interesting resource for any data miner striving to evaluate their methods on these datasets.

5 Conclusions

In gamma ray astronomy, the separation of gamma and proton showers marks an important step in the analysis of astrophysical sources. Better classifier performance leads to less dilution of the interesting physics results and improves the statement of results of the astrophysical source. The result set will more frequently contain the infrequently appearing gamma showers, which should increase the effective observation time. Due to the importance of the separation in this field, understanding why the classifier does not perform as desired is extremely valuable. The SCaPE model class for EMM helps to understand how to improve the classifier performance.

Beyond its importance within astrophysics, SCaPE is agnostic of the domain of the dataset it analyzes. In fact, it can be used to assess the performance of any soft classifier on any dataset when a ground truth is available. This makes SCaPE an invaluable tool for any data miner who wants to learn where his/her classifier works well and where its performance can be improved. What one could practically *do* with this knowledge depends on the task at hand. Our FACT experiments teach us at which settings the telescope delivers the best results, which allows us to improve the effectiveness of future observations. One could imagine the benefits of oversampling difficult regions, or learning a more expressive classifier on only the difficult regions of the input space. SCaPE provides you with the understanding where your classifier does (not) work: feel free to reap the benefits of that knowledge in any way you see fit.

Acknowledgments. This research is supported in part by the Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 “Providing Information by Resource-Constrained Analysis”, project C3.

References

1. Anderhub, H., Backes, M., Biland, A., et al.: Design and Operation of FACT - the First G-APD Cherenkov Telescope, arXiv:1304.1710 (astro-ph.IM)
2. Bretz, T., Anderhub, H., et al.: FACT – The First G-APD Cherenkov Telescope: Status and Results, arXiv:1308.1512 (astro-ph.IM)
3. Duivesteijn, W.: Exceptional Model Mining, Ph.D. thesis, Leiden University (2013)
4. Duivesteijn, W., Thaele, J.: Understanding where your classifier does (Not) work – the SCaPE model class for EMM. In: Proc. ICDM, pp. 809–814 (2014)
5. Duivesteijn, W., Thaele, J.: Understanding Where Your Classifier Does (Not) Work – the SCaPE Model Class for Exceptional Model Mining, technical report 09/2014 of SFB876 at TU Dortmund (2014)
6. Henelius, A., Puolamäki, K., Boström, H., Asker, L., Papapetrou, P.: A peek into the black box: exploring classifiers by randomization. Data Mining and Knowledge Discovery **28**(5–6), 1503–1529 (2014)
7. Leman, D., Feelders, A., Knobbe, A.J.: Exceptional model mining. In: Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008, Part II. LNCS (LNAI), vol. 5212, pp. 1–16. Springer, Heidelberg (2008)
8. Morik, K., Boulicaut, J.F., Siebes, A. (eds.): Local Pattern Detection. Springer, New York (2005)
9. Tsoumakas, G., Katakis, I., Vlahavas, I.P.: Mining multi-label data. In: Data Mining and Knowledge Discovery Handbook, pp. 667–685. Springer (2010)
10. Vanschoren, J., Blockeel, H.: Towards understanding learning behavior. In: Proc. BENELEARN, pp. 89–96 (2006)
11. Vilalta, R., Drissi, Y.: A Perspective View and Survey of Meta-Learning. Artificial Intelligence Review **18**(2), 77–95 (2002)