Active Learning using Dirichlet Processes for Rare Class Discovery and Classification

Tom S. F. Haines
thaines@eecs.qmul.ac.uk
Tao Xiang
txiang@eecs.qmul.ac.uk

School of Electrical Engineering and Computer Science
Queen Mary, University of London
London, UK

Classification is a classic problem. In many real-world problems the proportion of exemplars in different classes is highly imbalanced - the majority of the examples belong to uninteresting background classes whilst the interesting classes have few examples. For example, in the Sloan Digital Sky Survey most of the survey images of galaxies and quasars capture known phenomena, whilst unusual phenomena that could be evidence of new science constitute only 0.001% of the total dataset [3]. Crucially, those interesting rare classes are often not known a priori and need to be discovered. Typically to both discover and classify rare classes you would exhaustively label the entire dataset, to obtain sufficient instances of each rare class. Such a manual labelling process is often prohibitively expensive, rendering a supervised learning approach impractical.

Active learning offers a solution by minimising the manual labelling requirement. There are two competing goals - to find all the rare classes, and to refine the boundaries between the known classes. Most existing active learning methods either assume that all classes are known and thus focus on the classification problem or focus on the class discovery problem only. The approaches that try to meet both goals simultaneously [2, 4] are heuristic, and have free parameters that need tuning for each scenario.

A novel active learning approach is presented, which automatically balances the two competing goals and has no tunable free parameters. It takes the standard form of iteratively selecting a problem *instance* from a *pool* of instances. For each iteration the selected instance is given to the *oracle*, which provides the class label, and then used to update the model. The selection proceeds in three steps.

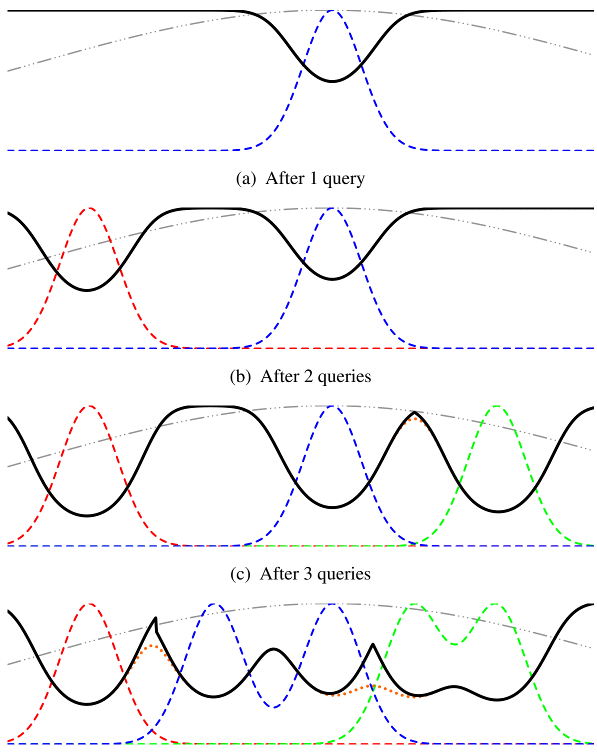
1. For each instance in the pool the probability of assigning it to each existing class, and also the probability of it belonging to a new class, is calculated under a Dirichlet process (DP) assumption. An explanation of the DP may be found in the paper. To continue it may be denoted as $DP(\alpha, \beta)$, where α is its *concentration parameter* and β is its *base measure*. Its marginal posterior, the Chinese restaurant process [1], is all that needs to be calculated for active learning, such that we obtain the probability of the instance belonging to each existing class and to a currently unknown class:

$$P_n(c \in C \cup \{\text{new}\}|d) \propto \begin{cases} \frac{m_c}{\sum_{k \in C} m_k + \alpha} P_c(d|c) & \text{if } c \in C \\ \frac{\alpha}{\sum_{k \in C} m_k + \alpha} P(d) & \text{if } c = \text{new} \end{cases} \tag{1}$$

where d is the data for the considered instance, C is the set of known classes, m_c the number of instances labelled with class c and α is the concentration parameter for the DP. Once normalised this provides a distribution for each instance that consists of the probability of the instance belonging to each of the known classes, as well as to an unknown class.

2. The probability that the instance will be misclassified is calculated. Normally this would be an *uncertainty* method of active learning, and would only improve the boundary between existing classes, but by considering the possibility of the instance belonging to a new class when the classifier can only assign known classes this metric also achieves the goal of class discovery. The balance between the two goals is decided by the concentration parameter of the DP, which is automatically inferred. Two assumptions are made - firstly that the classifier will select the class to which it has assigned the highest probability, noting that this only includes known classes, and secondly that the calculated distribution is an accurate estimate of what the true class of the instance could be, noting that it includes the possibility of a new class. It is then a simple matter to calculate the probability of incorrectly classifying an instance,

$$P(\text{wrong}|\text{data}) = 1 - P_n(c|\text{data}), \qquad c = \underset{c \in C}{\text{argmax}}\, P_c(c|\text{data}) \tag{2}$$



(a) After 1 query

(b) After 2 queries

(c) After 3 queries

(d) After 5 queries

Figure 1: $1D$ demonstration of the approach with 3 classes, showing the probability distributions. The prior is constant, indicated by the dot-dash grey line, whilst the distributions for the 3 classes use the 3 primary colours, dashed. Interest, as determined by $P(\text{wrong})$, is given in black.

where $P_n(c|\text{data})$ is the probability of the instance belonging to the selected class as calculated above, whilst $P_c(c|\text{data})$ is the probability calculated by the classifier, typically using Bayes rule with a $P(c)$ term.

3. A single instance is selected, based on the estimated chances of misclassification. In the paper itself results are presented with random selection, where the instances are weighted using $P(\text{wrong}|\text{data})$.

The key contribution of the paper is the above active learning criterion, $P(\text{wrong})$, which is specifically designed to balance the two competing goals of discovery and classification. Implementation is simple, it lacks tuneable parameters and is agnostic to the specific classifier being used, all whilst providing state of the art results as demonstrated in the paper. A quick demonstration of its behaviour is given in figure 1. It clearly shows the various behaviours expected - an interest in areas where either new classes could be or the boundary could be refined, with the latter gaining dominance as it loses interest in finding new classes.

[1] D. Blackwell and J. B. MacQueen. Ferguson distributions via Polya urn schemes. *Annals of Statistics*, 1(2):353–355, 1973.

[2] T. M. Hospedales, S. Gong, and T. Xiang. Finding rare classes: Adapting generative and discriminative models in active learning. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 15, 2011.

[3] D. Pelleg and A. Moore. Active learning for anomaly and rare-category detection. *Advances in Neural Information Processing Systems*, 17:1073–1080, 2004.

[4] J. W. Stokes, J. C. Platt, J. Kravis, and M. Shilman. ALADIN: Active learning of anomalies to detect intrusion. Technical Report 2008-24, Microsoft Research, 2008.