Challenging users' perceptions of decision boundaries in machine learning systems (WIP)

Rob Bowman
robert.bowman@cantab.net

Luke Church
Computer Laboratory
University of Cambridge
luke@church.name

Abstract

We investigate the design of smart systems that drive the exploration of decision boundaries created by Machine Learning (ML). We are interested in users' perceptions of decision boundaries, particularly in scenarios that involve the categorisation of non-categorical data. We present a system for colour categorisation using active learning with decision boundary visualisation and functionality for investigating users' perceptions of the decision boundaries they are teaching the system.

1. Introduction

One of the ways of writing programs is providing examples. This was studied early on as 'programming by example' (Lieberman 2001). The broad adoption of statistical machine learning techniques has changed this from an important, but niche, research question to a field with broad social implications.

Historically, it has often been asserted that Excel was the most popularly used programming language (Scaffidi et al 2005), this has now been displaced by labelling of machine learning systems through everyday interaction with ML systems such as Google Search, Spam/No-spam filters or other every day labelling tasks.

However, labelling is subject to a wide set of biases, and is an active area of research (Blackwell 2018). In this paper we describe a work in progress to build a tool for probing labelling behaviour in uncertain cases, with the intention of using it as a tool for providing creative interventions in tools used by artists.

2. Approach

2.1. Colour Categorisation System

We are implementing a prototype system that will be used in user studies to capture empirical results.

The premise of our system is a colour categorisation task. The system is initialised with points from the HSL (Hue, Saturation, Lightness) colour space with the objective of classifying each point as 'blue', 'green' or 'neither'. The motivation for our choice of task is that it requires the user to categorise non-categorical data, there is existing work which uses this task to study the phenomena of categorical perception that we can draw on, and we anticipate that it will lend itself well to decision boundary visualisation.

We frame the task as an active learning task. Active learning is a form of machine learning where the machine learning system learns to label an unlabelled dataset through interactively querying an information source such as a human user. Active learning is typically applied in situations where unlabelled data is abundant and manual labeling is expensive. A key challenge of active learning is the selection of the next data point to query the label of. There are a wide range of algorithms for this, the choice of which depends on the objective of labeling the system is learning.

In our active learning system the unlabelled data is a set of HSL points and the objective of the system is to learn how to classify each point through querying a user to label points. Our system uses support-vector machines to learn a probabilistic classification of the dataset. We select the next point

for labelling by computing the point with minimum difference in belief between its two most likely classifications. The support vector machine will learn decision boundaries which are partitions of the input space where it will classify points on either side differently.

We expect decision boundaries to form partitioning 'blue' from 'neither', 'green' from 'neither' and 'green' from 'blue' during experiments. We will analyse where decision boundaries form to study if biases occur, if the decision boundary visualizations can decrease bias in the learning process and the effects of factors such as the order of data points during labelling.

2.2. Research question 1: Do decision boundary visualizations improve users perceptions of machine learning models?

During uses of this system we are interested in users' perceptions of what the machine learning component is doing. We are interested in how users expect the ML system to label unlabelled points at different stages during an experiment:

- Before any points have been labelled
- When a small number of points have been labelled
- When a large number of points have been labelled

The system will be able to query the user to capture their belief as to how the system will classify new points which can then be compared with the ML systems actual behaviour to record measure of the error in users' perceptions.

We will record this error under the system in two modes.

1. A minimal system which presents the user with a visualization of only the target colour to label or predict the systems labelling of.
2. A system augmented with decision boundary visualizations.

We will look to address the research question by contrasting the recorded perception error under these modes.

2.3. Research question 2: Can decision boundary visualizations be used to reduce the effects of hysteresis?

We are interested by hysteresis effects whereby the users perception of how a datapoint should be labeled is affected by the preceding interactions with the system.

An approach we will take to exploring if there are such effects in this system is to use a system mode that queries the user to label colours in a sequence that transitions across a decision boundary. We will investigate how this impacts the positions of decision boundaries and users' perceptions of what the system is learning.

3. Current Status

We have implemented a system that enables the querying of users' to label points and express a belief about how the system would label a point. The system uses support vector machines to learn a classification over the colour space. We are currently considering options for the visualisation of decision boundaries. Figure 1 shows a screenshot of the interface in a labelling mode. The upper graphic within Figure 1 shows a decision boundary visualisation.
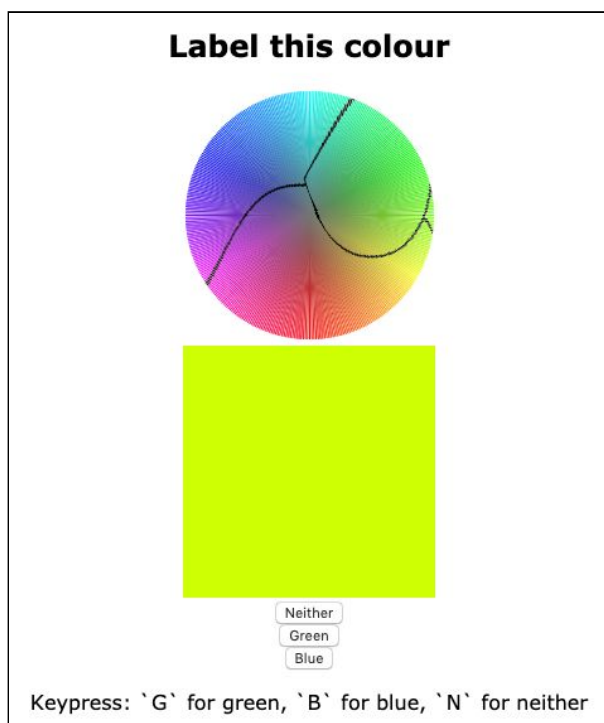
Figure 1 - Screen shot of the current work-in-progress system showing the labelling mode with decision boundary visualisation.

The system provides us with the functionality to run a range of experiments that are characterised by different sequences of interactions.

4. Forthcoming Work

We will be performing pilot studies test the setup of our experiment and options for decision boundary visualisation. We will then perform larger scale user studies to collect the empirical results which we will analyse in the context of our research questions.

5. References

Blackwell, A. F. (2018). Interaction with Machine Learning.
        https://www.cl.cam.ac.uk/teaching/1819/P230/

Lieberman, H. (2001). Your Wish is My Command: Programming by Example. Morgan Kaufmann.

Scaffidi, C., Shaw, M., and Myers, B. (2005) Estimating the numbers of end users and end user programmers. In Proceedings of *IEEE Visual Languages and Human-Centric Computing (VL/HCC)* 2005.