Edinburgh Research Explorer

Active Learning in Changing Environments

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
As humans, we have a remarkable ability to adapt our beliefs and strategies when we are confronted with new problems. One reason for our successes is that we are not passive observers. We direct our attention and choose actions in ways that allow us to update our beliefs more efficiently than if we had to learn through observation alone (Gureckis & Markant, 2009; Oaksford & Chater, 1994; L. E. Schulz & Bonawitz, 2007). This is true even when our goals are pragmatic, rather than being about learning per se. We strike a balance between exploring options and gaining information to serve future goals, and exploiting knowledge to obtain immediate rewards (Mehlhorn et al., 2015).

However, active learning only goes so far by itself: efficient exploration and exploitation depend on having a high-level model of the environment or the task to be solved. Without a sense of how actions relate to one another or what variables might be relevant to actions’ outcomes, learning will be aimless and rewards will be comparatively scant. The distinctively human talent is not active learning itself – which has been studied extensively and yielded valuable algorithms in many domains – but active learning in a world where the problems we must solve are constantly changing. These changes may be gradual, requiring minor adaptation to the environment, but there may also be structural changes that require drastic adaptation. Sensitivity to these structural changes allows an agent to transfer knowledge across domains, select informative actions, and bootstrap learning without requiring extensive domain-specific knowledge.

Our project seeks to understand active learning in settings where the structure of the task is subject to abrupt change, using a combination of computational models and psychological experiments. To assess the predictions of our models and gather data about human active learning more generally, we introduce a flexible new experimental framework. Participants are presented a sequence of grids made of tiles with different observable features, and are told to select tiles in order to maximize their score. Each click reveals a reward, providing indirect information about the problem’s structure, such as whether brightness is relevant, or whether there is a hidden spatial pattern that can be exploited.

To observe how adaptive people are in their decision making and learning process, we introduced the grids in different orders to two different groups. The sequences of grids are explained visually in Figure (a) below. Each type of grid was presented three times in a row.

Participants’ behaviour and performance in our experiment demonstrate that people, in the case of the brightness-first condition, behave adaptively, acquire a new rule, detect a change in the the underlying structure of the environment and re-use a previously seen rule. The most noteworthy learning qualities amongst human participants are incremental progress over sequential trials on the same grids, the ability to detect change across grid types and the transfer of structural knowledge across grid types (see Figure (b) ).

Despite the clear group effects that were observed in the experiment, there was a great amount of differences between participants. Studying the rich diversity in learning and decision making processes is a necessary step to the study of human learning and other cognitive processes, yet has often been ignored, psychologists preferring cleaner data and modelling of group behaviour. We believe that this approach can help design flexible adaptive models that can outperform current models. In the next part, we will describe the theoretical framework devised to confront these multiple challenges.

Our approach builds on recent developments in active learning and sequential decision making algorithms, and allows us to understand group-level and individual human behaviour. We use Bayesian optimization (BO), a family of techniques that has proven useful in solving practical active learning problems (Snoek, Larochelle, & Adams, 2012). To explore how participants learn...
in a dynamical setting, we combine these tools with sequential Monte Carlo algorithms that have shed light on human judgments in causal reasoning and categorisation (Abbott, Griffiths, et al., 2011; Sanborn, Griffiths, & Navarro, 2006), and show how a resource-constrained agent might make approximately optimal inferences in practice. Using these methods, we can explicitly compare different theories of human active learning, as expressed in priors over problem structures and environmental change as well as mechanisms underlying belief revision.

In BO, we are interested in finding the optimal sequence of observations to find the maximum of an unknown function. With our observations, we can construct a Bayesian posterior over the unknown function. To do this, BO uses Gaussian Processes, a type of non-parametric Bayesian model (Williams & Rasmussen, 2006). Non-parametric Bayesian models are an attractive framework for modelling human learning, since they do not require any arbitrary postulations about the parametric form of the function (or rule) to be inferred. A Gaussian Process model relies on the covariance function to express a rich distribution on functions, as a prism that expresses the assumptions — e.g. smoothness or stationarity — about the the unknown function. Overall, BO is an attractive framework with transparency about structure representation and explore-exploit policies, enabling us to test and compare theories explicitly. Furthermore, GPs have been used to successfully explain a range of human learning phenomena (Lucas, Griffiths, Williams, & Kalish, 2015; E. Schulz, Konstantinidis, & Speekenbrink, 2015).

An acquisition function is required to resolve the explore-exploit trade-off when trying to find the global maximum in few queries (Mockus, Tiesis, & Zilinskas, 1978). The sampling strategies of the model are the policies it follows when making its decision about the next tile to select, given the representation of the grid given by the GP. Selecting the appropriate strategy is a difficult problem given the lack of information about the function. A strategy that works very well for a certain class of problems offers no guarantee to be efficient when faced with a different class of problems.

Our framework enables us to explore different acquisition functions and test the ones that humans might use to trade off exploration and exploitation. In this work, we use Thompson sampling, often described as the probability matching strategy, because of its long history in the psychology literature. It has been recently used in recent computational accounts of people’s decision making strategies (E. Schulz et al., 2015), and shown to be asymptotically optimal in general environments (Chapelle & Li, 2011).

Having an accurate model of the environment is crucial for being able to predict events and a level of control necessary to achieve desired goals. An environment may change gradually over time, which requires minor adaptation to the environment, however, there may also be abrupt changes that require drastic adaptation, and a revision of the structural assumptions about the environment and of the animal’s behaviour. One of the focal points of this study is to understand how people identify new environment structures and adapt their beliefs. We explore different kinds of models with different hypotheses about participants’ behaviour in our experiment. We compare our models against the approach presented in Snoek et al. (2012) and argue for some of theirs advantages.

To implement the dynamical assumptions for our model, we use Particle Filters (PF), a sequential Monte Carlo method, that can be used to approximate a sequence of posterior distributions, when performing Bayesian inference repeatedly in response to a sequence of observations. These models make the Bayesian computations tractable and produce order effects that are con-
sistent with human behaviour. Sanborn et al. (2006) showed that PFs were a plausible model to explain human categorization. Furthermore, by varying the number of particles we can explore the interaction between the cognitive constraints of the agent and the statistical inference problem. This makes them a natural framework in which to define models that are rational both in defining the computational problem, but also the approximation to solution of that problem given the resources of the agent.

Our results show that participants are able to learn efficiently and adaptively in changing environments, but can also suffer from garden-path effects and sometimes fail to adapt. Our models successfully predict human decisions, and characterise the learning behaviours and inductive biases of individual participants. In so doing, they provide a means to mimic the best learners’ adaptive behaviour, and simultaneously provide insights into how and why human learners occasionally fail. Initial results show that our human inspired models can outperform standard blackbox bayesian optimization algorithms on sequences of tasks that share structural similarities whilst being sensitive to change and still being able to adapt to unforeseen tasks.

References


