Learning Machines

Introduction

This document is a proposal for the Learning Machines project. The project is funded by the Turing Health programme, and the Turing Criminal Justice programme.

The first part of this document contains our understanding of the project background. The second part defines the research challenges to be addressed by this project. The third part outlines possible approaches and risks to addressing the research challenges.

1. Context

Humans make decisions based upon past experiences. These experiences are gained from personal, social and professional interactions.

In a professional setting, a human expert such as a clinician decides on treatment regimens while considering patient socioeconomic profiles, medical histories and current disease and comorbidities states. In another example, a judge decides if a prisoner can be released on parole based on criminal histories, current behaviours and psychological profiles.

In both examples, clinicians and judges build, retain and update their experiences of disease and criminal behaviours. These experiences are referred to as domain knowledge or domain models. Experts update their models from interactions' outcomes; clinicians update their models when they receive updates about treatment outcomes. Judges update their models when they encounter (possibly again) prisoners in court.

Domain knowledge/models contain complex representations of patients (or prisoners), which must include sufficient information such as variables which may confound predicted or desired outcomes.

There are good reasons for formalising domain knowledge/model, which has traditionally been referred to as an expert system. Access to a shared model has the capacity to enable objective decisions because it is trained with a diverse range of interactions. A formal representation of a model when built with particular ML techniques to maximise transparency also allows identification of biases.

Machine learning techniques are good at finding patterns from large datasets. Data collection is expensive. In the real world (such as in the medical or criminal justice domain as in the examples above), data collecting and labelling under real world circumstances specifically for the application of machine learning is rare.

Clinicians and judges are generating relevant data every time they interact with the medical or criminal system. When data is stored and labelled over a period of time, they can be used to develop and iteratively update a model. For example, the diversity of patients treated by multiple clinicians over time generates heterogeneous patient profiles which is not experienced by a single clinician.

When a machine is retrained or iteratively updately with new data, its performance and parameters may change to reflect the new variations contained in the new data.

2. The Problem

Changes in model behaviours are inevitable as models are updated (or retrained) with newly labelled diverse cases. This is true during initial model development, but in this project we are particularly interested in changes in model behaviour as a result of iterative updates over the long term.

For one example patient demographics may change. In the case of cystic fibrosis, new treatments mean that patients can survive longer without the need for a lung transplant; this means that an algorithm that predicts when a patient requires a transplant will be gradually dealing with older patients. Another example is when new treatments are introduced or new data such as genomic profiles become more accessible.

For another example, an uprise in anti-immigrant sentiments has lead to an increase in reported harrassment and convicted cases. An algorithm that predicts the probabilities of reoffending will need to take into account different criminal profiles.

The research question posed by the Learning Machine project is:

What issues will arise when models are retrained iteratively over a period of years? How should these issues be addressed to provide users with both new information and a sense of consistency?

Examples of issues that Learning Machines can address are:

1. How do we measure model changes
2. What can examination of the model change tell us about new data?
3. What guidance will we provide users when they are confronted with new model behaviours?

We propose to address these issues from three perspectives:

2.1 Uncertainty

Uncertainty information is usually provided as part of a decision or classification outcome. It is typically in the form of a probabilistic, or paired upper and lower bound value. During initial development process, uncertainty values are expected to be large. After a number of training epochs, a model which has been trained with sufficient information, and without too much noise is expected to produce decreasing uncertainty values. Uncertainty is not a measure of the trained models' capabilities, although it is often intepreted as such.

Uncertainty about outcome is an important piece of information to be provided to users such as clinicians. A measure of uncertainty enables clinicians to exercise professional judgement about accepting or declining an algorithm's recommended outcome. Measures of uncertainty also enables a form of ranking in regards to algorithm outcome options.

This project will survey how uncertainty values, as well as confidence intervals change over time, for a number of different datasets.

2.2 Intepretability

Intepretability or explainability is an important topic in machine learning. A system built for the medical or criminal justice domain has to be transparent. This means that users receive justifications or explainations on how decisions are reached, in order to be able to decide if they will accept or reject the recommended outcome. The ease of intepretation and the closeness to which model parameters map to real-life features are essential components for tools built to augment an expert's decision making process.

When a machine is updated iteratively over a period of years, it is inevitable that its behaviour will change. This can be due to changes in the data. Changes in data can be due to many things, such as gradual shifting of patient demographic or they can be due to new data collection practices. When this occurs, capabilities such as transparency and explainability becomes foremost in deciding if a machine is valid in producing different results over time.

This project will address a two aspects of intepretability or explainability; that is 1) how to provide explainations for new algorithms and 2) develop measurements that would provide information about algorithm parameter change over time.

2.3 Heterogeneous and/or high dimensional longitudinal data

The study of heterogeneous, high dimensional data for training machine learning algorithms is well established. As mentioned above, data must contain sufficient information and this includes variables that may confound outcomes. Typically, a fix set of datatypes is used to train a model, but in real life, more diverse

data types are being collected than before, with the purpose of addressing any information from possibly confounding variables. For example, in the medical domain there may be more lab results, imaging reports, 'omic'-type profiles and quality of life questionnaires.

This project will address the challenge of introducing new data types into an existing system. There is a further challenge here because while a completely new model may be developed to incorporate new data types, the impression communicated to users should be one that the system is being extended, rather than completely rebuilt.

This project will also research the concept of a 'recency bias' (i.e. should new data be weighted more heavily than older data) so that old data will not have to be discarded completely.

3. Goal

A demonstration of a machine/model that: * predicts outcomes / recommends actions or treatments * automated updated or retrained over time with new data * meets the interpretability, safety and ethical requirements of sensitive domain areas

4. Example datasets

- UK Cancer registry
- (Symbolic metamodelling paper) Predict 5 year mortality risk of breast cancer patients using age, number of nodes, tumour size, tumour grade, Estrogen-receptor status.
- Meta-analysis Global Group in Chronic heart failure database (MAGGIC)
- Reference work (AutoPrognosis paper)
- Cystic Fibrosis Trust
- Reference work (AutoPrognosis paper)
- United Network for Organ Sharing (UNOS) database
- Reference work (AutoPrognosis paper)
- UNOS-I: pre-transplant, UNOS-II post-transplant
- Surveillance, Epidemiology, and End Results (SEER) cancer registries
- Reference work (AutoPrognosis paper)

- Comorbidities - predict cardiac deaths in patients diagnosed with breast (SEER-I), colorectal (SEER-II), Leukemia (SEER-III), respiratory (SEER-IV), digestive (SEER-V), urinary (SEER-VI)cancers.
- MIMIC critical care database
- deidentified health data associated with ~60,000 intensive care unit admissions. It includes demographics, vital signs, laboratory tests, medications, and more

5. Risks

The risks around this project are centered around the issue of datasets. There are a number of requirements for the datasets which can be used for this project. These are:

- Datasets must be collected over a long period of time, in order to show how algorithms' outputs can change over time.
- Use cases must show that some actions can be taken in response to the machines' outputs, in order for this system to be useful
- Some evidence that the data has changed over time (eg. patient demographic), in order to show the value of retraining/redevelopment.

6. Work breakdown

We anticipate undertaking the following activities. It is not possible to say with certainty which activities will take more or less time. For example, more time may be required to format data into the appropriate structures. Therefore some activities may be removed or changed, and other added if it would be appropriate for the project to do so.

6.1 Data management

- We anticipate working within "Safe Haven" environments in order to reproduce some existing publications.

6.2 Scoping for use cases

- Work with domain experts to identify use cases within datasets which would meet the requirements of this project. ### 6.3 Automated Evaluation Platform
- Develop automated pipelines for re-evaluating updated models
- Develop measures and visualizations of model parameter changes over time

- Develop an interactive system to enable users to assess model outcome with different inputs

tags: `Learning Machines Documentation`