PRECISION MEDICINE:

Identifying Cancer Cells with Machine Learning using Oscar: Data Science

OSCAR
DATA SCIENCE

Personalisation of treatment

The early diagnosis and prognosis of a cancer type has become a necessity in cancer treatment. Recognising the specific subtype of cancer and whether the patient belongs to high or low risk groups can make a substantial difference to life expectancy. As such cancer research is constantly looking for quicker and more accurate ways of identifying different types of malignant and benign tumours.

Massive Analytic Limited (MAL) in conjunction with University College London (UCL) through a Knowledge Transfer Partnership (KTP), has developed new capabilities within it's product *Oscar:DataScience* (Oscar) for precision medicine. Taking shape as new machine learning algorithms, these algorithms can distinguish between the thousands of features that identify the different types of cancerous cells.

The research focused on microarray gene expression datasets; these are high-dimensional and noisy datasets that contain genetic data typically used for distinguishing between benign or malicious tissues or classifying different types of cancer. The end goal was personalised treatment for cancer patients, based on the individual features of their cancers.

A Novel Solution to Feature Selection

Datasets with thousands of features are the new norm, but how do you tell which features have the predictive power you're looking for? Feature selection is a central problem in machine learning, where the aim is to derive a representative set of features from which to construct a classification (or prediction) model for a specific task. By discarding information that is irrelevant (i.e. has little predictive or classification power), noisy and redundant, and considering only a subset of the features – it's possible to identify a set of classifiers (or predictors) that have strong correlations with the task you're trying to accomplish.

MAL developed a novel approach to feature selection using multiagent reinforced learning. Reinforcement learning is a model in which an active decision-making agent interacts with its environment and learns from reinforcement (a numeric feedback) in the form of reward or punishment. The feedback received

is used to improve the agent's actions. The central idea of the approach was to "assign" a learning agent to each feature. Each agent controlled a single feature and learned whether it would be included in or excluded from the final relevant feature subset; we refer to this approach as Multi Agent Reinforcement Learning (MARL).

Applying this to microarray datasets created new challenges, namely a large multiagent coordination problem between the thousands of learning agents. To address the scalability challenge CLEAN rewards were applied, a form of reward shaping, that removes the exploratory noise caused by other learning agents. This achieved excellent empirical results in a variety of domains, enabling scalability for data sets with even millions of dimensions.

The result of this research is a set of algorithms that can select the most important and predictive features out a dataset that includes thousands. Applied to precision medicine it enables Oscar to identify different types of cancer automatically and with exceptional accuracy.

Autonomous Diagnosis and Predictions

MAL's mission is to bring autonomous process automation to all industries with it's AI technologies. Through the KTP fully autonomous predictive powers was brought to precision medicine, where none existed before. The project addressed the challenges of fully automating data science with the goal of predicting outcomes with the absolute minimum human input, by combining multiple skill sets in a single platform. The work created new algorithms for dimensionality reduction, resulting in a novel approach to autonomous feature selection using MARL.

The deployment of these algorithms is automated inside of Oscar and can deployed inside of a workflow for autonomous decision making in precision medicine and also other fields where large numbers of features pose challenges in creating prediction models.

MassiveAnalytic
The Precognition Company
www.massiveanalytic.com