When Computers Decide: Understanding Uncertainty in Data-driven Systems

Conferencista: Lynda Hardman

The term "AI", or Artificial Intelligence, has become a familiar term in the media. Nations across the globe are investing in it, companies are anticipating earning large huge amounts of money with it, some are worried about the consequences for our privacy and even our humanity, others are profiting from stories of societal havoc. But what exactly are we up against?

An "AI" is a computer program. It was created by a human. So what is the difference between a "normal" computer programme and "an AI"?

While in the academic field of AI a number of different technical approaches exist, the term nowadays tends to refer to data-driven AI: training data is provided to a computer programme, a machine learning system, that creates a model based on the data. The model is incorporated into another computer programme that can be used on new input data, e.g. to make decisions.

Examples of learned models are classifiers that can distinguish between healthy or cancerous cells, or deciding whether a product in a production line is damaged.

A classifier is based on good "old-fashioned" statistics, but applied to huge sets of data at lightening speed.

When applying learned classifiers in the real world we need to be aware of a number of things, including who supplies the data, how the data is analysed and what the consequences are of using the learned model in its application area. Politicians, decision makers and the general public need to be aware of the advantages and limitations of classifiers both to enhance the economic benefits and to curb potential misuse. Through basic education of how classifiers work we can increase successful application of the technology while reducing potential negative effects.

A classifier is only as reliable as its input data but sometimes even experts don't, or perhaps can't, agree on whether an example belongs to one category or another. This is not necessarily a problem as long as we can inform the end-user how reliable the classifier is.

When training a classifier, sometimes we want the system to err on the side of caution (for example when trying to identify cancer cells) or perhaps be more lenient to save money, in the case of not throwing away too many damaged goods. Users need to understand that they can adjust the error rates of the classifier according to their application needs.

Sometimes the training process occurs in multiple stages, where the output from one stage is used as input to another. The question is then how do "errors" or inconsistencies in the output of one stage affect the results in a stage downstream.

While those who train classifiers endeavour to select examples that are representative of the data that will be used in the field, it is possible that different proportions of classes will be present in the field. This also has consequences for interpreting the results of the classifier.

The talk will explain these points in an intuitive way, avoiding the need for mathematics and jargon, allowing the audience to understand the processes behind data-driven AI and allow them to ask the right questions when purchasing, applying or being the target of data-driven systems.