tsg insights
November 2021

Testing of artificial intelligence (AI) and machine learning (ML) – Supervised learning

This whitepaper is the third in a series which act as companion pieces:

- An overview of artificial intelligence (AI) and machine learning (ML).
- Introduction to the testing of AI and ML.
- Testing of AI (artificial intelligence) and ML (machine learning) – supervised learning.
- Testing of AI (artificial intelligence) and ML (machine learning) – unsupervised learning.
- Testing of AI (artificial intelligence) and ML (machine learning) – reinforcement learning.
- AI and machine learning – algorithmic bias – the cruel mirror AI and ML reflects back at us.

In this paper, we look at the approaches to and challenges regarding supervised learning.

What is supervised learning?

Supervised learning is example-based. A computer algorithm is trained on input data labelled for a particular output. Similar to test automation, the test cases have expected results. This is straightforward enough, but comes with challenges:

- There must be sufficient training data to start with.
- Data must be selected and most likely whittled down.
- There needs to be an expected result.
- An evaluation scoring system must assess the success of the machine learning.

Because humans have provided the basis for decisions, supervised learning does have the advantage of being more likely to come up with answers that humans can relate to.

Key approaches for supervised learning

The main approaches for supervised learning are regression training and classification training.

The regression training approach

There are two basic types of regression training:

- Simple regression analysis ((also known as linear regression analysis) estimates the relationship between a single independent variable and one dependent variable. For example: network capacity and file transfer speed.

- Multiple regression analysis estimates the relationship between multiple independent variables and one dependent variable. For example: speed achieved using different gear ratios and engine horsepower output. This analysis is obviously more complex than simple regression, but tends to produce more realistic results.

Either of these forms of regression analysis can lead to statistical bias as well as algorithmic bias.

The classification training approach

In machine learning, classification training is the process of assigning the correct classes based on observations. There is a close relative called “clustering” which is used for unsupervised learning.

The algorithm used to perform classification is called a classifier. Both classification and clustering perform pattern recognition.

Classification also strays into the realm of probability analysis. Essentially, incoming unlabelled data is classified by observed features. These are then allocated to vectors based on combinations of these features.

At its most basic, it can be used in binary ways – for example spam or non-spam for email, or positive or negative feedback. More involved scenarios might include medical scenarios – for example BMI, or blood pressure, which might be classified as low, ideal, pre-high or high.

Moving up the scale, one example would be credit scoring. This involves what might be called an educated guess following correlation of multiple factors such as income, age, family status, outgoing expenses, home ownership and ratio of existing debt to income. Even then, it’s usually not just simply a case of outright approval or rejection, but acceptance for certain loans only, or credit offers with more or fewer conditions.

Features are defined, then unlabelled data is statistically analysed and grouped into vectors (defined by the financial institution concerned) and credit decisions are thus triggered.

The more vectors there are, the more features that are added and the more complex the vectors, the more there’s a risk of combinatorial explosion. In other words, there are so many factors involved that the answers can become unmanageable or unacceptably biased.

Car navigation apps, for example, will limit the number of possible options and therefore might send you by a less than optimum route on occasion. The problem is potentially more serious in applications such as facial or handwriting recognition, or with medical imagery and biometrics.

As with regression analysis, unintended bias can also occur. So it’s not surprising that in sensitive fields such as these, the law is taking a close interest in how decisions are reached.

Algorithmic bias

As we’ve seen, supervised learning is particularly at risk of algorithmic bias. There might well be – quite possibly unintentionally and even not recognised – bias in the training data. Our sixth whitepaper in this series, Algorithmic Bias – the Cruel Mirror, details actual cases where this is happening.

Those cases highlight the efforts made to counter this bias by the organisations concerned. But even major corporations were sometimes forced to admit defeat, despite spending a lot of money and effort on countering the bias. What happens with organisations which lack their resources? Or where the bias isn’t even recognised as such?

Under current and proposed algorithmic accountability (AA) legislation, this could have serious legal consequences. Even before any applicable decision is being made, reputational damage due to negative traditional media, social media and community condemnation could also be significant.

Consequences can be severe if potential algorithmic bias is not recognised in time, either during the selection and whittling down phase prior to training or, at the very latest, after the training and the review of the results.

Other issues

There can be issues of both too little and too much data with supervised learning. Because it typically requires large amounts of quality data that's been correctly labelled, sometimes there's simply not enough data.

On the other hand, in some supervised learning scenarios, the sheer amount of potential data is a problem in itself. The challenge of selecting and whittling down training data also has an effect on the amount and effectiveness of testing.

In addition, overusing the same training data could reinforce biases within ML, making them more difficult to overcome in the long term.

If only recent data is used if AIs and MLs, algorithmic bias will be less of a risk. However, where do you draw the line? And how far back do you need to go to have sufficient data? The answer here would seem to be employing a judicious combination of historic and recent data, with one caveat. Any organisation using historic data should have a policy of pausing at the outset to consider the risk of algorithmic bias being present and, if so, to what degree.

TSG provides expert guidance on AI and ML, as well as assurance and testing services. We make change happen, safely and predictably. If you have any question about issues covered in this whitepaper or would like to know more about how we can help you, please contact us now. **Call: +44 (0) 207 469 1500** **Email: info@tsgconsulting.co.uk** **www.tsgconsulting.co.uk**