Maximizing Smiliness

Machine learning and the Emojibots

Louise Dennis

THE BASIC PREMISE BEHIND THE THREAT POSED BY THE EMOJIBOTS IN *SMILE* is that they have been programmed to maximize the happiness of the colonists. When faced with a situation (grief) where their normal mechanisms for increasing happiness proved ineffective they realised that an alternative solution was to eliminate unhappy colonists.
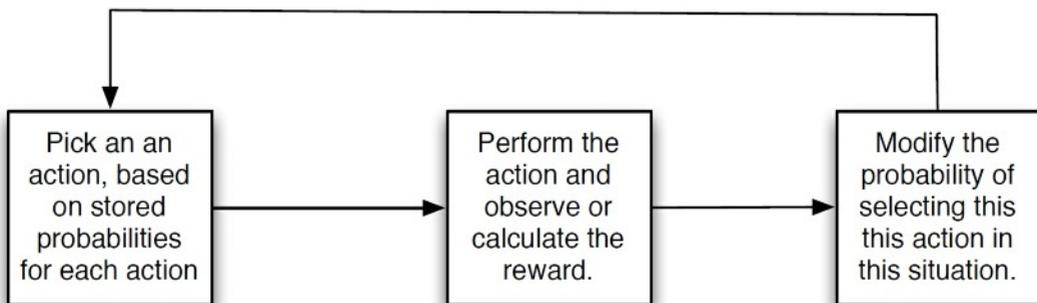
There is (understandably) relatively little dialogue in the episode about how the Emojibots (actually the micro Vardy bots) work. I've taken the following words of the Doctor from the episode.*

The Vardies. Well, their job was to maintain happiness. At first, that meant making sure there was enough oxygen and water. That's what the badges are meant to communicate. Satisfaction, a positive mental state. But the Vardy are smart. They learn, try to be good servants, so they expand the definition of happiness until...

This strongly implies that we are looking at a machine learning process. Broadly speaking machine learning can be applied to do one of two things. Firstly it can classify things. Image recognition is a classic example of the application of machine learning as a classification process. The algorithm is trained on a set of images and adjusts its classification process based on information provided by the user about what each image shows, thereafter it can identify new images correctly (usually).

The second application of machine learning is as an optimisation process and that appears to be the version in use here. In optimisation problems a machine learning algorithm works with a **reward function** and it seeks to maximize the return of the reward function. Most algorithms work by trying various actions (initially all actions have the same chance of being chosen) observing the reward returned, and then tweaking some internal probabilities. If the algorithm received a good reward then it is more likely to try that action again in a situation it recognises as similar. Over time the algorithm is less likely to adjust the probabilities much even when actions do not perform as expected and so is more likely to stick to some set of actions it has previously learned are the most effective in a given situation. There are a wide variety of such algorithms which vary in the mechanisms used for representing situations (where machine learning as a classification process may also come into play), representing actions and tweaking probabilities.

One of the big challenges for machine learning is the construction of good reward functions. In particular when the desired outcome has several competing components, constructing the function to balance out the various factors can be difficult, particularly since it often isn't obvious how that balancing should occur. So, for instance, it is relatively easy to construct a machine learning algorithm to figure out how to move a robot as quickly as possible, but it is much harder to construct one to move a robot as quickly as possible while showing consideration to humans and other entities in its environment.

```
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│ Pick an an      │     │ Perform the     │     │ Modify the      │
│ action, based   │ ──▶ │ action and      │ ──▶ │ probability of  │
│ on stored       │     │ observe or      │     │ selecting this  │
│ probabilities   │     │ calculate the   │     │ this action in  │
│ for each action │     │ reward.         │     │ this situation. │
└─────────────────┘     └─────────────────┘     └─────────────────┘
```

From a machine learning perspective, what went wrong on Gliese 581d** was that someone attempted to construct a reward function that was supposed to achieve a harmonious balance between basic human needs (oxygen, water, etc.) and less tangible considerations such as (presumably) a pleasant living environment, sufficient leisure time, etc. Whoever it was apparently settled on a reward function based on maximizing smiliness on the badges in use and failed to realize that if it only accounts for badges in use then one solution to a frowny badge was simply to eliminate the badge.***

This points to an additional challenge in current research on machine learning, that of **safe learning**. Really you wouldn't want the option to eliminate a human to be in the range of actions considered as part of Vardy learning, any more than you want a driverless car to experimentally crash into something to see what sort of reward that brings.

In short, whoever designed the Vardy and the Emojibots failed in two respects: firstly they supplied a reward function that could be maximized by reducing the number of emoji badges in existence and secondly they failed to place sufficient constraints on the machine learning process to prevent it taking unsafe actions as part of its experimentation process. There are a variety of solutions to this problem. I should note however, that in my opinion none of them involve paying the Vardy rent.

Endnotes

* Chrissie's Transcripts Site, www.chakoteya.net

** Wikipedia says this is the name of the planet in *Smile*, though it is never mentioned on screen.

*** A Watsonian (in-universe) explanation for this might be that if the reward function was based on total smiliness then lots of slightly grumpy colonists might score more highly than a small number of happy ones and so one solution would be to wake everyone in the colony ship up before the colony was ready for them.

The City of Arts and Sciences, Valencia, Spain, which portrayed the colony city in Smile.
Image: Maybelline71, (CC BY 2.0)