Math 189R expectation maximisation paper summary

Adam Guo 2020-03-09

Relational Neural Expectation Maximization: Unsupervised Discovery of Objects and Their Interactions
Sjoerd van Steenkiste, Michael Chang, Klaus Greff, Jurgen Schmidhuber
`https://arxiv.org/pdf/1802.10353v1.pdf`
`https://github.com/sjoerdvansteenkiste/Relational-NEM`

The goal of this paper is to create an unsupervised algorithm that is able to discover physical objects and infer their real-world interactions by analysing images. Many "common-sense physical reasoning" algorithms make use of external information about objects in the environment, such as prior physics simulations, to learn their interactions. As a result, these supervised algorithms have limited applicability to arbitrary real-world scenarios. Other unsupervised algorithms exist that use neural networks to learn interactions on a pixel level (analysing the graphical data itself), but fail to model the underlying behaviour of the objects themselves. The authors of this paper propose an algorithm called *relational neural expectation maximisation (R-NEM)*, an iteration of the existing neural expectation maximisation algorithm, that can discover both object representations and their interactions.

R-NEM builds on neural expectation maximisation (NEM), an algorithm that groups pixels that represent the same object, thereby identifying discrete objects in an image. Each object is represented by a component $\boldsymbol{\theta}_k$, and the goal is essentially to compute $P(\boldsymbol{x} \mid \boldsymbol{\theta}_k)$, a mixture statistical model of the K components, with which we can infer the component that each pixel most likely belongs to. To compute the components $\boldsymbol{\theta}_k$, a neural network is used to first transform $\boldsymbol{\theta}_k$ into parameters $\psi_{i,k}$, which are then fed into the expectation maximisation algorithm to maximise $P(\boldsymbol{x} \mid \boldsymbol{\theta}_k)$ via $\psi_{i,k}$.

R-NEM develops NEM further by adding a recurrent neural network that takes a new parametrised interaction function $\Upsilon_k^{\text{R-NEM}}$, which transforms the parameters $\boldsymbol{\theta}_k$ and computes the pairwise effects of objects $i \neq k$. This approach models interactions between objects without compromising the learned object representations, overcoming the limitations of previous RNN-NEM models.

The paper tests R-NEM on simple graphical objects such as bouncing balls and video capture of the arcade game *Space Invaders*. Trained on input videos, the algorithm is able to discern individual objects (such as individual balls) and predict their movement with less loss than the standard RNN and RNN-NEM algorithms, even when external factors such as occluders that block certain parts of the frame are introduced. This suggests that the algorithm models human learning patterns (identifying objects and learning from their interactions) more closely. However, higher-level behaviour, such as understanding that groups of objects can behave similarly (for instance, the large block of aliens in *Space Invaders*), is still missing.