

Use chaos theory, dynamical systems or game theory to analyze an artificial intelligent system.

In this task, i want to use game theory to analyze an artificial intelligent system.

Game theory is the study of mathematical models of strategic interaction among rational decision-makers.

It has applications in all fields of social science, as well as in logic, systems science and computer science. Originally, it addressed zero-sum games, in which each participant's gains or losses are exactly balanced by those of the other participants.

Game Theory is a branch of mathematics used to model the strategic interaction between different players in a context with predefined rules and outcomes.

Game Theory can be applied in different ambit of Artificial Intelligence:

- Multi-agent AI systems.
- Imitation and Reinforcement Learning.
- Adversary training in Generative Adversarial Networks (GANs).

Because I have relatively little knowledge of Multi-agent AI systems and RL, I just did some research and then introduced them.

This time, I want to focus on the game theory used in the GANs.

Nash Equilibrium

For example, a Classification algorithm such as SVM (Support Vector Machines) can be explained in terms of a two-player game in which one player is challenging the other to find the best hyper-plane giving him the most difficult points to classify. The game will then converge to a solution which will be a trade-off between the strategic abilities of the two players (eg. how well the fist player was challenging the second one to classify difficult data points and how good was the second player to identify the best decision boundary).

The Nash Equilibrium is a condition in which all the players involved in the game agree that there is no best solution to the game than the actual situation they are in at this point. None of the players would

have an advantage in changing their current strategy (based on the decisions made by the other players).

Following our example of before, an example of Nash Equilibrium can be when the SVM classifier agrees on which hyper-plane to use classify our data.

One of the most common examples used to explain Nash Equilibrium is the **Prisoner's Dilemma**.

Let's imagine two criminals get arrested and they are held in confinement without having any possibility to communicate with each other.

- If any of the two prisoners will confess the other committed a crime, the first one will be set free while the other will spend 10 years in prison.
- If neither of them confesses they spend just one year in prison for each.
- If they both confess, they instead both spend 5 years in prison.

In this case, the Nash Equilibrium is reached when both criminals betray each other.

Adversary training in Generative Adversarial Networks (GANs)

GANs consists of two different models: a generative model and a discriminative model.

Generative models take as input some features, examine their distributions and try to understand how they have been produced.

Discriminative Models instead take the input features to predict to which class our sample might belong.

In GANs, the generative model uses the input features to create new samples which aim to resemble quite closely the main characteristics of the original samples. The newly generated samples are then passed with the original ones to the discriminative model which has to recognise which samples are genuine and which ones are fake.

An example application of GANs can be to generate images and then distinguish between real and fake ones.

This process resembles quite closely the dynamics of a game. In this game, our players (the two models) are challenging each other. The first one creates fake samples to confuse the other, while the second player tries to get better and better at identifying the right samples.

This game is then repeated iteratively and in each iteration, the learning parameters are updated in order to reduce the overall loss.

This process will keep going on until Nash Equilibrium is reached (the two models become proficient at performing their tasks and they are not able to improve anymore).

Because the GAN framework can naturally be analyzed with the tools of game theory, we call GANs “adversarial.”

But we can also think of them as cooperative, in the sense that the discriminator estimates this ratio of densities and then freely shares this information with the generator.

From this point of view, the discriminator is more like a teacher instructing the generator in how to improve than an adversary. So far, this cooperative view has not led to any particular change in the development of the mathematics.

Reference

1. [Game Theory in Artificial Intelligence](#)
2. [Game theory\(wikipad\)](#)
3. Goodfellow I. Nips 2016 tutorial: Generative adversarial networks[J]. arXiv preprint arXiv:1701.00160, 2016.