Cogisen Cognitive AI platform

White paper October 2018

Contents

Introduction

Cogisen has developed a machine learning platform that has set it apart from the current trends in the field.
This white paper presents the philosophy behind our company and a high-level technological overview of our approach.

Company philosophy

We believe the industry's quest for faster machine learning with temporal models and multimodal data integration has largely been left unsatisfied by the state of the art of deep learning.

We know deep learning has proven to be a successful tool for a broad range of applications, yet looking at it as the only paradigm to frame problems is too narrow.

We feel the need to take a step back and use physics and different mathematics tools as the foundations for the next generation of explainable and composable machine learning models.

How: quantum approach

Much of machine learning is based on probabilistic assumptions. Yet, the Bayesian approach is not guaranteed to provide the best description of a phenomena. Quantum probabilistic models give a more insightful framework.

Data can be purposely remapped in a much more structured space, where mathematical models exploited in quantum mechanics can effectively manipulate information.

We aim to obtain a general model with the key properties of composability and explainability. That is, a model made of independent parts that can be pruned depending on the task, without re-training. Each part of the model should have a meaning, hence the explainability.

Why: cognitive neuroscience

The successful integration of multimodal data, fast computation and the benefits of temporal information is far from just an AI industry problem; nature has a very sophisticated example - our mind.

Humans perceive the world as structures of information and represent it symbolically. Inputs coming from different senses are meaningfully merged and complex, abstract relations are inferred. Temporal (diachronic) conditions are as important as the spatial (synchronic) ones to obtain a complete neuroscience description of perception.

Analogous with the human mind, we know that the ability to extract meaning from data and infer relations among elements dramatically reduces the amount of data required to recognise and understand information. Cognitive models of AI decision making are an example of this.

We believe cognitive neuroscience has a key role to lead us from a narrow AI perspective to a general artificial intelligence philosophy.

Core technology overview

Our platform has been conceived to be a general data manipulation framework.

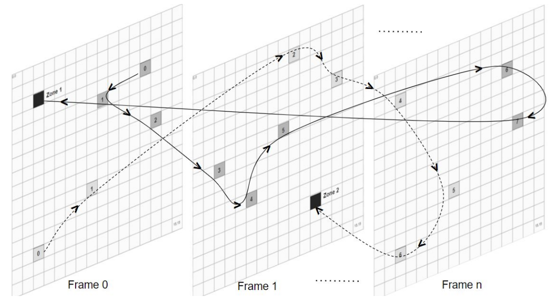
Data input

Information can be carried by different data types, each one with specific properties. The platform has already been proven on images, audio, video and log files.

Input processing is completely data type agnostic, no prior assumption on the modality is embedded to improve the performance. Likewise, no task-related cues are used to empower the representation.

As per the supervised framework, at the training phase there is the need to provide ground truth information.

Remarkably, temporal information is preserved. This means that video frames are not processed as still images like most deep learning counterparts do.



Example of sparse selection of temporal information. Data can be taken over *m* successive entries (here frames).

Data representation

Multimodal or monomodal inputs are non-linearly combined. Vectors describing this combination are the result of a representation learning procedure at training time.

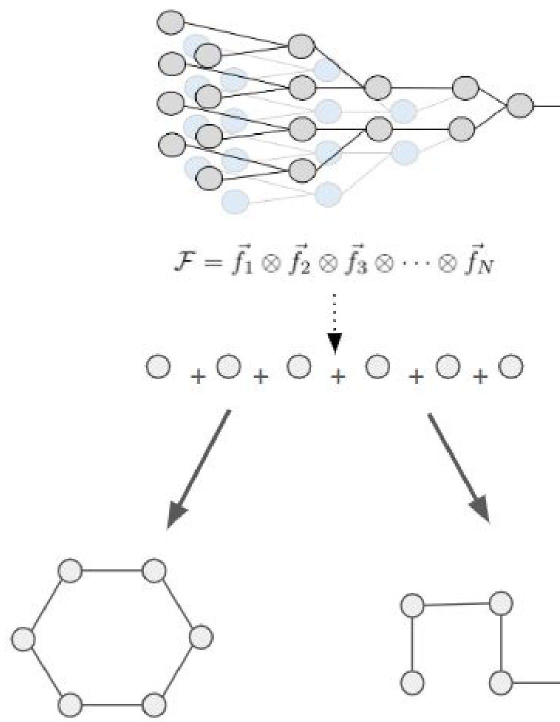
Our data remapping has key properties: each vector is a quantum state encoded in a Riemann vector and represented in the Hilbert space. They are further combined in a tensorial representation.

Tensor networks are generally used for contraction and decomposition of high dimensional complex networks. Likewise, our final model is composable: from the same wider model it is possible to extract sub-models without retraining.

Training

The training process can be delicate and time-consuming in deep learning. We have demonstrated our data remapping can have substantially less parameters involved and a reduced dataset to train on.

The objective of the learning algorithm is to sparsify and linearize the input and contextually select the frequencies best suited for the problem. Finding the optimal solution is a combinatorial problem, but a proprietary heuristic allows investigation of a reduced solution space. We have found a way to avoid a combinatorial explosion without diminishing the strength of the obtainable models.



Pictorial representation of tensor networks methods. From a general model it is possible to derive sub-models.

Use-case examples

Compression

Our platform has been successfully applied to video compression where we were able to reduce the video bitrate by 40% with no quality loss.

Our platform has trained a model of visual complexity. It shows where information can be removed from video without causing noticeable compression artefacts. Temporal data is a key element of video compression.

We developed a new video quality metric that better matches subjective quality scores.

The product is a plugin for video codecs (H264 and H265) and new codecs will be added soon (VP9 and AV1).

Action recognition

Temporal information is naturally suited for action recognition. The implementation of our solution is so light-weight that it can be used on edge components, without the need for cloud computation.

For example, the detection of actions like violence can help videostream platforms to block inappropriate content, limiting or avoiding the use of people in the reviewing process. Thanks to a cpu-only implementation, action detection can run real-time directly on CCTV cameras.

Gaze estimation

Gaze estimation is an important plus for the automotive industry and in every Human Computer Interaction scenario. The interaction through gaze is intuitive - no learning curves - and can augment touch and speech.

Sensor Fusion

The platform is able to handle and combine data from different modalities. It can find an appropriate combination of audio-visual (video) information for a given task. An example is the improvement of speech recognition by combining video and audio in a single model.

Cybersecurity

Cybersecurity is a challenge that is proving hard to solve with current approaches and represents the ideal ground to prove how we perform in a multi-dimensional unsupervised scenario.

Conclusions

The Cogisen platform has been conceived at the cross-section of quantum probability and neuroscience and it is designed to address crucial industry problems in machine learning such as the capturing of temporal data, the speed of models and multimodal integration.