THESIS INTRODUCTION

SPRINT 16

PAPER ONE

Token Shift Transformer for Video Classification

Link to Paper

This paper present Token Shift model (i.e. TokenShift), a novel zero-parameter, zero-flops operators for modeling temporal relations within each transformer encoder. Specifically, the tokenshift barely temporarily shift partial (class) token features back-and-forth across adjacent frames. They uses the model in an encoded plain 2D vision transformer for learning 3D video representation. The tokenshift transformer is a pure convolutional-free transformer plot with computational efficiency for video understanding. The experiments on standard benchmarks to verify its robustness, effectiveness, and efficiency, particularly with impact clips of 8/12 frames, the tokenshift transformer achieves SOTA precision: 79.83% | 80.40% in the kinetics - 400 66.56% on EGTEA-Gazet, and 96.80% on Uaf-101 datasets.

PAPER TWO
IDENTIFIABLE ENERGY-BASED REPRESENTATIONS: AN APPLICATION TO ESTIMATE HETEROGENEOUS CAUSAL EFFECTS.

Link to Paper

Conditional average treatment effects (CATEs) allows us to understand the effect heterogeneity across a large population of individuals. This research is put together because typical CATEs learners assume all confounding variables measured in order for the CATE to be identifiable. Often this requirement is satisfied by simply collecting many variables, at the expense of increased sample complexity for estimating CATEs. To solve this the paper propose an energy-based model(EBM) that learns a low-dimensional representation of the variables by employing a noise constructive loss function. With the EBM they introduce a preprocessing step that alleviates the dimensionality curse for any existing model and learning developed for estimating CATE. The EBM keeps the representations particularly identifiable up to some universal constant, as well as having universal approximation capability to avoid excessive information loss from model misspecification, these properties combined with no loss function, enables the representations to coverage kind keep the CATE estimation consistent.

PAPER THREE

UNIFYING NONLOCAL BLOCKS FOR NEURAL NETWORKS

Link to Paper

The nonlocal-based blocks are designed for capturing long-range spatial-temporal dependencies in computer vision tasks. Even after solving excellent pre performance, they still lack the mechanizm to encode the rich, structured information among elements in an image or video. To theoretically analyse the property of these nonlocal-based blocks this paper provides a new perspective to interpret them, where we view them as a set of graph filters generated on a fully - connected graph. Specifically the chebyshev graph filter was chosen, a unified formation can be derived for explaining and analyse the existing nonlocal - based block. (eg. nonlocal block, nonlocal stage, double attention block). Furthermore, by concerning property of spectral nonlocal block, which can be more robust and flexible to catch long-range dependencies when inserted into deep neural networks than existing nonlocal blocks.

PAPER FOUR
THE AI ECONOMIST: OPTIONAL ECONOMIC POLICY DESIGN VIA TWO-LEVEL DEEP REINFORCEMENT LEARNING

Link to Paper

AI and reinforcement learning has not yet widely adopted in economic policy design, mechanism design, or economic at large. The current economic methodology is limited by a lack of counterfactual data, simplistics behavioural models and limited opportunities to experiment with policies and evaluated behavioural responses. This paper show that machine learning - based economics simulations is a powerful policy and mechanism design framework to overcome these limitations. The AI economist is a two level, deep RL framework that trains both agents and social planner who co-adapt, providing a tractable solution to the highly unstable and novel two-level RL challenges. These two-level deep RL can be used for understanding and as a complement to theory for economic design, unlocking a new computation learning-based approach to understand economic policy

PAPER FIVE
ICECAP: INFORMATION CONCENTRATED ENTITY-AWARE IMAGE CAPTIONING

Link to Paper

Most current image captioning systems focus on describing general image, content and lack background knowledge to deeply understand the image , such as exact named entities or concrete events. This paper focuses on the task which aims to generate information captions by leveraging the associated news articles to provide background knowledge about the target image. To overcome the limitations of the previous works, an information concentrated entity-aware news image captioning (ICECAP) model, which progressively concentrates on relevance textual information within the corresponding news article from sentence level to the world level. The model first create coarse concentration on relevance sentences using a cross-modality retrieval model and then generates captions by further concentration on relevents words within the sentences. Extensive experiments on both Breaking News and Good News dataset demonstrate the effectiveness of the proposed method.