Research Summary – Hieu M. Vu

This document briefly summarises research projects that I've worked on, which I categorise into two sections:

- **Information Extraction from Business Documents**
  - Extracting key information (e.g., names, dates, quantities) from business documents (e.g., invoices, ID cards, contracts)
  - Including works that were done at my industry job.

- **Personal Exploration in Explainability**
  - Some small projects I worked on in my free time.
  - My attempts to expand my breadth of knowledge and satisfy my curiosity.

A more detailed summary is available here.

Additionally, a full list of my publications is available on my Google Scholar.

Information Extraction from Business Documents

Related to this problem are the following publications:

A Span Extraction Approach for Information Extraction on Visually-Rich Documents

DIL-ICDAR 21 (Oral), Best Paper Award

Motivation:

- The need to apply LayoutLM on a low-resource language.
- Traditional span extraction setup only predicts a fixed number of spans.

The work proposed:

- A cross-lingual transfer learning method that combines the LayoutLM weights of a source language with the BERT weights of a target language. The hypothesis was positional features are reusable across languages with similar reading orders. (Fig. 1)
- A recursive relation predicting scheme for multi-span extraction. (Fig. 2)

Jointly Learning Span Extraction and Sequence Labeling for Information Extraction from Business Documents

IJCNN 22 (Oral)

Motivation: Information extraction can be formulated as sequence labelling or span extraction, but

- Sequence labelling is affected by data imbalance when the target text is only a small portion of the document.
- Span extraction cannot extract multiple spans.

The work proposed: (Fig. 3)

- A query-based span extraction scheme that can process multiple queries simultaneously.
- Jointly learning span extraction and sequence labelling on two different branches, then combining the predictions.

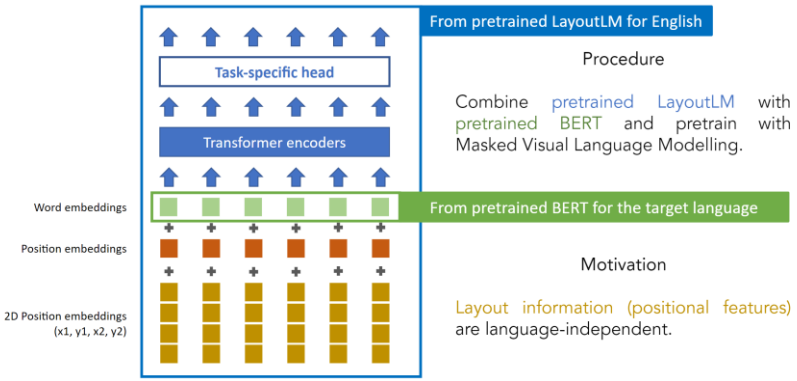

Figure 1: Adapting LayoutLM (English) to Japanese.

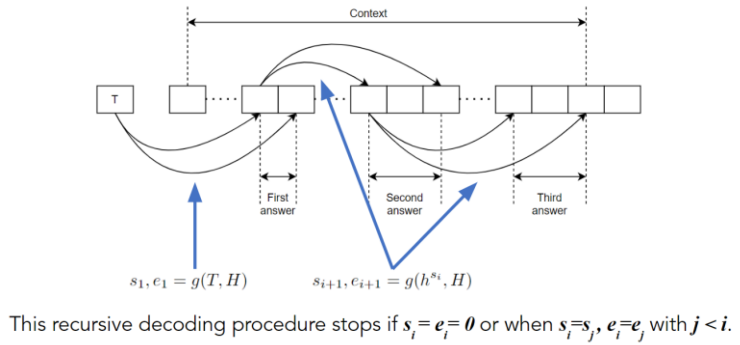

Figure 2: The recursive span extraction mechanism.

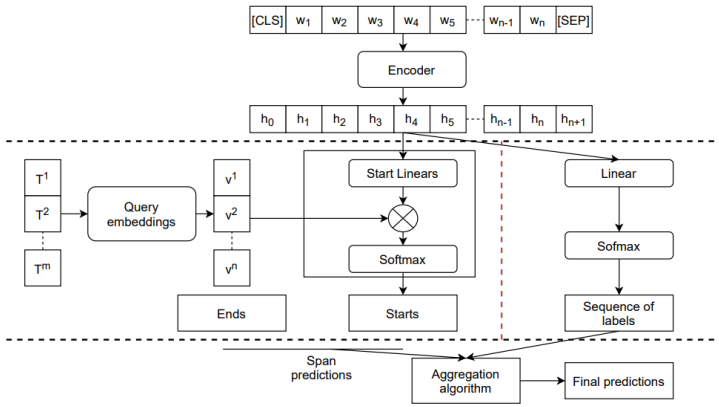

Figure 3: Jointly learning span extraction and sequence labelling.

Personal Exploration in Explainability

Learning Causal Inference and Improving DECAF

Motivation:

- I was curious about Causal Inference. So, I picked up a paper called DECAF and tried to improve it as an exercise during my learning. *DECAF is a method that generates synthetic (and debiased) tabular data using a GAN-based model with an assumed causal DAG.*
- I found an opportunity to improve the data utility (how similar the synthetic data is to the original) of DECAF without sacrificing fairness.

The work proposed:

- A new relaxed DAG (with different relaxation logics) to generate the target attribute.
- Alternating between the DECAF's relaxed DAG and the new relaxed DAG to maximize the information flow to the target attribute while not violating any d-separation constraints.
- Metrics to measure the trade-off between data utility and fairness.

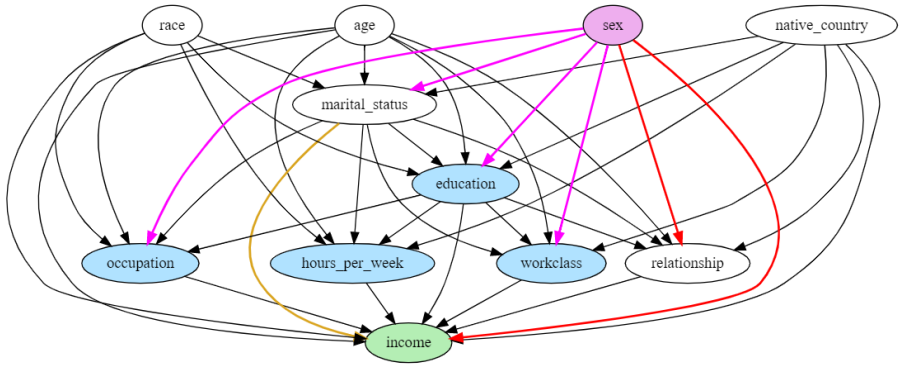

Figure 1: DAG for generating income in the Adult dataset.
Target variable: protected attributes; allowed Conditional Fairness variables.
Demographic Parity is achieved by removing:
Conditional Fairness is achieved by removing:

Inferring Properties of Graph Neural Network and Defending Against Backdoor Attacks

An *under-review journal paper*

Motivation: Information extraction can be formulated as sequence labelling or span extraction, but

- I wanted to gain knowledge on GNNs and backdoor attacks.
- My friend, who was a PhD Candidate, needed help with his research.

The work proposed:

- GNN-Infer - a new method for automatic property inference technique of GNNs. The method comprises 4 stages (Fig. 2).
- Detect and defend against backdoor attacks on GNNs using the method :
  - Infer properties on a poisoned GNN and its training data.
  - Decides which properties are likely to be triggers based on their noticeability, stealthiness and effectiveness.
  - Prune those properties from the GNN.

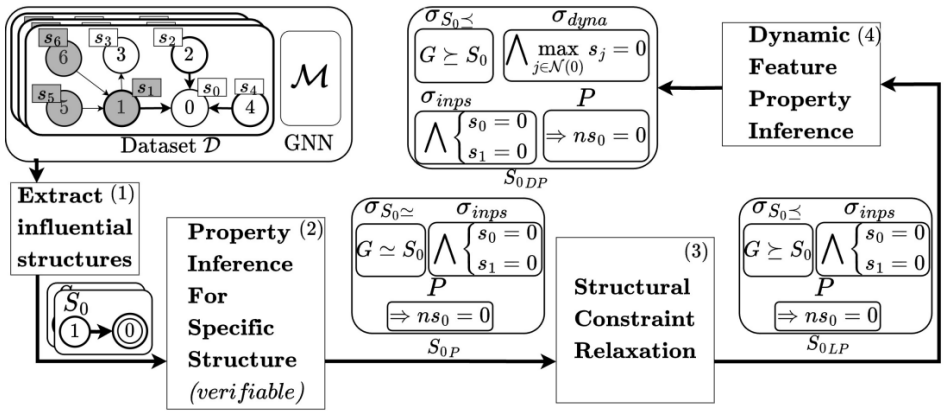

Figure 2: GNN-Infer overview.