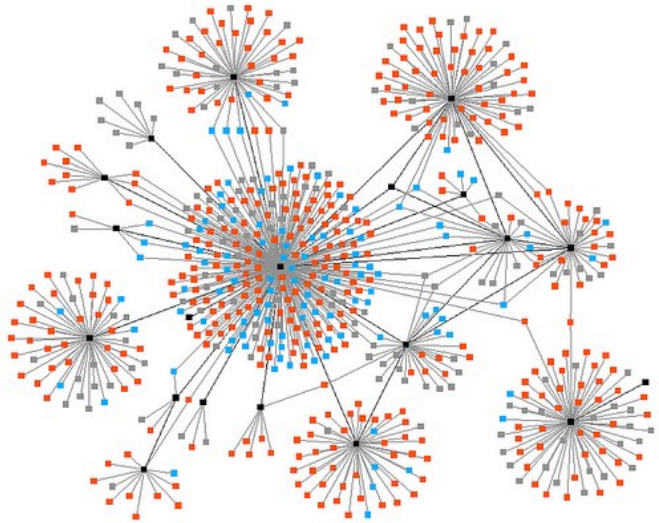
# Learning Sequence Patterns in Knowledge Graph Triples to Predict Inconsistencies

Mahmoud Elbattah, Conor Ryan
mahmoud.elbattah@u-picardie.fr
Université de Picardie Jules Verne (UPJV), France

https://www.insticc.org/Primoris/Resources/PaperPdf.ashx?idPaper=84946

# What is a Knowledge Graph(KG)?

- KGs are large networks of entities, their semantic types, properties, and relationships connecting entities (Kroetsch and Weikum, 2015).
- The concept was reinforced by Google in 2012, which developed a vast KG to process its web queries (Singhal, 2012).
- Other major companies (e.g. Facebook, Microsoft) pursued the same path and created their own KGs to enable semantic queries and smarter delivery of data.

# Motivation: Predicting the Consistency of KB Triples

- KGs are largely constructed by extracting contents using web scarpers, or through crowdsourcing.

- Our goal: Training a classification model to learn the sequence patterns of the triples.

# Data Source Description:

Freebase

- **Freebase:** A huge knowledgebase -> (a structured version of Wikipedia and other data sources)
- Entities (Topics) about people, places, and things.
- Data about: Science & Technology / Arts / Sports/ Society /Products Transportation / Time & Space.
- **≈ 57 million** entities.
- Accessible via data dumps.
- 22 GB / 250 GB (uncompressed).

Bollacker, Kurt, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. "Freebase: a collaboratively created graph database for structuring human knowledge." In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 1247-1250. ACM, 2008.
www.freebase.com

# Freebase Data Model

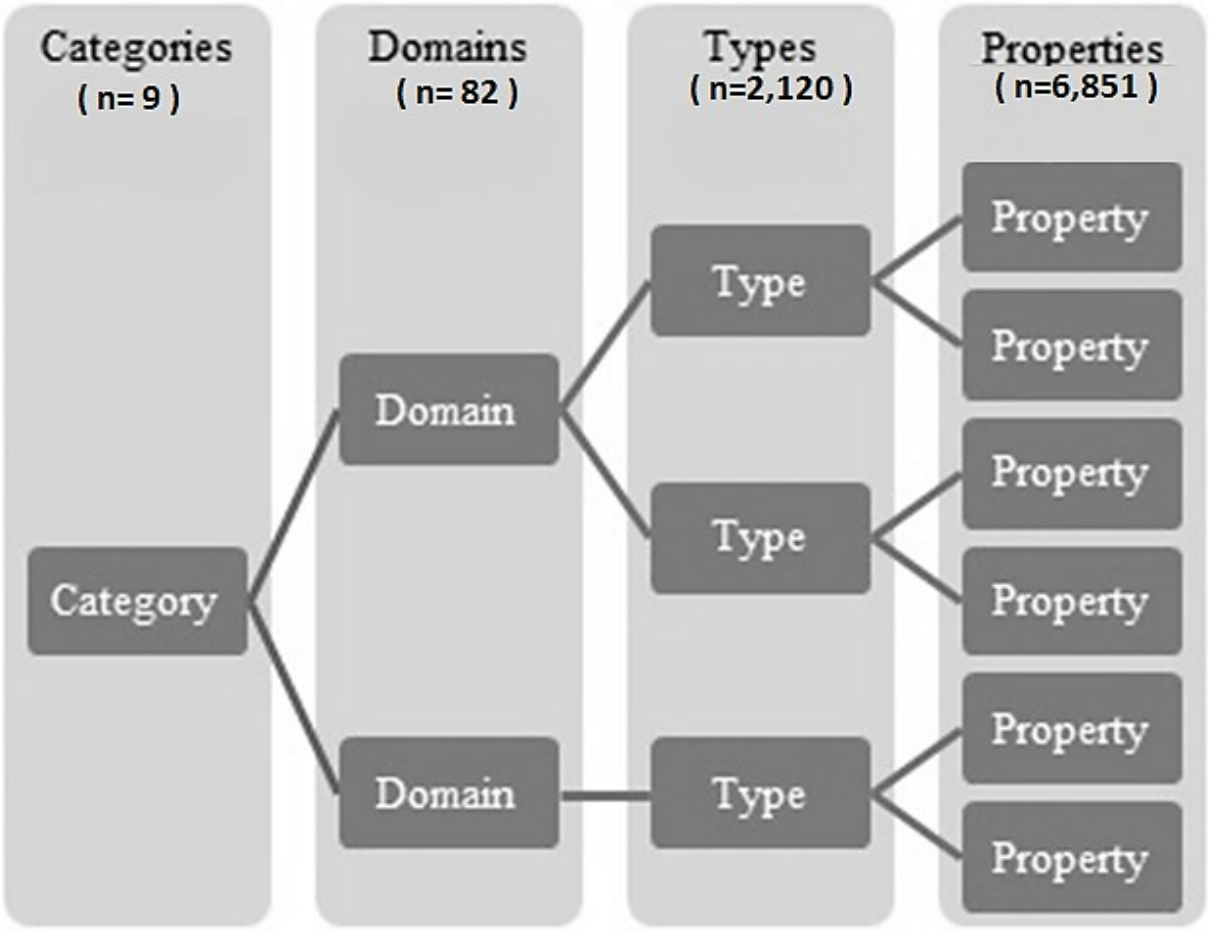
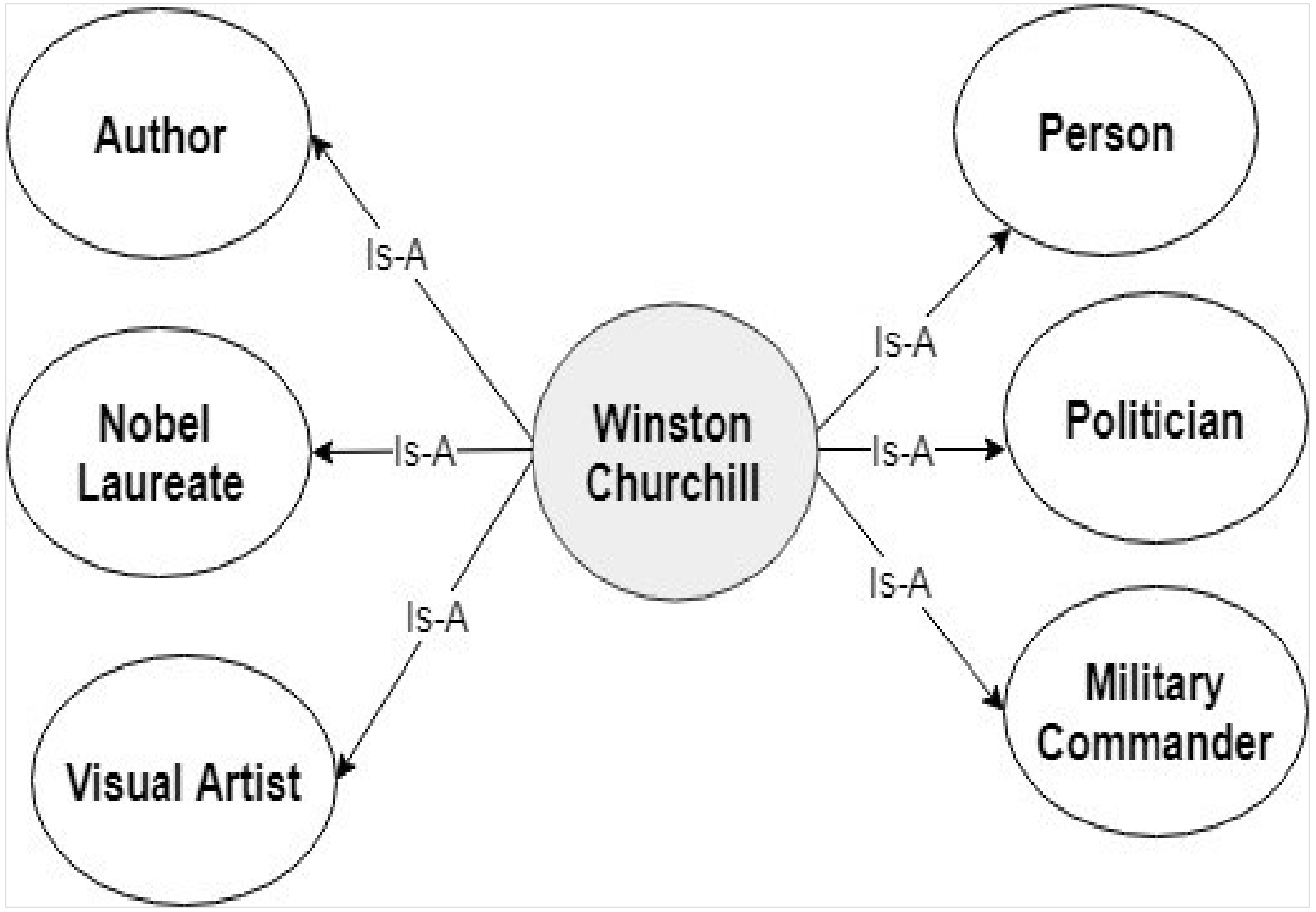



Figure 4: Freebase data model.

# Example


Author
Person
Is-A
Is-A
Nobel Laureate
Is-A
Winston Churchill
Is-A
Politician
Is-A
Is-A
Visual Artist
Military Commander

UNIVERSITÉ de Picardie Jules Verne

# Dataset Statistics

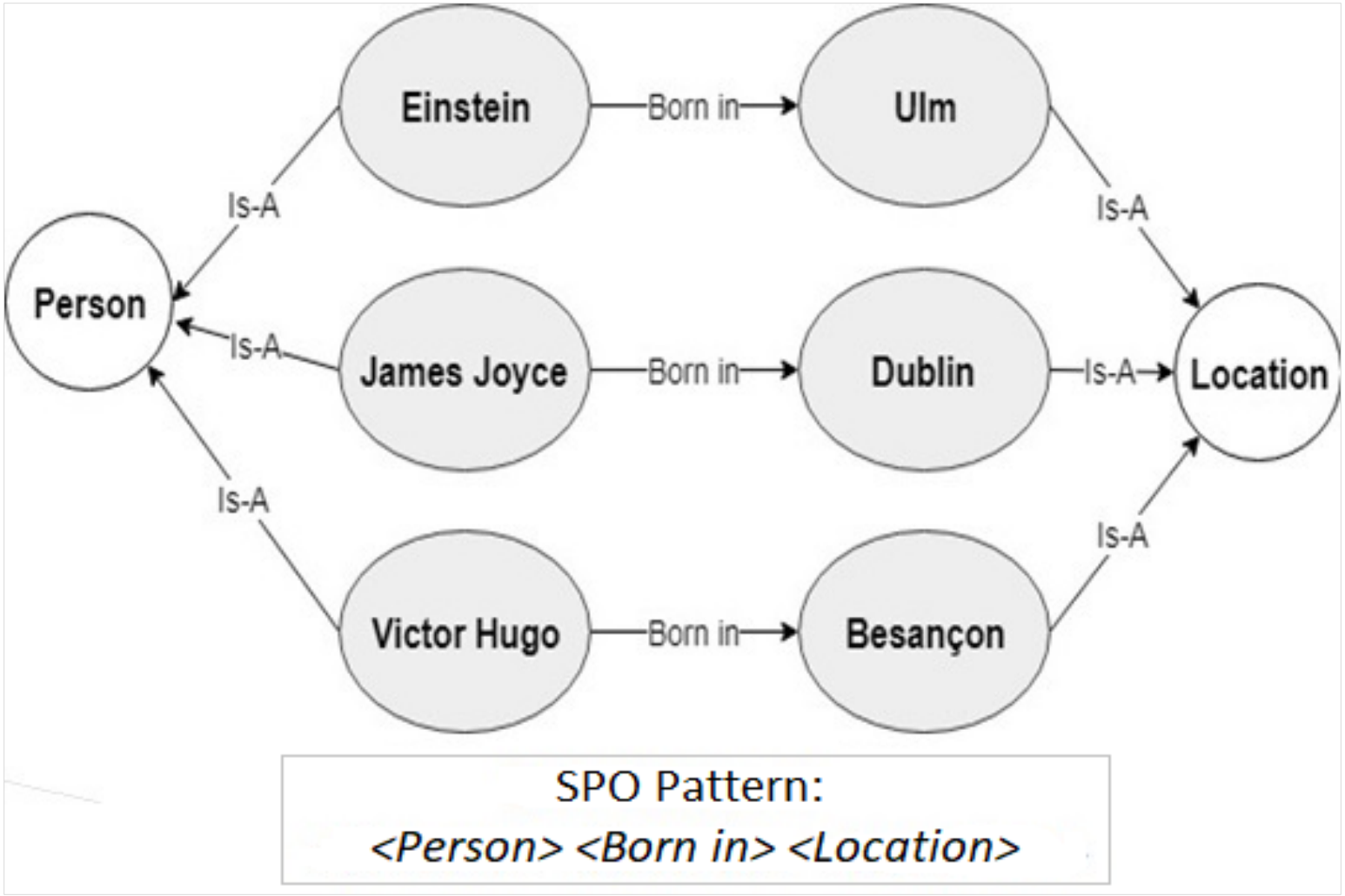
| Freebase Category | Count of Triples |
|---|---|
| #1 Arts & Entertainment | 2.45M |
| #2 Time & Space | 1.5M |
| #3 Society | 650K |
| #4 Science & Tech | 580K |
| #5 Products & Services | 485K |
| #6 Special Interests | 178K |
| #7 Transportation | 116K |
| #8 Sports | 37K |

# Our Approach: Key Ideas

- Key Idea I: Extracting Unique SPO Patterns in KG

- Key Idea II: Generating Syntenic False Patterns

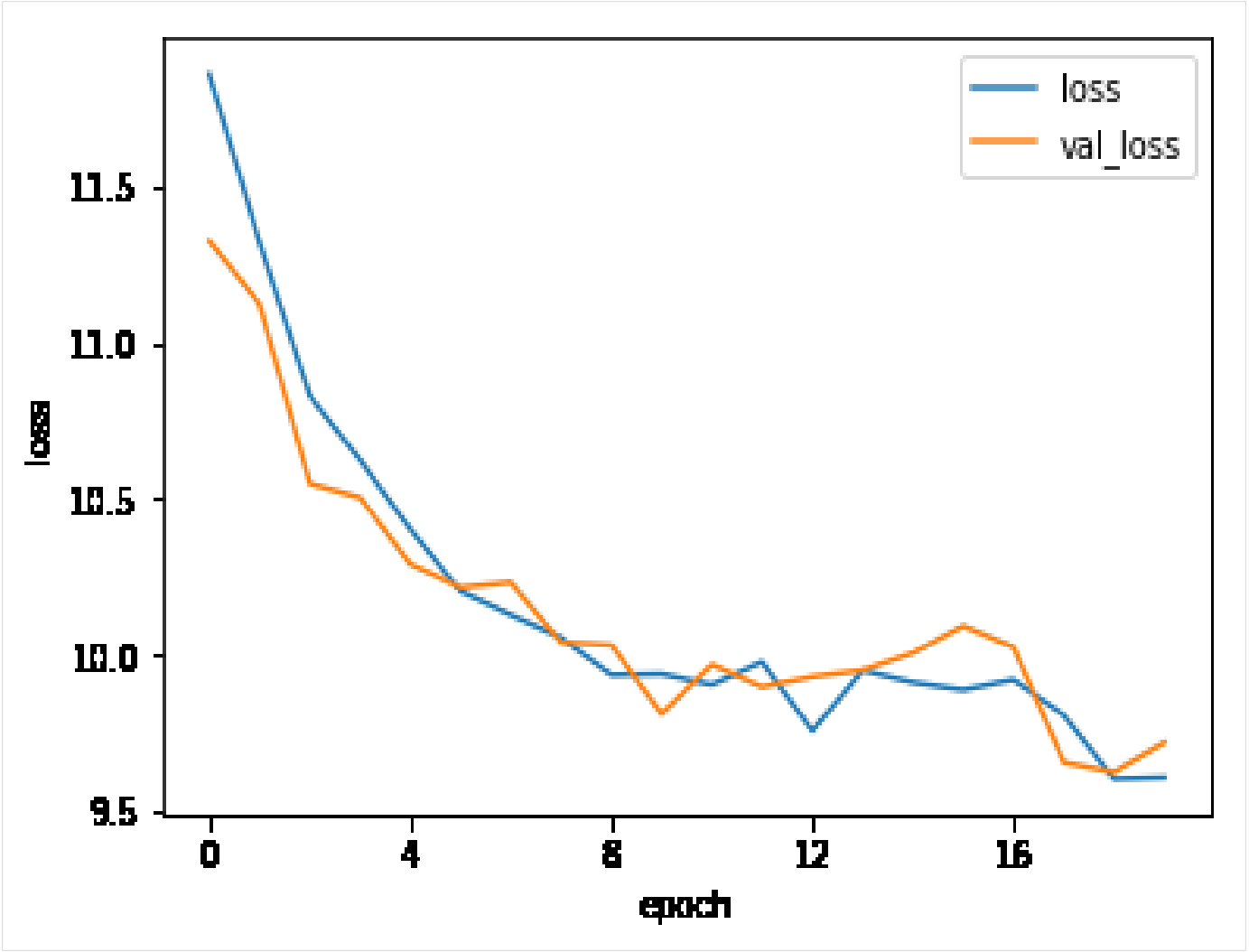
# Example


Einstein
Born in
Ulm
Is-A
Is-A
Person
Is-A
James Joyce
Born in
Dublin
Is-A
Location
Is-A
Is-A
Victor Hugo
Born in
Besançon
SPO Pattern:
<Person> <Born in> <Location>

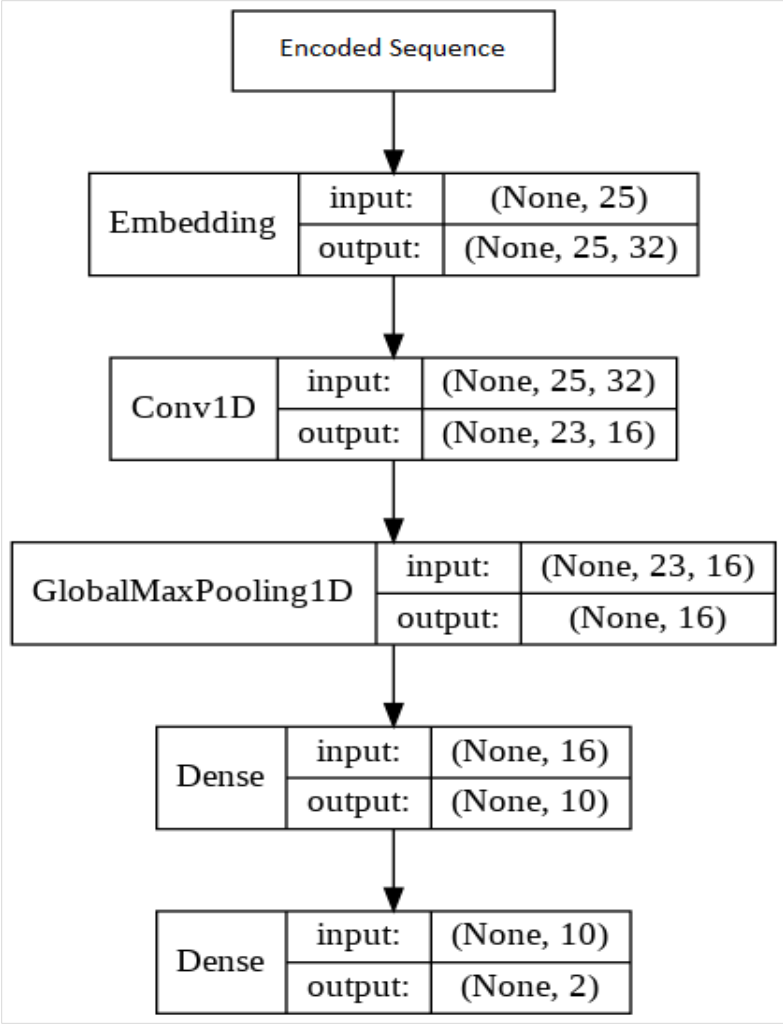
# Experiments

- The dataset initially contained 6M triples, while the unique patterns were only about 124K.
- Data Science Virtual Machine (DSVM) provided by the Azure platform.
- Double GPU-VM (NVIDIA Tesla-K80 GPUs).

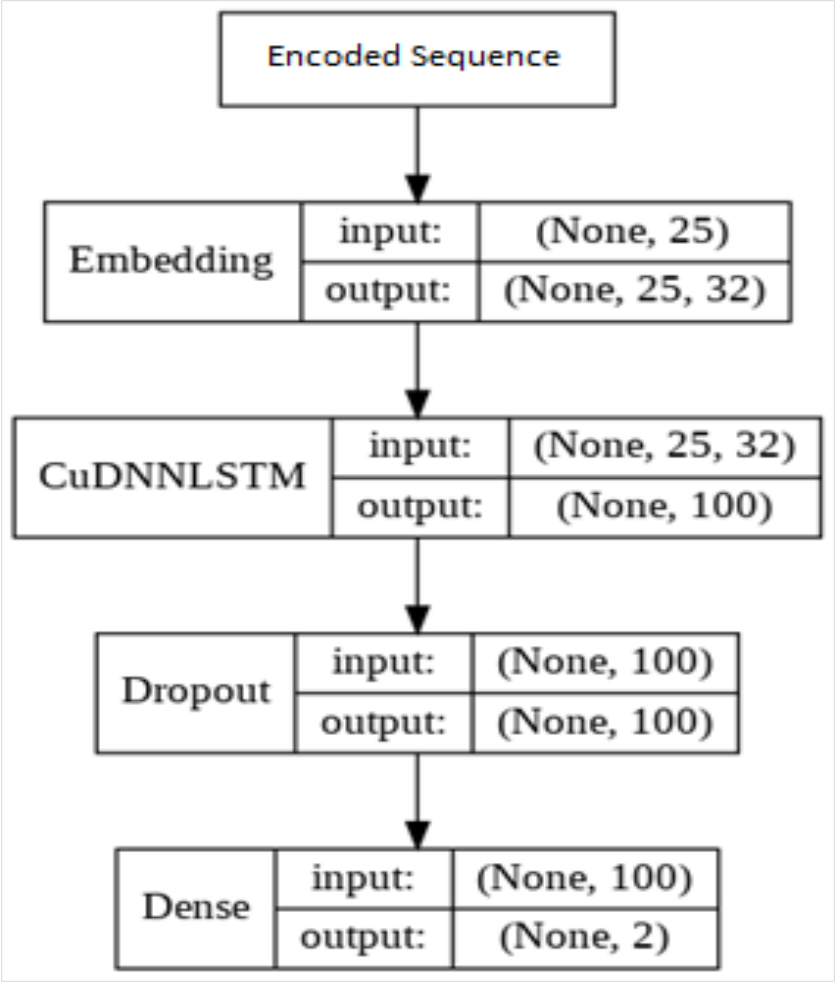
# Generative Model (LSTM)



# Classification Model (CNN)

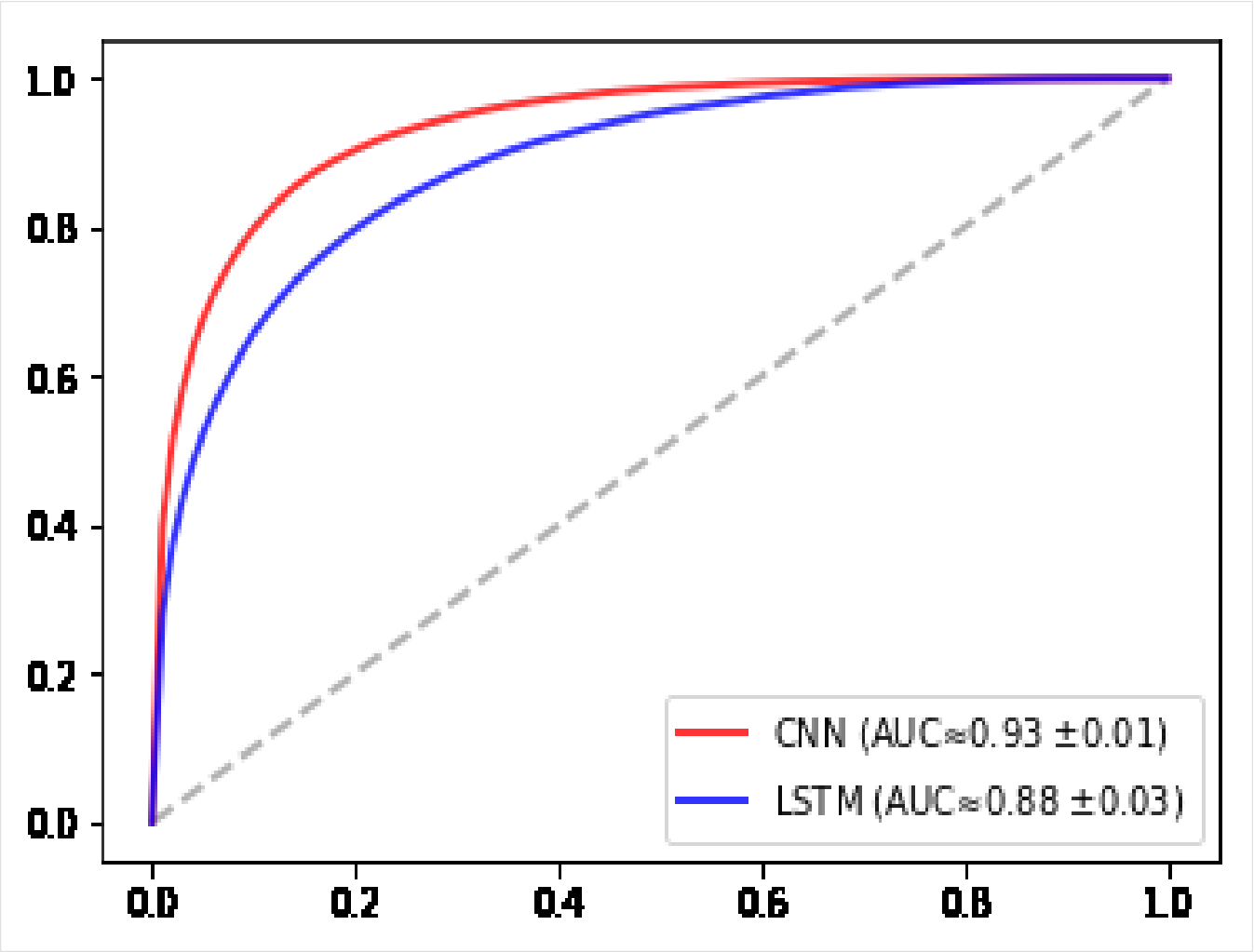


# Classification Model (LSTM)



UNIVERSITÉ de Picardie Jules Verne

# Results (3-Fold CV)



UNIVERSITÉ de Picardie Jules Verne

# THANK YOU!