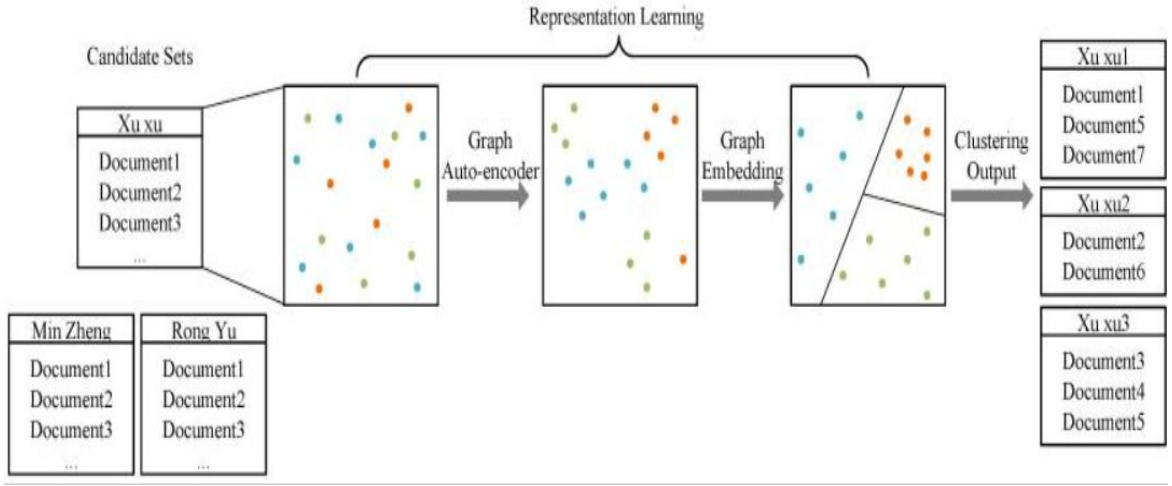
Assignment 3 Data Science In Practice
Instructor : Dr Parthiban Srinivasan

Topic : Author Name Disambiguation , a literary review of data science techniques deployed

Rita Abani 19244
Department of EECS, IISER-B

Anushka Shreyam 20050
Department of Data Science, IISER-B

Case study 1: Representation Learning through Graph Neural Networks, Ma et al.



Representation learning works by reducing high-dimensional data to low-dimensional data, making it easier to discover patterns and anomalies while also providing a better understanding of the data's overall behaviour.

Graph Neural Network is a type of Neural Network which directly operates on the Graph structure. A typical application of GNN is node classification. Essentially, every node in the graph is associated with a label, and we want to predict the label of the nodes without ground-truth.

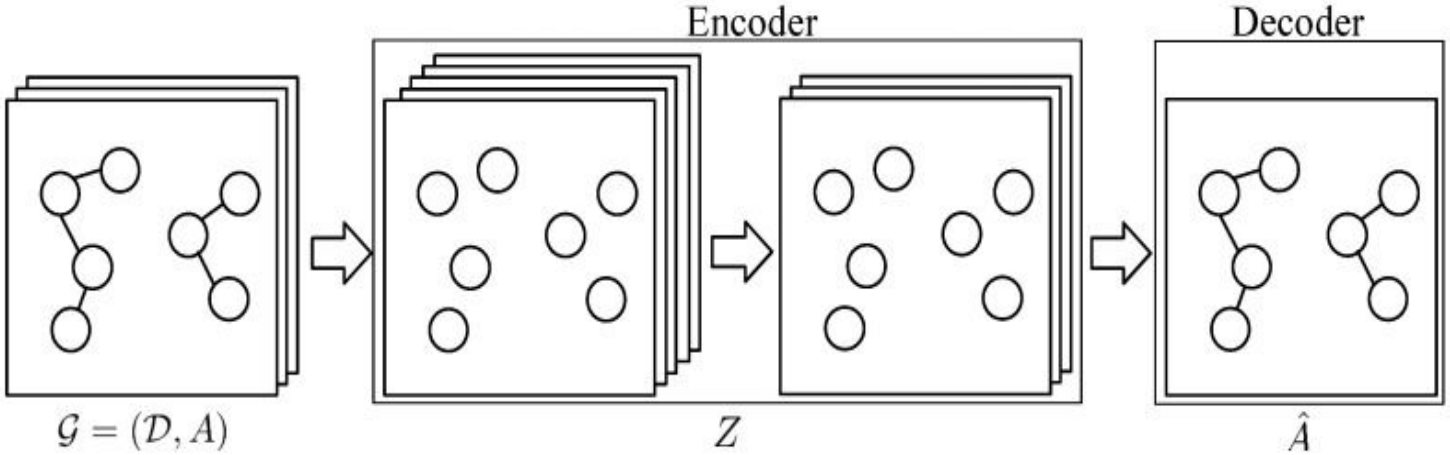
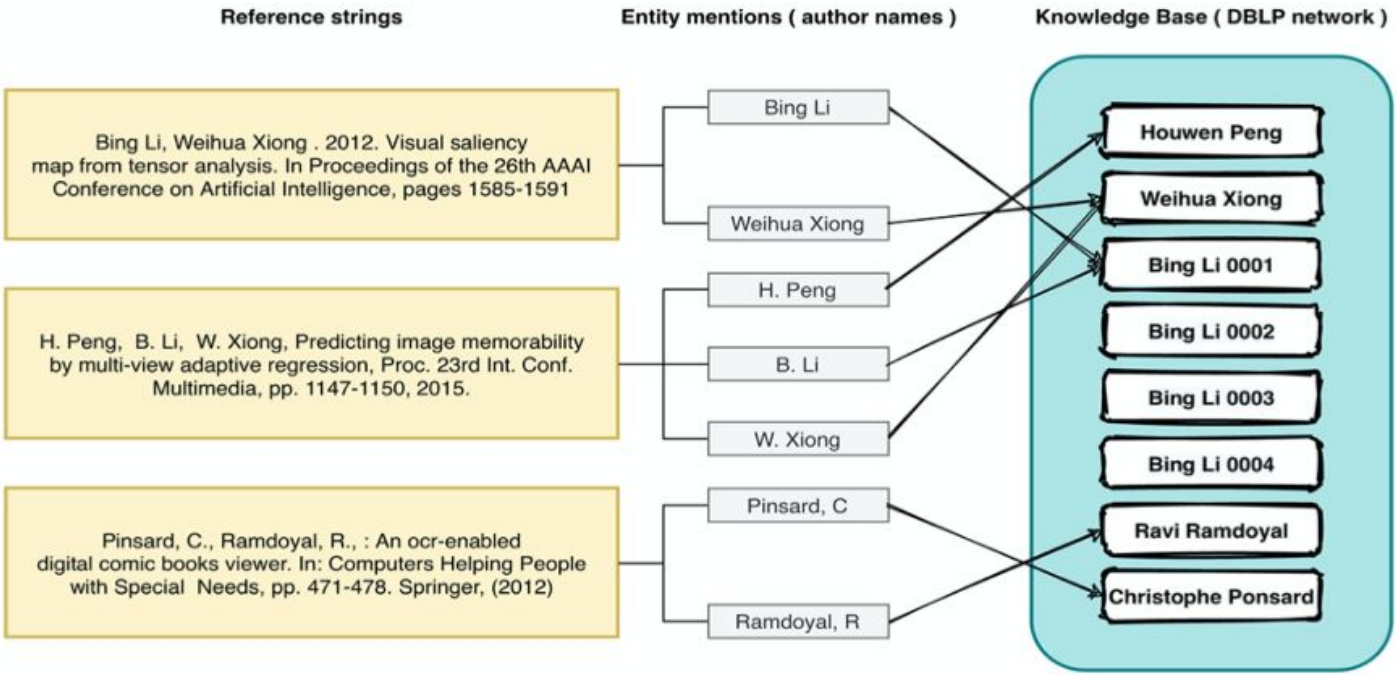
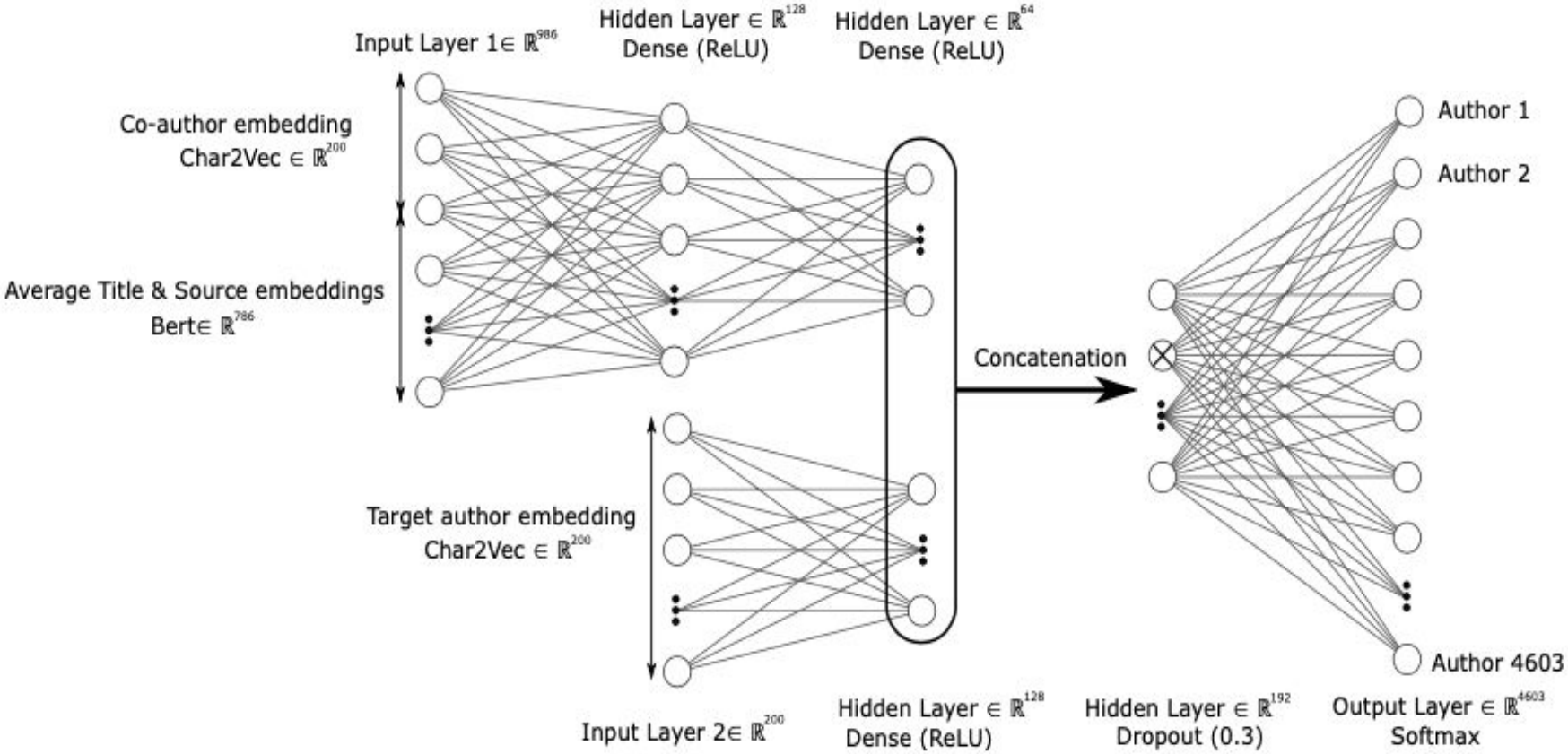


Figure 2 : Variational Graph Auto-Encoder is used in the model to improve the generalization ability of the model. Let X=[dT1,dT2,…] be the representation matrix of documents associated with an author name. The encoder is a two-layer graph convolutional network. D={d1,d2…,dn} is the document representation output Di. The document embedding vectors could represent graph nodes and the adjacency matrix A here represent edges between these nodes.

Case Study 2 : Deep Learning Approach for Author Disambiguation using Bibliographic Data




Input Layer 1∈ ℝ^986
Hidden Layer ∈ ℝ^128
Dense (ReLU)
Hidden Layer ∈ ℝ^64
Dense (ReLU)
Co-author embedding
Char2Vec ∈ ℝ^200
Average Title & Source embeddings
Bert∈ ℝ^786
Concatenation
Author 1
Author 2
Author 4603
Target author embedding
Char2Vec ∈ ℝ^200
Input Layer 2∈ ℝ^200
Hidden Layer ∈ ℝ^128
Dense (ReLU)
Hidden Layer ∈ ℝ^192
Dropout (0.3)
Output Layer ∈ ℝ^4603
Softmax

Case study 1

Case study 2

Table 4

Clustering Results of Each Component.

	Prec	Rec	F1
Feature Embedding	72.29	50.14	59.21
Graphh Auto-Encoder	75.53	58.01	65.62
Graph Embedding	77.71	54.46	64.04
Overall	78.10	67.47	72.40

	Macro	Micro
Precision	0.989	0.975
Recall	0991	0.975
F1-Score	988	0.975