Admin

- Teaching evaluation is closing soon, do it on NuKu
- Project full code due in week 6
- Project marking in week 7
- Time management
  - Research, Open ended,
  - Block some days for project

Today

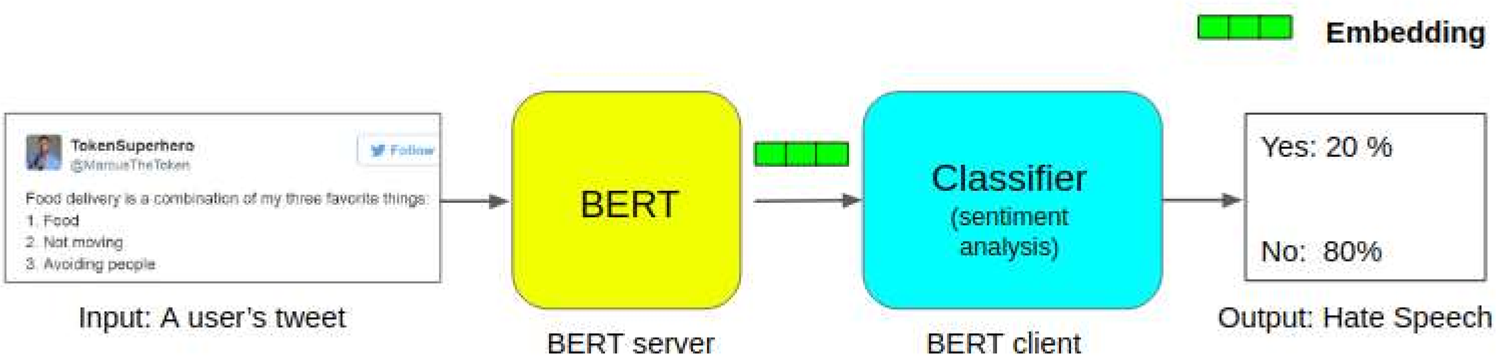
- Last lecture on the first half of this course
- BERT for text classification
- Some of my thesis students' projects
  - Representation, Multi-view representation
  - Learning the network architecture and meta parameters

BERT and text classification:

- The pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of NLP tasks.

Typically,

- Each sentence is changed to a vector
- List of sentences is changed to a matrix, can be directly fit into any classifier.
- Many extensions: Sentence-BERT is much faster



BERT for text classification

- Use CLS token as sentence representation
- Use the output of the final encoder layer
- Use the output of any encoder layer
- Use the pretrained token embeddings from BERT

- You may try simple classifiers
  - LR
  - MLP

- Or CNN

Python implementation

- Code from Tobias Tuan Ha
- Sentence embedding using RoBERTa
- https://colab.research.google.com/drive/13ZKL3b18j3IvLXtGqYY4VuGIwRnUFKmL

- BERT for text classification tutorial (with IMDB)
  - https://www.tensorflow.org/tutorials/text/classify_text_with_bert
  - BERT encoder as a kerasLayer

- More recent tutorial
  - https://curiousily.com/posts/sentiment-analysis-with-bert-and-hugging-face-using-pytorch-and-python/

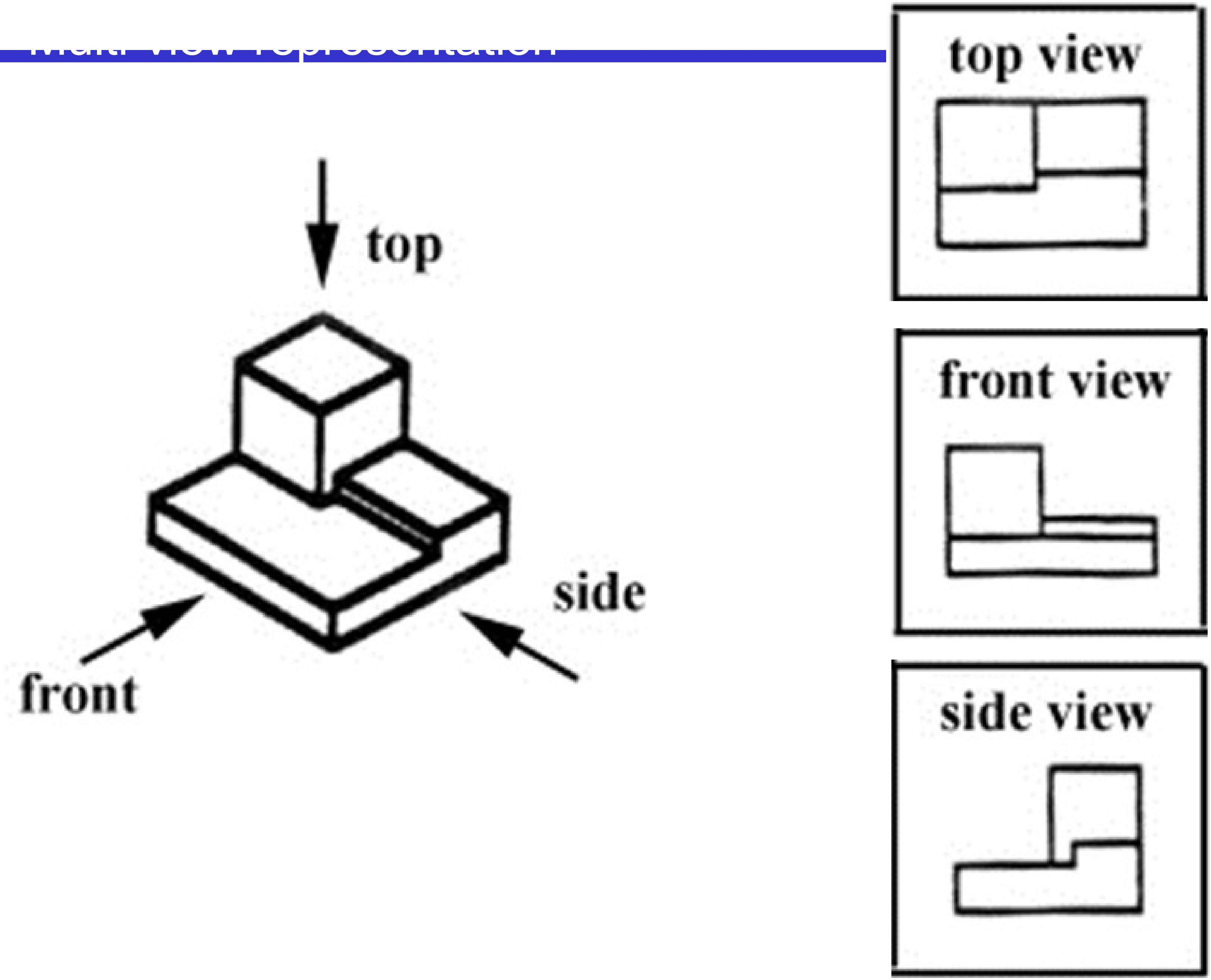
Features (Representation)

mass	width	height	color_score	label
192	8.4	7.3	0.55	1
180	8.0	6.8	0.59	1
86	6.2	4.7	0.80	2
176	7.4	7.2	0.60	1
90	7.1	5.6	0.75	2

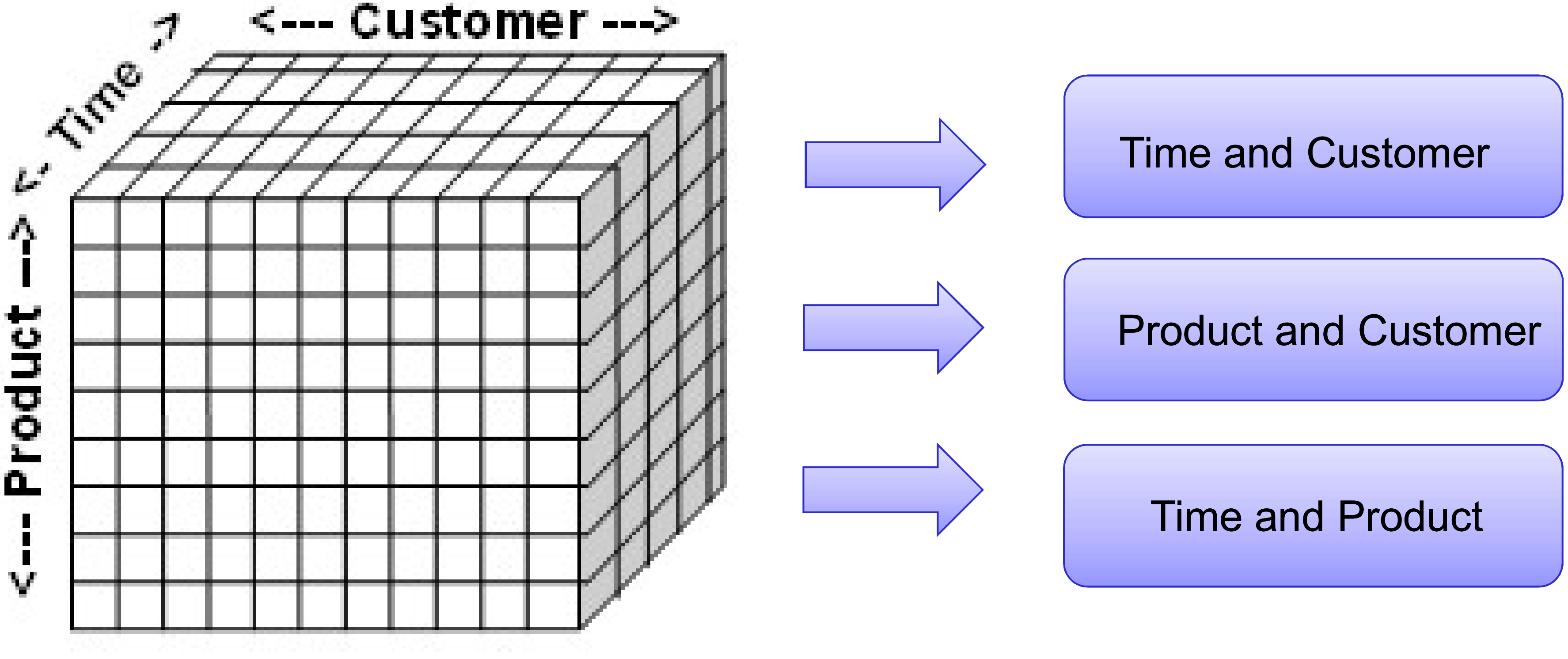
Text representation

- TF.IDF
- Word embedding
- Multiple sense embedding
- Sentence representation
- Higher level
  - Document representation
- Multi-view representation

Multi-view representation



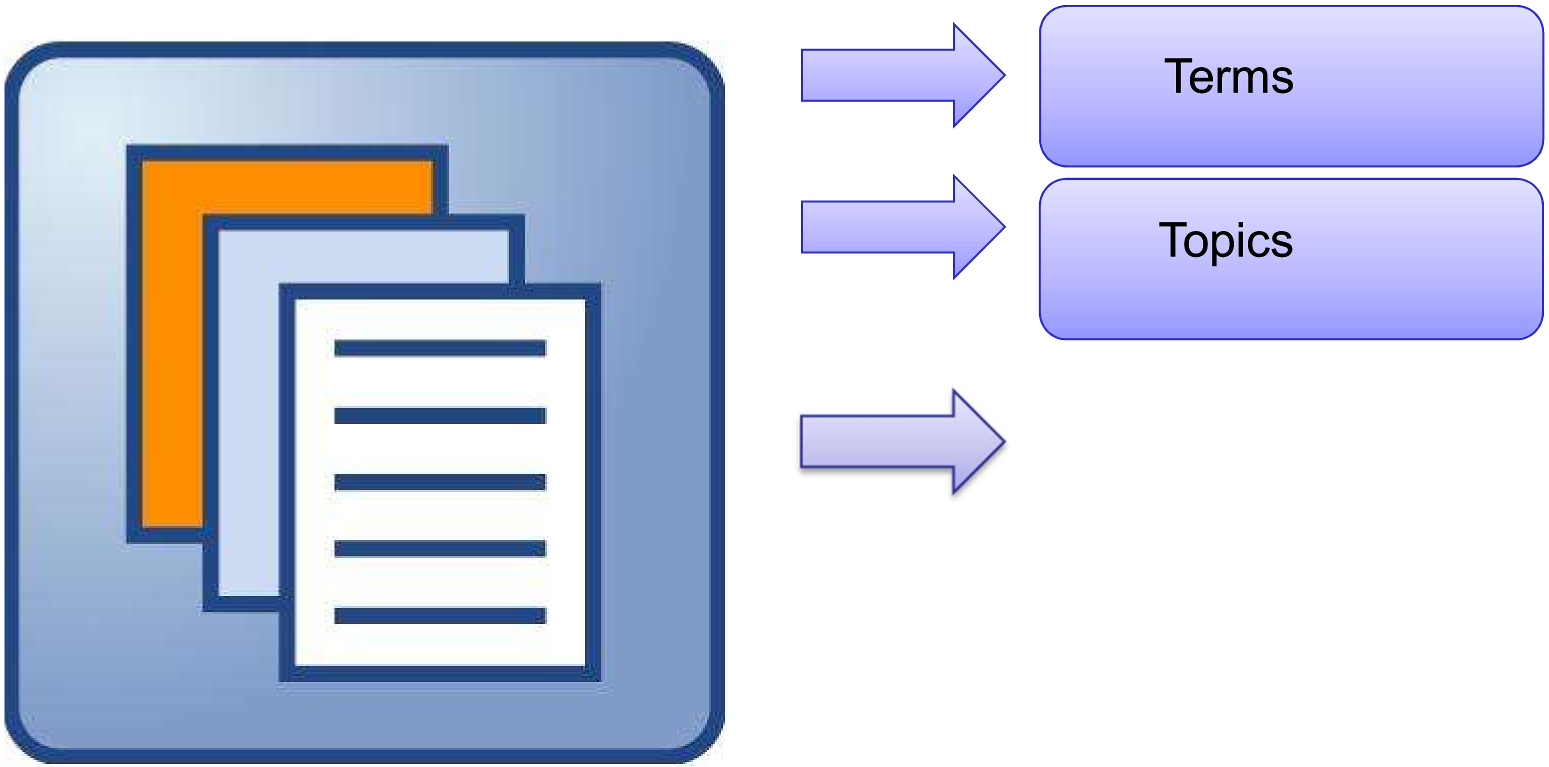
Data Representation



Data

Views

Multiple Views of Documents



Abdul Wahid, Xiaoying Gao, Peter Andreae:
Multi-view clustering of web documents using multi-objective genetic algorithm. IEEE Congress on Evolutionary Computation 2014: 2625-2632

Ontology based representation (Domain knowledge)

An example:

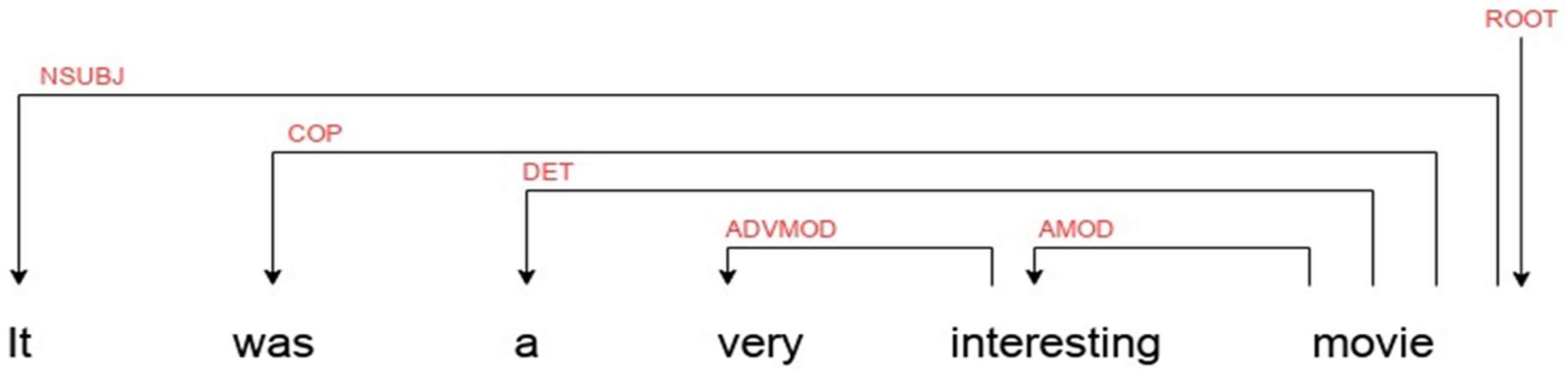
- *"Hyperlipidemia: The patient's Lipitor was increased to 80 mg q.d. A progress note in the patient's chart from her assisted living facility indicates that the patient has had shortness of breath for one day."*

Detected Phrases	Extracted Concepts	Scientific Name	Selected
hyperlipidaemia	[Disease or Syndrome]	hyperlipidemia	√
	[Finding]		×
80	[Quantitative Concept]	80	×
shortness of breath	[Sign or Symptom]	Dyspnea	√
	[Clinical Attribute]		×
	[Intellectual Product]		×
one day	[Temporal Concept]	One day	×

Mahdi Abdollahi, Xiaoying Gao, Yi Mei, Shameek Ghosh, Jinyan Li, Michael Narag: Substituting clinical features using synthetic medical phrases: Medical text data augmentation techniques. Artif. Intell. Medicine 120: 102167 (2021)

Term dependency

Dependency Parse Trees



Dependency based word embeddings: context contains the words with dependency, may be far away in a sentence.

Normal word embedding uses a sliding window of fixed size to get the context.

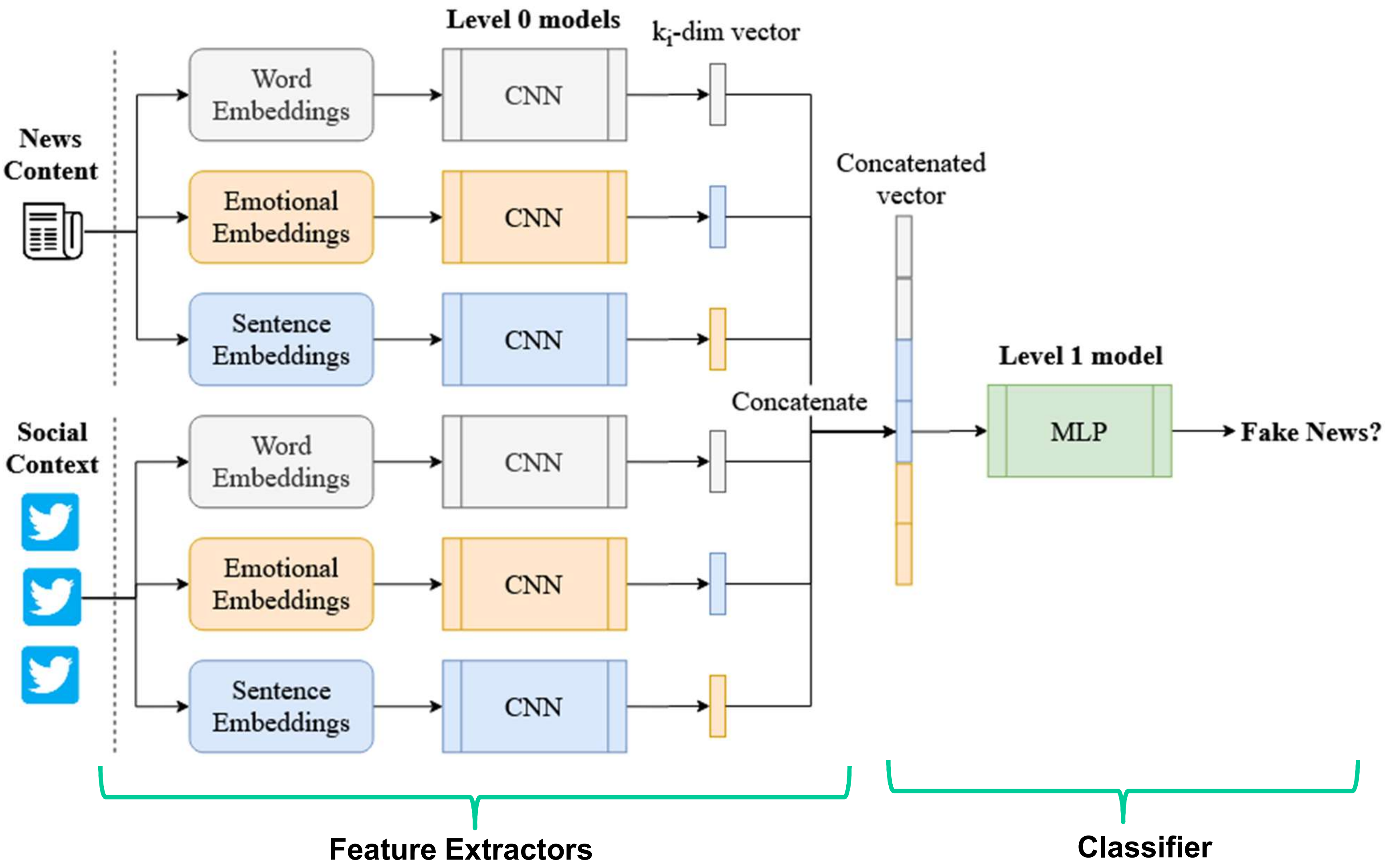
Kosisochukwu Judith Madukwe, Xiaoying Gao, Bing Xue:
Dependency-Based Embedding for Distinguishing Between Hate Speech and Offensive Language. WI/IAT 2020: 860-868

Emotion embedding

- NRC Word-Emotion Association Lexicon (EmoLex) [Mohammad2013]

	anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust
I	0	0	0	0	0	0	0	0	0	0
love	0	0	0	0	1	0	1	0	0	0
this	0	0	0	0	0	0	0	0	0	0
movie	0	0	0	0	0	0	0	0	0	0

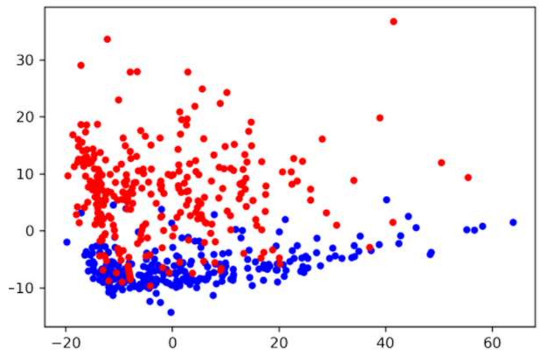
Multi-view representation for fake news detection



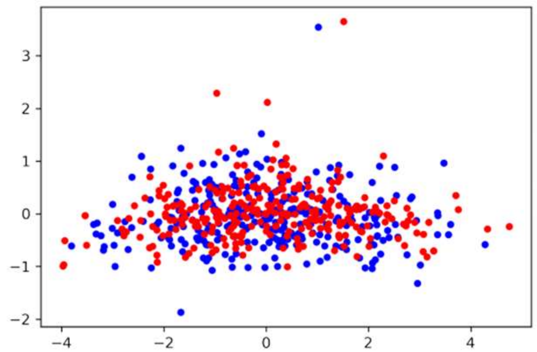
Tuan Ha, Xiaoying Gao: Fake News Detection Using Multiple-View Text Representation. PRICAI (2) 2021: 100-112

Comparison of Features

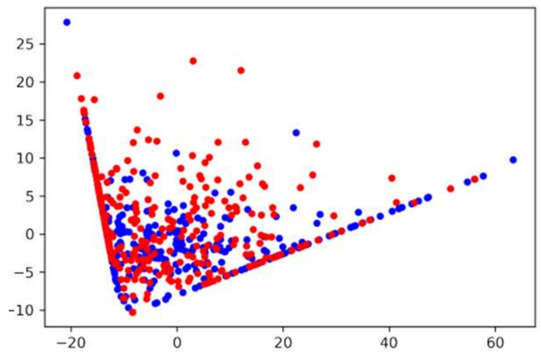
- Projection of text representations (Red: fake news, Blue: real news)



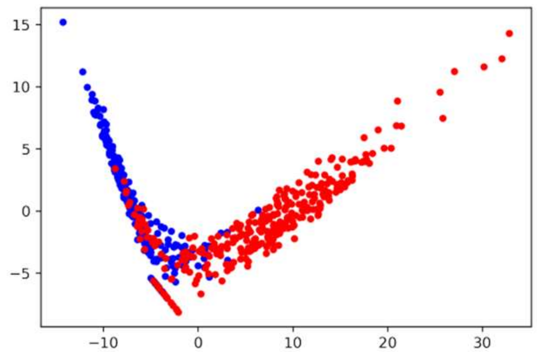
(a) Multiple-view Text Representation.



(b) Word-level features.



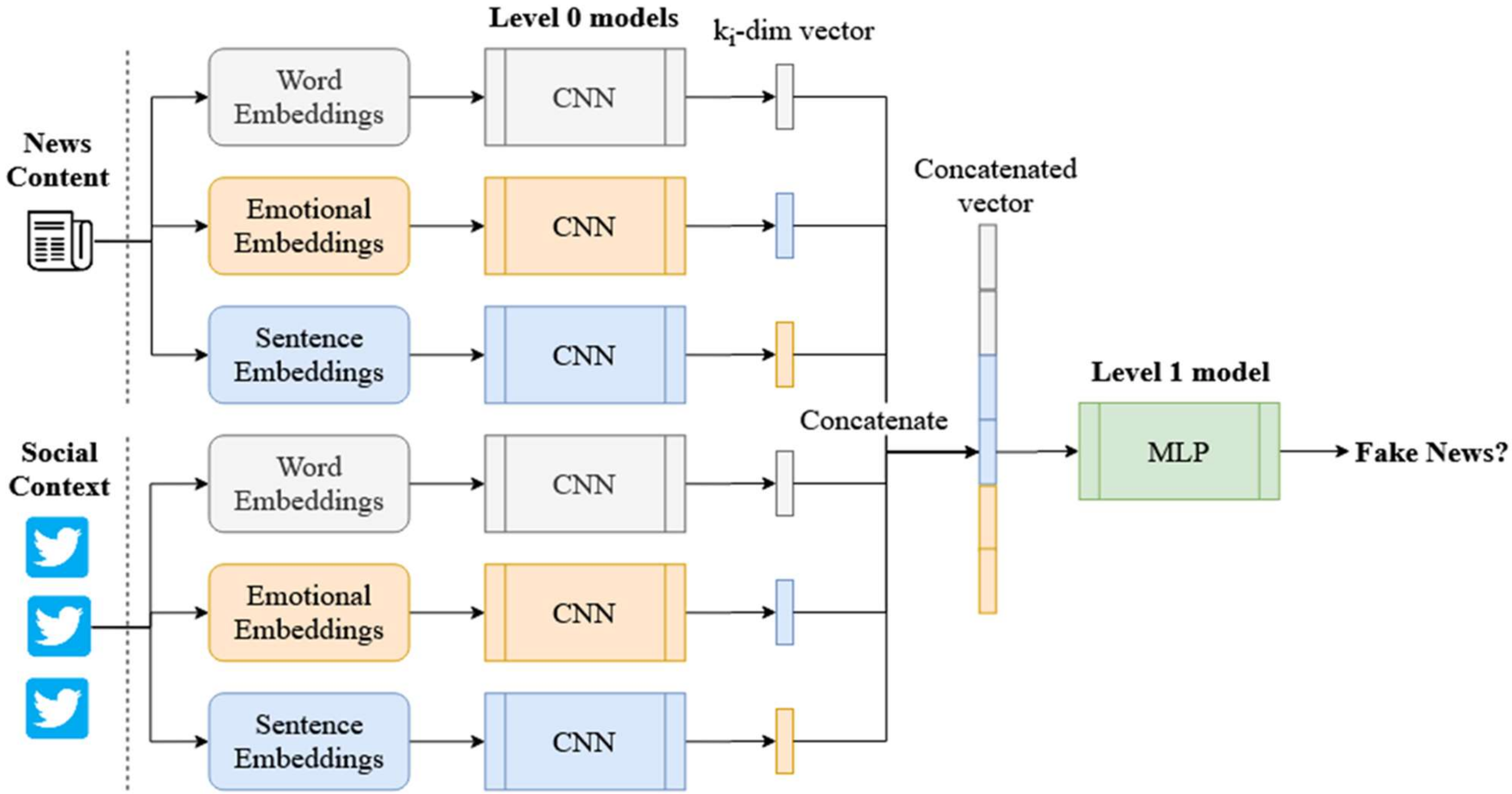
(c) Emotional features.



(d) Sentence-level features.

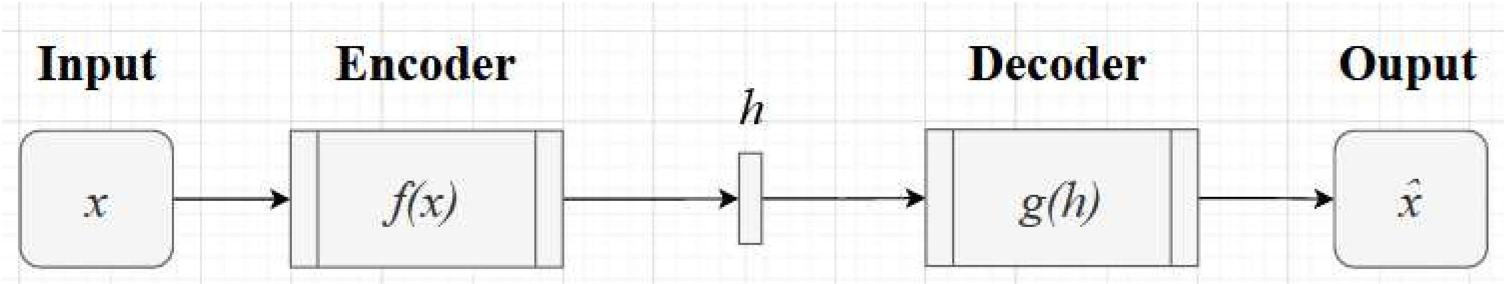
How to learn the multi-view representation?

Tobias Tuan Ha and Xiaoying Gao Evolving Multi-view Autoencoders for Text Classification, WI-IAT21,



Auto-encoders

- A class of neural networks, which are trained to attempt to approximately copy their input x to their output $\hat{x}$.

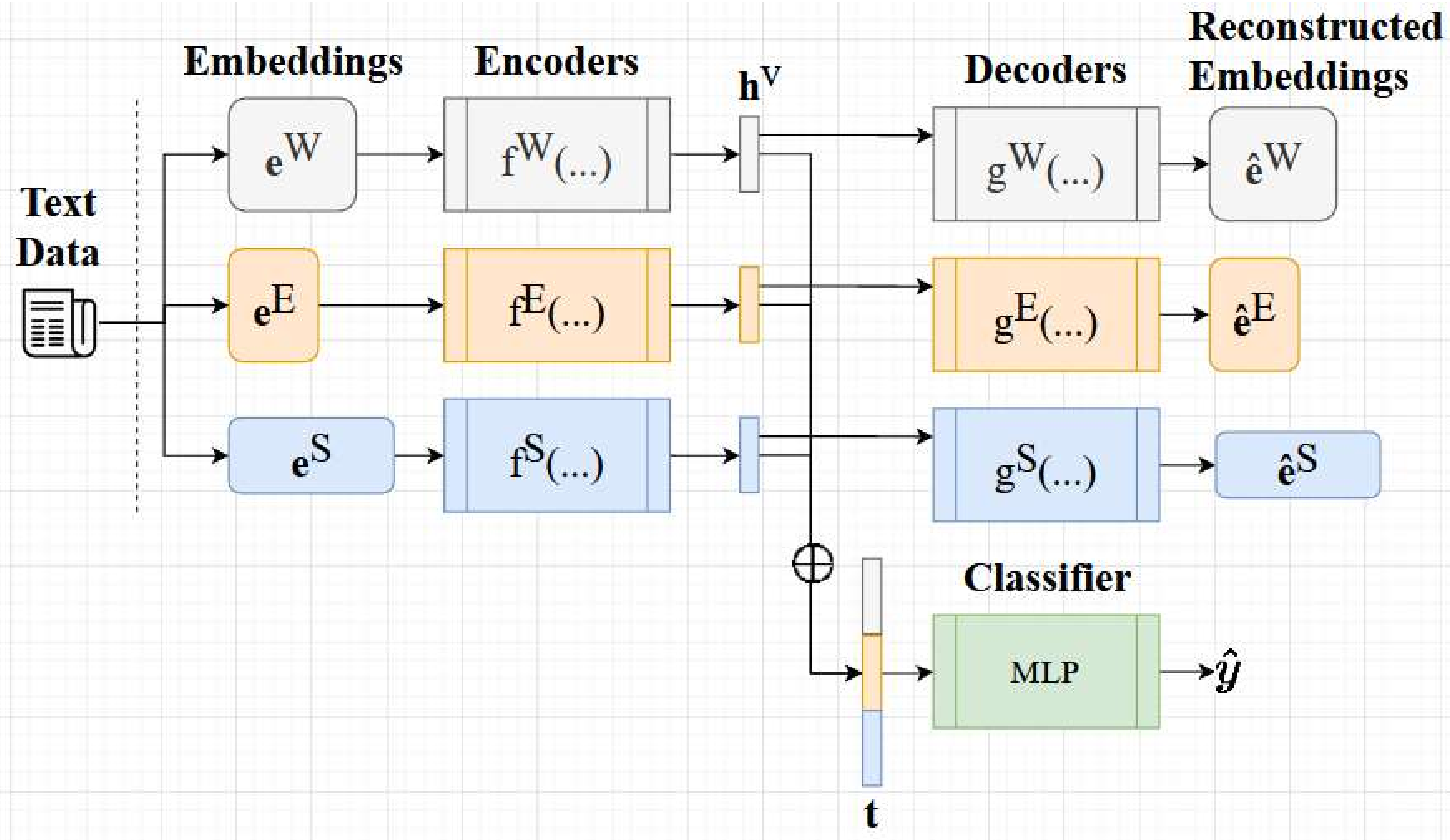


With the simplest Autoencoder, the learning process is described as minimising the loss function

$$L(x, g(f(x)))$$

where L is a loss function such as Mean Squared Error (MSE)

Autoencoder to learn multi-view representation



Loss function for learning simultaneousl

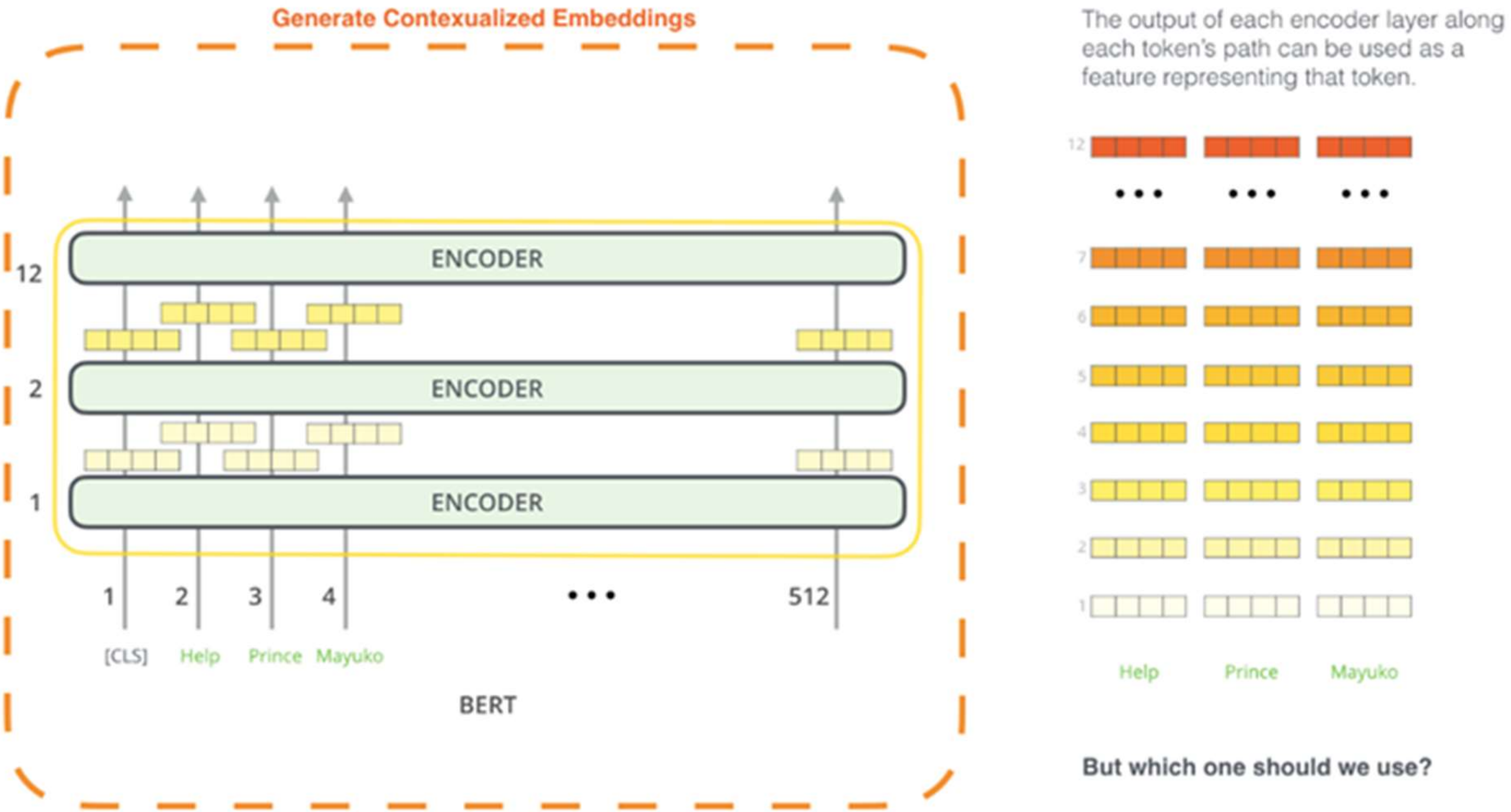
- In order to train the Autoencoders and the classifier, we introduce an objective function containing two types of losses:
  - the Mean Squared Errors (MSE) as the reconstruction errors of the Autoencoders, and
  - the Cross Entropy as the loss of the classifier.

$$\frac{1}{m}\sum_{i=1}^{m}\left(\alpha\frac{1}{|V|}\sum_{v\in V}(\boldsymbol{e}_i^v - \hat{\boldsymbol{e}}_i^v)^2 + \beta(-\sum_{j=1}^{c}(\boldsymbol{y}_i^j \log(\boldsymbol{p}_i^j)))\right)$$

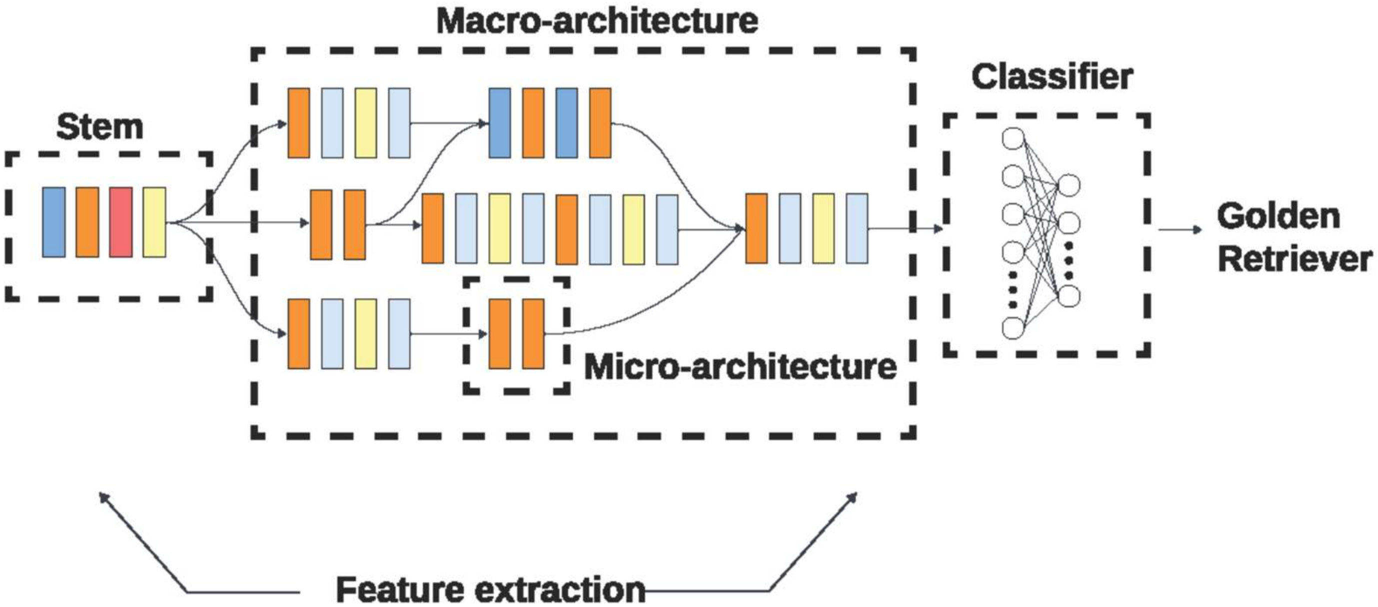
Leaning the network architecture: NAS

- Learning methods
  - Evolutionary Computation
    - Genetic algorithms
    - Genetic programming
- Character level model
  - Trevor Londt, Xiaoying Gao, Peter Andreae:Evolving Character-Level DenseNet Architectures Using Genetic Programming. EvoApplications 2021: 665-680
- Word level model
  - Hayden Andersen, Sean Stevenson, Tuan Ha, Xiaoying Gao, Bing Xue:
  - Evolving Neural Networks for Text Classification using Genetic Algorithm-based Approaches. CEC 2021: 1241-1248
    Hayden Andersen, Xiaoying Gao, Bing Xue, Mengjie Zhang:
    Evolving network structures for text classification using genetic algorithms. GECCO Companion 2020: 109-110
- BERT based model
  Kosisochukwu Judith Madukwe, Xiaoying Gao, Bing Xue:
  A GA-Based Approach to Fine-Tuning BERT for Hate Speech Detection. SSCI 2020: 2821-2828

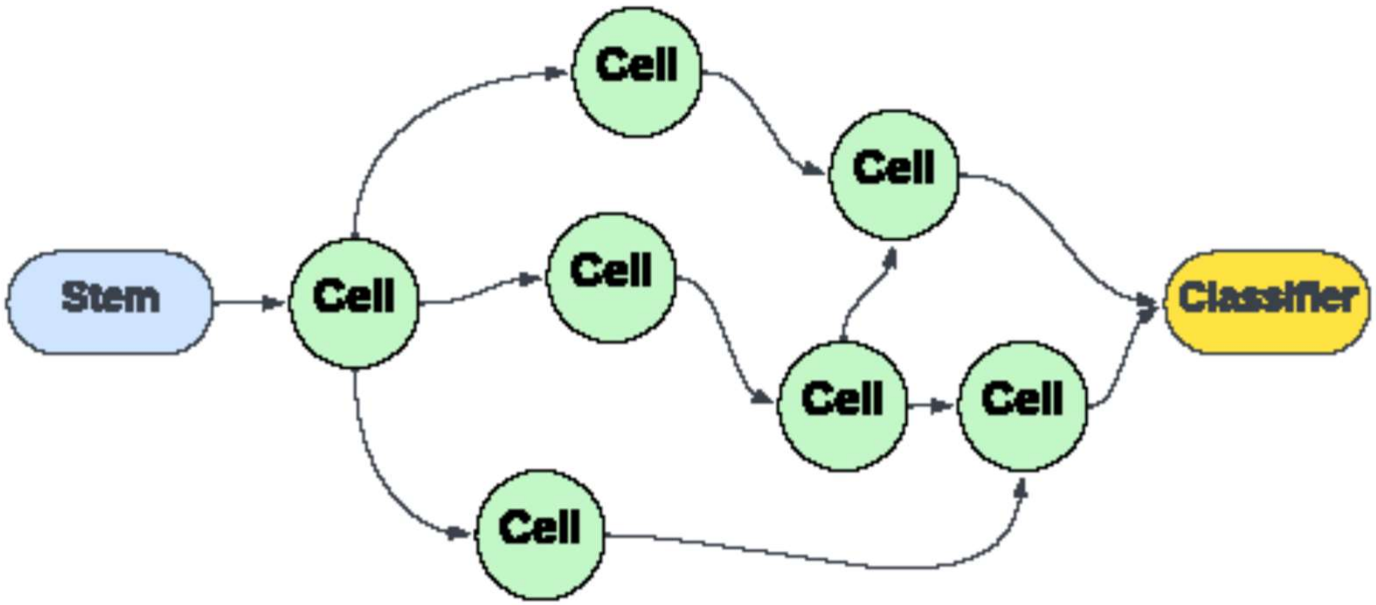
CLS token, Layers, which layer?



CNN architecture search



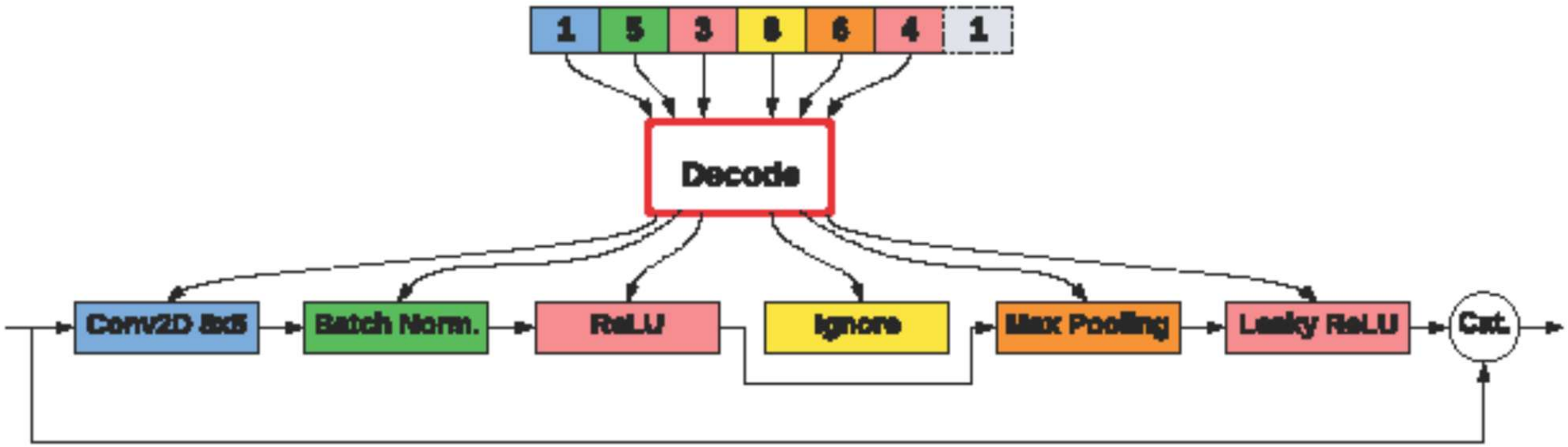
Cellular encoding for NAS



- Sequential
- Parallel

Co- evolution

- Cellular encoding + GP for macro-architecture
- GA for micro-architecture, stem, classifier



- Co- evolution