Text Sentiment Analysis

P. Giannouris, D. Karamouzas, Dr. Ioannis Mademlis,
Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.3

Artificial Intelligence &
Information Analysis Lab

Affect, Sentiment or Emotion?

- *Affect* refers to the **internal experience** of affective states (e.g. fear, joy) and may even be non-conscious. When they become conscious, they are called **feelings**.

- *Emotions* are the **observed** reactions to affective states. They can be genuine or feigned.

- *Sentiment* is an enduring, **conscious disposition** in a personality to respond with a positive or negative affect towards an entity.

Affective Computing

What?: Affective Computing aims at assessing the emotional response of a person. Imprecise term better replaced with emotional computing.

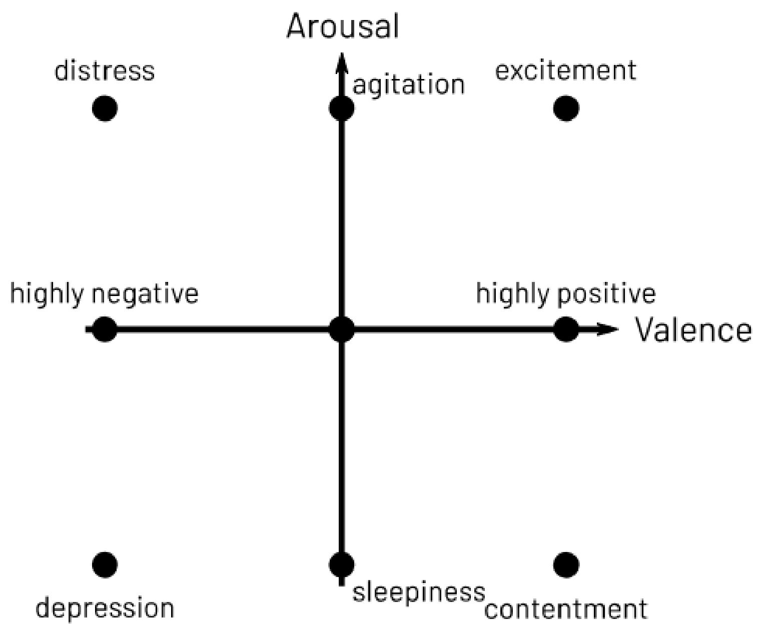
Where?: Images (facial expressions, posture), audio (speech), video (movement, gestures), **text** (tweets, letters)

Why?: Get a feel of public opinion behind certain topics, adapt strategies to win preference of customers/voters.

Artificial Intelligence & Information Analysis Lab

Russell’s model

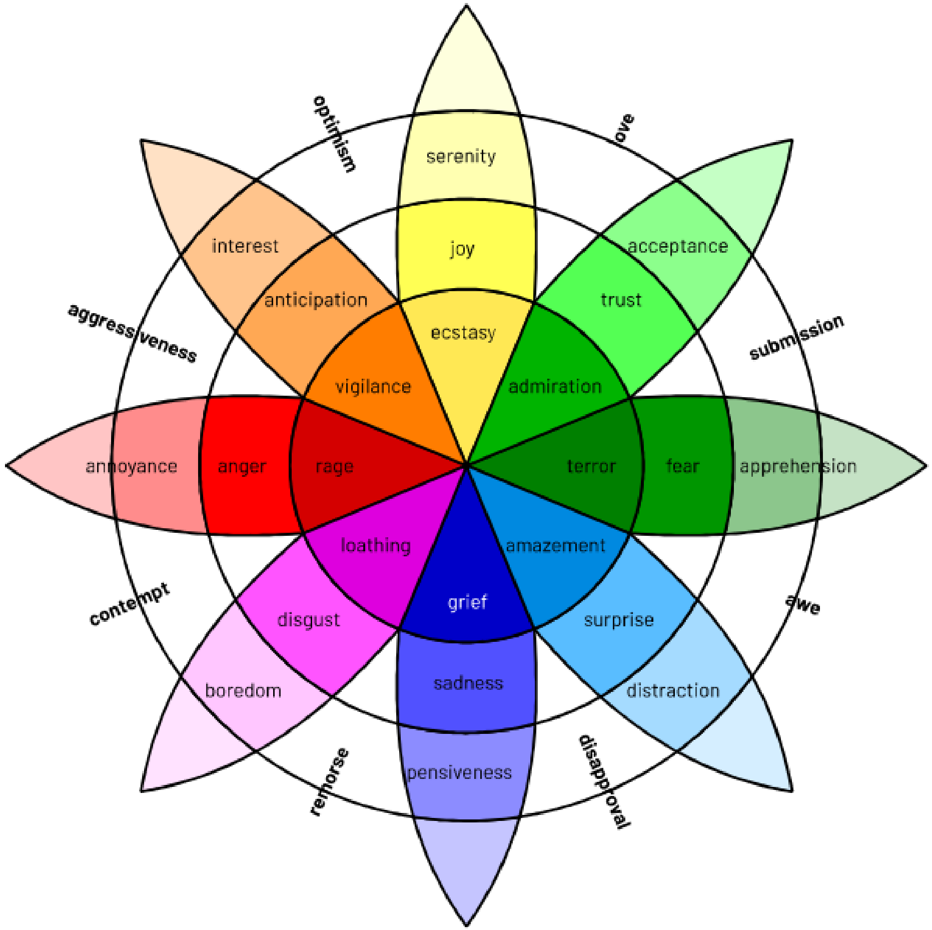
- Emotions are **dimensional**: they can be located in a continuous space, with each of its axes defining an emotional spectrum. [RUS1980]
- Thus, every emotion can be described as a 2D vector.
- The 2 axes, or parameters, are **valence** (positive/negative – level of pleasantness) and **arousal** (agitation/sleepiness – level of energy).
- Later, the third dimension of **dominance** (in control or not) was added.



2D representation of Russell’s model [DAR2006].

Plutchik’s wheel

- 3-dimensional model. Emotions are a continuous spectrum like color. [PLU1970]
- Arranges emotions in concentric circles.
- Inner circles = strong emotions.
- Outer circles = weak emotions.



Plutchik’s wheel of emotion in 2D [PLU2001].

Artificial Intelligence & Information Analysis Lab

Ekman’s model

- Categorical model (no spectral aspect). [EKM1992]
- Six emotions (anger, disgust, fear, happiness, sadness, surprise).
- Each emotion originates from a different neural system.
- List can be extended: E.g., amusement, shame, relief.

OCC model

- Disagree with the idea of “basic emotions”. [ORT1988]
- Emotions vary in terms of intensity.
- Adds 16 emotions to Ekman’s 6 for a total of 22.
- Often preferred since it spans a larger emotional field.

Model implications

- Depending on the dataset used ED can be approached either as a **classification** (categorical models) or a **regression** (dimensional models) problem.
- Basic emotions are not equidistant (fear is closer to disgust than joy).
- No consensus on the best model.
- Limited research on mappings between models.

About SA

What?: *Sentiment analysis* (belongs to NLP) is the interpretation and classification of emotions (positive, negative and neutral) within textual data, using text analysis algorithms.

Where?: Reviews (products, movies), social media (Twitter, Facebook), articles, etc.

Why?: Give machines emotional intelligence, get a feel of public opinion behind certain topics, infer effect of product on user.

SA vs ED

Sentiment Analysis

1. Concerned about polarity (e.g., positive-negative, objective-subjective, figurative-literal).
2. Categorizes opinionated text.
3. Task-specific.

Emotion Detection

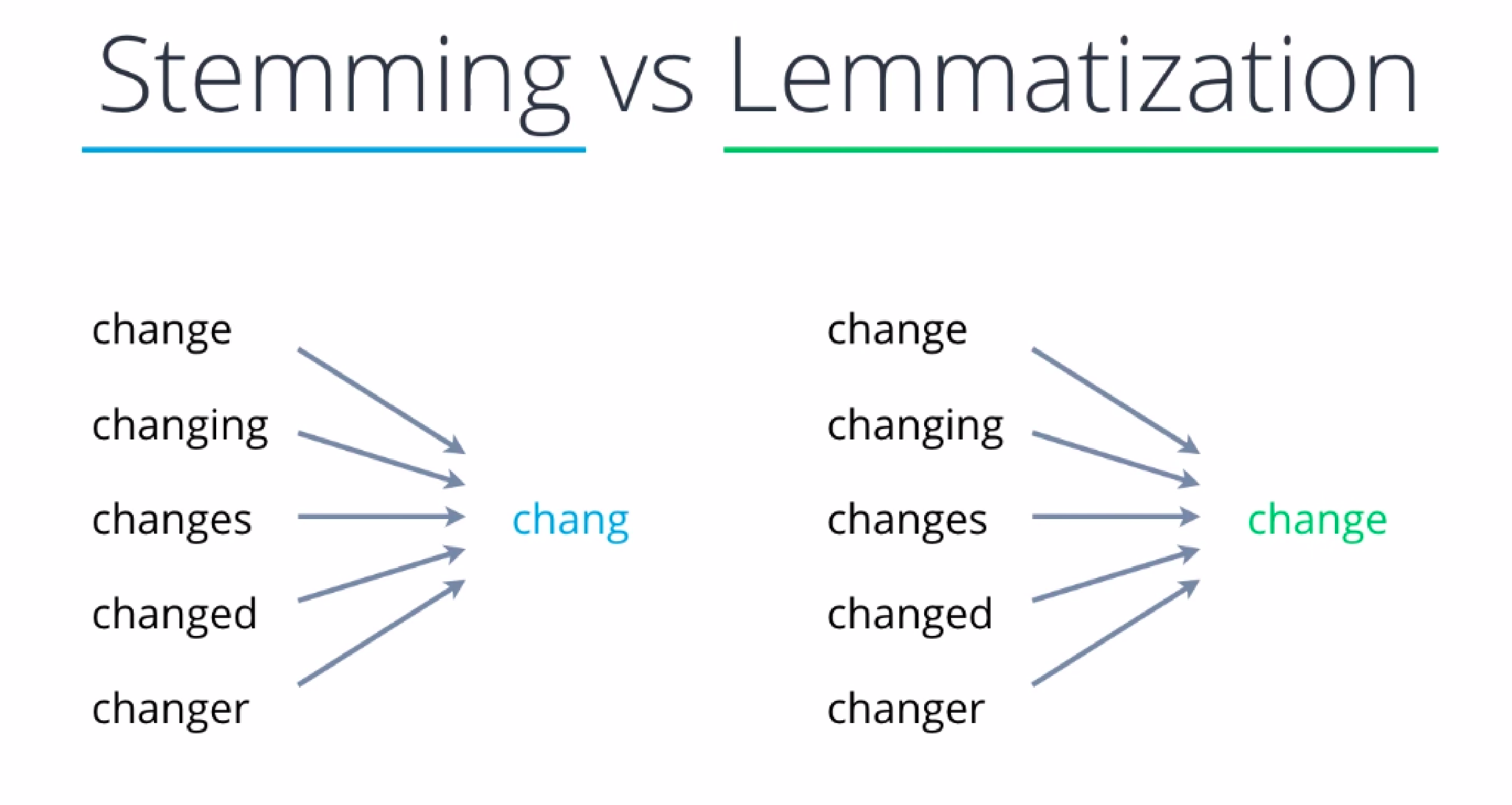
1. Concerned about emotional or psychological state.
2. Follows a certain psychological model.
3. More objective.

Three levels of granularity

- Sentence level.
- Paragraph level.
- Document level.

Common preprocessing

- **Text cleaning**: Remove all unnecessary words/characters (URL, hashtags, stopwords).
- **Lowercase** all words.
- **Stemming** or **Lemmatization**.
- **Tokenization**.
- **Vectorization**.



Stemming vs Lemmatization
Source: https://itnext.io/what-is-nlp-an-introduction-to-natural-language-processing-f48ff68a2e90

Tokenization

Usually, most frequent words appear higher in the token dictionary (“the”, “of”, “so”)

vocabulary - all unique words in a source of text
token - an integer value assigned to each word in the vocabulary

token dictionary

{‘the’: **0**, ‘of’: **1**, ‘so’: **2**, ‘then’: **3**, ‘you’: **4**, … ‘learn’: **3191**, … ‘artificial’: **30297**… }

sample text		**tokenized text**
“the pettiness of the whole situation”	→	[0, 121241, 1, 0, 988, 25910]

Tokenization of a sample text
Source: https://cezannec.github.io/CNN_Text_Classification/

Text Vectorization

- We cannot work with text directly when using machine learning algorithms.
- **Bag-of-Words**
  - focuses on the occurrence of words in a document.
- **Words Counts**
  - count for the number of times each word appeared in the document.
- Word Frequencies with **TF-IDF**
  - **Term Frequency**: summarizes how often a given word appears within a document.
  - **Inverse Document Frequency**: downscales words that appear a lot across documents.
- PROBLEM: **cannot account for the similarity between words.**

Solution: Word Embeddings

- Words are represented with **real-valued vectors** that are learned in an unsupervised training process.
- Words with **same meaning** have **similar representations.**
- We can represent a sentence as a list of word embeddings and use it as input to a recurrent or convolutional neural network.

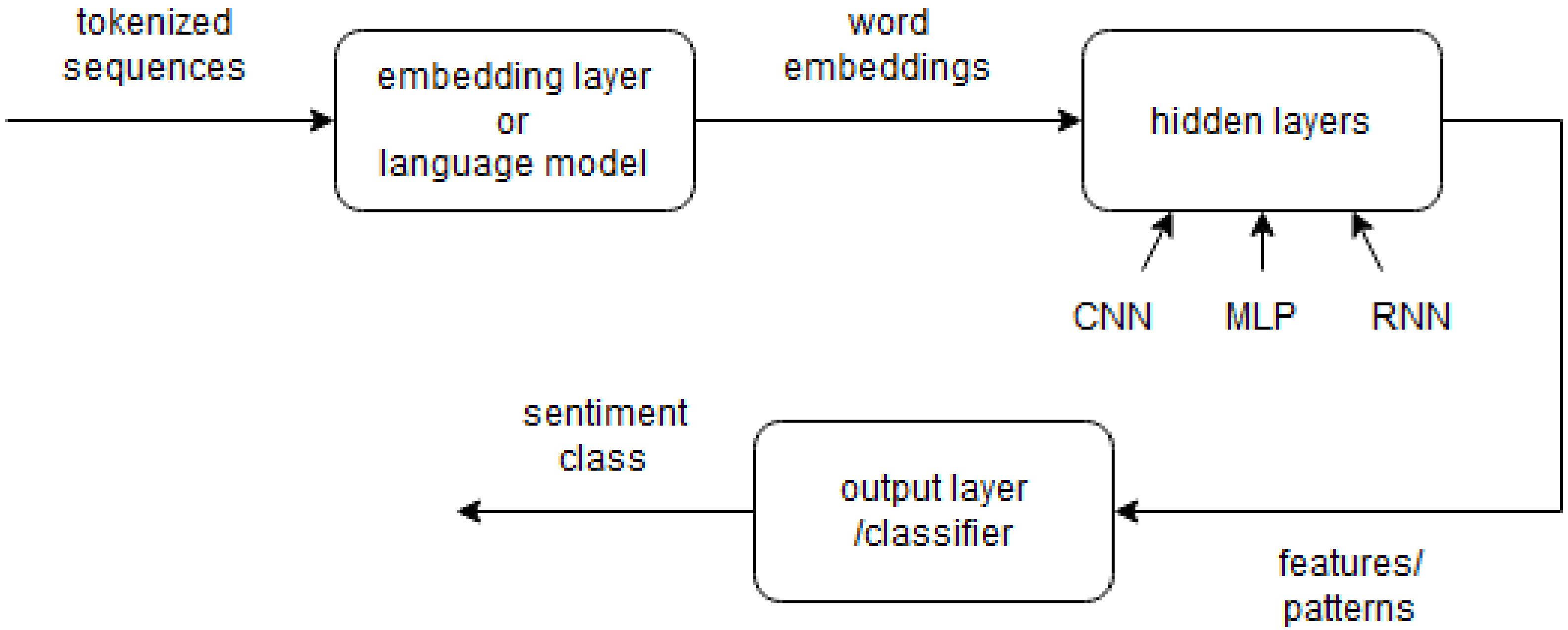
Simple example of word embedding for 3 words. Source: https://cezannec.github.io/CNN_Text_Classification/

Supervised ML methods

- In this approach, machine learning is utilized to train models with labeled data (supervised learning).

- Given the data, the only remaining thing is to choose the proper model and specify its hyper-parameters in order to obtain high prediction accuracies while avoiding overfitting.

Classification schema

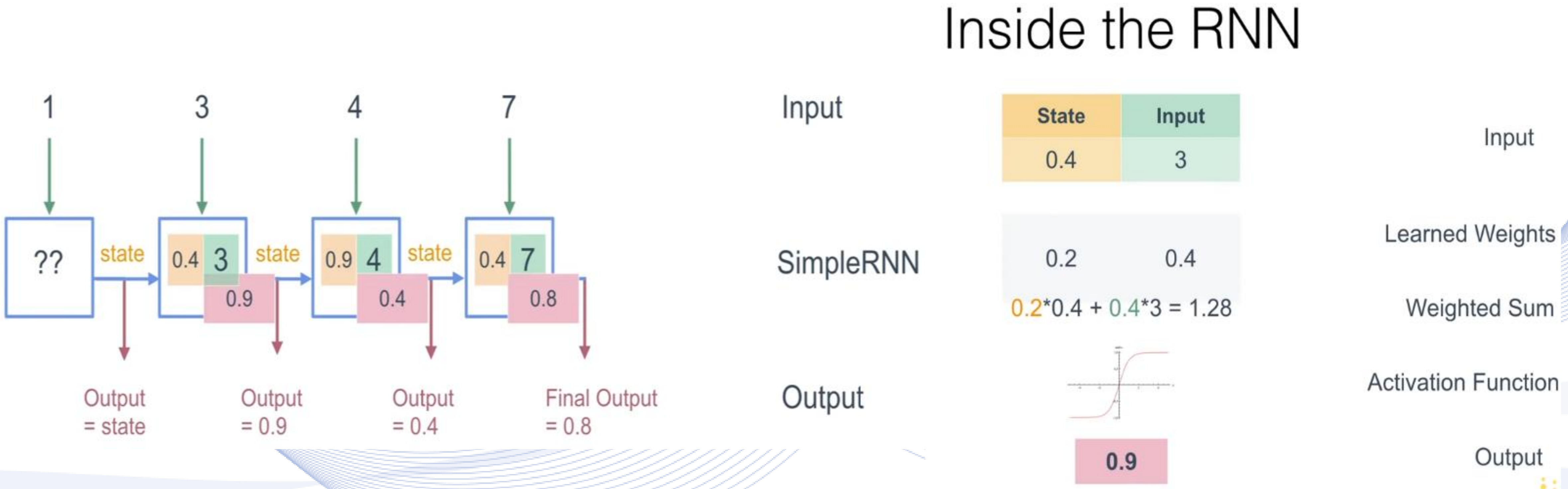


Take the **matrix representation** of a sentence as input and **extract high level features with hidden layers**.

Feed these features to the **output layer/classifier** (fully-connected layer) to **classify the input sample** into one of the sentiment classes.

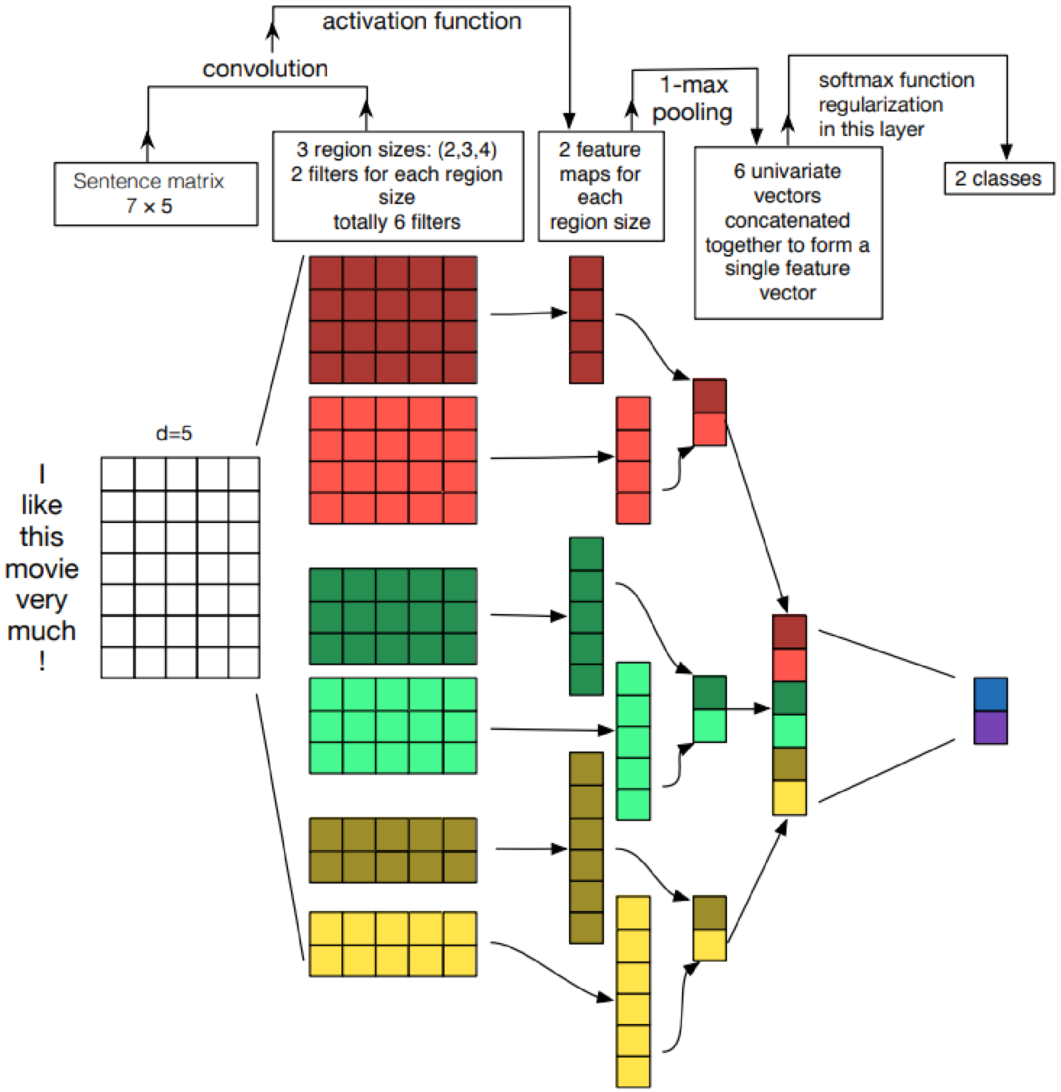
Binary classification -> **sigmoid** output , **Multiclass** classification -> **softmax** output

Classification example with RNN



Use of RNN for a simple classification task (*Source*: https://wandb.ai/site/tutorial/long-short-term-memory-lstm).

Classification example with CNN



Text classification using CNN with 1 convolutional layer [KUR2020].

CNN/LSTM Hybrid

- **CNN** will **detect local features** (word groupings) and **LSTM** will **model long-distance temporal relationships** in the input sequence of embeddings. [REH2019]
- CNN is usually used first to **reduce the dimensionality** of the input by producing the feature vectors. These vectors are input to the LSTM.
- Therefore, the **LSTM will have to process smaller inputs** than the full input representations, making its **training a lot faster** and efficient.
- As always, a fully-connected output layer is needed at the end of the model for the classification result.

Classification with BERT

- A **language model** can be used instead of the embedding layer to **produce the word embeddings** from the tokenized sequences.

- BERT [DEV2018] is a state-of-the-art language model that leads to top results due to its high-quality **contextualized embeddings**.

Text classification with BERT

- Load the **pretrained BERT** and add a classifier on top of it.
- The **classifier** can be a **CNN, LSTM, MLP** or just a single fully connected layer.
- BERT **feeds the classifier with the word embeddings** to produce the final output.
- Since BERT is pretrained, we **just need to train the classifier** for the classification task using our labeled dataset.
- We can fine-tune BERT along with the classifier on the specific task but this is computationally expensive and usually brings small benefit.

Sentiment Analysis in Texts with Figurative Language

- Existing solutions face issues with non-literal text.

- Figurative language is difficult to analyze.
    - It is inherently semantically ambiguous.
    - The DNN may have trouble understanding whether a figurative phrase implies negative or positive opinion.

- However, sarcasm and irony tend to prevail in many texts, such as social media posts.

Knowledge Distillation for improved Sentiment Analysis on Figurative Language

- Teacher-student architecture to enrich the student model with the knowledge of a pretrained figurative language (sarcasm, irony and/or metaphor) recognizer - teacher. [KAR2022b]

- Knowledge distillation from a pretrained binary recognizer of figurative language, employed as an auxiliary task while training a multiclass sentiment analysis neural model under a multitask setting.

Teacher - F

- A DNN-based binary text classifier F has been pretrained under a regular supervised setting on a database containing two classes: “figurative”, “literal/non-figurative”.

- The teacher’s output F(x) for a respective input data point x would lie in the interval [0, 1] with 0/1 being interpreted as figurative/literal, respectively.

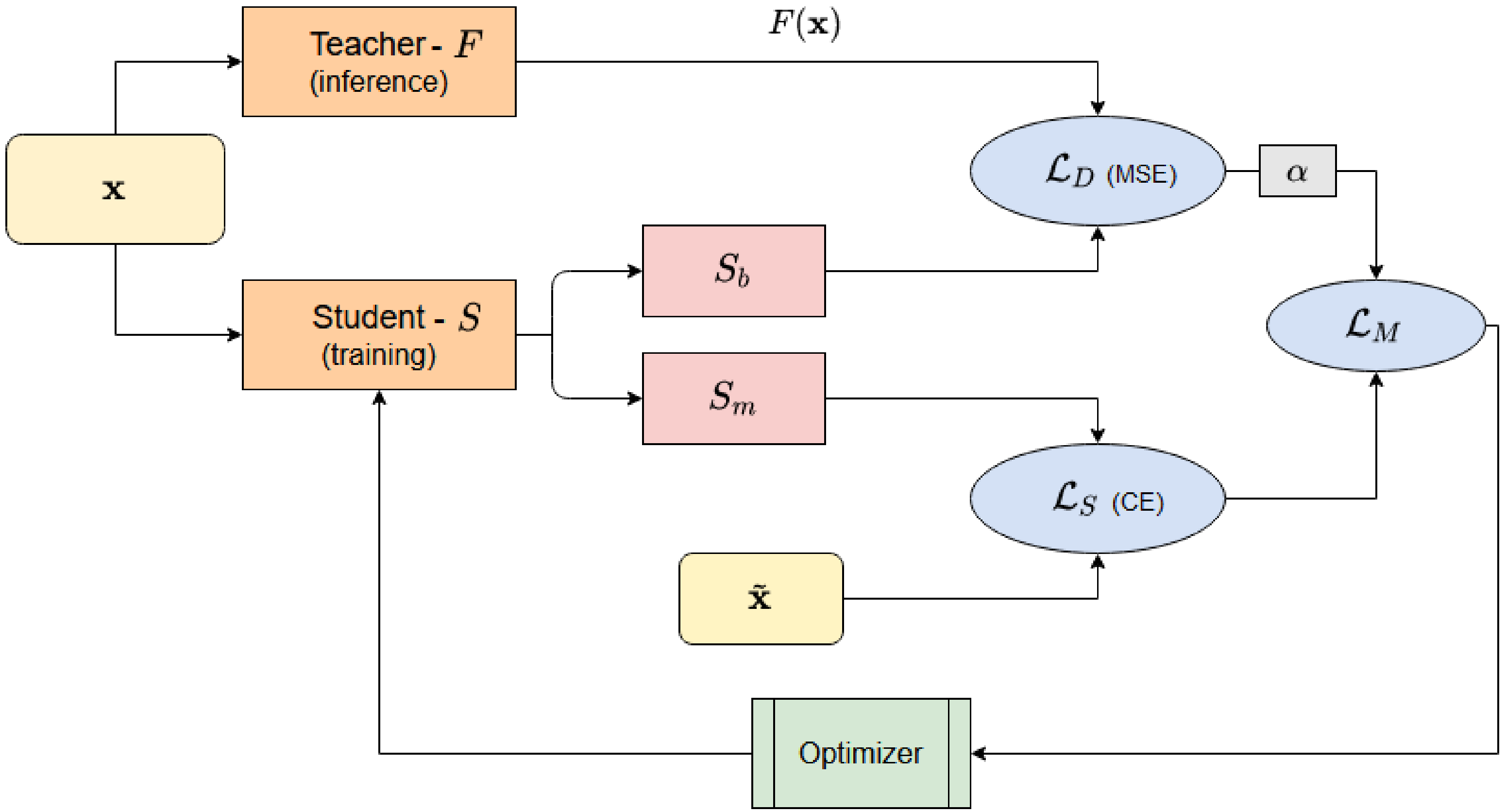
Student - S

- The student S is the neural model we want to optimize; on a different, sentiment-annotated dataset.

- Typically, $N \geq 3$ classes are employed for the sentiment analysis/opinion mining task (“positive”, “neutral”, “negative”, etc.) .

- S is trained by a regular, suitable loss function LS , such as Cross-Entropy (CE).

Distillation schema



Block diagram of the training process [KAR2022b].

Conclusions

- Short text sentiment analysis has progressed a lot thanks to deep learning architectures.
- Modern word embeddings provide semantically meaningful token representations to the neural classifier.
- Figurative language (sarcasm, metaphor, irony) significantly increases the difficulty of the sentiment analysis task.
- Estimations about the existence of figurative language in an input text can boost the accuracy of a sentiment classifier, by helping it internally resolve semantic ambiguities.

Bibliography

[HUT2014] Hutto, Clayton, and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Proceedings of the international AAAI conference on web and social media*. Vol. 8. No. 1. 2014.

[ESU2006] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. 2006.

[HOC1997] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

[ANS2020] Ansari, Mohd Zeeshan, et al. "Analysis of political sentiment orientations on twitter." *Procedia Computer Science* 167 (2020): 1821-1828.

[KUR2020] Kurniasari, L. I. L. I. S., and A. R. I. E. F. Setyanto. "Sentiment analysis using recurrent neural network-lstm in bahasa Indonesia." *J. Eng. Sci. Technol* 15.5 (2020): 3242-3256.

Bibliography

[GAI2019] Gaikar, Dipak, et al. "Twitter Sentimental Analysis for Predicting Election Result using LSTM Neural Network." *IRJET* (2019).

[SAH2020] Saha, Baidya Nath, Apurbalal Senapati, and Anmol Mahajan. "LSTM based deep RNN architecture for election sentiment analysis from Bengali newspaper." *2020 International Conference on Computational Performance Evaluation (ComPE)*. IEEE, 2020.

[WAN2019] Wang, Mei, et al. "Sentiment analysis based on attention mechanisms and Bi-Directional LSTM fusion model." *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2019.

[PAS2013] Pascanu, Razvan, et al. "How to construct deep recurrent neural networks." *arXiv preprint arXiv:1312.6026* (2013).

[SCH1997] Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional recurrent neural networks." *IEEE transactions on Signal Processing* 45.11 (1997): 2673-2681.

Bibliography

[ZHA2015] Zhang, Ye, and Byron Wallace. "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." *arXiv preprint arXiv:1510.03820* (2015).

[GAN2020] Gan, Chenquan, et al. "Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis." *Knowledge-Based Systems* 188 (2020): 104827.

[POT2018] Pota, Marco, et al. "A subword-based deep learning approach for sentiment analysis of political tweets." *2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 2018.

[REH2019] Rehman, Anwar Ur, et al. "A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis." *Multimedia Tools and Applications* 78.18 (2019): 26597-26613.

[KAL2021] Kaladevi, P., and K. Thyagarajah. "RETRACTED ARTICLE: Integrated CNN-and LSTM-DNN-based sentiment analysis over big social data for opinion mining." *Behaviour & Information Technology* 40.9 (2021): XI-XIX.

Bibliography

[UME2021] Umer, Muhammad, et al. "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model." *Computational Intelligence* 37.1 (2021): 409-434.

[MIN2019] Minaee, Shervin, Elham Azimi, and AmirAli Abdolrashidi. "Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models." *arXiv preprint arXiv:1904.04206* (2019).

[ABI2019] Abid, Fazeel, et al. "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter." *Future Generation Computer Systems* 95 (2019): 292-308.

[LIN2020] Ling, Mingjie, et al. "Hybrid neural network for Sina Weibo sentiment analysis." *IEEE Transactions on Computational Social Systems* 7.4 (2020): 983-990.

[CAI2020]Cai, Yi, et al. "Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: A multi-task learning approach." *Knowledge-Based Systems* 203 (2020): 105856.

Bibliography

[KAR2022a] Dionysios Karamouzas, Ioannis Mademlis & Ioannis Pitas, “Public opinion monitoring through collective semantic analysis of tweets”, Social Network Analysis and Mining, 12(1), 91, Springer, 2022.

[KAR2022b] Dionysios Karamouzas, Mademlis & Ioannis Pitas, "Neural Knowledge Transfer for Sentiment Analysis in Texts with Figurative Language“, Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2022.

[DEV2018] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[RUS1980] Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

[PLU1970] Plutchik, R. (1970). Emotions, evolution, and adaptive processes. In *Feelings and emotions: the Loyola Symposium* (pp. 1-14). New York: Academic Press.

[PLU2001] Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, *89*(4), 344-350.

Q & A

Thank you very much for your attention!

Contact: Ioannis Pitas
pitas@csd.auth.gr