

We're hiring Internship student for a Phd funding

- Subject : **Improving the Automatic Detection of Word Meaning Negotiation Indicators in Conversation through Active Learning**

Place of work: INRIA Paris, ALMAnaCH project-team

Starting date: From April 2026 (flexible)

Duration: 6 months

Supervision: Aina Garí Soler, Chloé Clavel,

Request for Phd funding with the call of PRAIRIE IA cluster.

https://github.com/ainagari/ainagari.github.io/blob/master/documents/Internship_call_indicator_detection_2026.pdf

Sentiment Analysis

Chloé Clavel, Inria Paris, ALMAaCH

What is sentiment?



Sentiment analysis in English not exactly the same as *Analyse de sentiments* in French

Définitions proposées par : [Oxford Languages](#) · [En savoir plus](#)

sentiment

/'sentim(ə)nt/

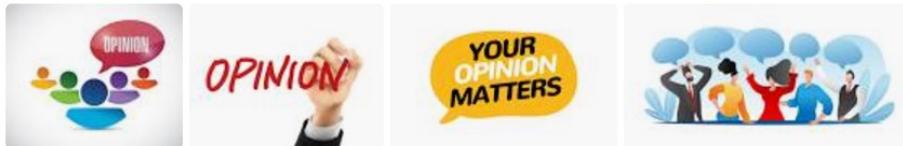
noun

noun: **sentiment**; plural noun: **sentiments**

1. a view or opinion that is held or expressed.

"I agree with your sentiments regarding the road bridge"

Synonyms : [view](#) [point of view](#) [way of thinking](#) [feeling](#) [attitude](#) [thought](#) [▼](#)



- general feeling or opinion.
"public sentiment was on the side of reform"
- a feeling or emotion.
"an intense sentiment of horror"

Synonyms : [feeling](#) [emotion](#)

• ARCHAIC
the expression of a view or desire especially as formulated for a toast.

2. exaggerated and self-indulgent feelings of tenderness, sadness, or nostalgia.
"many of the appeals rely on treacly sentiment"

Synonyms : [sentimentality](#) [mawkishness](#) [over-sentimentality](#) [emotionalism](#) [▼](#)

Définitions proposées par : [Dictionnaires Le Robert](#) · [En savoir plus](#)

sentiment

nom masculin

1. Conscience plus ou moins claire, connaissance comportant des éléments affectifs et intuitifs.
Un sentiment de solitude.

Synonyms : [impression](#)

2. **LITTÉRAIRE**

Capacité d'apprécier (un ordre de choses ou de valeurs).
Le sentiment de la beauté.

Synonyms : [sens](#)

3. **LITTÉRAIRE**

Avis, opinion.
J'ai le sentiment que...

4. État affectif complexe, assez stable et durable.
Manifester ses sentiments.

Synonyms : [émotion](#) [passion](#)

- **Les (bons) sentiments**

les sentiments généreux, les inclinations altruistes.

5. **Le sentiment**

la vie affective, la sensibilité.

6. Expression de la sensibilité.
Elle a chanté avec beaucoup de sentiment.

What is sentiment?

-  • *Sentiment analysis* in English not exactly the same as *Analyse de sentiments* in French
- Sentiment has several senses in English :
 - feeling: *She experiences a sentiment of nostalgia*
 - opinion: *My sentiment is that this project isn't working and we need to rethink it.*
 - exaggerated emotion: *That film has too much sentiment for my taste*
- Sentiment has several - but different - senses in French
 - Impression, feeling: *Dans cette affaire j'ai le sentiment que tu te trompes.*
 - Intention: *C'est une personne qui a de très bon sentiments*
 - Avis, opinion: *Il est difficile de partager le même sentiment que toi sur cette affaire.*

Examples from wordreference

Sentiment terms or equivalent in the literature

- Sentiment analysis is used broadly in literature to analyze
 - emotions, opinions, stances, affects, mood, etc.
- Equivalent terms:
 - opinion mining, subjectivity analysis, stance detection, affect sensing, emotion detection, ...



But they are different phenomena...

- Ex. Scherer's definitions from psychology [Scherer, 2005]
 - **Emotion:** short phenomenon, physiological reaction, appraisal of a major event (stimulus)
 - **Mood:** diffuse non-caused low-intensity long-duration change in subjective feeling
 - **Interpersonal stances:** affective stance toward another person in a specific interaction
 - **Attitudes:** enduring, affectively colored beliefs, dispositions towards objects or persons
 - **Personality traits:** stable personality dispositions and typical behavior tendencies
- PRACTICE: link the following terms to the most relevant phenomenon
 - liking, gloomy, contemptuous, jealous, sad

Other work related to sentiment analysis

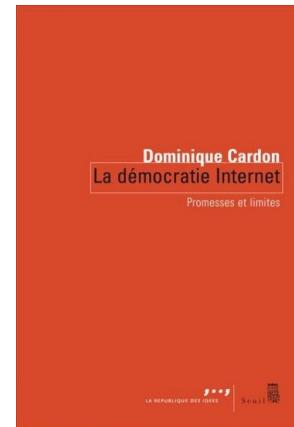
- Study of specific categories
 - Fear (emotions)
 - Hate-speech (an extremely strong dislike: attitude or interpersonal stance)
 - Condescension (attitude or interpersonal stance)
 - Sarcasm detection (interpersonal stance)

Applications of sentiment analysis

Social Data analysis
Human-agent interaction (ex: chatbot)

Social data analysis

- Social data: text data from social networks
 - Expressions of the citizens on the web
- Context
 - opportunities for criticism and action via the Internet



Social data analysis : applications

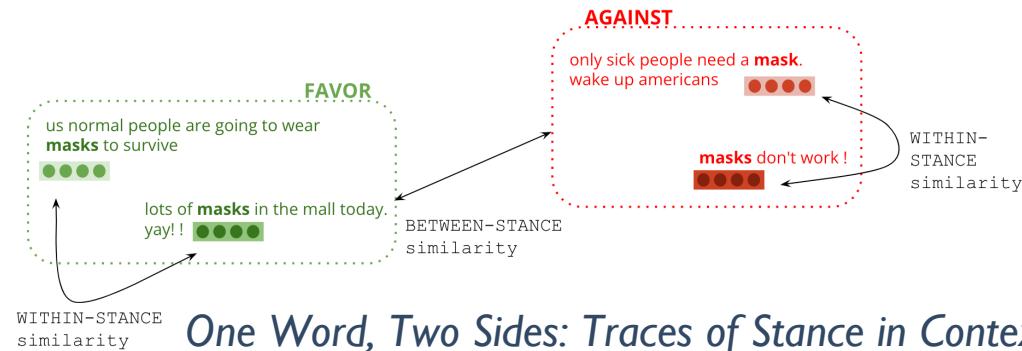
- **Computational Social Sciences :**
 - Detection of misinformation, cyberbullying, hate speech, etc.
 - Analysis of societal trends (e.g. environmental engagement)
 - Analysis of citizens' opinions on candidates in elections/ Evaluation of the success of communication campaign
- Analysis of **movie/product/restaurant reviews**
 - Analysis of the opinions of Internet users on a product
 - Analysis of the e-reputation of a brand, a product
 - Identify target clients / recommendation systems

Computational social sciences

- 2 fields/2 types of methodologies
 - Sociology:
 - qualitative / manual / sociological analysis of small corpora selected to form a panel of studies
 - Computer science :
 - development of automatic large corpus analysis methods

Examples of our research for computational social sciences

Automatic analysis of stances
(in favor/against vaccinations)
on twitter



One Word, Two Sides: Traces of Stance in Contextualized Word Representations, [Gari Soler, Labeau & Clavel COLING 2022]

Automatic analysis of fallacies
Chaire ANR NoRDF



Context

Fallacies - erroneous or invalid ways of reasoning
- play a significant role in spreading **fake news** and **propaganda**. Their **automated detection and classification** is vital yet challenging due to subjectivity and the lack of unified approaches.

[Helwe, Paris, Calamai, Suchanek & Clavel, NAACL 2024]

Examples of our research for computational social sciences

ACL 2025

Graphically Speaking: Unmasking Abuse in Social Media with Conversation Insights

Célia Nouri^{1,2} Jean-Philippe Cointet² Chloé Clavel^{1,3}

¹INRIA, ALMAnach

²Sciences Po, médialab

³Télécom Paris

{celia.nouri, chloe.clavel}@inria.fr, jeanphilippe.cointet@sciencespo.fr

Abstract

Detecting abusive language in social media conversations poses significant challenges, as identifying abusiveness often depends on the conversational context, characterized by the content and topology of preceding comments. Traditional Abusive Language Detection (ALD) models often overlook this context, which can lead to unreliable performance metrics. Recent Natural Language Processing (NLP) approaches that incorporate conversational context often rely on limited or overly simplified representations of this context, leading to in-

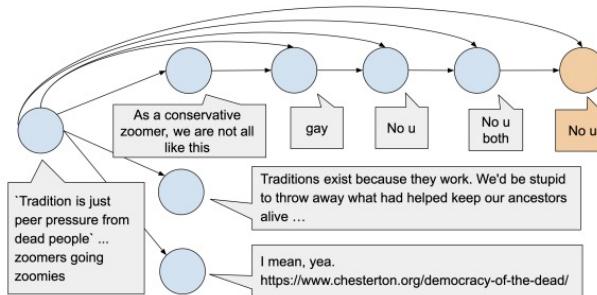


Figure 1: Example conversation from the Contextual Abuse Dataset (CAD), the graph was generated from our Affordance-based method. The target node is labeled abusive and colored in orange.

Human-agent interaction

Analyze and reproduce human behaviors to interact socially with humans

Virtual characters
(Embodied Conversational Agents)



[GRETA Platform, Pelachaud]

Robots



[Softbank robotics]

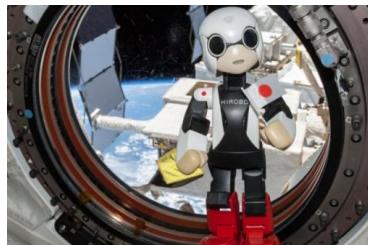
Vocal assistants
(google home, siri)



Chatbots



Applications of sentiment analysis in conversational systems



Kirobo: the Japanese robot who left 18 months in space to keep an astronaut company

Agents work of Bickmore and colleagues combines these



Figure 1. SimSensei, the Virtual Human Interviewer



Customer relationship management



Health
(agent for motivational interviews) Ex: SimSei

Detecting distress signals for “safeguards” in GPT5

Examples of our research for conversational systems

Education : Robots as learning partners

Global aim: build intelligent autonomous social robots that can promote children's learning (for example in computational thinking) by assisting teachers through complementary activities

Machine learning models for the automatic analysis of child's self-confidence

H2020 Marie Skłodowska-Curie Innovative Training Network ANIMATAS

QTrobot

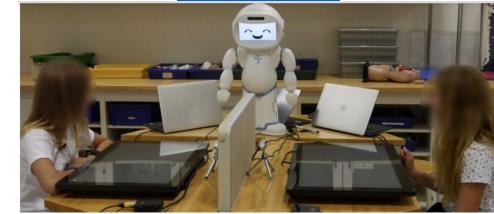
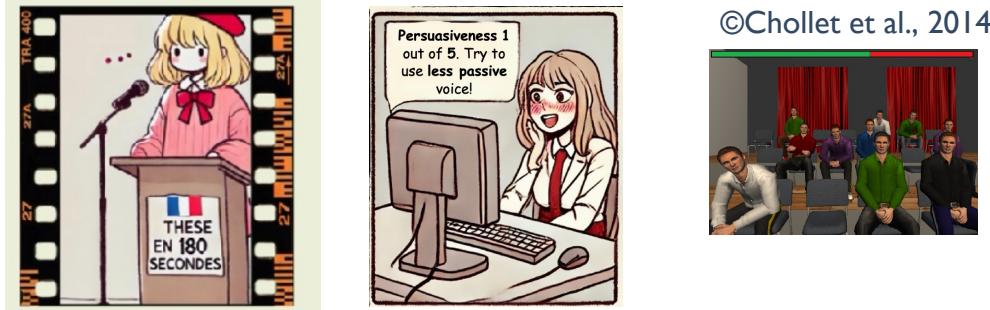


Figure 1: The JUSTthink activity setup.



Dinkar, T., Vasilescu, I., Pelachaud, C., & Clavel, C. *How confident are you? Exploring the role of fillers in the automatic prediction of a speaker's confidence.* ICASSP 2020

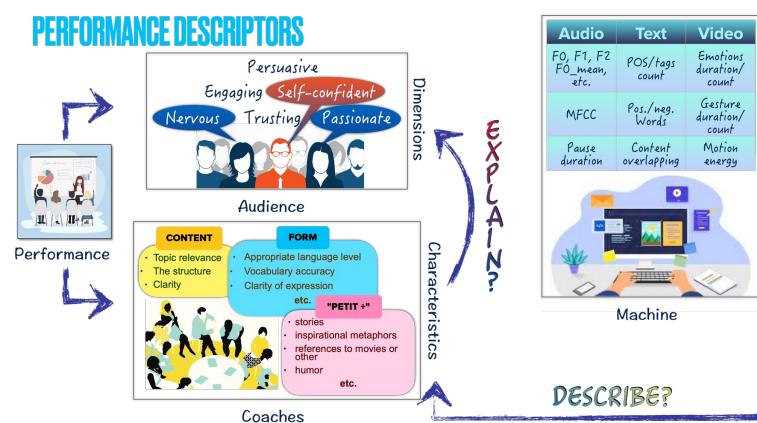
Interactive systems: applications



Public speaking training systems : ANR (National Project) Revitalise

Global aim: virtual environment with a virtual audience for public speaking training

Our work: Machine learning models for the automatic analysis of persuasiveness from speech, **Explainable approaches** to give feedback



Barkar, A., Chollet, M., Biancardi, B., & Clavel, C. (2023). Insights Into the Importance of Linguistic Textual Features on the Persuasiveness of Public Speaking. ICMI

Importance of societal impact

Key question: *Why are we building these models, and who will use them?*

The research duality: a same research can be

beneficial to human well-being:
socio-conversational systems
supporting well-being and human empowerment

vs.

detrimental to human-being:
socio-conversational systems
enabling manipulation

Engagement strategies:
encouraging use

Creating addiction/dependence

Ben-Youssef, A., Clavel, C., & Essid, S. (2019). Early Detection of User Engagement Breakdown in Spontaneous Human-Humanoid Interaction, *IEEE Transactions on Affective Computing*.

Adam, D. (2025). Supportive? Addictive? Abusive? How AI companions affect our mental health. *Nature*, 641(8062), 296-298.

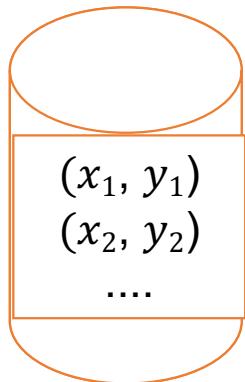
Importance of Societal Impact

- « *Researchers need to be aware of the real-world applications of their work and consciously decide to choose to help marginalized communities via technology* » [Hovy and Prabhumoye, 2021]
- « *Build beneficial AI by supporting human well-being* » Keynote Talk Verena Rieser at ACL 2025

Overview of sentiment analysis tasks

1/Sentiment Classification

- Supervised learning of sentiment classes



- Training a model :
- Support Vector Machine, (end-to-end) transformers
- Fine-tuning **foundation models**

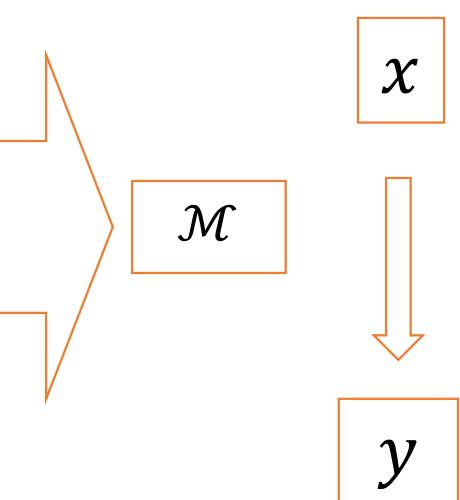
Training corpus: set of (x_i, y_i)

Documents: x_i

- Short (tweets) to long (newspaper articles)
- Monologues (presentation) or dialogues (call-centres, chats)

Categories: $y_i \in \mathcal{L}$

- Label set \mathcal{L} depending on the application for sentiment analysis:
frustration, satisfaction, fear, etc.



Example

X = the movie review

y = the score attributed by a user (1 to 5)

 4,0 Publié le 2 août 2023

Ce long "court métrage" vaut le détour, on a le droit à du Dupieux dans le texte, scénario un peu loufoque et WTF (il faut aimer ce type d'humour) mais toujours des dialogues hilarants et super bien construit avec un petit message sous-jacent un peu plus profond.

C'est barré, c'est décalé, on passe une bonne heure de cinéma

Ps = le film ne dure qu'une heure, ça permet d'aller le voir sans trop de difficulté à la pause déjeuner ;)

Various types of classification tasks

Binary classification

ex: $y_i \in \{\text{positive}, \text{negative}\}$

Multi-class classification

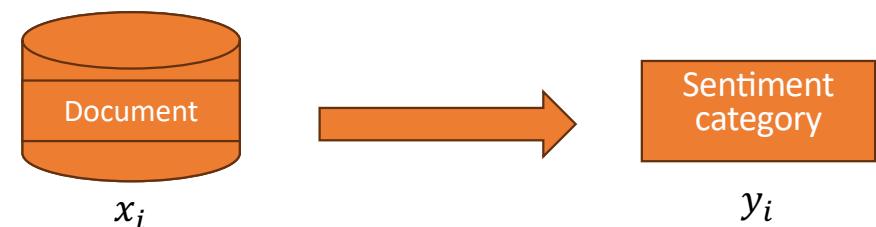
ex: $y_i \in \{\text{fear}, \text{anger}, \text{joy}, \text{disgust}, \text{sadness}\}$

Multi-label classification

ex: $y_i = (l_1, l_2, l_3)$ with $l_1, l_2, l_3 \in \{\text{set of hashtags}\}$ (prediction of the set of hashtags of a tweet : (#party #FRIDAY #fail))

Regression task

ex: emotional intensity $y_i \in [0; 5]$



2/Labelling sentiment-related words

« Je suis satisfait des contacts que j'ai eus avec le service client mais pas des tarifs pratiqués »

- Task#1- multi-label classification – x_i : document; $y_i = (l_1, l_2, l_3)$
 - Sentiment: satisfaction or dissatisfaction?
 - Topics: *contact* and *price*
- Task#2 - fine-grained analysis – x_i : word; ; $y_i \in \{\text{speaker}, \text{satisfaction}, \text{contact}, \text{etc.}\}$
 - Retrieve sentiment-related units
 - And link sentiment to their targets and sources
 - Satisfaction of the speaker towards *contact*
 - Dissatisfaction of the speaker towards *price*

Also known as **Aspect-based sentiment analysis (ABSA)**

Source-target sentiment model

- PRACTICE : Source ? Target?
 - I liked the plot of this movie.
 - This travel was a mistake.
 - It sounds like a great plot.
 - He is frightened of spiders.
- Note: identifying the source and the target can be tricky
 - “il les adore”
 - Requires coreference solving

Labelling sentiment-related words: B.I.O. labels

x_i : word; y_i = B.I.O. tag

B - beginning of a span

I - inside a span

O - out of a span (indicates the token doesn't belong to any span)

Ex : OTE (Opinion Target Extraction), retrieve target spans using BIO annotations

OTE (Opinion Target Extraction), with BIO annotations

The	O
onion	B-TARGET
rings	I-TARGET
are	O
great	O
,	O
and	O
the	O
sushi	B-TARGET
very	O
tasty	O
...	

(figure from Agerri, R. and Rigau, G., 2019. Language independent sequence labelling for opinion target extraction. Artificial Intelligence, 268, pp.85-95.)

PRACTICE : given the following categories : source terms (S), target terms (T), evaluation terms (E), provide a BIO annotation of the following sentence:

“What I wanted to say is that I liked the plot of this movie.”

A short story of sentiment analysis methods

From the simplest to the most complex/opaque ones

0/By hand

- The sentiment class is attributed by a human
 - E.g. Yahoo in the old days

“This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can’t hold up.”



?

- ✓ Very accurate and consistent assuming experts
- ✗ Super slow, expensive, does not scale

1/ Rule-based

- ✓ Accuracy: high if rules are suitable
- ✓ Transparent: explicit decision process
- ✗ Need to manually build and maintain rule-based system.

- Retrieve words and linguistic/syntactic patterns featuring the class
- Class knowledge explicitly modelled
- Ex1: advanced search criteria using advanced regular expressions to retrieve sentiment expressions

$text_i$

(manque|~negation-patt|(il/#NEG/y/avoir/~negation-patt))/(#PREP_DE)?/ (conseil|contact|~services-lex)*

Pattern retrieving for example « manque de conseils » or
« il n'y a pas eu de contact » and assigning the text
« dissatisfaction » class

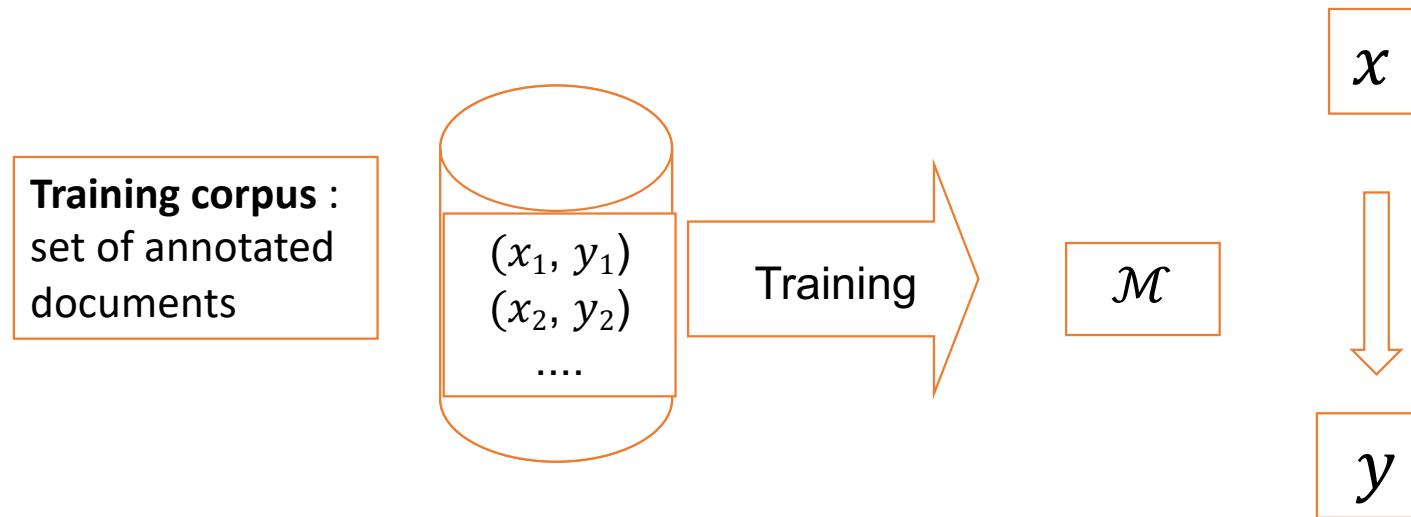
y_i

Dissatisfaction

TASK: SENTIMENT
classification

2/ Supervised MACHINE/DEEP learning

Task : text classification

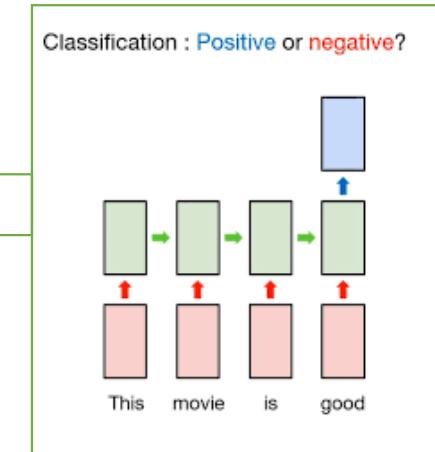


2/ Supervised Machine/deep learning

Task : text classification

Models

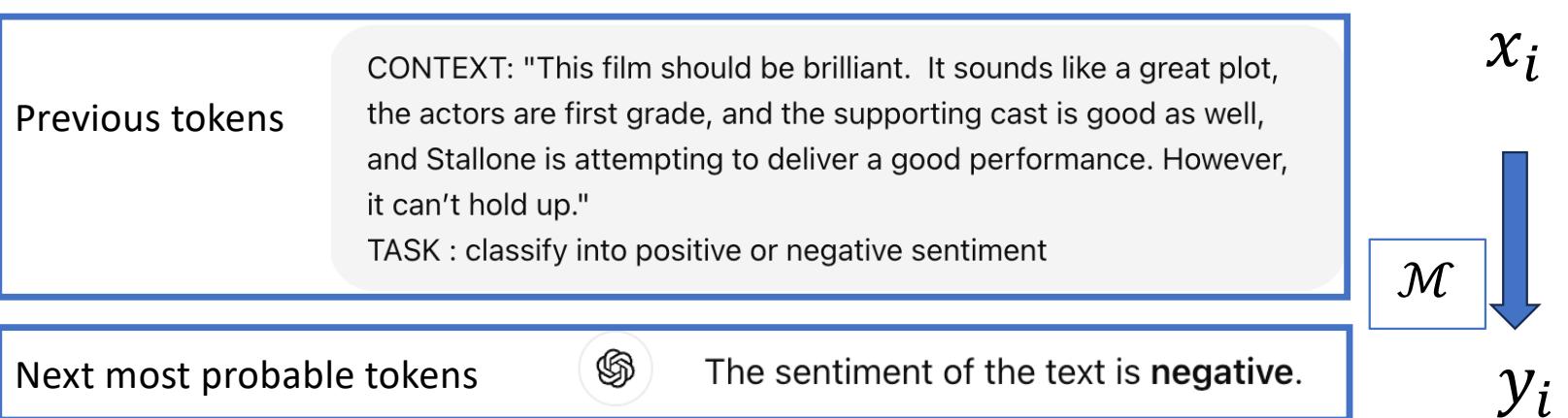
- Naive Bayes
- Logistic regression
- Multi-Layer Perceptron,
- Support Vector Machine
- Convolutional Neural networks
- Recurrent neural networks in a many-to-one configuration
- Transformers
- Possible to be trained in an end-to-end fashion
- Fine-tuning foundation models



From :
https://goodboychan.github.io/python/deep_learning/tensorflow-keras/2020/12/06/01-RNN-Many-to-one.html

3/Prompt-based using generative foundation models

Do Language Models Enjoy Their Own Stories? Prompting Large Language Models for Automatic Story Evaluation. (Chhun et al., To appear TACL)



Generative AI models: models capable of generating texts

Transform text classification task into **a task of generating an answer to a question over a context**

Prompt: question over a context

Can be used in a **zero-shot/few-shot setting or fine-tuned**

From Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods
in Natural Language Processing, <https://arxiv.org/pdf/2107.13586v1.pdf>

Focus on the transparency issue

For automatic analysis

- Rule-based
- Supervised machine/deep learning
- Prompting large generative foundation models



From the oldest to the most recent

From lowest to highest performance

But ...

From the lowest to the highest computational costs

From the least to the most opaque

The Transparency Challenges in Sentiment analysis

As AI is everywhere and accessible to everyone ...

Need for transparent methods that **make explicit**:

1. the decision-making processes behind the **automated analysis of socio-emotional behavior**
2. the underlying **biases of the agent** when it interacts with humans

=> In the context of socio-conversational systems, socio-emotional behaviors are complex and we need to be transparent on how they are encoded in models

Focus on supervised machine learning models

For sentiment analysis

Linguistic challenges

PRACTICE: Are these reviews positive or negative? According to you? Highlight the expressions corresponding to the expression of an opinion. Do they seem positive or negative in general?

- “This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can’t hold up.”
- Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is not so good either, I was surprised.

Linguistic challenges

- negative • “This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.”
- negative • Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is not so good either, I was surprised.

Linguistic challenges

More complex than a simple positive vs. negative word counts

- Requires modelling complex linguistic phenomena
 - conditional tense
 - This film **should be** brilliant
 - discourse markers
 - **However**, it can't hold up
 - negation processing
 - Laurence Fishbourne is **not** so good either
 - modifiers
 - Laurence Fishbourne is not **so** good either
 - dealing with metaphors and connotations
 - **global warming** vs. climate change (example from [Ahmad et al., 2011])

A simple sentiment classification pipeline with word embeddings

$$\text{linguistics} = \begin{bmatrix} 15 \\ 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{bmatrix}$$

Ex: average the representation
of the words of the document

Ex: with logistic regression

Training corpus :
set of
labelled documents:
Movie reviews
labelled into
pos/neg



...



Convert each
document into a
vector



Learn the models
corresponding to
each class

Practice

A simple way to compute a distributed representation for a sentence is to add up distributed representations of the words in that sentence. Consider a sentence $s = (w_1, \dots, w_M)$ of M words, and a linear sentiment analysis model parametrized by the vector \mathbf{w} in which the predicted sentiment is

$$\psi(s) = \mathbf{w} \cdot \left(\sum_{m=1}^M \mathbf{r}_{w_m} \right)$$

with \mathbf{r}_w being the distributed representation of word w .

1. Prove that, in such a model, the following two inequalities cannot both hold :

$$\psi(\text{good}) > \psi(\text{not good})$$

$$\psi(\text{bad}) < \psi(\text{not bad})$$

2. What are the simplifying assumptions made by our classification model that lead to this result ?
Propose two very simple changes that would solve the issue.

Two solutions

1/ Play with the neural architectures

2/ Integrate external information about sentiments into neural models (hybrid ML)

The answer of neural architectures

doc=Seq. of char./words(input) —— NN ——> sentiment

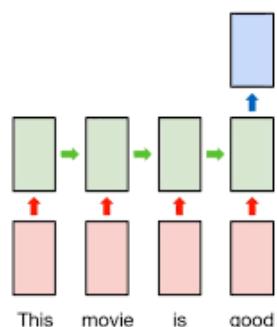
- Multilayer stack of simple modules
 - that computes non-linear input-output mappings :
 - distort the input space to make the classes of data linearly separable
 - First layers: build general purpose learning procedures to automatically learn representations
 - Final layer: specific to the task
- If the weights are set correctly, a neural network with enough neurons and a non-linear activation function can approximate a very wide range of mathematical functions
- With a depth of 5 to 20 non-linear layers, a system can implement extremely intricate functions of its inputs that are simultaneously :
 - sensitive to minutes details
 - insensitive to large irrelevant variations

To go further: Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

Classical NN architecture for NLP classification task

- **Pre-trained word vectors/model :**

Classification : Positive or negative?



- representations already learned via BERT, word2vec, glove, etc.
- on big unlabelled datasets such as wikipedia,
-> not specific neither to the data nor to the task
- **Inputs of a NN classifier** that is then trained on your task/data
- This can be considered a **basic transfer learning**

Drawback: pre-trained models have difficulties to deal with **data that is very different** from the one they have been pre-trained on (ex : wikipedia data to twitter data).

But now pre-trained models have been trained on a huge amount of data with high variety.

And there are **specialized pre-trained models** (ex: SentiBERT)

<https://arxiv.org/abs/2005.04114>

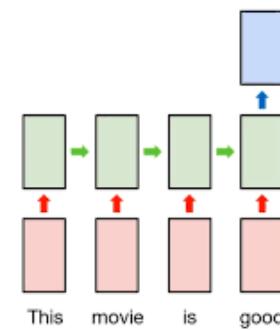
From :
https://goodboychan.github.io/python/deep_learning/tensorflow-keras/2020/12/06/01-RNN-Many-to-one.html

Classical NN architectures for NLP classification task

The pre-trained model and the classifier can rely on **various NN classical architectures**:

- Recurrent Neural Networks,
- Convolutional Neural Networks,
- Transformer
- Graph neural networks

Classification : Positive or negative?

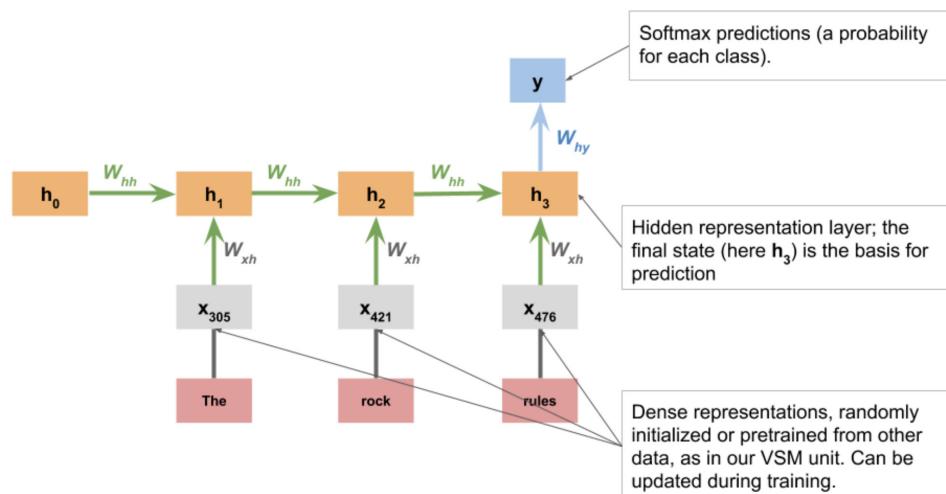


From :
https://goodboychan.github.io/python/deep_learning/tensorflow-keras/2020/12/06/01-RNN-Many-to-one.html

1/Recurrent neural networks in a many-to-one configuration

Read inputs x_t to accumulate state h_t and predict outputs y_t .

- ❖ ex. use for sentiment analysis :
 - ❖ input = sequence of tokens corresponding to one or more sentences (ex: review)
 - ❖ output = sentiment category



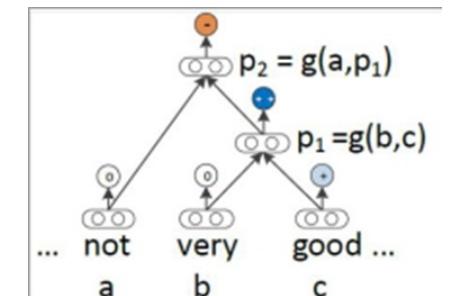
h_t contains information about the whole past sequence $h_t = f(h_{t-1}, x_t, \theta)$

the network learns to use h_t as a kind of **lossy summary of the sentiment-relevant aspects** of the past sequence of inputs up to t

RNN Variants

- GRU (Gated Recurrent Units) & LSTM networks (Long Short Term Memory Networks)
 - RNN using **gating mechanisms**
 - in order to better control the flow of information
 - and to solve vanishing gradient issue
- **Recursive tensor network:**
 - learning deep tree structure rather than the chain-like structure of RNN
 - Illustration for sentiment analysis:

*R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts,
Recursive deep models for semantic compositionality over a sentiment
treebank, EMNLP 2013*

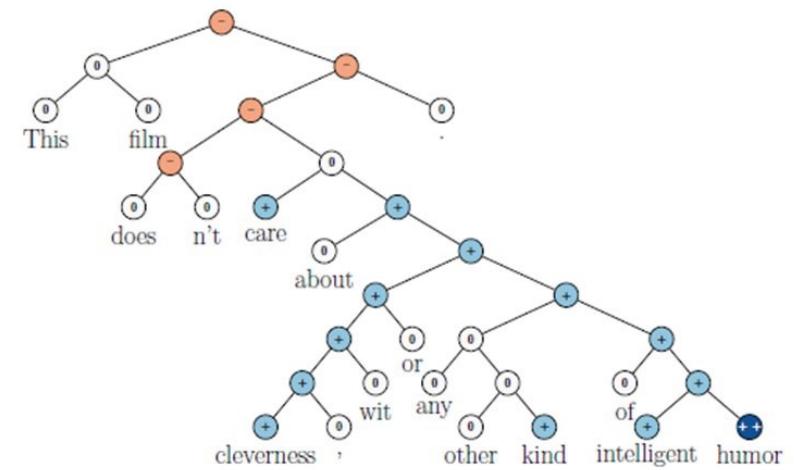


Recursive tensor network

Database :

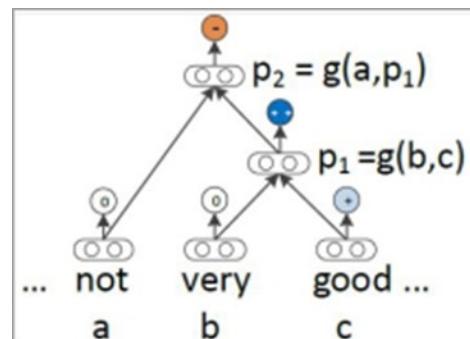
- Treebank Sentiment sentences of movie reviews represented by a tree
- parsed with the Stanford parser :
 - tool that automatically represent a sentence by a binary tree based on the structure of the sentence.
- Labels for each node of the tree :
 - opinions (-, +, 0)

=> provide the structure that is required for the training of a recursive model



Recursive tensor network

- Training step :
 - learning g function that computes the upper outputs in the binary tree
- Decision step :
 - recursively apply the activation functions :

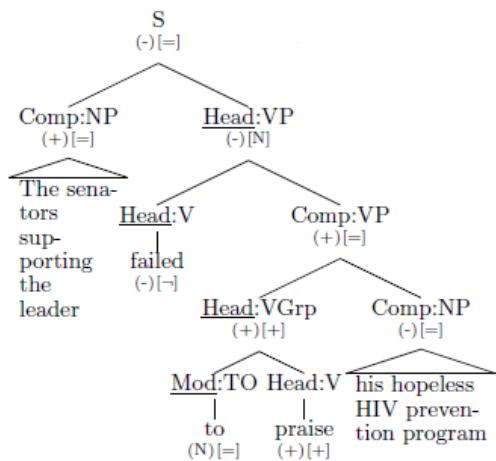


NOTE: Socher's method is mimicking compositional approach using the recursive neural networks

Compositional approach

- Representation of the sentence by constituents
- Calculates the overall polarity of an output constituent from the input constituents

[Moilanen 2007]

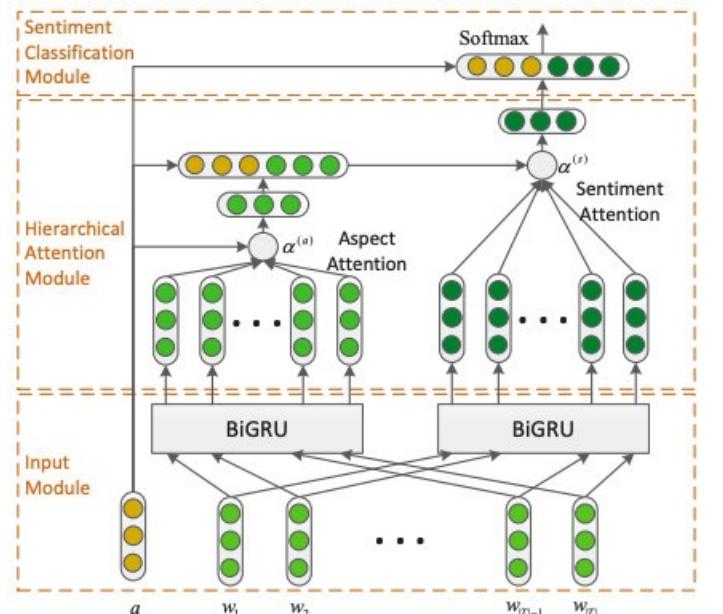


« The senators supporting the leader failed to praise his hopeless HIV prevention program »

- Propagation rules:
 - the polarity of a neutral constituent is "erased" by that of a non-neutral constituent
 - $\{(+)(N)\} \rightarrow (+)$
 - $\{(-)(N)\} \rightarrow (-)$
 - Ex: +(to admire) N(his behavior) -> + (to admire his behavior)
- Inversion rules:
 - $(+) \rightarrow (-)$; $(-) \rightarrow (+)$ in order to deal with negation
 - Ex: never +(succeed)-> -(never succeed)
- Polarity conflict resolution rules:
 - when the two polarities are conflicting at different levels of the syntactic structure

Recurrent Neural Networks for ABSA

Cheng, Jiajun, et al. "Aspect-level sentiment classification with heat (hierarchical attention) network."



Step 1: Encode the text and the target aspect into distributed vector representations

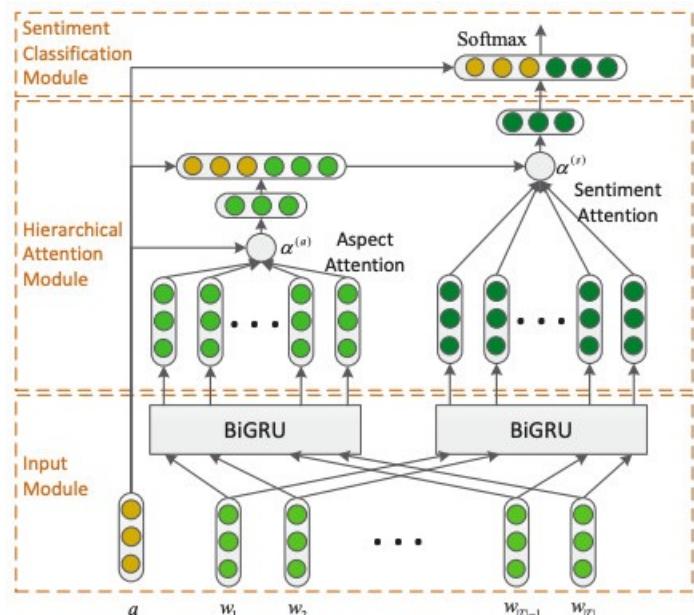
Inputs:

a given target aspect a (ex: food)

A text: ex: « the tastes are great, but the service is dreadful »

Recurrent Neural Networks for ABSA

Cheng, Jiajun, et al. "Aspect-level sentiment classification with heat (hierarchical attention) network."



Inputs:

a given target aspect a (ex: food)

A text: ex: « the tastes are great, but the service is dreadful »

Step 2: Use of bi-directional Gated Recurrent Units and Hierarchical Attention Modules

Under the guidance of the target aspect a : use of **aspect attention** to pay attention to the aspect information

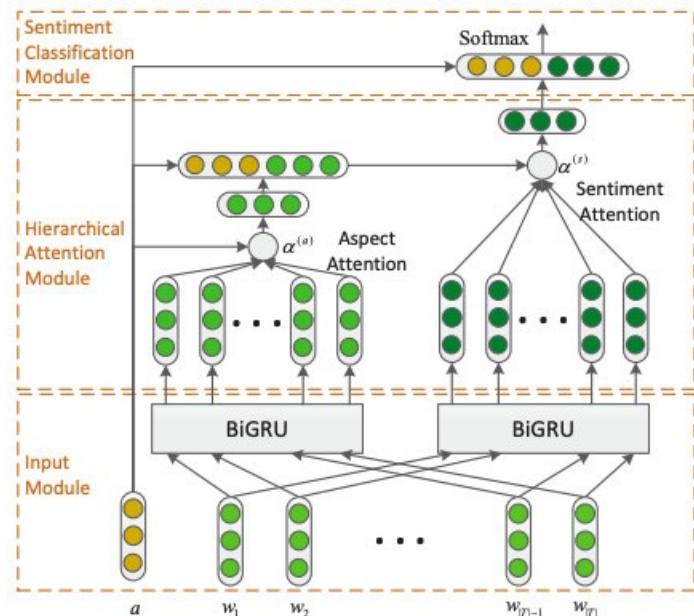
- better allocate aspect-specific sentiment word,
- i.e., aspect terms “tastes” (aspect term \neq aspect (food))

Under the guidance of extracted aspect information (“tastes”), use of **sentiment attention** to capture the sentiment feature of the text

- choose the sentiment word “great”

Recurrent Neural Networks for ABSA

Cheng, Jiajun, et al. "Aspect-level sentiment classification with heat (hierarchical attention) network."



Step 3: Softmax to assign the polarity of the sentiment corresponding to the aspect food.

Inputs:

a given target aspect a (ex: food)

A text: ex: « the tastes are great, but the service is dreadful »

2/Convolution neural networks for text classification tasks

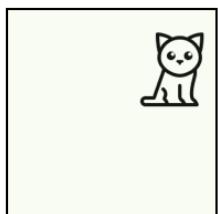
- Convolutional layer to detect sequence of words that are informative to classify a text (here as a negative review)

“This film **should be brilliant**. It **sounds like a great plot**, the actors are first grade, and the supporting cast is good as well, and Stallone **is attempting to deliver a good performance**. However, it **can't hold up**.”

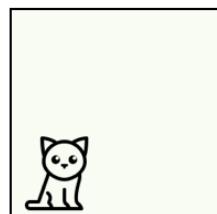
- + CNN can model semantic clues in contextual windows
- CNN have difficulty to preserve sequential orders and to model long-term contextual information

Analogy with CNN used for images

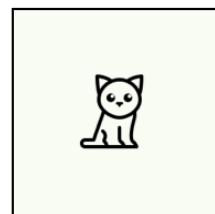
A convolutional layer performs a *position-invariant linear transformation* over local regions of the input sequence.



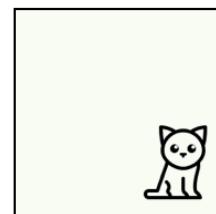
Label: **cat**



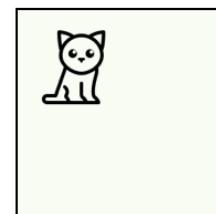
Label: **cat**



Label: **cat**



Label: **cat**

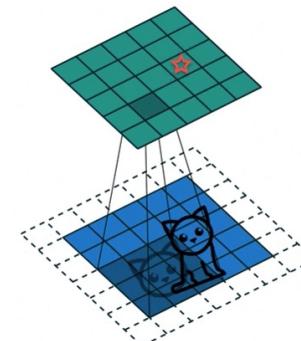


Label: **cat**

We don't care where the cat is,
we care that it is somewhere.

Then why don't we process all
these cats similarly?

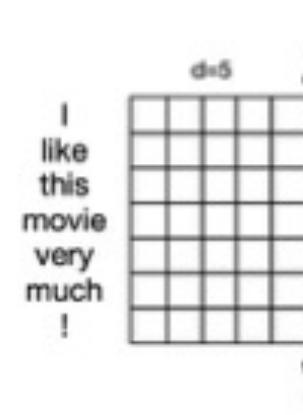
From https://lena-voita.github.io/nlp_course/models/convolutional.html



CNN inputs

- each word is represented by an embedding vector
- matrix representing a sentence,

$$(n_{\text{lines}} = n_{\text{words}}, n_{\text{columns}} = \text{size}_{\text{embeddings}})$$



Retrieve opinion patterns with filters

- A convolutional layer applies the same linear filter to every sliding window of the sentence

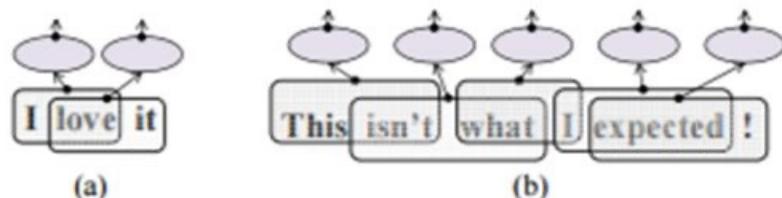


Figure 3: Convolution layer for variable-sized text.

Each window has a size equal to the filter's height:
Here, windows can be 2 words or 3 words in size,
with possible overlaps between the windows

The output of the convolution is therefore a sequence of feature vectors, one for each window, representing the responses of the filter to all local contexts in the sentence.

From Johnson, R., & Zhang, T. (2014). Effective use of word order for text categorization with convolutional neural networks.

Convolution layer: apply linear filters

Each window constitutes a matrix $X \in \mathbb{R}^{h \times d}$, where:

- h the number of tokens in the window
- d is the embedding dimension

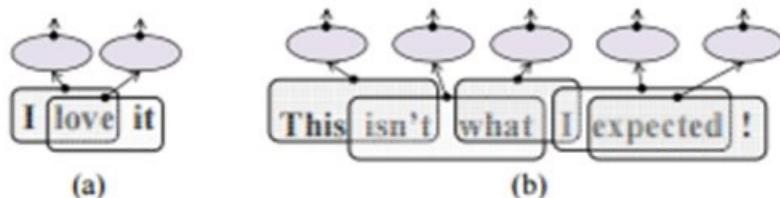
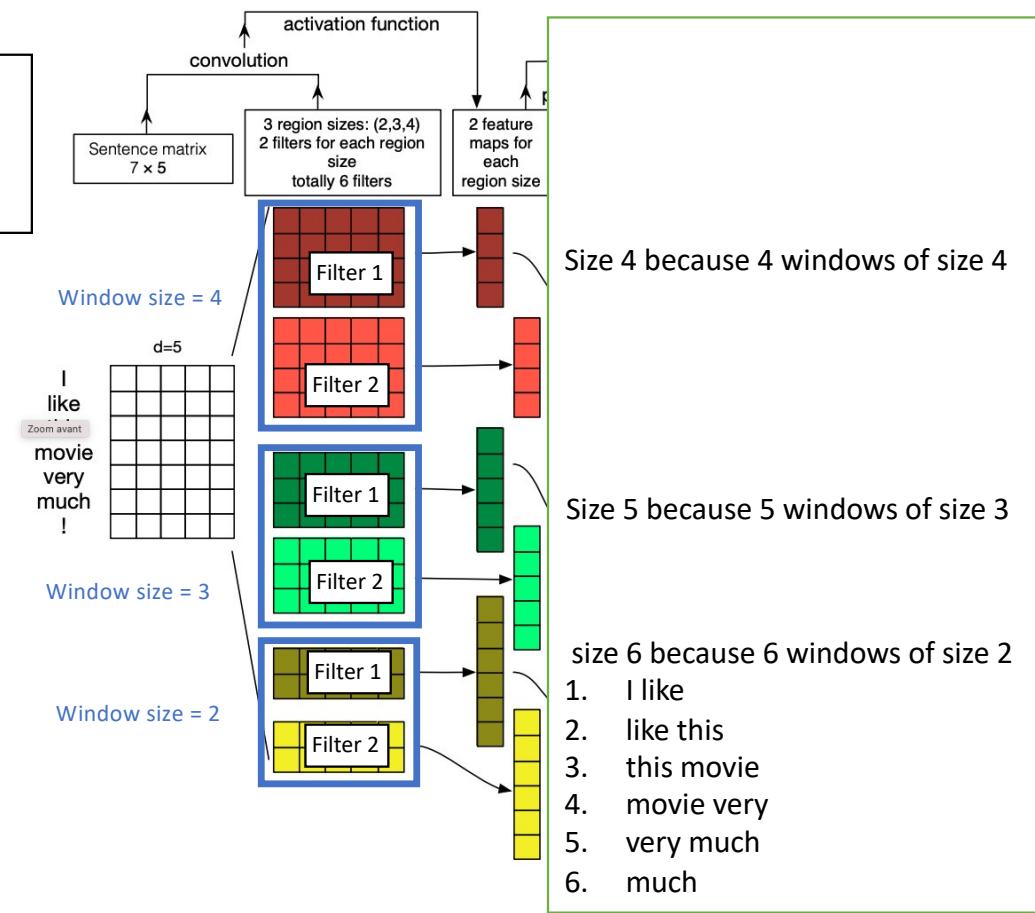


Figure 3: Convolution layer for variable-sized text.

Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.



Convolution layer: apply linear filters

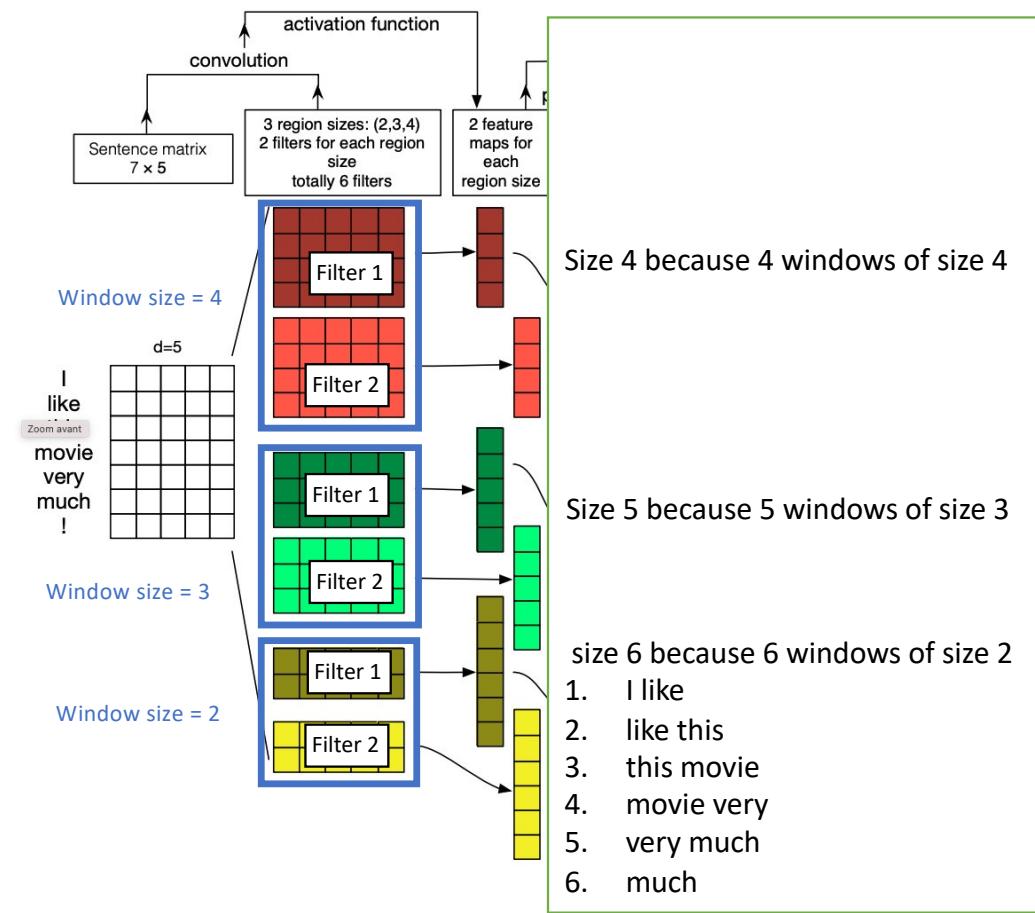
A same convolutional filter of height h is applied to every contiguous window.

Here 2 filters are considered : Filter 1 and Filter 2 with:

- A fixed width = d ;
- Varying heights $h = 4, 3, 2$ (window sizes also called kernel sizes)

Each filter computes a weighted combination of the word embeddings inside the window, producing one output vector per window.

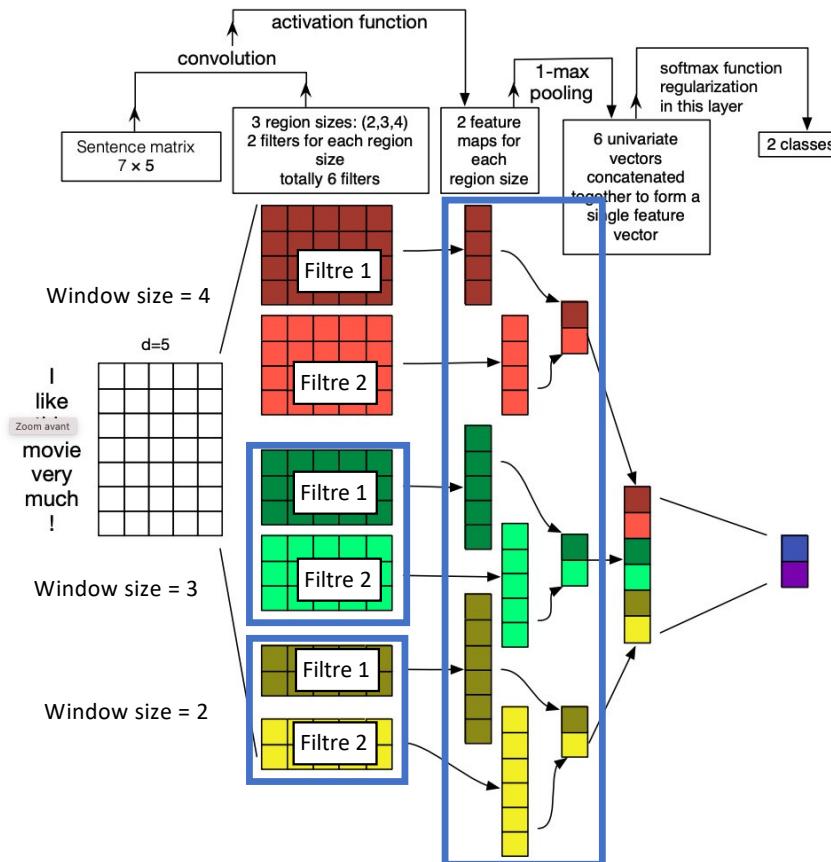
Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.



Pooling layer: summarize the information produced by each convolutional filter

Max-pooling: retain only the strongest activation produced by the filter

Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.

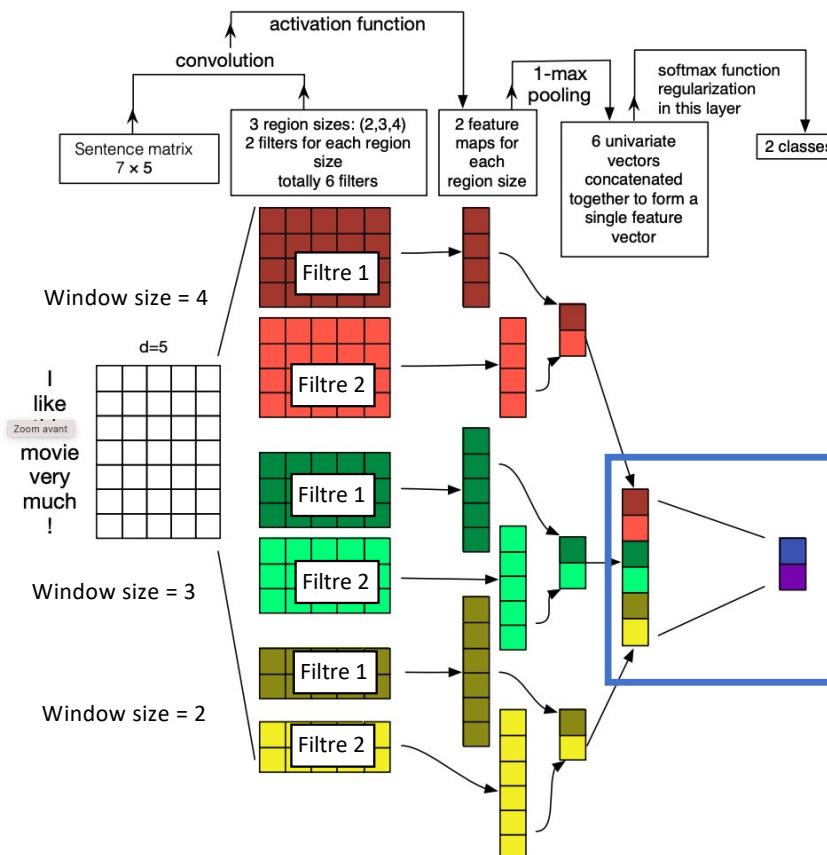


Classification layer

Classify with softmax function

Input: a vector of size 6 (one feature per filter and per height: 2×3)

Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.



CNN and Aspect-Based Sentiment Analysis

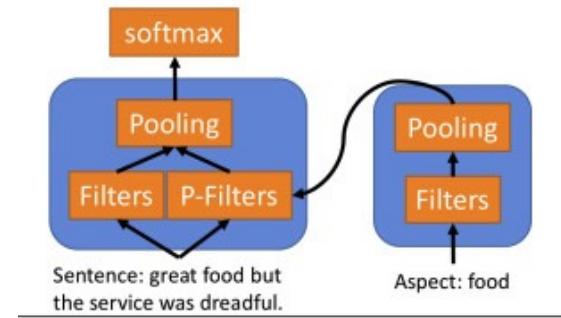
Here the task is : given ...

- a sentence $s = [w_1, w_2, \dots, w_i, \dots, w_n]$ and
- Aspect terms $t = [w_i, w_{i+1}, \dots, w_{i+m-1}]$.

... classify whether the sentiment towards the aspect in the sentence is positive, negative, or neutral (note that here the aspect terms are given).

Method : parametrizing filters using aspect terms

- the output of the CNN applied to aspect terms is used as a convolutional filter applied on the sentence



Huang, B., & Carley, K. M. (2019). Parameterized convolutional neural networks for aspect level sentiment classification.

Transformer-based architectures for sentiment analysis

- A model introduced in *Attention is All You Need in 2017*
- Improvement on recurrent and convolutional architectures:
 - learning **long-range dependencies** between words in the sequence thanks to the **self-attention mechanisms**
 - the representation of a token by the layer k relies on the representation of all tokens by the layer k-1, using an attention mechanism
 - **multi-head ..** : several heads of attention (concentrate simultaneously on several parts of a text)
 - low computational complexity, and parallelization,

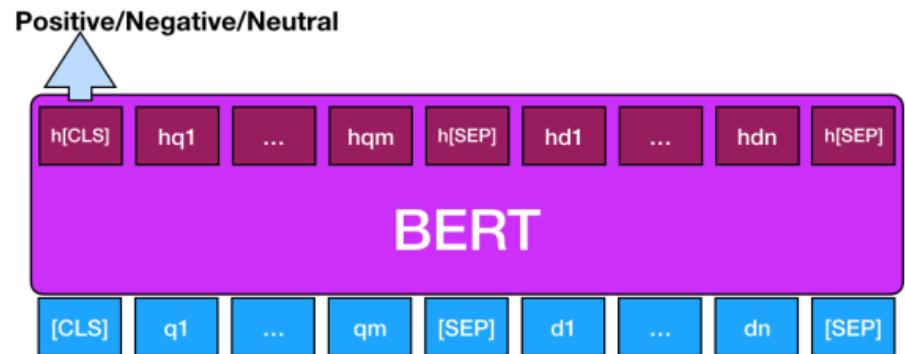
Transformer-based architectures for sentiment analysis

EXAMPLE : “It **sounds like a great plot**, the actors are first grade, and the supporting cast is good as well, and Stallone **is attempting to deliver a good performance**. However it can’t hold up”

- the **self-attention** mechanism
 - computes how important are other words of the sentence to understand the meaning of the current word (ex: “sounds”)
 - a high score means that a word is very important for understanding the corresponding part.
 - could assess the importance of ‘great’ compared with ‘like’, ‘plot’, ‘sounds’, etc.
- the **multi-head** self-attention mechanism
 - Different heads/angles to read a sentence, each **focusing on a different part**
 - E.g. retrieve positive expressions, analyse argumentative structure, etc.
 - To build complementary representations of the sentence
 - the model **combines** the information provided by each head **to obtain a complete view of the sentence.**
 - the model could capture that the first part of the text is a concessive argumentation

Transformer-based architectures for sentiment analysis

- using a pre-trained BERT or its variants to represent each token
- [CLS] token aggregates contextual information from the entire sequence
- fine-tuning it on your data:
 - the last hidden state of [CLS] is usually connected to a classification layer (e.g., a Dense layer with softmax activation) to predict labels.



From [BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis] (Xu et al., NAACL 2019)

Example of specialized BERT

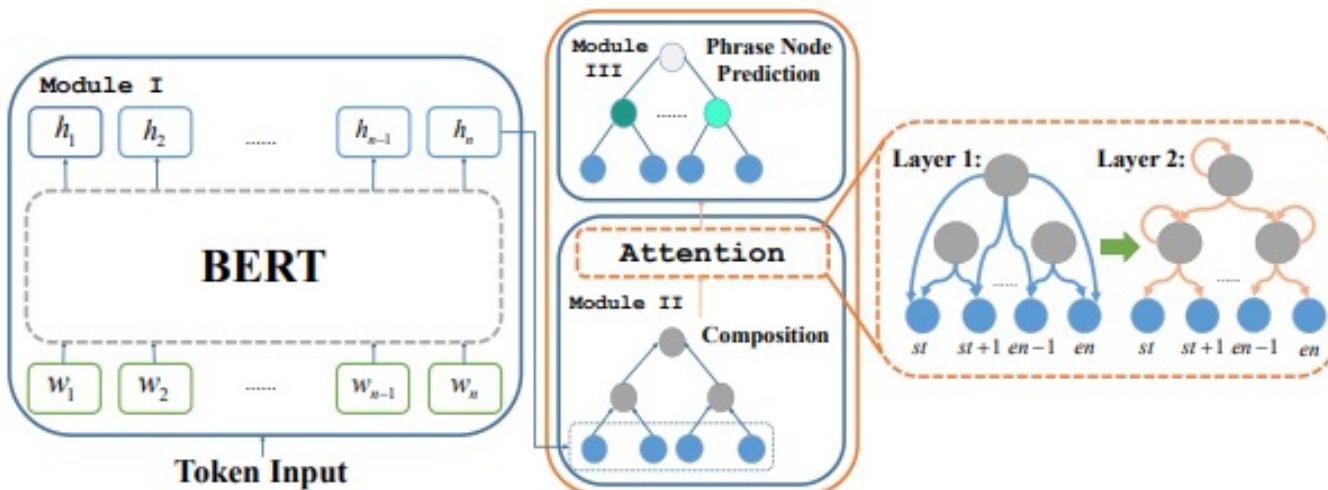
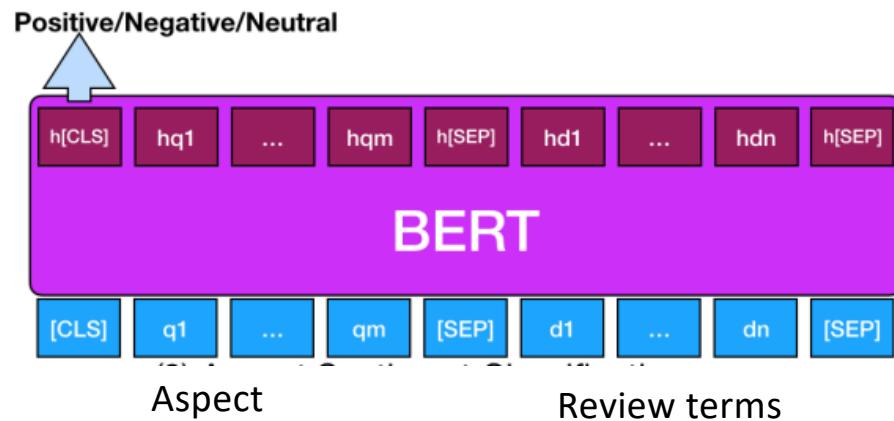


Figure 2: The architecture of SentiBERT. Module I is the BERT encoder; Module II denotes the semantic composition module based on an attention mechanism; Module III is a predictor for phrase-level sentiment. The semantic composition module is a two layer attention-based network (see Section 3.1) The first layer (**Attention to Tokens**) generates representation for each phrase based on the token it covers and the second layer (**Attention to Children**) refines the phrase representation obtained from the first layer based on its children.

SentiBERT – trained on sentiment treebank dataset <https://arxiv.org/abs/2005.04114>

Transformer-based architectures for Aspect-based sentiment analysis



Classify the review : classify the sentiment polarity (positive, negative, or neutral) expressed on a given aspect

From [BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis] (Xu et al., NAACL 2019)

Our research on ABSA using BERT

Deng, L. Panl and C. Clavel (2022), Domain Adaptation for Stance Detection towards Unseen Target on Social Media, 10th International Conference on Affective Computing and Intelligent Interaction (ACII)

And more recently....

Generative models with prompting

- Transform text classification task into a generation or a slot filling task
- ex: slot filling task for sentiment classification

Name	Notation	Example	Description
<i>Input</i>	x	I love this movie.	One or multiple texts
<i>Output</i>	y	++ (very positive)	Output label or text
<i>Prompting Function</i>	$f_{\text{prompt}}(x)$	[X] Overall, it was a [Z] movie.	A function that converts the input into a specific form by inserting the input x and adding a slot [Z] where answer z may be filled later.
<i>Prompt</i>	x'	I love this movie. Overall, it was a [Z] movie.	A text where [X] is instantiated by input x but answer slot [Z] is not.
<i>Filled Prompt</i>	$f_{\text{fill}}(x', z)$	I love this movie. Overall, it was a bad movie.	A prompt where slot [Z] is filled with any answer.
<i>Answered Prompt</i>	$f_{\text{fill}}(x', z^*)$	I love this movie. Overall, it was a good movie.	A prompt where slot [Z] is filled with a true answer.
<i>Answer</i>	z	“good”, “fantastic”, “boring”	A token, phrase, or sentence that fills [Z]

Table 2: Terminology and notation of prompting methods. z^* represents answers that correspond to true output y^* .

From Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing,
<https://arxiv.org/pdf/2107.13586v1.pdf>

Sentiment analysis as a generation task

C

What is the opinion of the writer of this movie review? "This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."



Based on the given movie review, the writer's opinion is that the film had the potential to be brilliant. They express positive expectations about the plot, the actors (including Stallone), and the supporting cast. However, the writer ultimately concludes that the film does not live up to those expectations and falls short in some way, stating, "it can't hold up." It seems that the writer found the film disappointing or lacking despite its promising elements.

C

Is the movie review positive or negative?



The movie review can be considered negative. While the writer initially expresses positive expectations for the film, stating that it "should be brilliant" and highlighting various positive aspects such as the plot and the actors, the overall tone of the review becomes negative when they conclude that the film "can't hold up." This indicates that the film ultimately falls short of the writer's expectations and fails to deliver a satisfactory experience.

More recently...

Generative models with prompting

From ChatGPT, 7 June 2023

How do generative models work? Using large language models

- **Transformer**-based pre-trained language models
- **Generative pre-training objective**: predict the next token from a long previous context (ex: 2048 tokens in GPT3)
- GPT3:
 - trained on a lot of data: 300 billion tokens : Wikipedia, books corpora, and crawled webpages (CommonCrawl, WebText2)
 - 175 billions of parameters, (vs. 150 billion for GPT-2)
 - **LLMs increase their capability as the size of their input datasets and parameter space increase.**
 - Example of GPT-n models :
 - <https://beta.openai.com/docs/guides/embeddings/use-cases>

Sentiment analysis as a generation task

Previous tokens

C

What is the opinion of the writer of this movie review? "This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

Next most probable tokens



Based on the given movie review, the writer's opinion is that the film had the potential to be brilliant. They express positive expectations about the plot, the actors (including Stallone), and the supporting cast. However, the writer ultimately concludes that the film does not live up to those expectations and falls short in some way, stating, "it can't hold up." It seems that the writer found the film disappointing or lacking despite its promising elements.

C

Is the movie review positive or negative?



The movie review can be considered negative. While the writer initially expresses positive expectations for the film, stating that it "should be brilliant" and highlighting various positive aspects such as the plot and the actors, the overall tone of the review becomes negative when they conclude that the film "can't hold up." This indicates that the film ultimately falls short of the writer's expectations and fails to deliver a satisfactory experience.

More recently...
Generative models with prompting

From ChatGPT, 7 June 2023

What is behind ChatGPT(3)?

- Large language models used alone do not provide so relevant answer
- Idea: incorporate human supervision
 - STEP 1: Fine-tuning GPT 3 with supervised learning
 - STEP 2: feedback for the training **using reinforcement learning by human feedback**

Step 1

**Collect demonstration data,
and train a supervised policy.**

A prompt is
sampled from our
prompt dataset.

A labeler
demonstrates the
desired output
behavior.

This data is used
to fine-tune GPT-3
with supervised
learning.



STEP 1- SUPERVISED Fine-tuning GPT 3

- Data used for the supervision :
 - 13,000 input / output samples = prompts collected from the API + labellers providing human answers to the prompt
 - Supervised fine-tuning of GPT 3
 - After this step, the model generates better aligned responses to user prompts.

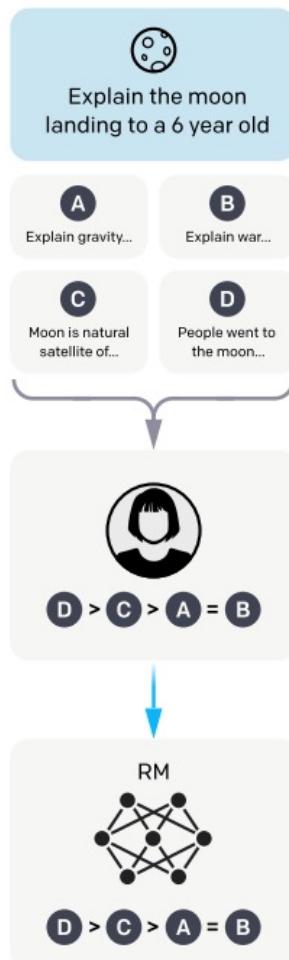
<https://arxiv.org/pdf/2203.02155.pdf>

Collect comparison data,
and train a reward model.

A prompt and
several model
outputs are
sampled.

A labeler ranks
the outputs from
best to worst.

This data is used
to train our
reward model.



STEP 2 – reinforcement learning from human feedback

STEP 2a: REWARD MODEL

Collect comparison data:

Given the SFT model of step 1, generate the n^{th} best answers (with n from 4 to 9) to a same prompt

Ask a labeler to rank the outputs

Train a **reward model** (RM) on this data:

a model able to provide a score to each (prompt,answer) pair that could reproduce the labeller ranking



Note: the reward model is trained to reproduce the preference of labellers

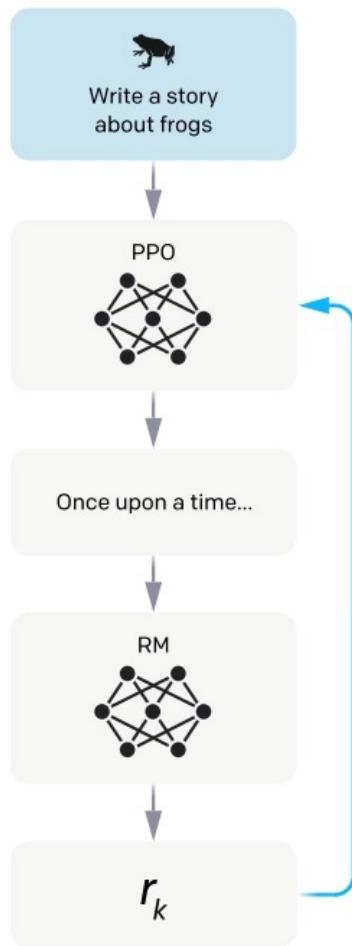
Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy



STEP 2 – reinforcement learning from human feedback

STEP 2b: Proximal Policy Optimization (PPO)

Policy : a strategy that the machine learns to use to achieve its goal

Policy optimisation: updating the model's policy as each response is generated according to

***A Kullback-Leiber penalty** to reduce the distance of the response from the SFT model output of step 1

***The reward value computed in step 2a.**

SECOND SOLUTION :
Integrate sentiment knowledge
within neural models

How to encode sentiment-related information?

Integrating lexical, semantic and syntactic knowledge of sentiment phenomenon:

1/ Use resources

- word-level : sentiment lexicon storing lexical/semantic information (ex: SentiWordNet)
- sentence-level: linguistic/syntactic patterns

2/ Integrate these resources into neural models

Word-level resources

SENTIMENT LEXICONS

WordNet: Retrieve information on word meaning/sense

- Core idea :
 - A word can have several meanings (ex: « bat »)
 - groups English words into *synsets*
 - *Synsets* : set of synonyms

WordNet

- Synsets : set of synonyms
- Synonyms : words that are interchangeable in some context without changing the truth value of the proposition
- Synsets include simplex words as well as collocations like "eat out" and "car pool."
- The meaning of a synset is further clarified with a short definition and one or more usage examples

Example :

good, right, ripe – (most suitable or right for a particular purpose;
"a good time to plant tomatoes"; "the right time to act"; "the
time is ripe for great sociological changes")

WordNet

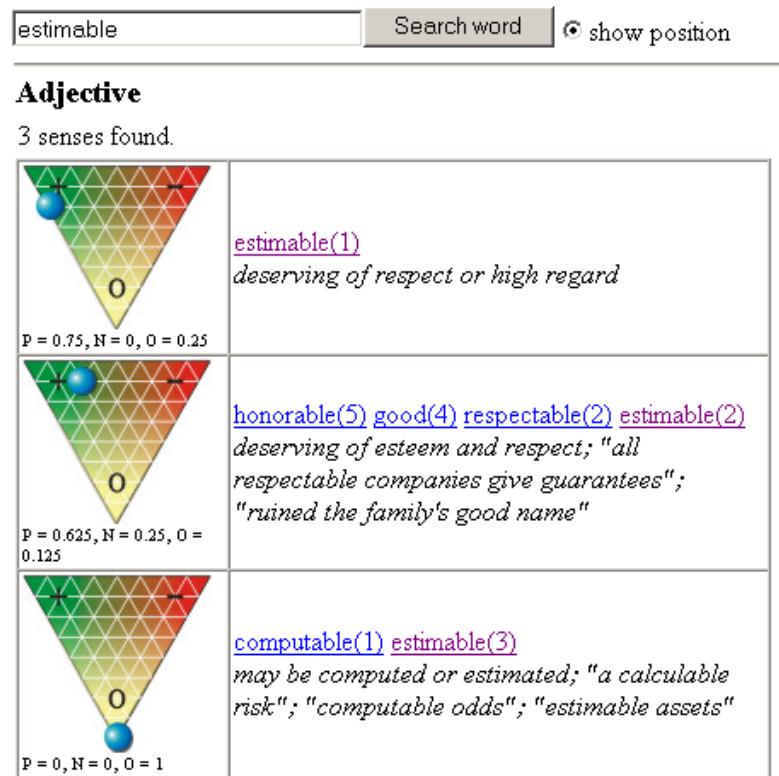
- All synsets are connected to other synsets by means of semantic relations:
 - Ex: canine is a hypernym of dog
 - Ex: window is a meronym of building

Version française : Wordnet Libre du Français (WOLF) :
<http://alpage.inria.fr/~sagot/wolf.html>

SentiWordNet

- <http://sentiwordnet.isti.cnr.it/>
- Principle : add to each synset a positive score, a negative score AND an objective score between 0 and 1
- [estimable(J,3)] “may be computed or estimated”
Pos 0 Neg 0 Obj 1
- [estimable(J,1)] “deserving of respect or high regard”
Pos .75 Neg 0 Obj .25

SentiWordNet



[main page](#)

(c) Andrea Esuli 2005 - andrea.esuli@isti.cnr.it

WordNet Affect

- Selecting a subset of wordnet Affective label + valence

A-Labels	Examples
EMOTION	noun anger#1, verb fear#1
MOOD	noun animosity#1, adjective amiable#1
TRAIT	noun aggressiveness#1, adjective competitive#1
COGNITIVE STATE	noun confusion#2, adjective dazed#2
PHYSICAL STATE	noun illness#1, adjective all_in#1
EDONIC SIGNAL	noun hurt#3, noun suffering#4
EMOTION-ELICITING SITUATION	noun awkwardness#3, adjective out_of_danger#1
EMOTIONAL RESPONSE	noun cold_sweat#1, verb tremble#2
BEHAVIOUR	noun offense#1, adjective inhibited#1
ATTITUDE	noun intolerance#1, noun defensive#1
SENSATION	noun coldness#1, verb feel#3

Table 4: A-Labels and corresponding example synsets

LIWC

- LIWC (Linguistic Inquiry and Word Count) - Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007).
- Home page: <http://www.liwc.net/>
- 2300 mots, >70 classes
- One word= one sense (no disambiguation depending on the context)
- Version française : http://sites.univ-provence.fr/wpsycle/outils_recherche/liwc/FrenchLIWCDictionary_V1_1.dic

English LIWC

		Examples	#Words in category
Psychological Processes			
Affective processes	affect	happy, cried	1393
Positive emotion	posemo	love, nice, sweet	620
Negative emotion	negemo	hurt, ugly, nasty	744
Anxiety	anx	worried, fearful	116
Anger	anger	hate, kill, annoyed	230
Sadness	sad	crying, grief, sad	136



Ex. Useful for the identification of appraisal expressions and their corresponding psychological process

Linguistic Dimensions			
Total function words	funct	it, to, no, very	491
Total pronouns	pronoun	I, them, itself	153
Personal pronouns	ppron	I, them, her	93
1st pers singular	i	I, me, mine	24
1st pers plural	we	we, us, our	12
2nd person	you	you, your, thou	30
3rd pers singular	shehe	she, her, him	17
3rd pers plural	they	they, their, they'd	11
Impersonal pronouns	ipron	it, it's, those	59



Ex. Useful for the identification of the sources

Other lexicons

- MPQA lexicon (Kiritchenko et al. [2014](#)),
- Sentiment140 lexicon (Kiritchenko et al. [2014](#)),
- NRC Hashtag Sentiment Lexicon (Mohammad et al. [2013](#))

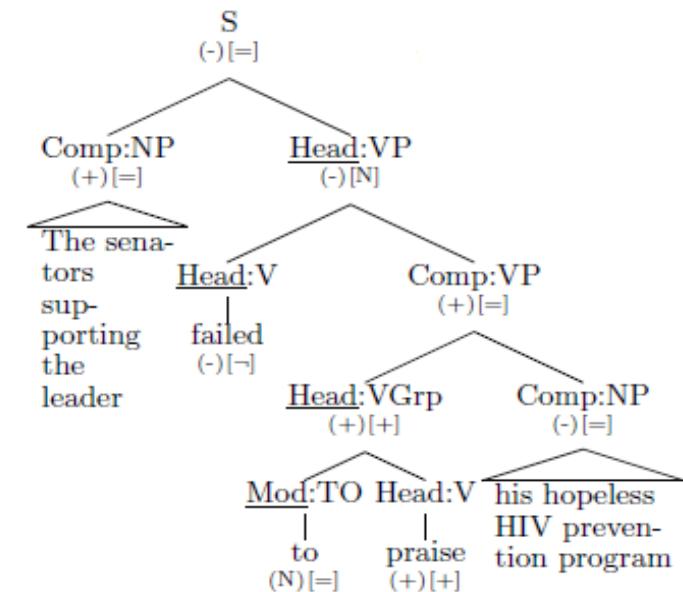
Sentence-level resources

Linguistic patterns/Rules

Syntactic-based rules for sentiment analysis

- Rules that are built on the syntactic structure of the sentence and on lexical resources

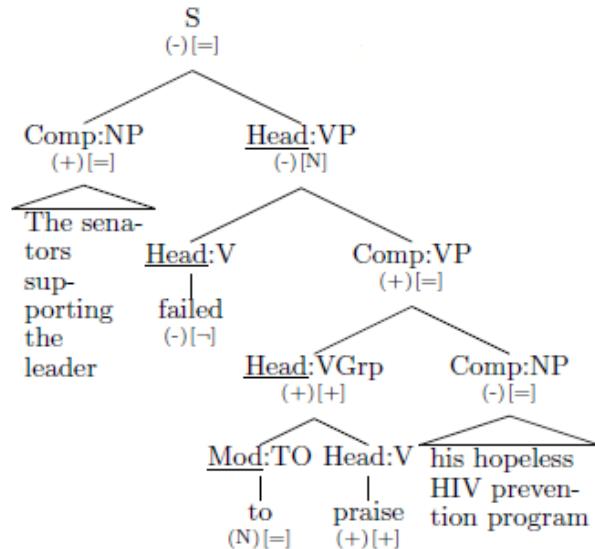
Require syntactic analysis :
Part Of Speech tagging,
Chunking
Dependency parsing



Rules with compositional approach

Representation of the sentence by constituents

[Moilanen 2007]

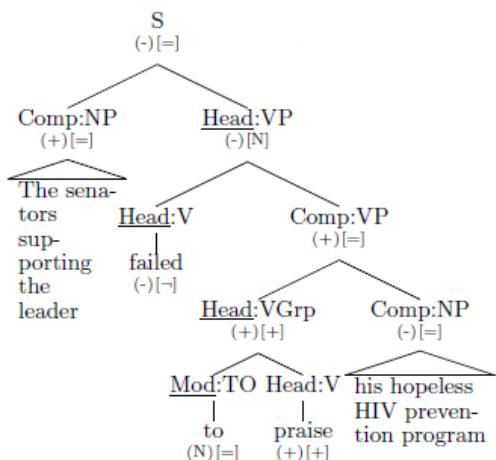


« The senators supporting the leader failed to praise his hopeless HIV prevention program »

Rules with compositional approach

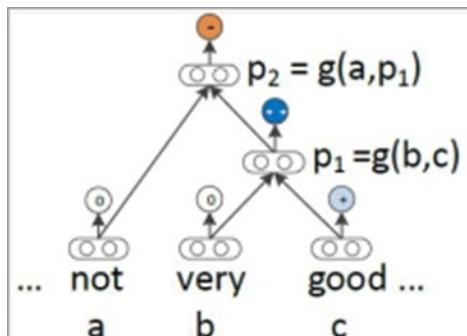
Calculates the overall polarity of an output constituent from the input constituents

[Moilanen 2007]



- Propagation rules:
 - the polarity of a neutral constituent is "erased" by that of a non-neutral constituent
 - $\{(+)(N)\} \rightarrow (+)$
 - $\{(-)(N)\} \rightarrow (-)$
 - Ex: +(to admire) N(his behavior) -> + (to admire his behavior)
- Inversion rules:
 - $(+) \rightarrow (-)$; $(-) \rightarrow (+)$ in order to deal with negation
 - Ex: never +(succeed)-> -(never succeed)
- Polarity conflict resolution rules:
 - when the two polarities are conflicting at different levels of the syntactic structure

Reminder: recursive tensor network and transformers mimicking the compositional approach



R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, EMNLP 2013

Use of recursive tensor networks and transformers to learn compositionality

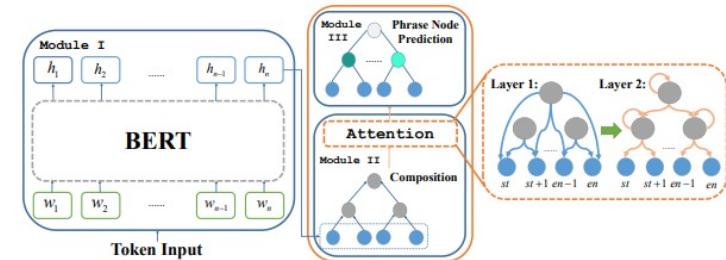


Figure 2: The architecture of SentiBERT. Module I is the BERT encoder; Module II denotes the semantic composition module based on an attention mechanism; Module III is a predictor for phrase-level sentiment. The semantic composition module is a two layer attention-based network (see Section 3.1) The first layer (**Attention to Tokens**) generates representation for each phrase based on the token it covers and the second layer (**Attention to Children**) refines the phrase representation obtained from the first layer based on its children.

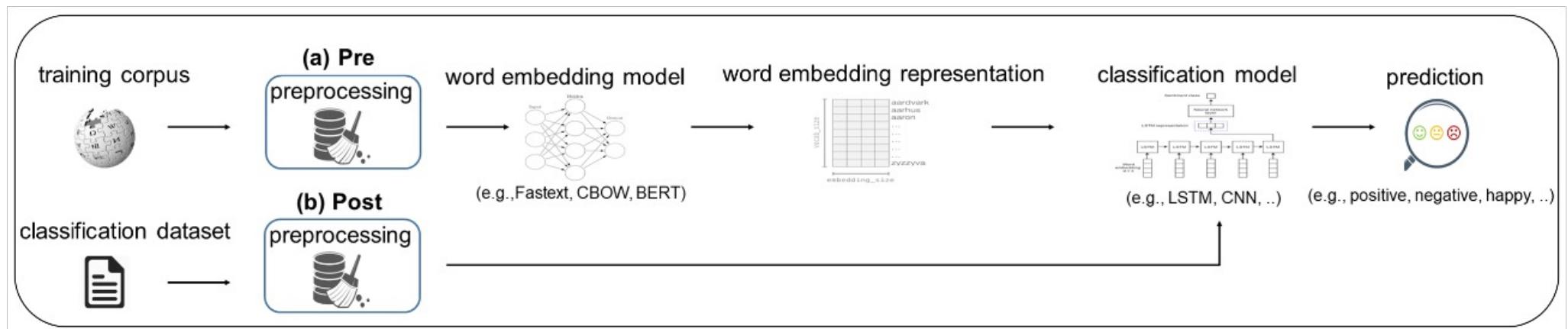
SentiBERT

How lexicons and rules are used
in neural models?

Seq. of words (input) -- **Pre-processing** -- ML Model --> sentiment

Integrate lexicons and rules at the preprocessing step

- Use of **lexicons and rules** to preprocess the data before the learning of word embeddings
 - Negation + negated word -> antonym of the negated word; ex: I am **not happy** -> I am **sad**



Babanejad, Nastaran, et al. "A comprehensive analysis of preprocessing for word representation learning in affective tasks." ACL 2020.

Seq. of words (input)–**knowledge-based features**—ML Model--> sentiment

Integrate lexicons and rules: when building feature vectors

Abdi A, Mariyam S, Hasan S, Piran J (2019) Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. Inf Process Manag 56(4):1245–1259

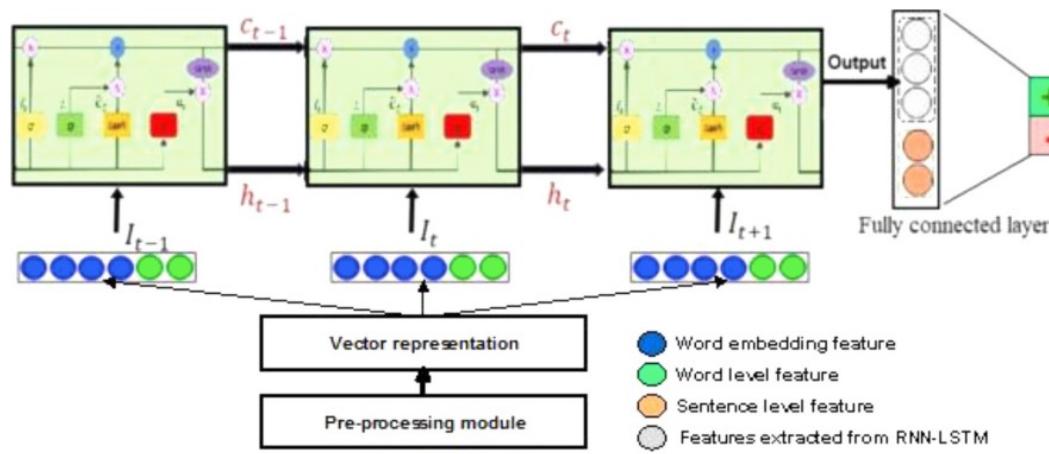


Fig. 1. The architecture of the proposed method RNSA.

Example of vector components at the sentence level:

The number of occurrences of a token in a certain category given

- a **lexicon** (ex: number of negative sentiment tokens in a sentence)
- a **grammatical category** (ex: frequency of nouns, adjectives, verbs, adverb)

Ignore part of a sentence for building the text vector using rules

- Ex: “If a sentence includes “however”, ignore the sentence before “however” and consider the sentiment of the sentence after “however”. “Stallone is attempting to deliver a good performance, however, it can’t hold up.”

Integrate lexicons and rules: when building feature vectors

*Seq. of words (input)--**knowledge-based features**—ML Model--> sentiment*

Sharing some similar work from our team

- Combine word embeddings representations and knowledge-based features (using lexicons and rules) to model :

- Opinions

*Barrière, V., Clavel, C., Essid, E., **Opinion Dynamics** Modeling for Movie Review Transcripts Classification with Hidden Conditional Random Fields, Interspeech 2017*

- Hedge

*"You might think about slightly revising the title": Identifying **Hedges** in Peer-tutoring Interactions (Raphalen et al., ACL 2022)*

Transforming syntactic rules into features

See: D.-T. Vo and Y. Zhang, “Target-dependent Twitter sentiment classification with rich automatic features,” in Proc. 24th Int. Joint Conf. Artif. Intell., Jun. 2015, pp. 1347–1353.

Identify the target terms (the target iPhone in “I love iPhone”) and define syntactic rules (“love” is a transitive verb and the target (Iphone) is its object)

-> *rich context extraction* (extraction of the relevant context of the target) for a better representation of the tweet.

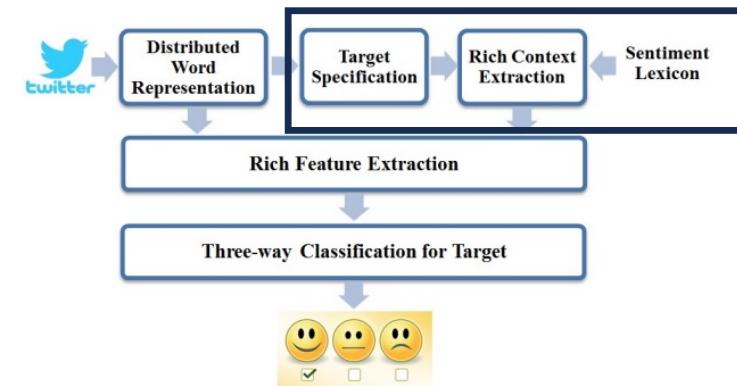


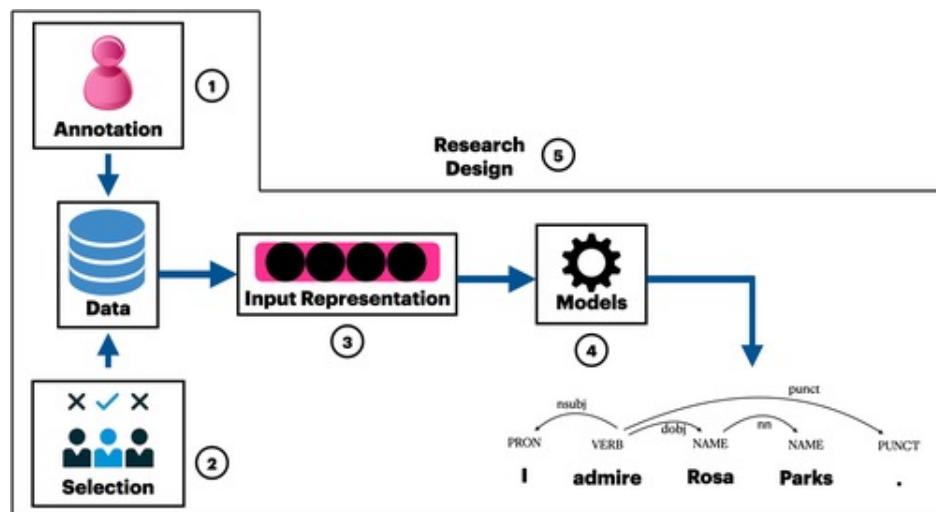
Figure 1: System Architecture.

The importance of the datasets in NLP

Example of Sentiment Analysis Datasets

Importance of the datasets in NLP

- Sufficient quantity and quality of data and annotations => quality of learned models



Biases in the data and in the annotation => biases in the models

five sources where bias can occur in NLP systems:

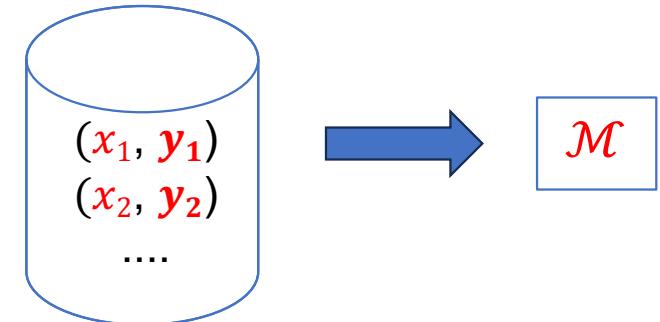
- (1) the data,**
- (2) the annotation process,**
- (3) the input representations,
- (4) the models
- (5) the research design (or how we conceptualize our research)

From [Hovy and Prabhumoye, 2021]

1st lever of transparency: the design of annotated datasets

Models inherently reflect...

- **their training data** → *How socio-emotional phenomena are expressed by writers/speakers*
- **annotation choices and biases** → *How the data was annotated*



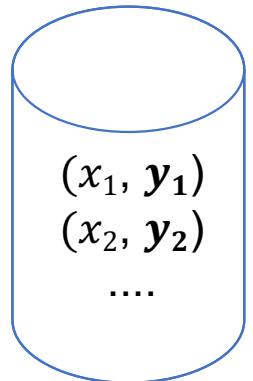
Keep in mind: Socio-emotional phenomena are complex, then...

- Data needs to reflect the high variety of socio-emotional expressions
- Annotation requires high human expertise
- Achieving consensus on labels is often difficult

1st lever of transparency: the design of annotated datasets



« Researchers have to be mindful of the entire research design: **data sets they choose, the annotation schemes or labelling procedures they follow, how they decide to represent the data, the algorithms they choose for the task and how they evaluate the automated systems.**» [Hovy and Prabhumoye, 2021]



\mathcal{M}

How to build a labelled dataset?

- The two steps
 - Data collection/choice
 - Data annotation

Data collection/choice

PRACTICE: consider these texts, which come from different data sets, and name the factors that distinguish them

lol, A +, mouhahaha

« I have a dream that one day this nation will rise up and live out the true meaning of its creed: "We hold these truths to be self-evident, that all men are created equal." »

Disfluences combinées Vous regardez les 5 derniers chiffres des chi des numéros gravés, pas les chiffres qui défilent hein

A: I'm **worried** about something.
B: What's that?
A: Well, I have to drive to school for a meeting this morning, and I'm going to end up getting stuck in rush-hour traffic.
B: That's **annoying**, but nothing to worry about. *Just breathe deeply when you feel yourself getting upset.*
A: Ok, I'll try that.
B: Is there anything else **bothering** you?
A: Just one more thing. A school called me this morning to see if I could teach a few classes this weekend and I don't know what to do.
B: Do you have any other plans this weekend?
A: I'm supposed to work on a paper that'd due on Monday.
B: *Try not to take on more than you can handle.*
A: You're right. I probably should just work on my paper. **Thanks!**

Figure 1: An example in **DailyDialog** dataset. Some text is shortened for space. Best viewed in color.

Data collection/choice

spoken (using transcripts)

Disfluences combinées Vous regardez les 5 derniers chiffres des chi des numéros gravés, pas les chiffres qui défilent hein

« I have a dream that one day this nation will rise up and live out the true meaning of its creed: "We hold these truths to be self-evident, that all men are created equal." »

written

lol, A +, mouhahaha

A: I'm **worried** about something.
B: What's that?
A: Well, I have to drive to school for a meeting this morning, and I'm going to end up getting stuck in rush-hour traffic.
B: That's **annoying**, but nothing to worry about.
Just breathe deeply when you feel yourself getting upset.
A: Ok, I'll try that.
B: Is there anything else **bothering** you?
A: Just one more thing. A school called me this morning to see if I could teach a few classes this weekend and I don't know what to do.
B: Do you have any other plans this weekend?
A: I'm supposed to work on a paper that'd due on Monday.
B: *Try not to take on more than you can handle.*
A: You're right, I probably should just work on my paper. **Thanks!**

Figure 1: An example in **DailyDialog** dataset. Some text is shortened for space. Best viewed in color.

Data collection/choice

w/o interactions

monologues

« I have a dream that one day this nation will rise up and live out the true meaning of its creed: “We hold these truths to be self-evident, that all men are created equal.” »

dialogues

A: I'm worried about something.
B: What's that?
A: Well, I have to drive to school for a meeting this morning, and I'm going to end up getting stuck in rush-hour traffic.
B: That's annoying, but nothing to worry about. *Just breathe deeply when you feel yourself getting upset.*
A: Ok, I'll try that.
B: Is there anything else bothering you?
A: Just one more thing. A school called me this morning to see if I could teach a few classes this weekend and I don't know what to do.
B: Do you have any other plans this weekend?
A: I'm supposed to work on a paper that'd due on Monday.
B: Try not to take on more than you can handle.
A: You're right. I probably should just work on my paper. Thanks!

lol, A +, mouhahaha

Figure 1: An example in **DailyDialog** dataset. Some text is shortened for space. Best viewed in color.

Disfluences combinées Vous regardez les 5 derniers chiffres des chi des numéros gravés, pas les chiffres qui défilent hein

Data collection/choice

Natural/collected through a real application

lol, A +, mouhahaha

Disfluences combinées Vous regardez les 5 derniers chiffres des chi des numéros gravés, pas les chiffres qui défilent hein

In real applications, corpora contain spontaneous expression and can be ‘*wild*’ [Schuller et al., 2016] (i.e, contain noisy text)

simulated/scripted/prepared

A: I'm **worried** about something.
B: What's that?
A: Well, I have to drive to school for a meeting this morning, and I'm going to end up getting stuck in rush-hour traffic.
B: That's **annoying**, but nothing to worry about. *Just breathe deeply when you feel yourself getting upset.*
A: Ok, I'll try that.
B: Is there anything else **bothering** you?
A: Just one more thing. *A school called me this morning to see if I could teach a few classes this weekend and I don't know what to do.*
B: Do you have any other plans this weekend?
A: I'm supposed to work on a paper that'd due on Monday.
B: *Try not to take on more than you can handle.*
A: You're right. I probably should just work on my paper. **Thanks!**

Figure 1: An example in **DailyDialog** dataset. Some text is shortened for space. Best viewed in color.

« I have a dream that one day this nation will rise up and live out the true meaning of its creed: “We hold these truths to be self-evident, that all men are created equal.” »

Data collection/choice : First question: what type of data is being used?

It impacts **the type of linguistic phenomena present in the data**

- Ex1: Spoken transcripts of call-centre data contain disfluencies
- Ex 2: written conversations contain typos, or chat features

lol, A +, mouhahaha

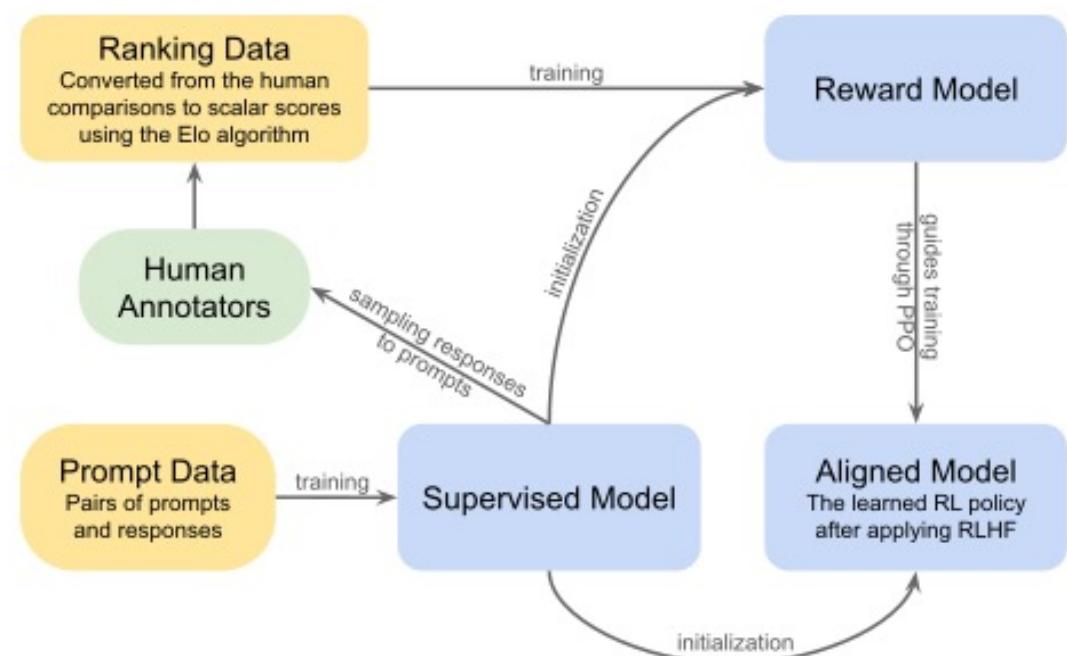
Disfluences combinées Vous regardez les 5 derniers chiffres des chi des numéros gravés, pas les chiffres qui défilent hein

Data collection/choice: Second question: what protocol for data collection?

- Data **collected via crowdsourcing platforms** (workers)
 - Ex: Empathetic Dialogues (Rashkin et al., 2018)
 - Some workers are asked to start the conversation following an emotional prompt.
 - Others have to reply.
- Open data **available on the web** and on social networks
 - Ex: GoEmotions retrieved from reddit
- Data **collected within companies**
 - Ex: call-center transcripts, chatbot interactions, complaint emails

2nd Step: Data annotation

- To supervise the training machine learning models + to evaluate them
- Or for ILM alignment (e.g. using reinforcement learning by human feedback)



Training language models to follow instructions with human feedback (Ouyang et al., Neurips 2022)

Data Annotation Example

TABLE I
EXAMPLES OF TWEET-TARGET PAIRS FROM SEMEVAL2016-T6 DATASET AND COVID-19-STANCE DATASET.

Dataset	Tweet	Target	Stance	Emotion
SemEval2016-T6	Job should always go to best candidate, regardless of gender. Gender shouldn't even matter anymore, it's 2015! #PaulHenry #SemST	Feminist Movement	In Favor	Positive
SemEval2016-T6	We are actually watching a video on radical feminism in history this is the funniest movie ive ever seen. #SemST	Feminist Movement	Against	Positive
COVID-19-STANCE	@realDonaldTrump @Mike_Pence What a disaster of a group. Not everyone is wearing masks. What are you people thinking?! Lead by example.	Wearing a Face Mask	In Favor	Negative
COVID-19-STANCE	It's amazing how many people just roll over and wear masks despite a preponderance of evidence that they dont help nor are they even necessary.	Wearing a Face Mask	Against	Negative

To define:

The annotation unit
(sentence, document, word)

The labels

Define the categories that will be relevant for
your application

RELEVANT LABELS DEPEND ON THE APPLICATIONS AND STUDIED DATA

- Ex: Stance (in favor or against a specific target) is used for opinion analysis in social networks

TABLE I
EXAMPLES OF TWEET-TARGET PAIRS FROM SEMEVAL2016-T6 DATASET AND COVID-19-STANCE DATASET.

Dataset	Tweet	Target	Stance	Emotion
SemEval2016-T6	Job should always go to best candidate, regardless of gender. Gender shouldn't even matter anymore, it's 2015! #PaulHenry #SemST	Feminist Movement	In Favor	Positive
SemEval2016-T6	We are actually watching a video on radical feminism in history this is the funniest movie ive ever seen. #SemST	Feminist Movement	Against	Positive
COVID-19-STANCE	@realDonaldTrump @Mike_Pence What a disaster of a group. Not everyone is wearing masks. What are you people thinking?! Lead by example.	Wearing a Face Mask	In Favor	Negative
COVID-19-STANCE	It's amazing how many people just roll over and wear masks despite a preponderance of evidence that they dont help nor are they even necessary.	Wearing a Face Mask	Against	Negative

Data Annotation: provide as much context as possible to the annotator

Societal context

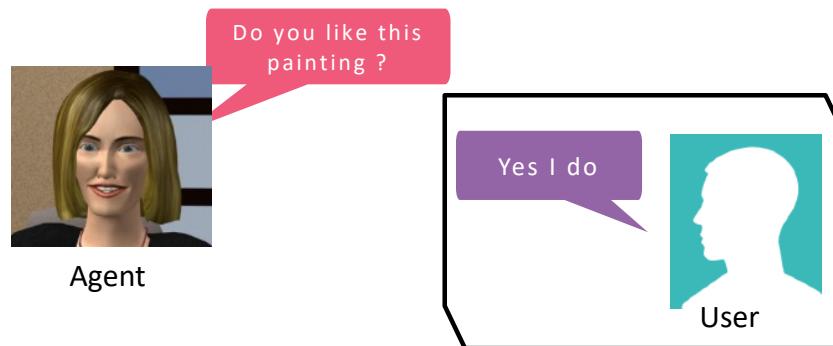
- who is the writer?
- what is its claimed political affiliation?
- What is the current political context?

⇒ help to better understand the opinion expressed

Data Annotation: provide as much context as possible to the annotator

- **Conversational context**

- Opinions can be very difficult to understand without at least the preceding speaker turn (Langlet et al., 2016)



- Founta et al. (2018) showed a decrease in the percentage of abusive labels from 18% to 10% when conversational context was provided

Founta et al. (2018). Large Scale
Crowdsourcing and Characterization
of Twitter Abusive Behavior.

⇒ help to better understand the opinion expressed

Provide as much context as possible to the annotator

- **Multimodal context**
 - Verbal content
 - Audio (prosody, voice quality)
 - Video (gesture, posture, facial expressions)



⇒ help to better understand the opinion expressed

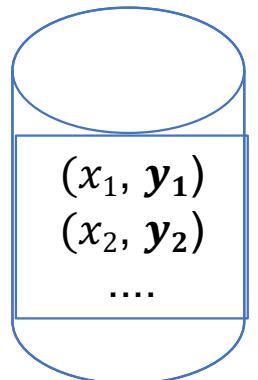
Data Annotation: Do not ignore human label variations



Classical Research practices :

Reflect consensus despite human label variation

- Ask several annotators and measure their agreement
- Build a consensus from majority
- Train a model from the labelled corpus



Data Annotation: ask several annotators and ...

Store their social variables, ex.:

- Personality questionnaires
 - (Big Five Model): Openness Conscientiousness Extraversion Agreeableness Neuroticism
 - EXAMPLE : POM dataset [Park et al., 2014] self-assessed personality of the workers (labellers)
 - Socio-demographic criteria : age, gender, education
 - EXAMPLE: [Introducing CAD: the Contextual Abuse Dataset] (Vidgen et al., NAACL 2021)

« The dataset includes annotations from 12 trained analysts. Age: 7 annotators were 18–29, 3 were 30–39 and 1 was 40–49. Gender: 4 were female and 7 were male. Ethnicity: 8 were white, 1 Latino, 1 of Middle Eastern ethnic origin and 1 was mixed. National identity: 7 were British, 1 American, 1 Ecuadorean, 1 Jordanian and 1 Polish. Social media use: 9 used social media more than once per day, and 2 use it once per day. Exposure to online abuse: All annotators had witnessed online abuse in the previous year, with 10 stating they had witnessed it more than 3 times and 1 stating they had witness it 2–3 times. »

Measure the reliability of annotations

Measures

- Cohen's kappa [Carletta, 1996]:
- agreement corrected for what it would be under the mere fact of chance

$$k = \frac{p_o - p_e}{1 - p_e}$$

- p_o is the proportion of agreement observed and p_e the probability that the annotators agree by chance

Measure the reliability of annotations

EXO

- Kappa values ?
 - When annotators agree as much as chance
 - When the annotators agree totally
- Exercice
 - 100 texts annotated by 2 people (Ann1 / Ann2) in 2 categories positive / negative
 - Calculate kappa between the two annotators

$$k = \frac{p_o - p_e}{1 - p_e}$$

Ann1\Ann2	Positive	Negative
Positive	50	10
Negative	30	10

Measure the reliability of annotations

- $P_o = (50+10)/100 = 0,6$
- Calculate P_e :
 - Ann1 uses positive label 60% of the time
 - Ann2 uses positive label 80% of the time
 - Probability that Ann1 and Ann2 use the positive label: $0.6 \cdot 0.8 = 0.48$
 - Probability that Ann1 and Ann2 use the negative label : $0.4 \cdot 0.2 = 0.08$
 - Probability to agree by chance : $0.48 + 0.08 = 0.56$
- Kappa computation:
 - $\text{Kappa} = (0.6 - 0.56) / (1 - 0.56) = 0.09$

$$k = \frac{p_o - p_e}{1 - p_e}$$

Measure the reliability of annotations

- Moderate agreement = standard for emotions [Landis & Koch, 1977]
- Other measures:
 - Fleiss Kappa : variant with random raters
 - ICC : Intraclass correlation coefficient
 - Useful existing variant with random raters (frequent when using crowdsourcing platforms)
 - Cronbach's Alpha [Cronbach, 1951] for dimensions

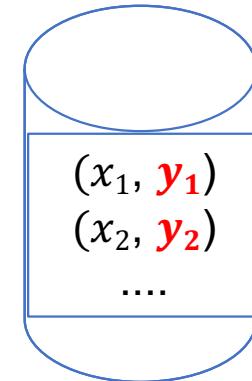
Kappa value	Interpretations
<0	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

To go further : Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial Kevin A. Hallgren

Examples from our research

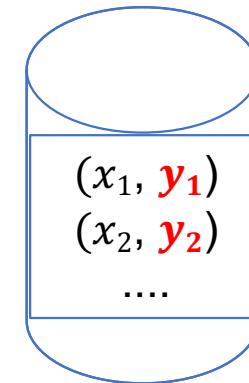
Learn a model from consensus labels

- Majority vote from
 - Expert annotations – examples:
 - Neurologists for **emotions of patients** (study of Huntington disease) [Chenain et al., MSA @ACII, 2024]
 - Recruiter for **social skills** of a candidate during a job interview [Hemamou et al., IEEE Trans. of Aff. Comp., 2021]
 - Non-expert annotations – examples:
 - **Persuasiveness** in public speaking [Biancardi et al., LRE 2022]
 - The reader's **empathy** with the characters and the **surprise** of the end of a story [Chhun et al., COLING, 2022]



But...

- Limitations of the majority vote approach
 - Inter-Annotator-Agreement is usually only moderate
 - Disagreements are interpreted as noise
 - For subjective task: equally valid annotations can coexist



Example from our research

- Learn a model from alternative labels for **fallacy** annotation
 - using a disjunctive annotation scheme at the span level

Let “*a b c d*” be a text where *a*, *b*, *c*, and *d* are sentences.

Suppose $S = \{a, b, d\}$ (i.e., the sentences *a* and *b* are one fallacious span, and *d* is a span of one fallacious sentence), *a b* has labels $\{l_1, l_2\}$, and *d* has label $\{l_3\}$. In that case, $G = \{(a, b, \{l_1, l_2\}), (d, \{l_3\})\}$

An example of prediction P could be $P = \{(a, l_1), (a, l_2), (b, l_3), (c, l_4), (d, l_1)\}$

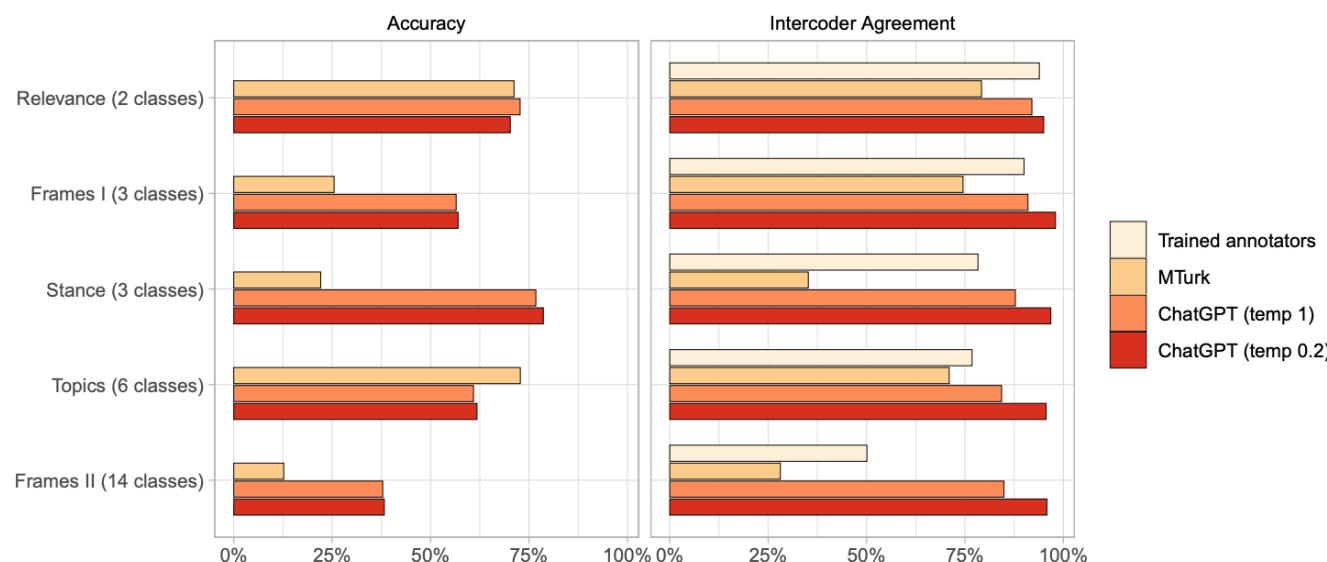


[Helwé et al., NAACL, 2024]

And very recently...

LLM as annotator - Automatic annotation with chatgpt

- « ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks »
<https://arxiv.org/pdf/2303.15056.pdf>



Temperature in order to control the creativity level :
Level close to 0 : more conservative and predictable outputs
Level close to 1 : more diverse and unpredictable outputs

Data Annotation : A few tips

- Build an annotation guide
 - Define the types and the categories that will be relevant for your application
 - Use questions to guide the annotation process and give examples of categories
- Use sufficiently-expert annotator and store their social variables
- Measure the reliability of annotations
- If you use automatic annotation: check the performance of the automatic by comparing to human annotation performance on a subset !

Materials to go further

Materials to go further

- NLP in general
 - <https://nlp.stanford.edu/IR-book>
 - From Miha Grčar “Text mining and Text stream mining tutorial”
 - Foundations of Statistical Natural Language Processing Christopher D. Manning and Hinrich Schütze
 - Lecture from Stanford http://cs224d.stanford.edu/lecture_notes/notes1.pdf

Materials to go further

- NLP and deep learning
 - Deep Natural Language Processing course offered in Hilary Term 2017 at the University of Oxford.
 - Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.
 - Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.

Materials to go further

- Tools :

- word2vec from Google <https://code.google.com/p/word2vec/> tutorial from
- tensorflow <https://www.tensorflow.org/tutorials/word2vec>
- Other representation : Glove <http://nlp.stanford.edu/projects/glove/>

- Sentiment analysis

- <https://web.stanford.edu/class/cs224u/slides/cs224u-2021-sentiment-part1-handout.pdf>
- Munezero M. D., Suero Montero C., Sutinen E., Pajunen J., "Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text", IEEE Transactions on Affective Computing, 2014.

QUIZZ

<https://forms.gle/Q1rzACzNqwsCtMpC8>

