

Machine Learning for social sciences

Natural Language Processing : Session 2

William Aboucaya

Natural Language Inference (NLI) is “the problem of determining **entailment** and **contradiction** relationships between a **premise** and a **hypothesis**”¹

1. Ankur Parikh *et al.* 2016. A Decomposable Attention Model for Natural Language Inference. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. <https://doi.org/10.18653/v1/D16-1244>

Natural Language Inference (NLI) is “the problem of determining **entailment** and **contradiction** relationships between a **premise** and a **hypothesis**”¹

⇒ Does the **hypothesis** logically follow from the **premise**?

1. Ankur Parikh *et al.* 2016. A Decomposable Attention Model for Natural Language Inference. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. <https://doi.org/10.18653/v1/D16-1244>

Natural Language Inference (NLI) is “the problem of determining **entailment** and **contradiction** relationships between a **premise** and a **hypothesis**”¹

⇒ Does the **hypothesis** logically follow from the **premise**?

Labels :

- **Entailment** : Hypothesis is true if the premise is true.
- **Contradiction** : Hypothesis is false if the premise is true.
- **Neutral** : Hypothesis is neither entailed nor contradicted by the premise.

1. Ankur Parikh *et al.* 2016. A Decomposable Attention Model for Natural Language Inference. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. <https://doi.org/10.18653/v1/D16-1244>

Applications of NLI

- Question Answering Systems
- Information Retrieval and Summarization
- Sentiment Analysis and Opinion Mining
- Verifying Claims in Automated Fact-Checking

Example :

- **Premise** : My cat is in the tree
- **Hypothesis** : There is at least one cat in the tree
- **Scores** :
 - **Entailment** : **0.98**
 - Contradiction : 0.00
 - Neutral : 0.02

Example :

- **Premise** : My cat is in the tree
- **Hypothesis** : My cat never left the house
- **Scores** :
 - Entailment : 0.00
 - **Contradiction** : **0.93**
 - Neutral : 0.07

Example :

- **Premise** : My cat is in the tree
- **Hypothesis** : It is winter
- **Scores** :
 - Entailment : 0.01
 - Contradiction : 0.01
 - **Neutral : 0.98**

What do NLI datasets look like ?

2. Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In Proceedings of the 2018 Conference of the North American Chapter of the ACL : Human Language Technologies. <https://doi.org/10.18653/v1/N18-1101>

What do NLI datasets look like ?

Example : the Multi-genre Natural Language Inference Corpus (MNLI)²

- **A premise** : At the other end of Pennsylvania Avenue, people began to line up for a White House tour.
- **A hypothesis** : People formed a line at the end of Pennsylvania Avenue.
- **A label** : Entailment

⇒ These datasets are used to fine-tune and evaluate language models

2. Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In Proceedings of the 2018 Conference of the North American Chapter of the ACL : Human Language Technologies. <https://doi.org/10.18653/v1/N18-1101>

- Handling ambiguity and polysemy
- Dealing with implicit knowledge and commonsense reasoning
- Managing syntactic and semantic variability
- Training robust models on diverse datasets (models tend to measure similarity rather than entailment³)

3. Tom McCoy, Ellie Pavlick, and Tal Linzen. 2019. Right for the Wrong Reasons : Diagnosing Syntactic Heuristics in Natural Language Inference. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P19-1334>

Challenges in NLI

- Handling ambiguity and polysemy
- Dealing with implicit knowledge and commonsense reasoning
- Managing syntactic and semantic variability
- Training robust models on diverse datasets (models tend to measure similarity rather than entailment³)

Let's see what we can do with it 🤓

3. Tom McCoy, Ellie Pavlick, and Tal Linzen. 2019. Right for the Wrong Reasons : Diagnosing Syntactic Heuristics in Natural Language Inference. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P19-1334>

Zero-shot learning (ZSL) is a machine learning paradigm where a model can make predictions for classes or tasks it has never explicitly seen during training using **knowledge transfer**.

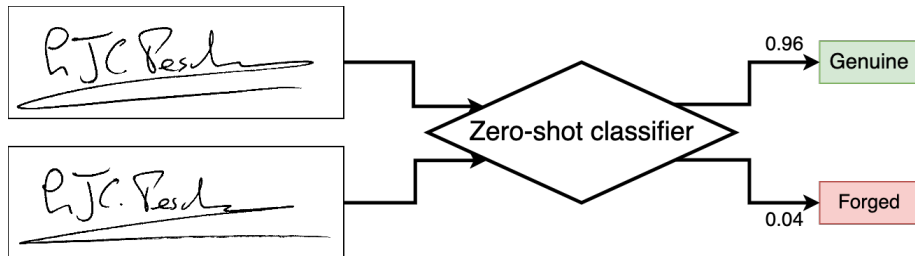
Zero-shot learning (ZSL) is a machine learning paradigm where a model can make predictions for classes or tasks it has never explicitly seen during training using **knowledge transfer**.

Key Components of ZSL :

- **Seen Classes** : Classes/tasks available during training.
- **Unseen Classes** : Classes/tasks not available during training.
- **Knowledge Representation** : Attributes or embeddings that describe both seen and unseen classes.
- **Model Objective** : Learn a mapping between input features and the shared knowledge space.

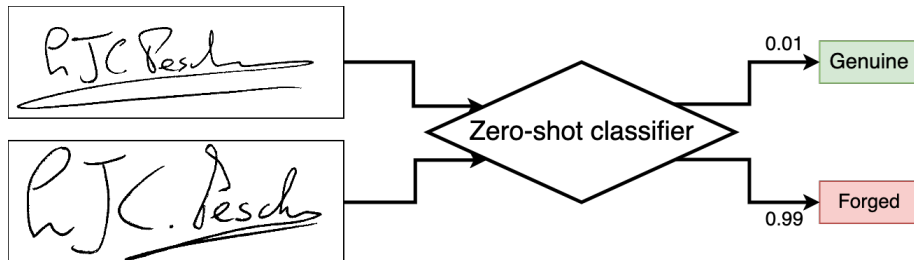
Zero-shot learning

Example : Signature genuineness evaluation



Zero-shot learning

Example : Signature genuineness evaluation



Zero-shot classification

What is it ?

- A NLP task where a model classifies documents into categories **without explicit training** on those categories
- Leverages **pre-trained language models** based on **prompting** or **NLI**

Zero-shot classification

What is it ?

- A NLP task where a model classifies documents into categories **without explicit training** on those categories
- Leverages **pre-trained language models** based on **prompting** or **NLI**

Advantages :

- **No Task-Specific Data Required** : Works directly with new labels
- **Rapid Prototyping** : Useful for tasks with little or no annotated data

Applications :

- Content moderation
- Sentiment analysis
- Topic classification
- Legal or medical text categorization
- etc.

Zero-shot classification

Applications :

- Content moderation
- Sentiment analysis
- Topic classification
- Legal or medical text categorization
- etc.

Example :

- **Input** : *“European Parliament adopts ‘historic’ AI Act”*
- **Labels** : Technology, Sports, Politics
- **Predictions** : Technology, Politics

. Results obtained using MoritzLaurer/deberta-v3-large-zeroshot-v2.0

Zero-shot classification

How to do it ?

- A first approach : Just ask ChatGPT 🤖

Zero-shot classification

How to do it ?

- A first approach : Just ask ChatGPT 🗣️
- No, really, the results obtained by asking GPT-4, Llama 3 or MISTRAL are... fine⁴

4. Bucher, M. J. J., & Martini, M. 2024. Fine-Tuned 'Small' LLMs (Still) Significantly Outperform Zero-Shot Generative AI Models in Text Classification. arXiv preprint. <https://doi.org/10.48550/arXiv.2406.08660>

Zero-shot classification

How to do it ?

- A first approach : Just ask ChatGPT 🤖
- No, really, the results obtained by asking GPT-4, Llama 3 or MISTRAL are... fine⁴

Issues of this approach :

- **Scalability** : The larger the model, the longer it takes to compute each classification
- **Price** : If you want to label hundreds of thousands of documents, asking a remote service to produce millions of tokens isn't cheap
- **Reliability** : Those models do not produce confidence intervals for the labels

4. Bucher, M. J. J., & Martini, M. 2024. Fine-Tuned 'Small' LLMs (Still) Significantly Outperform Zero-Shot Generative AI Models in Text Classification. arXiv preprint. <https://doi.org/10.48550/arXiv.2406.08660>

An approach to zero-shot classification based on NLI

- Define a set of candidate labels.
- Convert each label into a hypothesis (e.g., “This text is about *sports*.”).
- Use an NLI model to compute entailment scores : Higher score → Higher confidence in the relevance of the label
- Assign the label with the highest entailment score or a list of labels with a score higher than a threshold

An approach to zero-shot classification based on NLI

- Define a set of candidate labels.
- Convert each label into a hypothesis (e.g., “This text is about *sports*.”).
- Use an NLI model to compute entailment scores : Higher score → Higher confidence in the relevance of the label
- Assign the label with the highest entailment score or a list of labels with a score higher than a threshold

Let's see what we can do with it 🤓