Natural Language Inference

task overview

What is NLI?
Which datasets are available?
Which tools are the best?

Final project by Marina Kazyulina for DS School

Table of contents

Project description

Current work aims at developing an easy-to-use source for getting started with Natural Language Inference task.

The source provides comprehensive EDA for three large NLI datasets as well as for the dataset from corresponding Kaggle competition which serves here as a validation set.

Moreover, comparison of two transformer technologies is also present. Overall, during the process of development 2 * 3 models were trained.

The description of datasets and tre-trained model used is presented further.

The source itself is located here.

NLI: definition

Natural Language Inference (or Recognizing Textual Entailment), is the task of **determining the inference relation** between two (short, ordered) texts:

- entailment,
- contradiction,
- neutral

P R E M I S E

He came, he opened the door and I remember looking back and seeing the expression on his face, and I could tell that he was disappointed.

H1: entailment

Just by the look on his face when he came through the door I just knew that he was let down.

H2: neutral

He was trying not to make us feel guilty but we knew we had caused him trouble.

H3: contradiction

He was so excited and bursting with joy that he practically knocked the door off it's frame.

NLI: applications

- Question answering
  - evaluation whether the target question can be inferred from candidate answers extracted from the source document
- Semantic search
  - identification of approximate semantic equivalence between search queries and sentences in source documents.
- Automatic summarization
  - NLI can be used to ensure that the summary does not contain any sentences that can be inferred from the rest of the summary.
- Evaluation of machine translation systems.
  - assessment of approximate semantic equivalence between a candidate translation and a reference translation

NLI: brief history

- Shallow approaches
  - lexical or semantic overlap (Glickman et al. 2005),
  - pattern based relation extraction (Romano et al. 2006),
  - approximate matching of predicate argument structure (MacCartney et al. 2006, Hickl et al. 2006).
- Deep approaches
  - first-order logic (FOL),
  - natural logic (Lakoff 1970)

Data for NLI: SNLI

Stanford Natural Language Inference (SNLI) corpus is a collection of 570k human-written English sentence pairs manually labeled for balanced classification with the labels entailment, contradiction, and neutral.

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

Data for NLI: MNLI

The Multi-Genre Natural Language Inference (MultiNLI) corpus is a crowd-sourced collection of 433k sentence pairs annotated with textual entailment information. The corpus is modeled on the SNLI corpus, but differs in that covers a range of genres of spoken and written text

Fiction		
The Old One always comforted Ca'daan, except today.	*neutral*	Ca'daan knew the Old One very well.
Letters		
Your gift is appreciated by each and every student who will benefit from your generosity.	*neutral*	Hundreds of students will benefit from your generosity.
Telephone Speech		
yes now you know if if everybody like in August when everybody's on vacation or something we can dress a little more casual or	*contradiction*	August is a black out month for vacations in the company.
9/11 Report		
At the other end of Pennsylvania Avenue, people began to line up for a White House tour.	*entailment*	People formed a line at the end of Pennsylvania Avenue.

Data for NLI: XNLI

The Cross-lingual Natural Language Inference (XNLI) corpus is a crowd-sourced collection of 5,000 test and 2,500 dev pairs for the MultiNLI corpus. The pairs are translated into 14 languages. This results in 112.5k annotated pairs. Each premise can be associated with the corresponding hypothesis in the 15 languages, summing up to more than 1.5M combinations.

Language	Premise / Hypothesis	Genre	Label
English	You don't have to stay there. You can leave.	Face-To-Face	Entailment
French	La figure 4 montre la courbe d'offre des services de partage de travaux. Les services de partage de travaux ont une offre variable.	Government	Entailment
Spanish	Y se estremeció con el recuerdo. El pensamiento sobre el acontecimiento hizo su estremecimiento.	Fiction	Entailment
German	Während der Depression war es die ärmste Gegend, kurz vor dem Hungertod. Die Weltwirtschaftskrise dauerte mehr als zehn Jahre an.	Travel	Neutral
Swahili	Ni silaha ya plastiki ya moja kwa moja inayopiga risasi. Inadumu zaidi kuliko silaha ya chuma.	Telephone	Neutral
Russian	И мы занимаемся этим уже на протяжении 85 лет. Мы только начали этим заниматься.	Letters	Contradiction
Chinese	让我告诉你，美国人最终如何看待你作为独立顾问的表现。 美国人完全不知道您是独立律师。	Slate	Contradiction

Transformer models for NLI: BERT

BERT, or Bidirectional Encoder Representations from Transformers, is a language representation model pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers.

- As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, NLI included.
- In the present work *bert-base-multilingual-cased* model from huggingface was used. This model does not require language embeddings at inference time a it identifies the language used in the context and infer accordingly.

Transformer models for NLI: XLM-RoBERTa

XLM-RoBERTa is based on Facebook's RoBERTa model released in 2019. It was trained on 2.5TB of newly created clean CommonCrawl data in 100 languages.

- It provides strong gains over previously released multi-lingual models like mBERT or XLM on downstream tasks like classification, sequence labeling and question answering.
- In the current work *xlm-roberta-large* was used.

Sources (papers)

- Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. arXiv preprint arXiv:1508.05326.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
- Conneau, A., Lample, G., Rinott, R., Williams, A., Bowman, S. R., Schwenk, H., & Stoyanov, V. (2018). XNLI: Evaluating cross-lingual sentence representations. arXiv preprint arXiv:1809.05053.
- 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.
- MacCartney, B., & Manning, C. D. (2009). Natural language inference. Stanford: Stanford University.
- Williams, A., Nangia, N., & Bowman, S. R. (2017). A broad-coverage challenge corpus for sentence understanding through inference. arXiv preprint arXiv:1704.05426.

Sources (kernels)

- Tutorial Notebook by Ana Sofia Uzsoy
- Hands-on NLI w/ Transformers (M-BERT, XLM-RoBERTa) by Wasifa Chowdhury
- More NLI datasets - Hugging Face nlp library by Yih-Dar SHIEH