Profiling a Neural Language Model

Seminari ILC, 10/11/2022

Alessio Miaschi
ItaliaNLP Lab, Istituto di Linguistica Computazionale (ILC-CNR), Pisa
alessio.miaschi@ilc.cnr.it
https://alemiaschi.github.io/
http://www.italianlp.it/alessio-miaschi/

Outline

- The rise of Neural Language Models
- Interpretability of Neural Language Models
- Case Study: Profiling Neural Language Model
- Conclusion and Future Directions

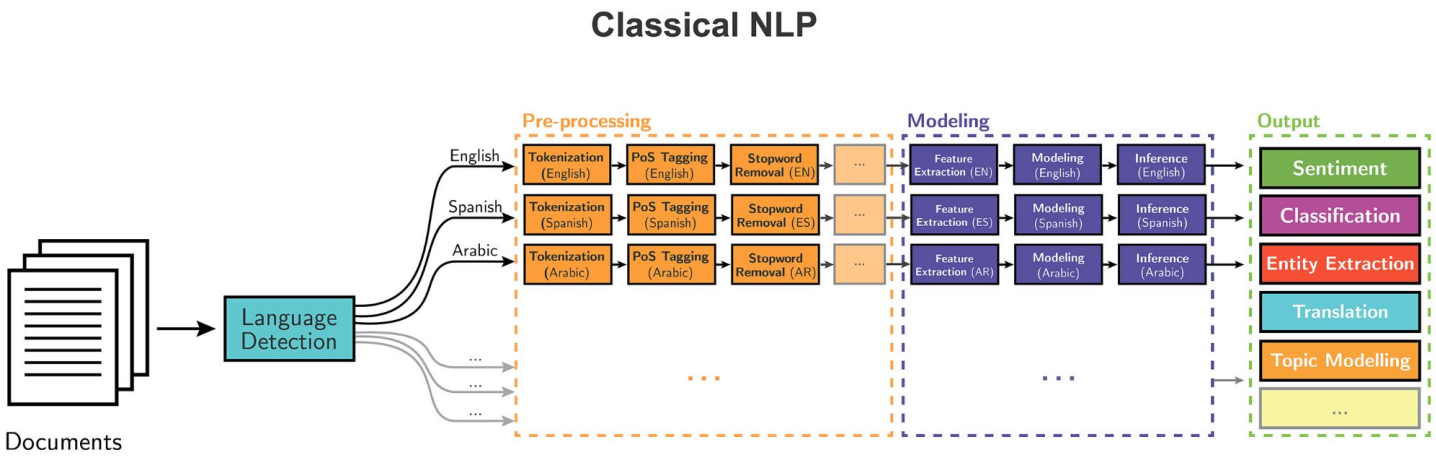
The rise of Neural Language Models

Introduction

- The field of NLP has seen an unprecedented progress in the last years
- Much of this progress is due to the replacement of traditional systems with newer and more powerful Deep Learning (DL) models

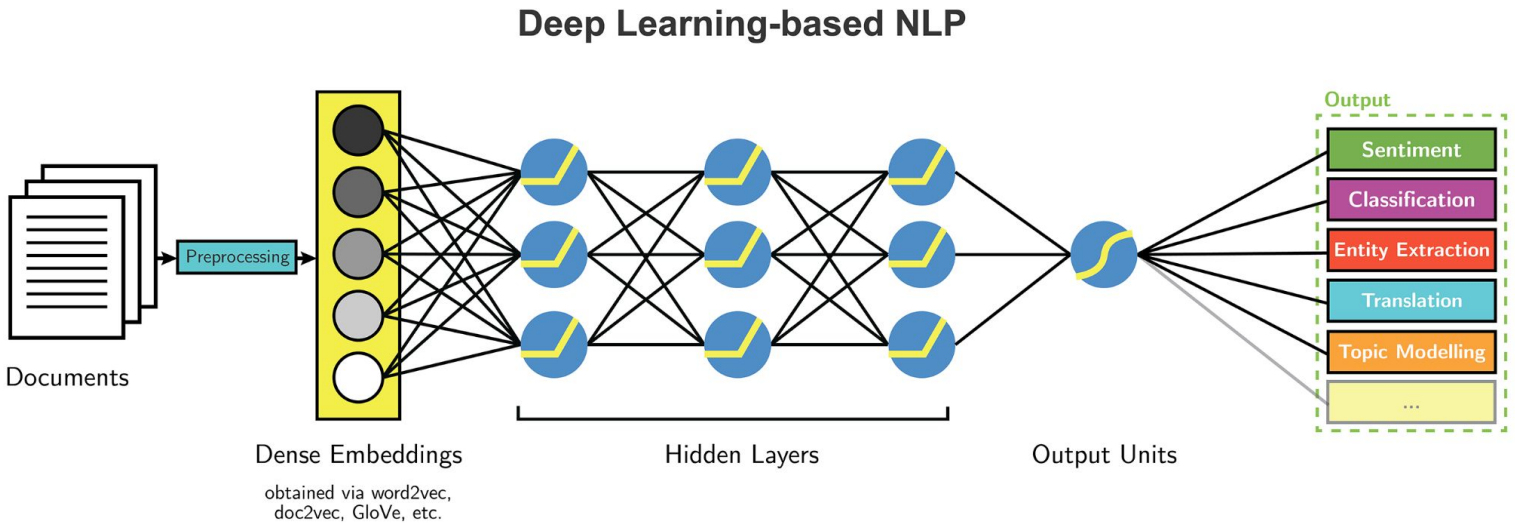
Introduction

- The field of NLP has seen an unprecedented progress in the last years
- Much of this progress is due to the replacement of traditional systems with newer and more powerful Deep Learning (DL) models



Introduction

- The field of NLP has seen an unprecedented progress in the last years
- Much of this progress is due to the replacement of traditional systems with newer and more powerful Deep Learning (DL) models



Neural Language Models

- Neural Network (NN) model trained to approximate the **language modeling** function

- A probabilistic language model (**LM**) defines the probability of a sentence $s = [w_1, w_2, \ldots, w_n]$ as:

$$P(s) = \prod_{i=1}^{N} P(w_i | w_1, w_2, \ldots, w_{i-1})$$

Neural Language Models

- Neural Network (NN) model trained to approximate the **language modeling** function

- A probabilistic language model (**LM**) defines the probability of a sentence $s = [w_1, w_2, \ldots, w_n]$ as:

$$P(s) = \prod_{i=1}^{N} P(w_i | w_1, w_2, \ldots, w_{i-1})$$

- Bengio et al. (2003) proposed a model that assigns a distributed vector for each word and then uses a NN architecture to predict the next word → **Neural Probabilistic Language Model**

Transformer Models

- Nowadays, the Transformer architecture has become the preferred solution for the development of state-of-the-art NLMs

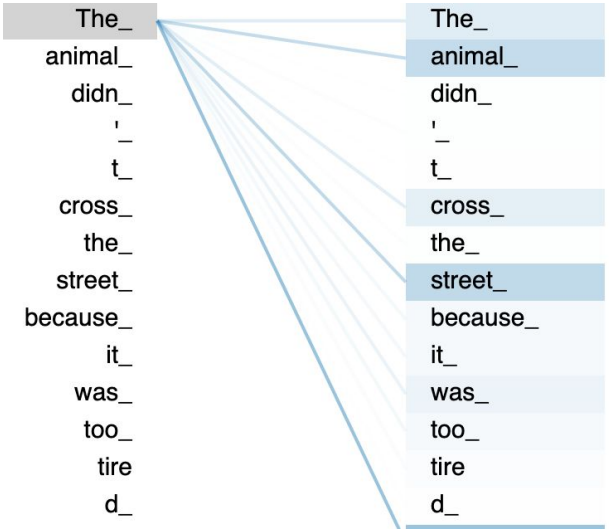
- Transformers (Vaswani et al., 2017) use only **attention** and fully connected layers to create highly scalable networks capturing distant patterns

Transformer Models

- Nowadays, the Transformer architecture has become the preferred solution for the development of state-of-the-art NLMs

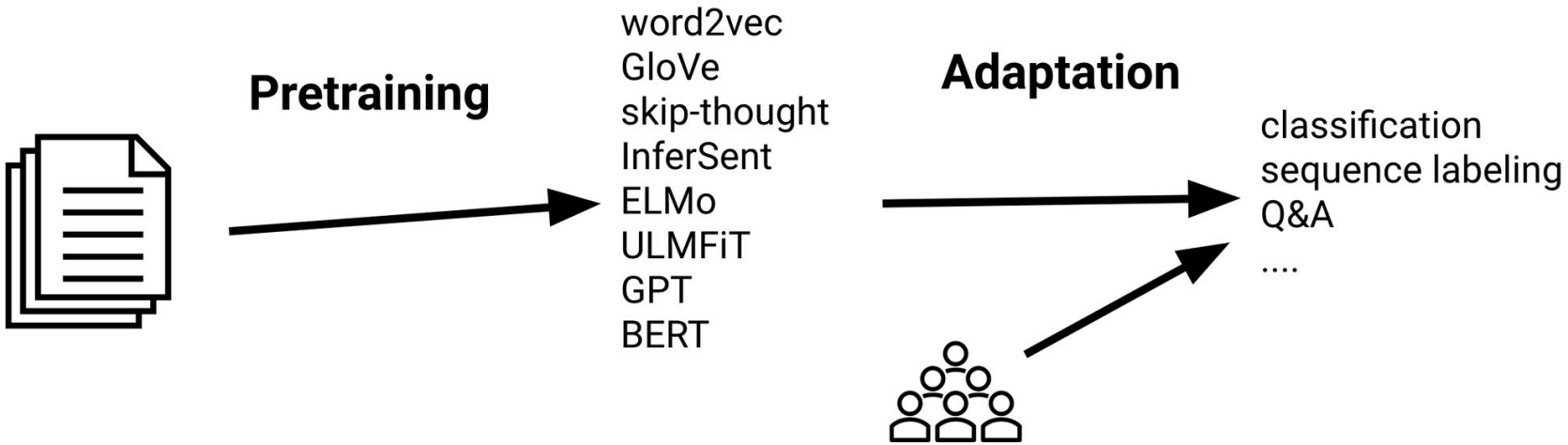
- Transformers (Vaswani et al., 2017) use only **attention** and fully connected layers to create highly scalable networks capturing distant patterns

- Attention is the method that allows the model to "attend" to different positions of the input sequence to compute a representation of that sequence



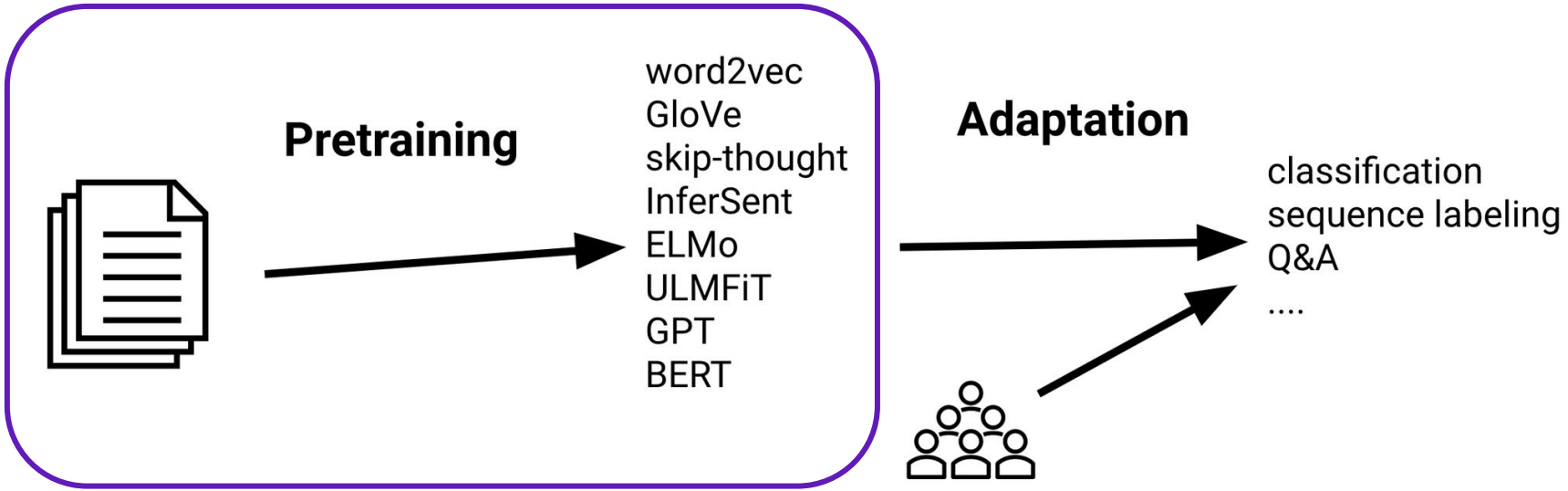
$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

Transfer Learning



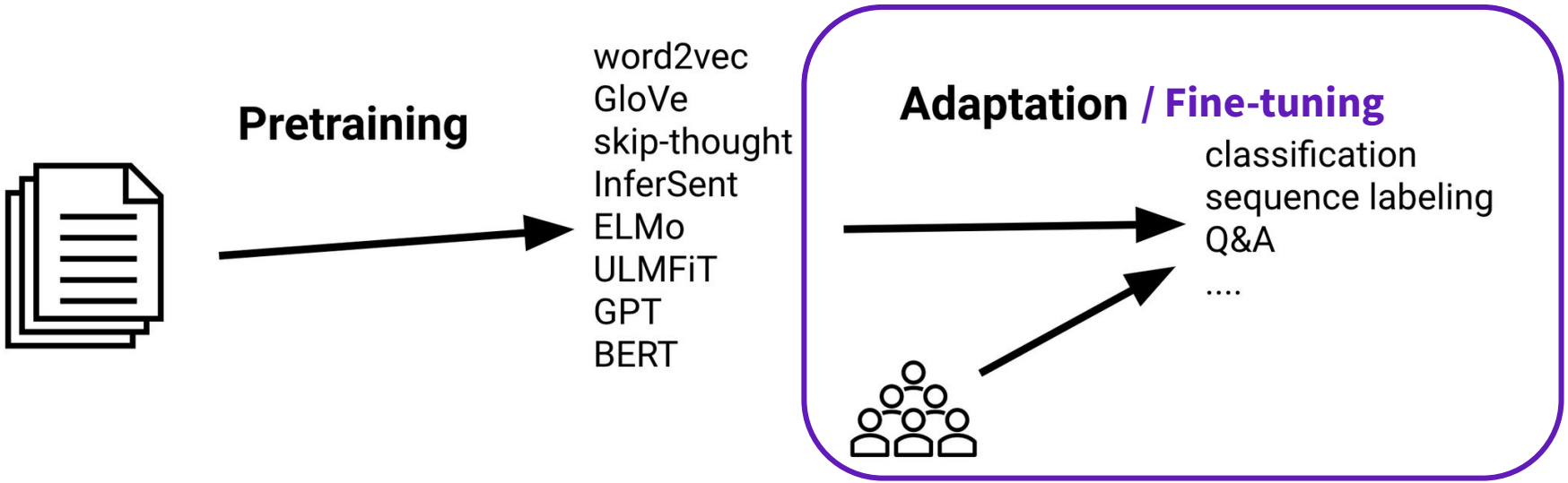
The State of Transfer Learning in NLP: https://ruder.io/state-of-transfer-learning-in-nlp/

Transfer Learning



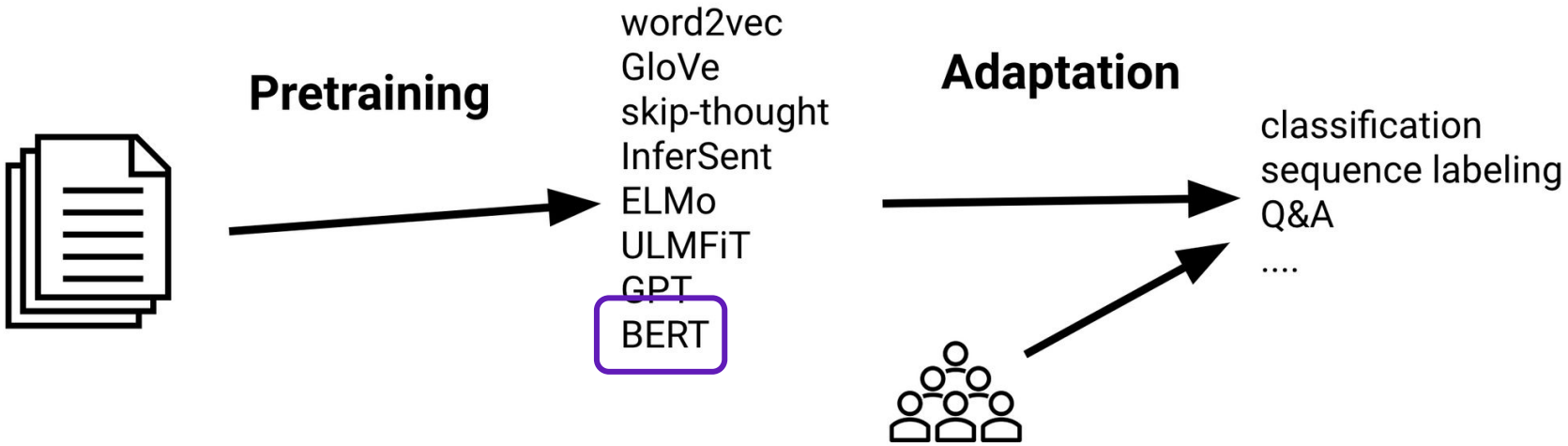
The State of Transfer Learning in NLP: https://ruder.io/state-of-transfer-learning-in-nlp/

Transfer Learning



The State of Transfer Learning in NLP: https://ruder.io/state-of-transfer-learning-in-nlp/

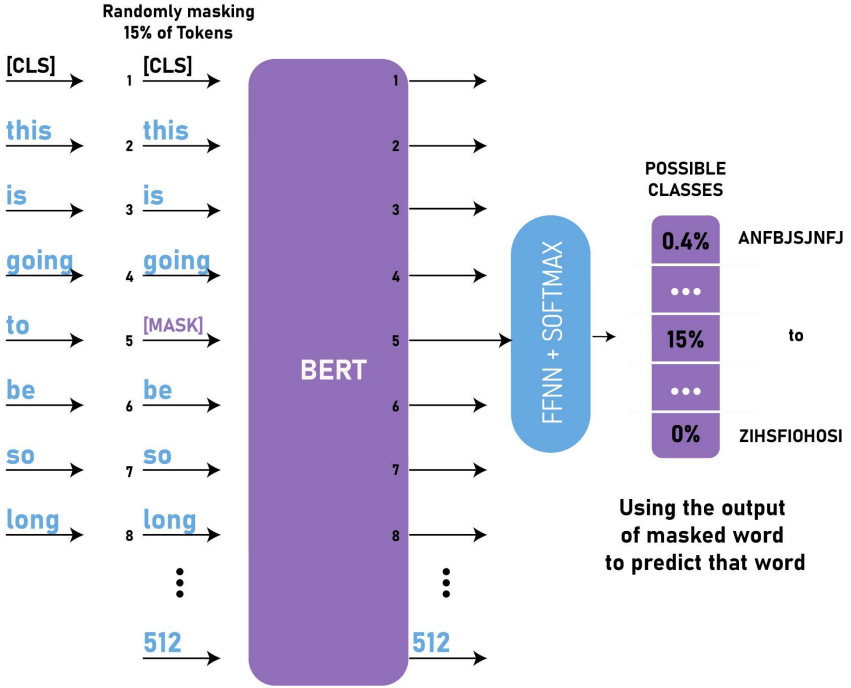
Transfer Learning



The State of Transfer Learning in NLP: https://ruder.io/state-of-transfer-learning-in-nlp/

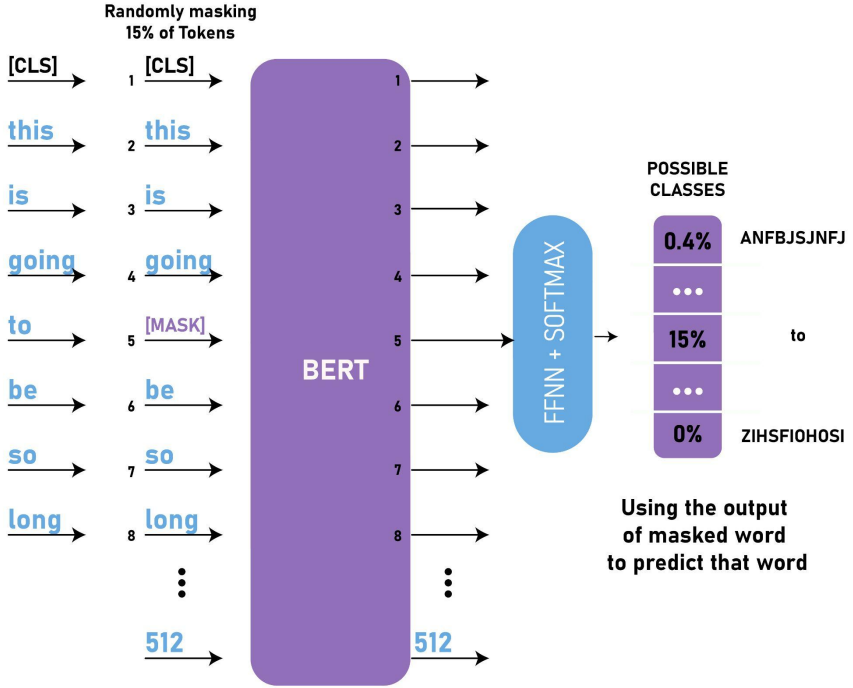
BERT (Devlin et al., 2019) 

- Encoder model (12/24 layers)
- Trained to approximate the **Masked Language Modeling** (**MLM**) function

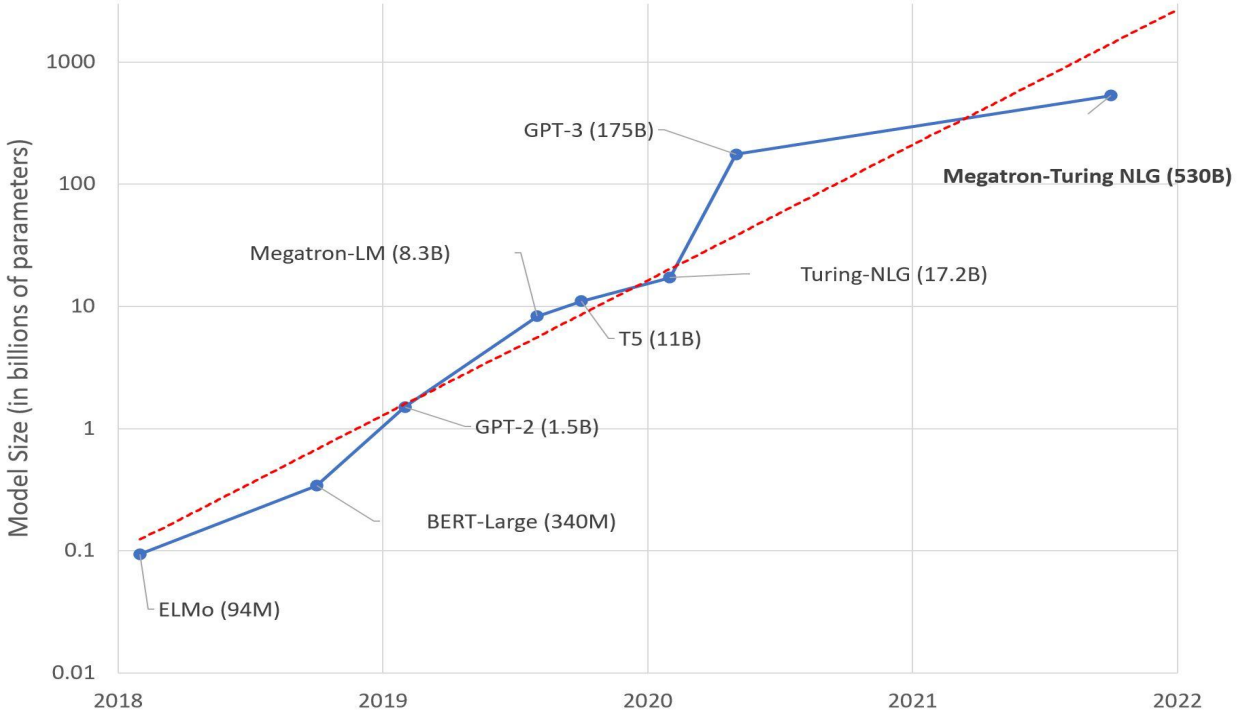


BERT (Devlin et al., 2019) 

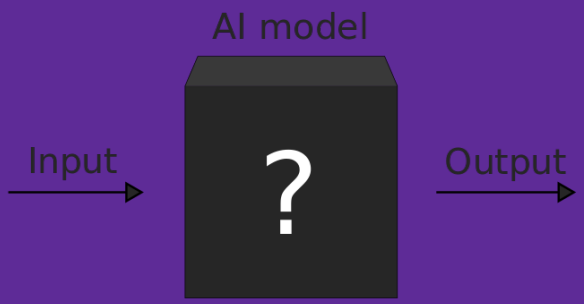
- Encoder model (12/24 layers)
- Trained to approximate the **Masked Language Modeling** (**MLM**) function
- The model can be fine-tuned in order to solve several NLP tasks:
  - Sentiment analysis;
  - Question answering;
  - Textual entailment;
  - etc.



Parameters Are All You Need



Interpreting Neural Language Models



Interpretability in NLP

“In the context of NLP, this question needs to be understood in light of earlier NLP work. [...] In some of these systems, features are more easily understood by humans. [...] In contrast, it is more difficult to understand what happens in an end-to-end neural network model that takes input (say, word embeddings) and generates an output.”

Belinkov and Glass, Analysis Methods in Neural Language Processing: A Survey (2019). In Transactions of ACL, Volume 7, pages 49-72.

Interpretability in NLP

"In the context of NLP, this question needs to be understood in light of earlier NLP work. [...] In some of these systems, features are more easily understood by humans. [...] In contrast, it is more difficult to understand what happens in an end-to-end neural network model that takes input (say, word embeddings) and generates an output."

Belinkov and Glass, Analysis Methods in Neural Language Processing: A Survey (2019). In Transactions of ACL, Volume 7, pages 49-72.

Research questions:

- What happens in an end-to-end neural network model when trained on a language modeling task?
- What kind of linguistic knowledge (i.e. features) is encoded within their representations?
- Is there a relationship between the linguistic knowledge implicitly encoded and the ability to solve a specific task?

Interpretability in NLP

- The analysis of the inner workings of NLMs has become one of the most addressed line of research in NLP

- Several methods have been implemented to obtain meaningful explanations and to understand how these models are able to capture syntax- and semantic- sensitive phenomena

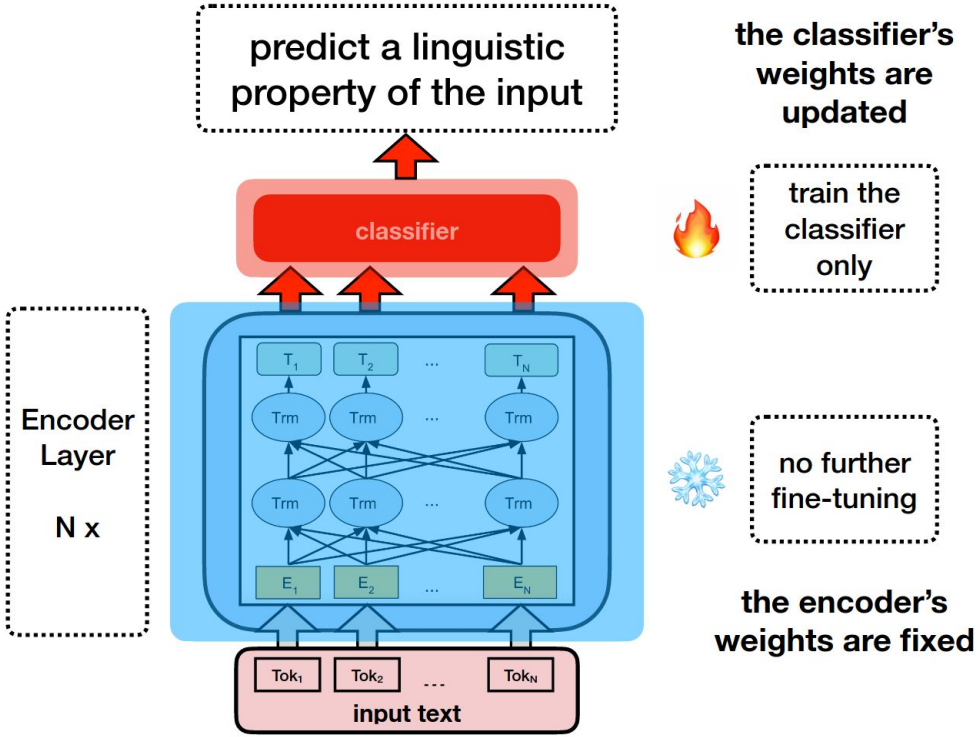
Interpretability in NLP

- The analysis of the inner workings of NLMs has become one of the most addressed line of research in NLP

- Several methods have been implemented to obtain meaningful explanations and to understand how these models are able to capture syntax- and semantic- sensitive phenomena

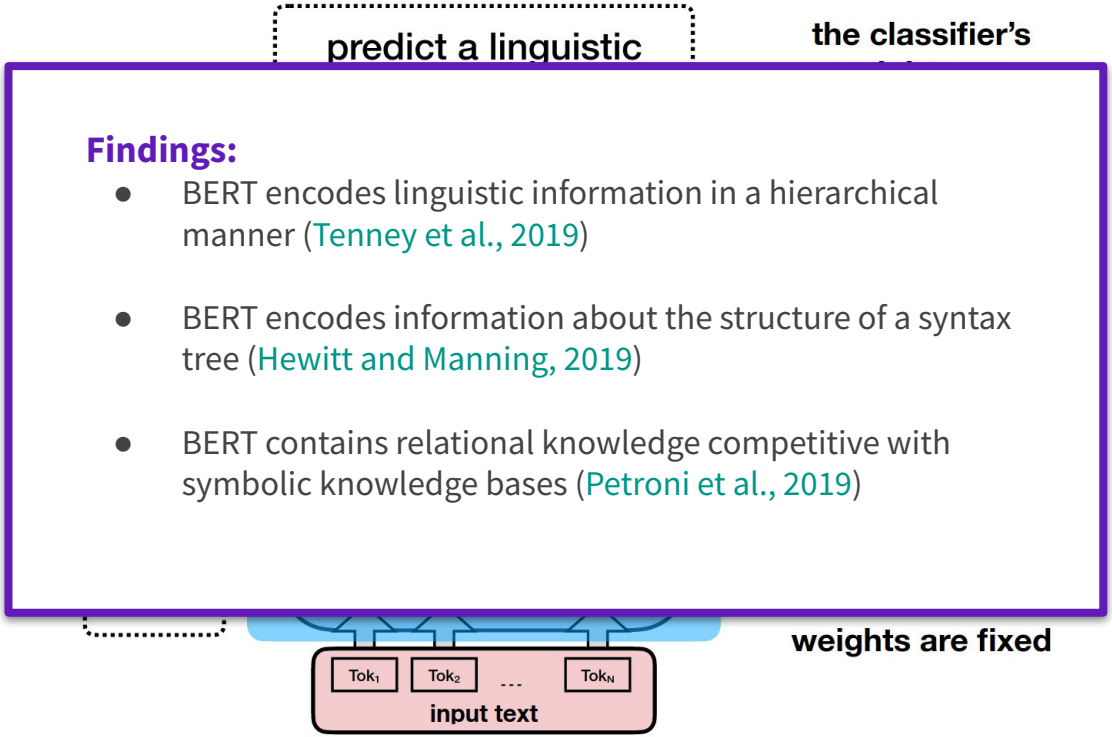
- Several approaches:
  - Probing tasks (e.g. Hewitt and Manning, 2019; Pimentel et al., 2020);
  - Analysis of attention mechanisms (e.g. Clark et al., 2019);
  - Explainability via Integrated Gradients (e.g. Ramnath, 2020);
  - Definition of diagnostic tests (e.g. Goldberg, 2019);

Probing Task Approach



Slide from: https://people.cs.umass.edu/~miyyer/cs685_f20/slides/19-probes.pdf

Probing Task Approach



Slide from: https://people.cs.umass.edu/~miyyer/cs685_f20/slides/19-probes.pdf

Case Study:
Profiling Neural Language Models

Profiling Neural Language Models

- The "*linguistic profiling*" methodology (van Halteren, 2004) assumes that wide counts of linguistic features are particularly helpful in the resolution of several NLP tasks, e.g.:
  - Text Profiling (e.g. text readability, textual genres)
  - Author Profiling (e.g. author's age and native language)

Profiling Neural Language Models

- The "*linguistic profiling*" methodology (van Halteren, 2004) assumes that wide counts of linguistic features are particularly helpful in the resolution of several NLP tasks, e.g.:
  - Text Profiling (e.g. text readability, textual genres)
  - Author Profiling (e.g. author's age and native language)

Research Question:

Could the informative power of these features also be helpful to understand the behaviour of state-of-the-art NLMs?

Profiling-UD: a tool for Linguistic Profiling of Texts

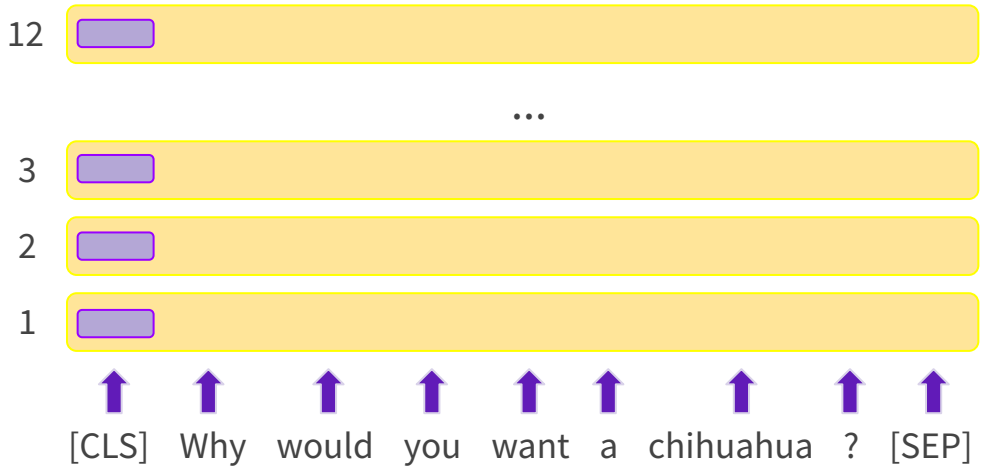
- ProfilingUD (Brunato et al., 2020) is a web–based application that performs linguistic profiling of a text, or a large collection of texts, for multiple languages

- It allows the extraction of more than 130 features, spanning across different levels of linguistic description

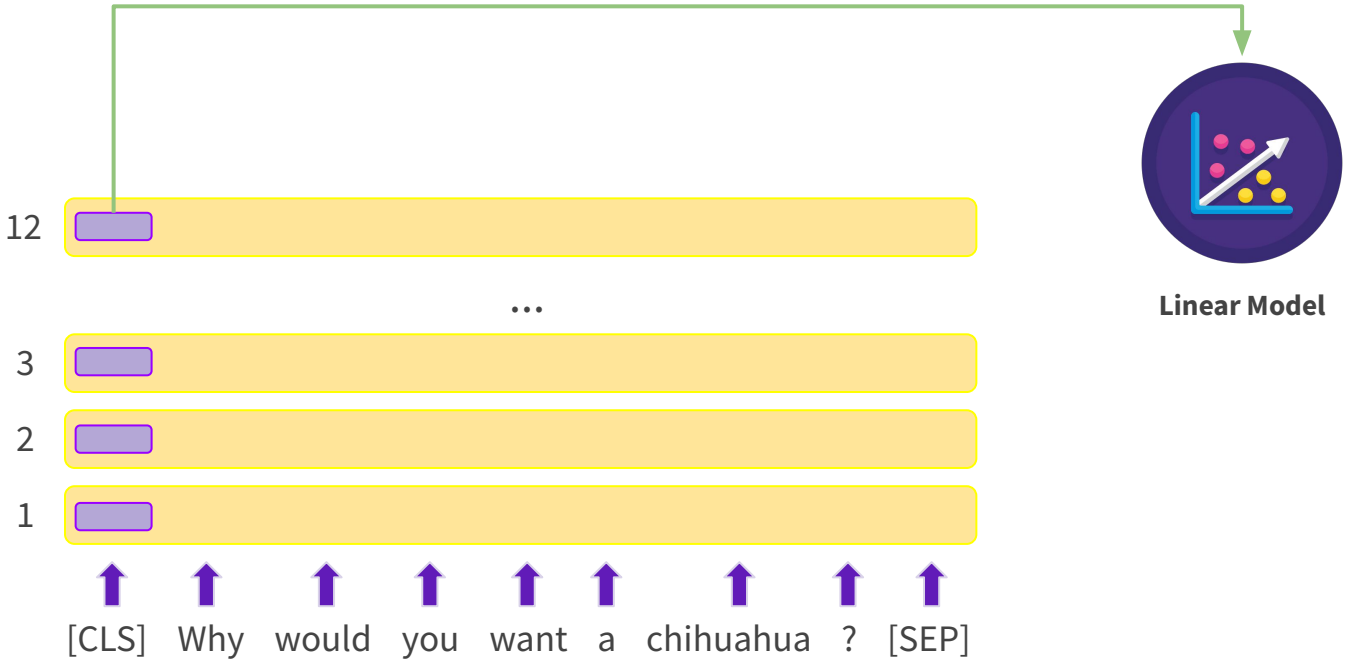
- Link: http://linguistic-profiling.italianlp.it/

Linguistic Feature
Raw Text Properties
Sentence Length
Word Length
Vocabulary Richness
Type/Token Ratio for words and lemmas
Morphosyntactic information
Distibution of UD and language–specific POS
Lexical density
Inflectional morphology
Inflectional morphology of lexical verbs and auxiliaries
Verbal Predicate Structure
Distribution of verbal heads and verbal roots
Verb arity and distribution of verbs by arity
Global and Local Parsed Tree Structures
Depth of the whole syntactic tree
Average length of dependency links and of the longest link
Average length of prepositional chains and distribution by depth
Clause length
Relative order of elements
Order of subject and object
Syntactic Relations
Distribution of dependency relations
Use of Subordination
Distribution of subordinate and principal clauses
Average length of subordination chains and distribution by depth
Relative order of subordinate clauses

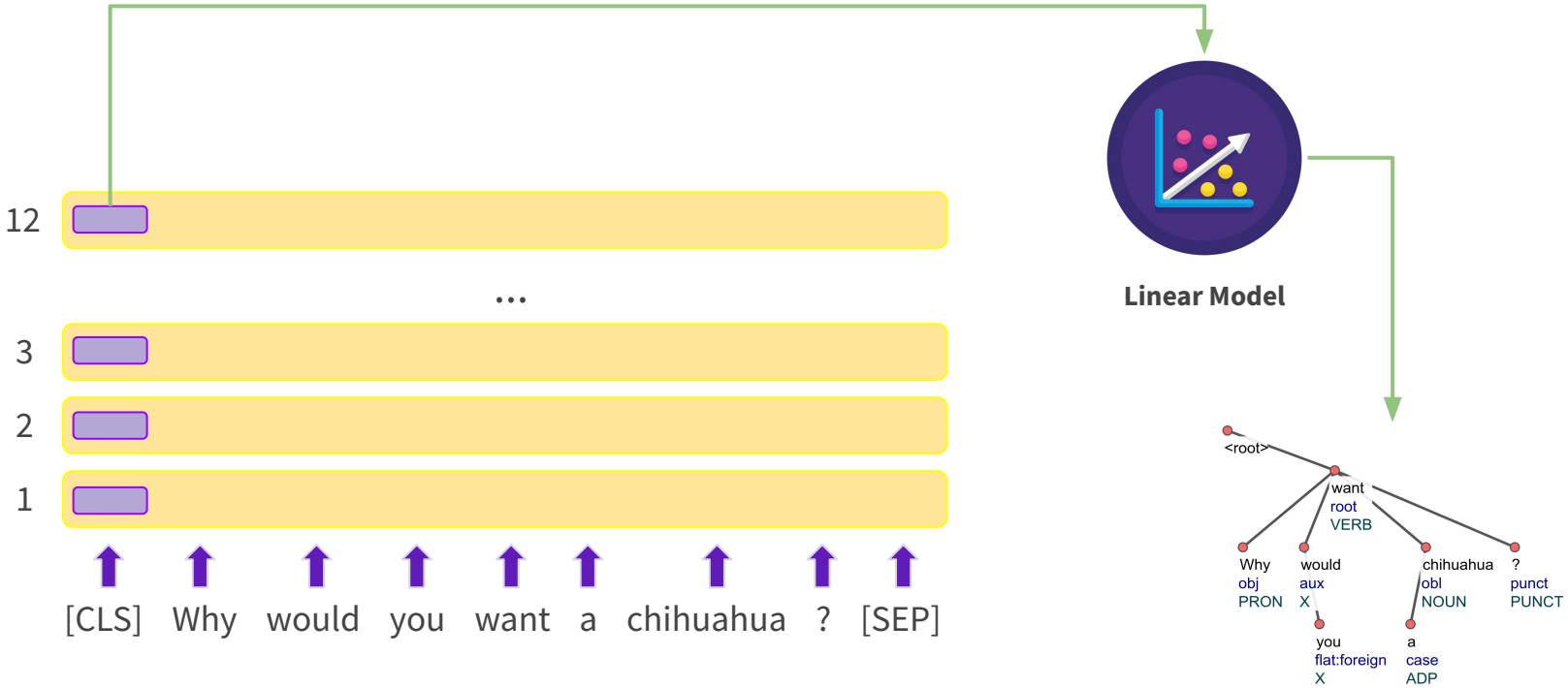
Profiling Neural Language Models



Profiling Neural Language Models



Profiling Neural Language Models



Profiling Neural Language Models

- Miaschi, A., Brunato, D., Dell'Orletta, F., and Venturi, G. (2020). Linguistic Profiling of a Neural Language Model. In Proceedings of the 28th International Conference on Computational Linguistics, pages 745–756, Barcelona, Spain (Online). International Committee on Computational Linguistics. [**Outstanding Paper for COLING 2020**]
- Miaschi A., Dell'Orletta F. (2020). Contextual and Non-Contextual Word Embeddings: an in-depth Linguistic Investigation. In Proceedings of the 5th Workshop on Representation Learning for NLP (ACL 2020, Online).
- Miaschi A., Sarti G., Brunato D., Dell'Orletta F., Venturi G. (2020). Italian Transformers Under the Linguistic Lens. In Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020).
- Miaschi A., Alzetta C., Brunato D., Dell'Orletta F., Venturi G. (2020). Is Neural Language Model Perplexity Related to Readability? In Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020).
- Puccetti G., Miaschi A., Dell'Orletta F. (2021). How Do BERT Embeddings Organize Linguistic Knowledge? In Proceedings of the 2nd Workshop on DeeLIO (NAACL 2021, Online).
- Miaschi A., Brunato D., Dell'Orletta F., Venturi G. (2021). What Makes My Model Perplexed? A Linguistic Investigation on Neural Language Models Perplexity. In Proceedings of the 2nd Workshop on DeeLIO (NAACL 2021, Online).
- Miaschi A., Alzetta C., Brunato D., Dell'Orletta F., Venturi G. (2021). Probing Tasks Under Pressure. In Proceedings of the Eighth Italian Conference on Computational Linguistics (CLiC-it 2021).
- Miaschi A., Sarti G., Brunato D., Dell'Orletta F., Venturi G. (2022). Probing Linguistic Knowledge in Italian Neural Language Models across Language Varieties. Italian Journal of Computational Linguistics (IJCoL), Vol 8., N. 1, June 2022, pages 25-44.

Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

- We investigated the linguistic knowledge implicitly encoded by BERT

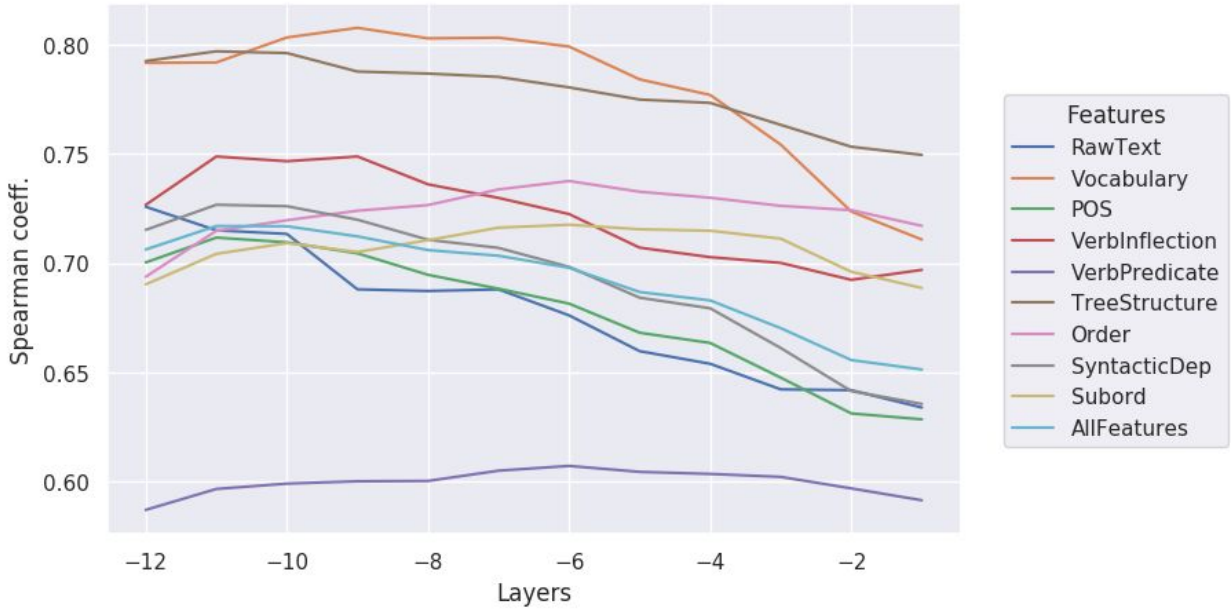
Research questions:

1. What kind of linguistic properties are encoded in a pre-trained version of BERT?

2. How this knowledge is modified after a fine-tuning process

3. Whether this implicit knowledge affects the ability of the model to solve a specific downstream task

Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)



Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

- Fine-tuning of BERT on the *Native Language Identification* (NLI)

> “No breakfast, coz you still have enough alcohol in your stomach.”

Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

- Fine-tuning of BERT on the *Native Language Identification* (NLI)

"No breakfast, coz you still have enough alcohol in your stomach."

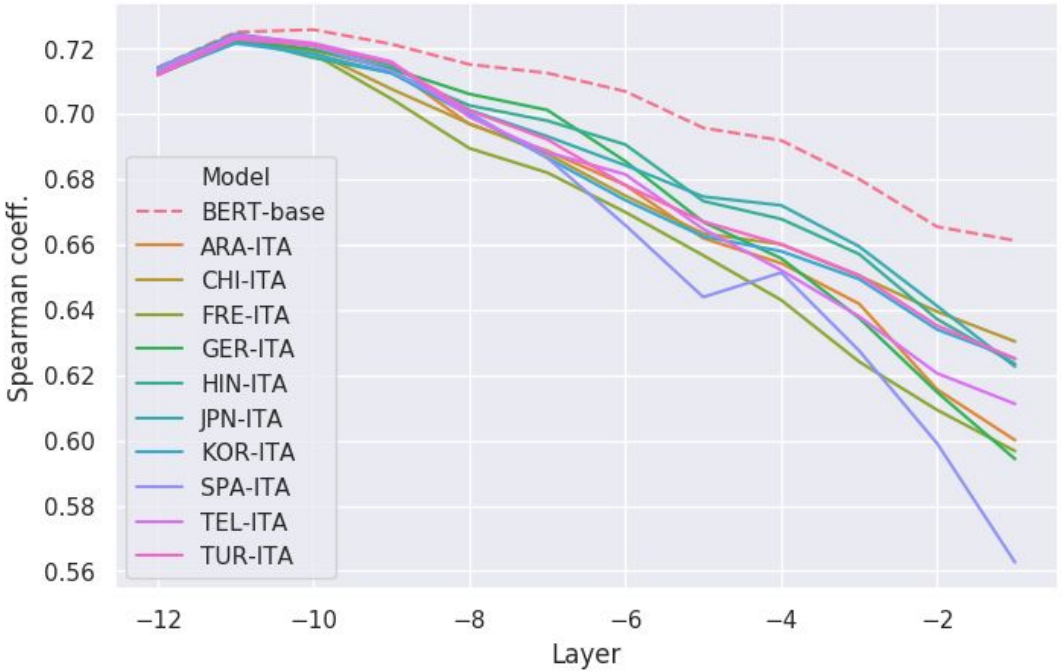
Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

- Fine-tuning of BERT on the *Native Language Identification* (NLI)

"No breakfast, coz you still have enough alcohol in your stomach."

- Probing tasks on the fine-tuned models (x10)

Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

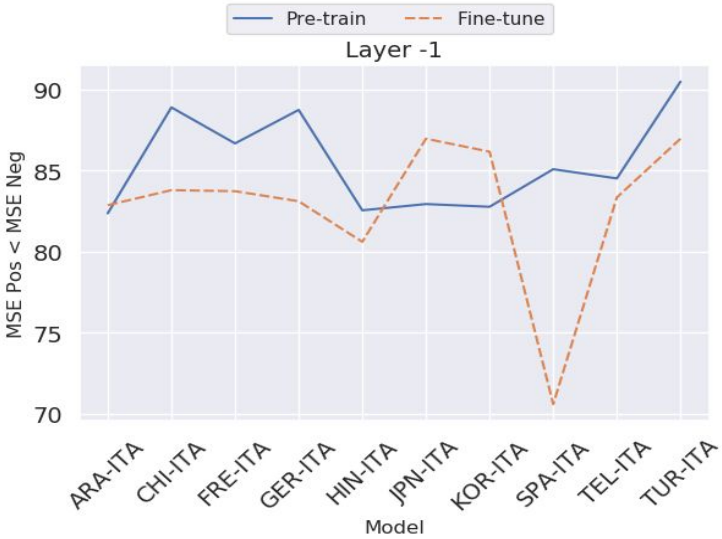


Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

- We have split each NLI dataset in sentences correctly and incorrectly classified by BERT
- We computed the MSE for each subset and each probing feature

Linguistic Profiling of a Neural Language Model (Miaschi et al., 2020)

- We have split each NLI dataset in sentences correctly and incorrectly classified by BERT
- We computed the MSE for each subset and each probing feature



Probing Linguistic Knowledge in Italian Neural Language Models

- How about Italian Transformers?

- In “Probing Linguistic Knowledge in Italian Neural Language Models across Language Varieties” (Miaschi et al. 2020), we applied our *profiling approach* on 7 different Transformer models available for the Italian language, in order to:
  - Compare the performances of the 7 pre-trained NLMs;
  - Investigate whether and how the knowledge encoded by these NLMs differs across textual genres and language varieties.

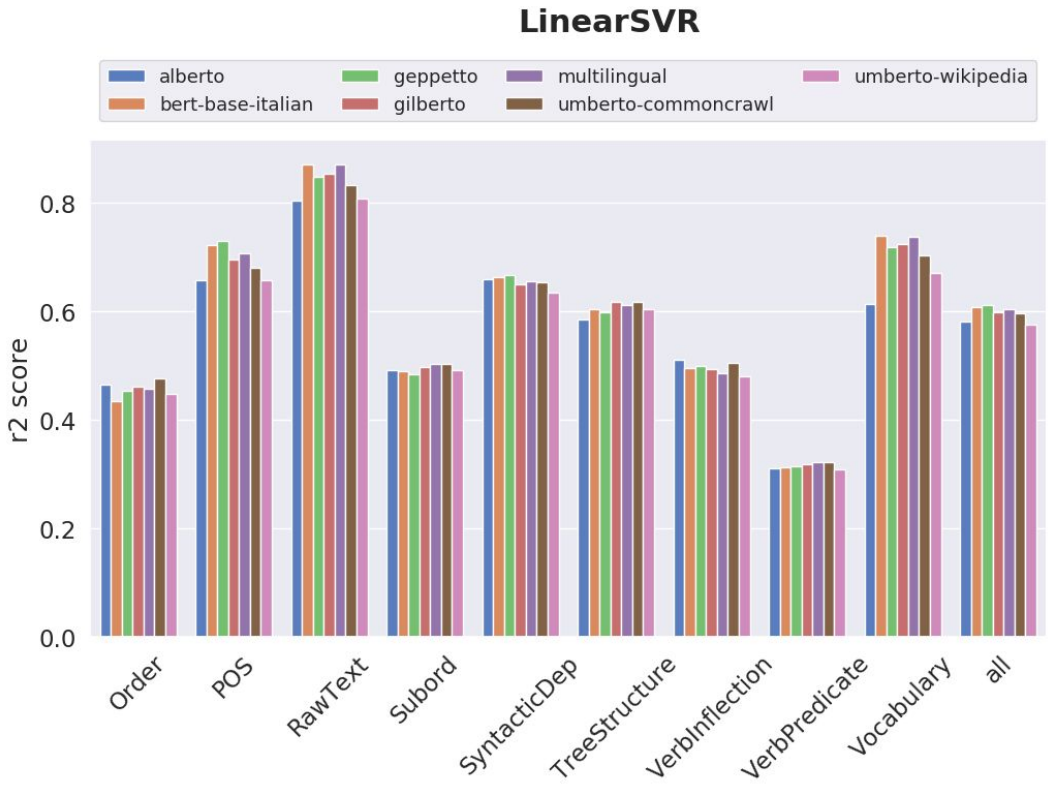
How about Italian Transformers?

Name	Training data
BERT Architecture	
Multilingual-BERT	Wikipedia
BERT-base-italian	Wikipedia + OPUS (13GB)
AlBERTo	TWITA (191GB)
RoBERTa Architecture	
GilBERTo	OSCAR (71GB)
UmBERTo-Commoncrawl	OSCAR (69GB)
UmBERTo-Wikipedia	Wikipedia (7GB)
GPT-2 Architecture	
GePpeTto	Wikipedia + ItWAC (14GB)

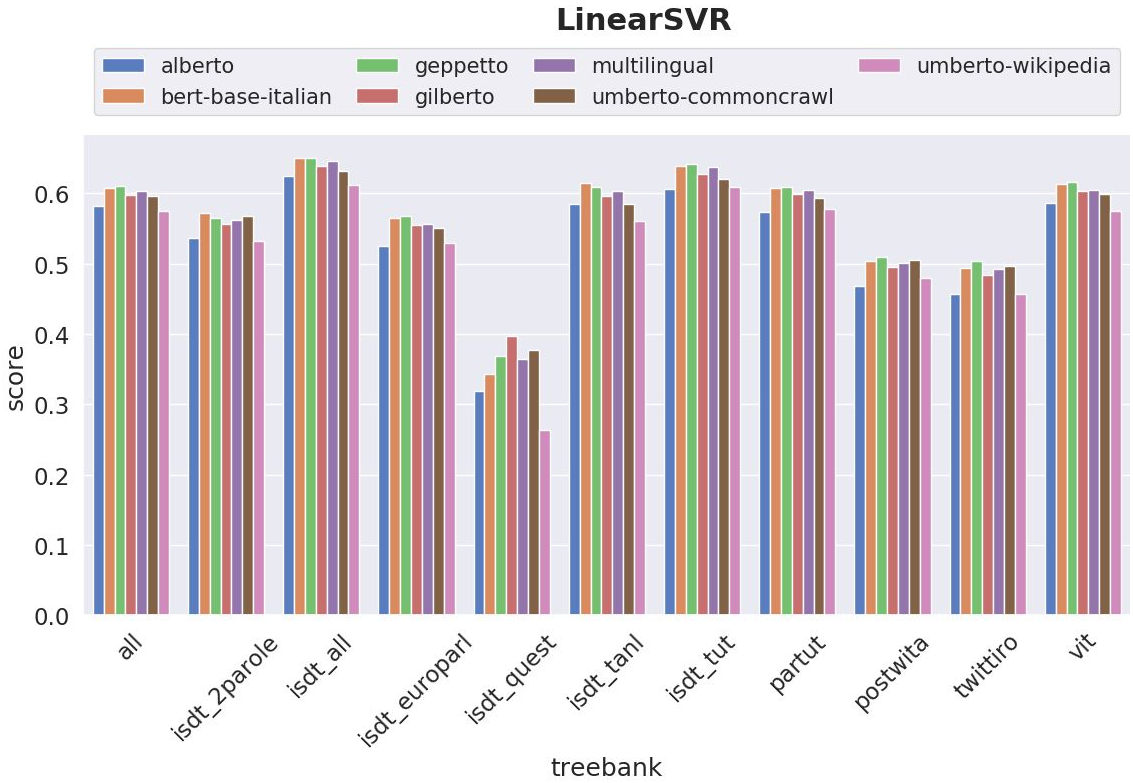
How about Italian Transformers?

Short Name	Types of texts	# sent
ParTUT	Multi-genre	2,090
VIT	Multi-genre	10,087
ISDT	Multi-genre	14,167
ISDT_tanl	Newswire	4,043
ISDT_tut	Legal/Newswire/Wiki	3,802
ISDT_quest	Interrogative sentences	2,162
ISDT_2parole	Simplified Italian news	1,421
ISDT_europarl	EU Parliament debates	497
PoSTWITA	Tweets	6,713
TWITTIRÒ	Ironic Tweets	1,424
Total		35,481

How about Italian Transformers?



How about Italian Transformers?



Conclusion and Future Directions

Conclusion and Future Directions

- NLMs have reached astonishing performance in almost all NLP tasks
- However, this improvement comes at the cost of **interpretability**
- Several methods have been implemented to understand the inner mechanisms and decision-making processes of these models
    - and it is an ever-evolving and exciting area of research (e.g. Li et al., 2022, Bensemann et al., 2022)

Conclusion and Future Directions

- NLMs have reached astonishing performance in almost all NLP tasks
- However, this improvement comes at the cost of **interpretability**
- Several methods have been implemented to understand the inner mechanisms and decision-making processes of these models
    - and it is an ever-evolving and exciting area of research (e.g. Li et al., 2022, Bensemann et al., 2022)

Future Directions:

- Study how the linguistic knowledge arise during the pre-training phase of a NLM and how it changes when dealing with different training objectives
- Improve the robustness of NLMs by e.g. selecting input data appropriately during the pre-training phase and thus strengthening their implicit linguistic competence
- …Prompting for linguistic competence? (Liu et al., 2021)

Thanks for the $softmax(\frac{QK^T}{\sqrt{D_k}})V$!

https://alemiaschi.github.io/

@AlessioMiaschi

http://www.italianlp.it/

@ItaliaNLP_Lab

References

- Bengio, Yoshua, et al. (2003). "A neural probabilistic language model." *The journal of machine learning research* 3, pages 1137-1155.
- Vaswani, Ashish, et al. (2017). "Attention is all you need." *Advances in Neural Information Processing Systems* (NEURIPS)
- Devlin, Jacob, et al. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.
- Belinkov, Yonatan, and James Glass. (2019). "Analysis methods in neural language processing: A survey." *Transactions of the Association for Computational Linguistics* 7, pages 49-72.
- Hewitt, John, and Christopher D. Manning (2019). "A structural probe for finding syntax in word representations." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.
- Clark, Kevin, et al. (2019) "What Does BERT Look at? An Analysis of BERT's Attention." *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*.
- Goldberg, Yoav (2019). "Assessing BERT's syntactic abilities." *arXiv preprint arXiv:1901.05287*.
- Pimentel, Tiago et al. (2020). "Information-Theoretic Probing for Linguistic Structure". In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4609–4622, Online. Association for Computational Linguistics.
- Ramnath, Sahana, et al. (2020). Towards Interpreting BERT for Reading Comprehension Based QA. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP), pages 3236–3242, Online. Association for Computational Linguistics.

References

- Tenney, Ian et al. (2019). “BERT Rediscovers the Classical NLP Pipeline”. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4593–4601, Florence, Italy. Association for Computational Linguistics.
- Petroni, Fabio et al. (2019). “Language Models as Knowledge Bases?”. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473, Hong Kong, China. Association for Computational Linguistics.
- van Halteren, Hans (2004). “Linguistic Profiling for Authorship Recognition and Verification”. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 199–206, Barcelona, Spain.
- Brunato, Dominique et al. (2020). “Profiling-UD: a Tool for Linguistic Profiling of Texts”. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 7145–7151, Marseille, France. European Language Resources Association.
- Miaschi, Alessio, et al. (2020) "Linguistic Profiling of a Neural Language Model." *Proceedings of the 28th International Conference on Computational Linguistics*.
- Miaschi, Alessio et al. (2021). “Probing Tasks Under Pressure”. In *CLiC-it 2021*.
- Hewitt, John and Liang, Percy (2019). “Designing and Interpreting Probes with Control Tasks”. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2733–2743, Hong Kong, China. Association for Computational Linguistics.
- Li, Jiaoda et al. (2022). “Probing via Prompting”. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1144–1157, Seattle, United States. Association for Computational Linguistics.

References

- Bensemann, Joshua et al. (2022). “Eye Gaze and Self-attention: How Humans and Transformers Attend Words in Sentences”. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 75–87, Dublin, Ireland. Association for Computational Linguistics.
- Liu, Pengfei, et al. (2021) "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing." *arXiv preprint arXiv:2107.13586*.