

The Role Of Domain Specific Language when Modeling Dutch Hospital Notes with Transformers

Stella Verkijk & Piek Vossen

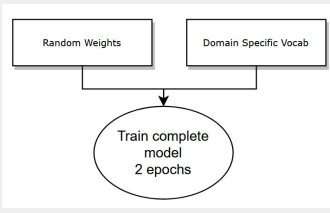


s.verkijk@student.vu.nl

Presenting the first domain-specific medical language model for Dutch

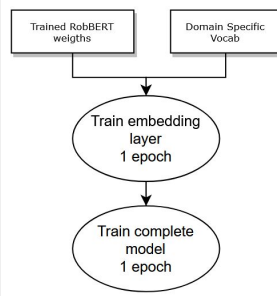
Building two models

From Scratch



NB: Both models are RoBERTa-based [2]

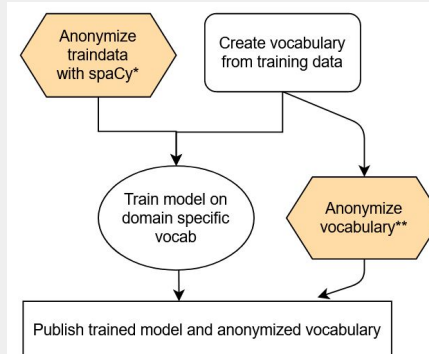
Extending RobBERT [1]



Anonymization

* replace names with *PERSON*

** remove tokens most similar to *PERSON*



Research Questions

1. Will a domain specific language model for Dutch hospital notes produce more accurate sentence embeddings within that domain than general language models for Dutch?

1. Will fine-tuning a domain specific language model for Dutch hospital notes lead to better performance at downstream tasks within that domain than when fine-tuning general language models for Dutch?

Data

		2017	2018	2020
AMC	GB	2.8	3.0	2.0
	notes	2.375.626	2.451.973	1.492.573
VuMC	GB	3.0	-	1.5
	notes	2.545.515	-	1.111.116

Evaluation of models

Testing the models' raw knowledge (assessing accuracy of sentence embeddings) by evaluating similarity judgements without fine-tuning

Intrinsic

Extrinsic

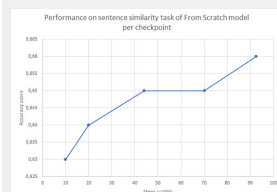
Testing the models' performance on a medical downstream task: classifying four domains of the WHO's International Classification of Functioning, Disability and Health (ICF)[5] in hospital notes. Annotated data set taken from the a-proof project [6].

Results on sentence level

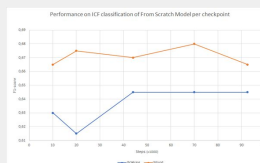
	RobBERT	BERTje	From Scratch	Ext. RobBERT
Walking & Moving	0.62	0.62	0.65*	0.63
Mood	0.66	0.69	0.67	0.66
Exercise Tolerance	0.42	0.45	0.45	0.45
Profession and Work	0.40	0.40	0.39	0.39
None	0.99	0.99	0.99	0.99
Macro Avg.	0.62	0.63	0.63	0.62

Scores averaged over 8 fine-tuned models per model type. * Wilcoxon Signed Rank Test: p-value of 0.1

Effect of training time



Effect of training time



“Inspanningsgebonden dyspneu en thoracale druk verdacht voor angina pectoris bij gering anteroseptale ischemie en bij eerdere laagnormale perfusie LADgebied met mix-plaque van D1/D2”

Conclusions

- Better to train from scratch
- Domain-specific model internalised more accurate semantic knowledge
- No big differences with other models when fine-tuned
- It is possible to create a competitive language model on limited computational power
- Domain-specific language models need less pre-training time

[1] Delobelle, P., Winters, T., & Berendt, B. (2020). Robbert: a dutch roberta-based language model. *arXiv preprint arXiv:2001.06286*. [2] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*. [3] 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*. [4] de Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., & Nissim, M. (2019). Bertje: A dutch bert model. *arXiv preprint arXiv:1912.09582*. [5] T. B. Üstün, S. Chatterji, J. Bickenbach, N. Kostanjsek, and M. Schneider. The international classification of functioning, disability and health: a new tool for understanding disability and health. *Disability and rehabilitation*, 25(11-12):565-571, 2003. [6] Unpublished