IDLab
INTERNET & DATA LAB
UNIVERSITEIT GENT

High-quality translation of clinical Dutch notes into fluent English

REMY FRANCOIS DE JAEGER PETER
DEMUYNCK KRIS

The ADAM Project

- Advanced Data-Augmented Medicine
  - Work package 1: extracting information from unstructured clinical notes
- Personalized healthcare
  - Bringing up medical records of similar patients for clinicians or models to look at
- Funding by AZ Delta and VLAIO
  - Project running in the next three years

IDLab INTERNET & DATA LAB

azdelta Uw ziekenhuis.

Usages for medical text

- Entity extraction and linking
  - MetaMap
  - SciSpacy
  - …
- Natural language inference
  - MedNLI
- Question answering
  - MedQA
- […]

“Up to 80% of the relevant information in electronic health records can only be found in the unstructured text typed by clinicians.”

- DAN RISKIN, MD

State of the art (Clinical NLP)

- Currently moving at a fast pace
  - Transformer models with high generalization capabilities make this possible
- Only exists in select few languages, like English
  - Existing multi-lingual models are not specialized in handling medical text
- Easiest route for re-use seems to be translation
  - Also useful for translating datasets, evaluation sets, etc…

“Google Translate has only 57.7% accuracy when used for medical phrase translations and should not be trusted for important medical communications.”

- PATRICK DAVIES, RESEARCHER

Challenges of medical translation

- Not a lot of medical data
  - But highly specific vocabulary
- Medical NLU in English is not trained on translated data
  - Correct but unnatural translations might not be understood
  - Required to stick to English terminology, not translate word for word
- There are many ways to say the same things
  - All ways of saying something must be translated properly
  - Lots of abbreviations and idiosyncrasies

Data scarcity: why?

- Privacy and data regulation in place
  - We cannot train on un-anonymized patient data
  - We usually need consent to use medical data, including EHR records
- Human translation of medical text is expensive
  - Rates for translation of medical text are up to 10x higher than regular text
  - Without the full hospital context, it is sometimes difficult to do anyway
- Solution:
  - Look for creative ways to add more data

“SNOMED CT is a standardized, multilingual vocabulary of clinical terminology that is used by physicians and other health care providers for the electronic exchange of clinical health information.”

- NATIONAL INSTITUTES OF HEALTH, VIA WIKIPEDIA

Parallel SNOMED CT corpus

"Every medical concept on earth"

- Medical terms are very specific; many phrases must be memorized
- Solution: add a parallel corpus of medical phrase translation to the training set in large quantity
- Method: using Google LaBSE to mine Snomed CT concepts for parallel translations

Medical concepts… in sentences!

- Poor generalization after seeing just the SNOMED dictionary translation
- Solution: create artificial sentences containing the concepts, and add to the training set
- Method: craft template-sentences with masks that will be replaced by terms in the ontology

Find parallel translations

SNOMED CT candidate matches

- Cat scratch fever
- Cat scratch disease
- Benign lymphoreticulosis
- Lymphoreticulosis benigna
- Kattenkrabziekte

Google LaBSE

- The model is a Transformer + Mean pooling (SimCLR loss)
- The embeddings of a phrase / sentence should be such that:
  - Sentences that are translations of each other have high cosine similarity
  - Sentences that are not translations of each other have low cosine similarity

Sentence templates

Knowing that SNOMED CT contains phrases, we can build sentences

- **NL:** Laten wij over <*pneumonie*> praten.
- **EN:** Let's talk about <*pneumonia*>.

Examples:

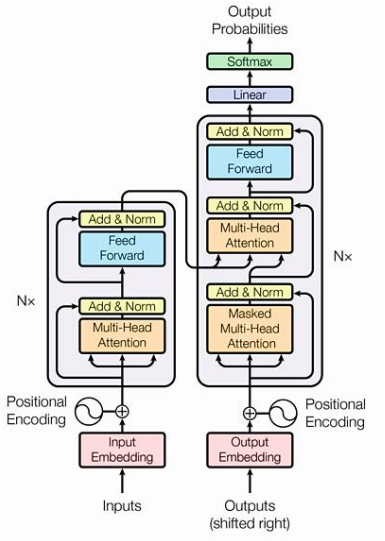
- Na zijn behandeling voor sputumvinding, werd een nieuw onderzoek gepland.
- Reden van opname: bloed klontert niet goed (2035-12-16)
- Hoe verklaar jij dat ongemak?
- …

Experiment

Train a state-of-the-art model (MarianNMT) using publicly available data and our novel augmented data sources, and compare translation performance both using BLEU and using performance on a downstream task as a metric

Transformer architecture
6-layer encoder + 6-layer decoder
Static positional embeddings

Vocabulary trained on OPUS-MT
So not relearnt based on medical domain, to be able to use existing models without retraining



SNOMED CT
EUROPEAN MEDICINES AGENCY
MARIANNMT

“Marian is a state-of-the-art Neural Machine Translation framework. Due to its self-contained nature, it is quite easy to optimize Marian for specific tasks.”

- MICROSOFT TRANSLATOR TEAM

Results

We were able to produce a model that:

1. is competitive with Google Translate (and possibly better in some cases)
2. without using any data from the hospital EHR systems, yet.

We show this using three metrics:

1. MetaMap entity extraction recall score
2. BLEU score on sentences extracted from medical records, unseen during training
3. GPT-2 perplexity score of the translations of the sentences above

Results: Metamap Score

Average of Recall	Column Labels											
Row Labels	Microsoft		MarianFT		Google		FT2		DeepL		Grand Total	
Metamap	69%		86%		82%		82%		80%		80%	
SciSpacy	49%		57%		56%		55%		57%		55%	
Grand Total	**59%**		**71%**		**69%**		**69%**		**68%**		**67%**	

Results: BLEU Score

	Google	DeepL	Bing	Ours
BLEU:	**64.29**	**61.76**	**58.54**	**69.13**
Precision x brevity:	67.87 x 94.72	64.81 x 95.29	59.68 x 98.09	72.02 x 96.00

Type	**1-gram**	**2-gram**	**3-gram**	**4-gram**	**1-gram**	**2-gram**	**3-gram**	**4-gram**	**1-gram**	**2-gram**	**3-gram**	**4-gram**	**1-gram**	**2-gram**	**3-gram**	**4-gram**
Individual	89.44	73.68	61.84	52.06	88.09	70.83	58.56	48.29	84.00	66.72	53.18	42.57	90.14	77.16	66.83	57.87
Cumulative	84.72	76.89	70.23	64.29	83.94	75.27	68.12	61.76	82.39	73.43	65.52	58.54	86.53	80.06	74.36	69.13

Results: GPT-2 Score

	Google	DeepL	Microsoft	Ours
Mean Perplexity	298.36	294.80	375.18	214.76
Mean Ppl. Z-score	-0.52	-0.23	1.03	-0.27

Table 3: Perplexity of the translation as evaluated by GPT2-Large (lower is better)

Results: Ablation studies

	Ours	NSS	NS
Mean BLEU score	69.13	62.38	56.63
MetaMap recall score	86%	83%	79%

adding the dictionary is useful!

NSS ***(No SnomedCT sentences)***: The first variant consists in a model trained on the same data as our best model, including the LaBSE-matched medical phrases from SnomedCT, but from which the computer-generated sentences including those matches were excluded. The results of this ablation show the gain in generalization obtained by generating sentences including the SnomedCT terms.

NS ***(No SnomedCT)***: The second variant consists in a model trained on the same data as our best model, but from which both the SnomedCT matches found by LaBSE and their augmented sentences have been excluded. The results of this ablation study show the gain in vocabulary obtained by the addition of the SnomedCT matches.

Results: Examples (Bonus)

Geen pericard- of pleuravocht. Levervenen niet gestuwd. Lichte aortaklepsclerose passend bij de leeftijd. Overige kleppen normaal. Linker en rechter ventrikel niet gedilateerd. Normale contractiliteit en wanddikte. Op doppler mitraalinflow en relaxatiestoornis passend bij de leeftijd. Mitraalinsufficiëntie graad 1. Tricuspiedinsufficiëntie graad 1 met normale pulmonaaldruk.

No pericardial or pleural effusion. Liver veins not stowed. Mild aortic valve sclerosis appropriate to age. Other valves normal. Left and right ventricles undilated. Normal contractility and wall thickness. On doppler mitral influx and age-appropriate relaxation disorder. Grade 1 mitral insufficiency. Grade 1 tricuspid insufficiency with normal pulmonary pressure.

No pericardial or pleural fluid. Hepatic veins not obstructed. Mild aortic valve sclerosis consistent with age. Other valves normal. Left and right ventricle not dilated. Normal contractibility and wall thicknesses. On doppler mitral inflow and relaxation disorder consistent with age. Mitral insufficiency grade 1. Tricuspid insufficiency grade 1 with normal pulmonary pressure.

Conclusion & QA

- Download the parallel corpus (NL/FR/ES/DA/SV):
  - https://github.com/FremyCompany/snomed-translate-dictionaries

- Contact me by email:
  - Francois.Remy@ugent.be

- Follow me on Twitter:
  - @FremyCompany

- To learn more about the ADAM Project:
  - https://www.azdelta.be/nl/geneeskunde-op-maat-van-de-patient-ai

- To learn more about IDLab:
  - https://idlab.technology/