A Machine Translation Pipeline for Medical Terminologies

K. Skianis[1], **T. Merabti**[2], **Y. Briand**[2], **M. Mary**[2], **T. Dart**[2]
[1]BLUAI, Athens, Greece
[2]Agence du Numérique en Santé, Paris, France

Presentation: https://y3nk0.github.io/presentations/medical_nmt_26-10-2022.pdf

26 October, 2022

Outline

Medical terminologies

- Essential for health institutions to store, organize and exchange all medical-related data generated in labs, hospitals etc.
- Arranged in dictionaries and lexicons, following specific structures and coding rules
- As the initial versions are created in English, there is an evident need for translation in other languages

Limitations

- Infeasible to model all structures and rules accurately
- Constantly updated
- Manual translation:
  - expensive in time and resources
  - number of medical terms may increase
  - requiring health professional efforts for evaluation

Machine translation to the rescue!

Related work

Statistical Machine Translation (SMT)

- Nyström et al. (2006): using ICD-10, ICF, MeSH, NCSP and KSH97-P for semi-automatic creation of an English-Swedish medical terminology via word alignment
- Deléger et al. (2009): automatically acquiring new translations of medical terms based on word alignment in parallel text corpora
  ↪ tested in English and French

Neural Machine Translation (NMT)

- Encode the input sequence and generate a variable length translated sequence using Recurrent Neural Networks (RNN) (Bahdanau et al. (2014);Sutskever et al. (2014))
- Khan et al. (2018): transfer learning on med-scientific publications
- Using NMT for medical terminologies remains unexplored!

Contributions

Generic and open approach for medical terminology translation:

- Speed-up the official translation process
- Created ground-truth datasets
- Introduced supervised and unsupervised metrics
- First baseline translations for multiple terminologies
- Paper published in **LOUHI 2020**
  Link: `https://y3nk0.github.io/papers/medical_translation_louhi2020.pdf`
- Poster in **WHO FIC 2021**
- Online demo, API and tools (available only internally for now)

Outline

Models

Neural machine translation methods were tested, using large parallel text corpora and medical terminologies to train translation models that learn correspondences between them:

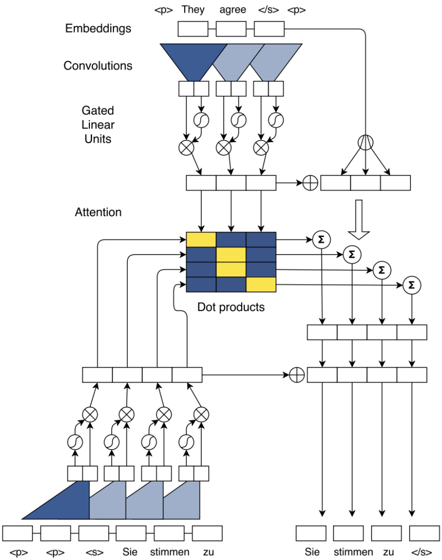
- Via deep neural networks and vector representations: continuous space representations for words (embeddings)
- fairseq (Ott et al., 2019): a sequence modelling toolkit that allows researchers and developers to train custom models for translation
- Finetune pre-trained models (transfer leaning on pre-trained model, Khan et al. (2018)):
  - 3 Convolutional Neural Networks (CNNs, Gehring et al. (2017))
  - 1 Transformer (Ott et al., 2018)
  - ensemble of CNNs

Convolutional Sequence to Sequence Learning

Compared to recurrent models:

- computations over all elements can be fully parallelized during training
- optimization is easier since the number of non-linearities is fixed and independent of input length
- use of gated linear units eases gradient propagation
- each decoder layer comes with a separate attention module

Paper: Gehring et al. (2017)



Rounds, datasets & setup

Rounds	Description
1st and 2nd rounds	2020 Datasets: ICD-10, CHU Rouen, ORDO, ACAD, MEDDRA, ATC, MESH, ICD-O, DBPEDIA, ICPC, ICF
3rd round	Cleaning to remove bilingual sentences leading to ambiguities (e.g. ICPC is not relevantly structured for use in a training set)
4th round	3rd round + PatTR corpus (patents database)
5th round	3rd round + Medline (training2), Scielo datasets
6th round	5th, with Transformer architecture
Ensemble	an ensemble of the 3 CNN models was created : 3rd, 4th, 5th rounds

Setup

- $\sim$1M labels for training
- NVIDIA RTX 2080 Ti (12GB)
- 30 epochs

Pre- & post-processing

- We may want custom solutions
- Not all rules can be be applied after inference
- Split them to pre and post

Pre-processing (before training)

- remove unwanted structures or even whole terminologies
- examine casing (upper-lower) for training

Pre-inference (before translation)

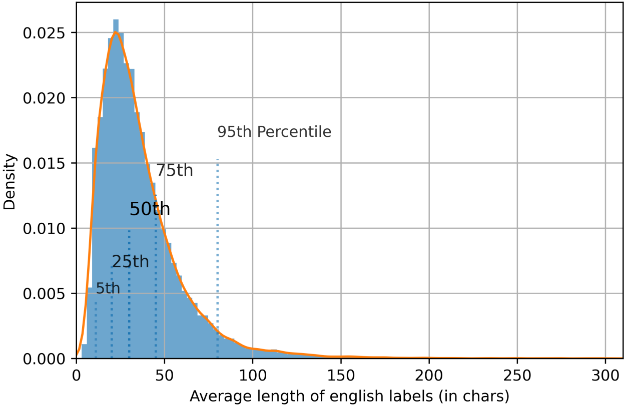
- examine casing for english label for inference
- structures, rules, punctuation removal

Post-inference (after translation)

- fastest but not easy to apply specific rules
- detect acronyms

Ground-truth datasets for testing-evaluation

(1) For ICD-11, since the French official version did not exist, we gathered terms from existing terminologies
(2) ATIH provided a human translated subset of ICD11 for reference corresponding of human validated translation preformed in 2019



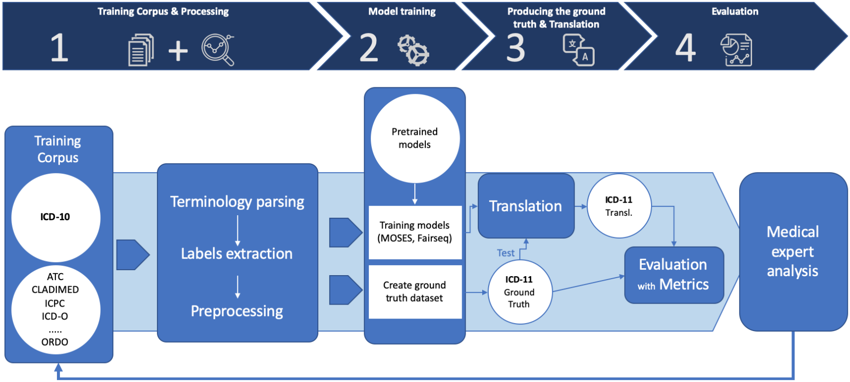
Distribution of chain length in ATIH reference set (75k validated terms)

Evaluation

Translation metrics

- BLEU (Bilingual Evaluation Understudy) (Papineni et al., 2002) is calculated for individual translated segments (n-grams) by comparing them with a dataset of reference translations - a dimensionless metric between 0 (possibly wrong translation) to 1 (exact match)
- BLEU is very harsh on penalizing sentences that may carry synonyms, which is applicable in cases where reference is limited
  $\hookrightarrow$ a relevant translation might get a very low BLEU score
- BLEU2VEC (Tättar and Fishel, 2017): a metric which utilizes word embeddings for taking under consideration similarity between translation and reference

Methodology

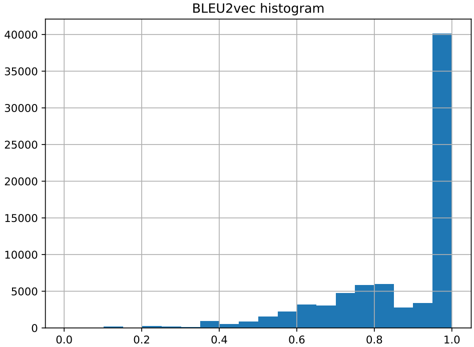


The proposed machine translation pipeline.

Outline

Results

Model	BLEU2VEC
3rd round	0.66808
4th round	0.67992
5th round	0.66948
6th round	0.67379
Ensemble CNNs	0.68254



Scores and distribution of BLEU2VEC scores on the ATIH reference set

Remarks

- fast training (with GPU) and inference (even with CPU)
- CNNs are effective enough for medical terminologies due to size
- casing is important

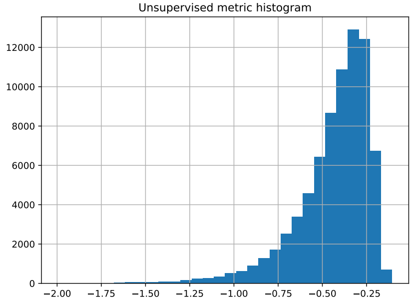
Examples

BLEU2VEC score range	English original label **ATIH human**	**Ensemble**	**Remark**
0-0,2	Glycogen storage disease type 3 **Maladie de stockage du glycogène de type 3**	Glycogénose type 3	Glycogénosis is part of index terms
	Fibular hemimelia **Raccourcissement (congénital) longitudinal de la fibula**	Hémimélie fibulaire	Automated translation was literal while human translator chose a description
0,21-0,9	mechanical prosthetic valve disease **atteinte de la valve prothétique mécanique valvulopathie mécanique**	valvulopathie prothétique mécanique	Automated translation more literal
	Aluminium bone disease, hand **Maladie osseuse provoquée par l'aluminium, à la main**	ostéopathie à l'aluminium, main	Automated translation more literal
$\geq$ 0,9	Small infarctions of cochlear, retinal and encephalic tissue **Petits infarctus cochléaire, rétinien et du tissu cérébral**	Petit infarctus cochléaire, rétinien et du tissu encéphalique	ATIH human translation and automated translation are very close.
	protozoal infection **infection à protozoaire**	infection à protozoaires	Very close, plural?

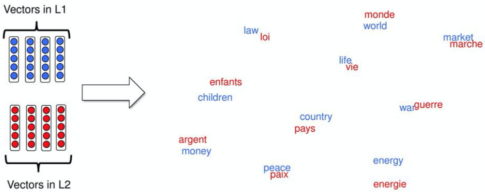
Examples of automated translation outputs compared with human translations.

Unsupervised metrics

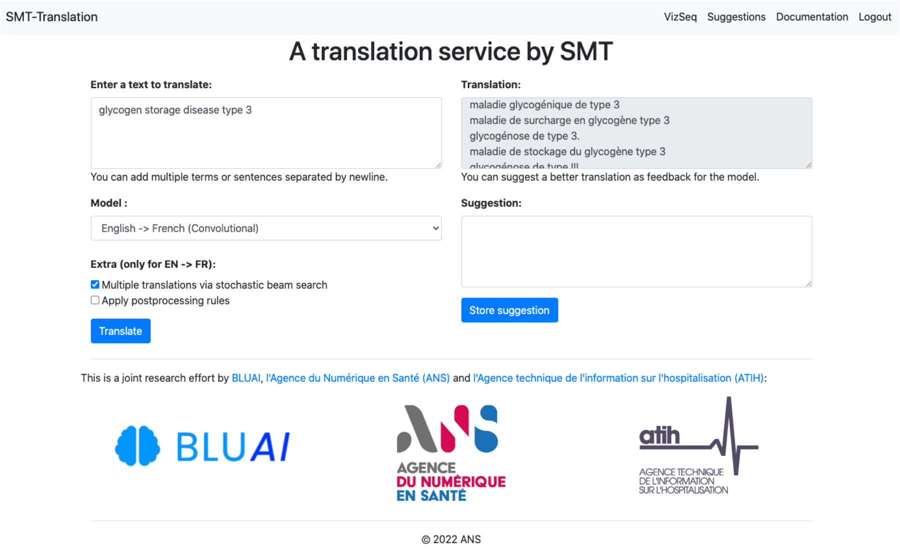
- Exploiting softmax output probability distribution + entropy of attention weights from the NMT model → leverage uncertainty quantification for unsupervised scoring



- Score multiple translations via cross-lingual word embeddings



Demo


SMT-Translation
VizSeq
Suggestions
Documentation
Logout
A translation service by SMT
Enter a text to translate:
glycogen storage disease type 3
You can add multiple terms or sentences separated by newline.
Model :
English -> French (Convolutional)
Extra (only for EN -> FR):
Multiple translations via stochastic beam search
Apply postprocessing rules
Translate
Translation:
maladie glycogénique de type 3
maladie de surcharge en glycogène type 3
glycogénose de type 3.
maladie de stockage du glycogène type 3
You can suggest a better translation as feedback for the model.
Suggestion:
Store suggestion
This is a joint research effort by BLUAI, l'Agence du Numérique en Santé (ANS) and l'Agence technique de l'information sur l'hospitalisation (ATIH):
BLUAI
AGENCE DU NUMÉRIQUE EN SANTÉ
atih
AGENCE TECHNIQUE DE L'INFORMATION SUR L'HOSPITALISATION
© 2022 ANS

Docs

ANS Translation Service

Search docs

CONTENTS:

» Welcome to ANS Translation Service's documentation!

View page source

Welcome to ANS Translation Service's documentation!

AGENCE DU NUMÉRIQUE EN SANTÉ

Our objectives can be summarized in the following points:

- An automated pipeline for translating and evaluating medical terminologies
- Generate ground truth datasets to evaluate our models
- Introduced supervised and unsupervised metrics
- Create online services to provide access to our tools
- Enable researchers and healthcare end-users globally with a jump start approach that allows fast and effective translation of newly updated versions of terminologies

Vizseq

 # 6 / 10 / 500 (376 / 500) fr

Source 0	Fracture, avulsion or collateral ligament rupture of medial malleolus with fracture of fibula above syndesmosis and fracture of posterior margin of distal tibia GTranslate
Reference 0	**Fracture,** arrachement ou **rupture** du **ligament** collatéral de la malléole interne avec fracture du péroné au-dessus de la syndesmose et fracture du bord postérieur du **tibia distal**
5th.round.cnn	**Fracture,** avulsion **ou rupture du ligament collatéral de la malléole médiale avec fracture du péroné au-dessus d'une syndesmose et fracture du bord postérieur du tibia distal**
6th.round.transformer	**fracture,** avulsion **ou rupture du ligament collatéral de la malléole interne avec fracture de la fibula au-dessus de la syndesmose et fracture du bord postérieur du tibia distal**
ensemble.round.cnns	**Fracture,** avulsion **ou rupture du ligament collatéral de la malléole interne avec fracture du péroné au-dessus de la syndesmose et fracture du bord postérieur du tibia distal**

Model	bleu
5th.round.cnn	*68.27*
6th.round.transformer	75.06
ensemble.round.cnns	**93.05**

Outline

Summary

- An automated pipeline for translating and evaluating medical terminologies $\rightarrow$ one of the first approaches to use deep learning for translating medical terminologies
- Multiple translations
- Provided ground truth datasets:
  - Generate automatically a test subset via existing terminologies
  - ATIH reference dataset
- Supervised and unsupervised metrics, pre- and post-processing
- Demo, tools & API available
- **Enable researchers and healthcare end-users globally with a jump start approach that allows fast and effective translation of newly updated versions of terminologies**
- **Beyond translation, the approach can be used to connect or enrich terminologies (e.g. using synonyms)**

Thank you! Questions?