Development of a Deep Learning-Based Automated Mapping Tool in the Conversion Process of OMOP-CDM

PRESENTER: **Yourim** Lee. MS[1],
Aelan Park, RN, CMD, Ph.D[1],
Hye Jin Kam. Ph.D[1]

Background

- Vocabulary mapping from the original EMR terms to OMOP-CDM vocabularies is time consuming and labor-intensive process.
- Usagi is quite useful to map between English vocabularies, there are barriers in vocabulary mapping with the same meaning of different forms.
- Because the basic principle TF-IDF of Usagi based on the bag-of-words model, it cannot capture position in text, semantics, co-occurrences (e.g. as compared to topic models, word embeddings).
- **We proposed a DEep learning based AUtomated mapping tool (DEAU) to reflect the information of the relationship between words and their semantic meanings.**

Methods

Data: Source codes from Korean diagnoses (n=83,113) that were mapped to condition domain concepts of OMOP.
Dataset Split: train : validation : test = 7 : 1 : 2, but It took some time to test Usagi, so we only used 500 rows of the testsets.
Adopted algorithms: FastText for word embedding, Infersent for sentence representation.

Results

- The DEAU found concept matches 77.37% and Usagi did 65.87% to the target OMOP concepts.
- At the Venn diagram of the concepts to be actually mapped in the top 100, we verified that the non-overlapping concepts were 18% of the whole test sets (overlaps were 69%).

Discussion

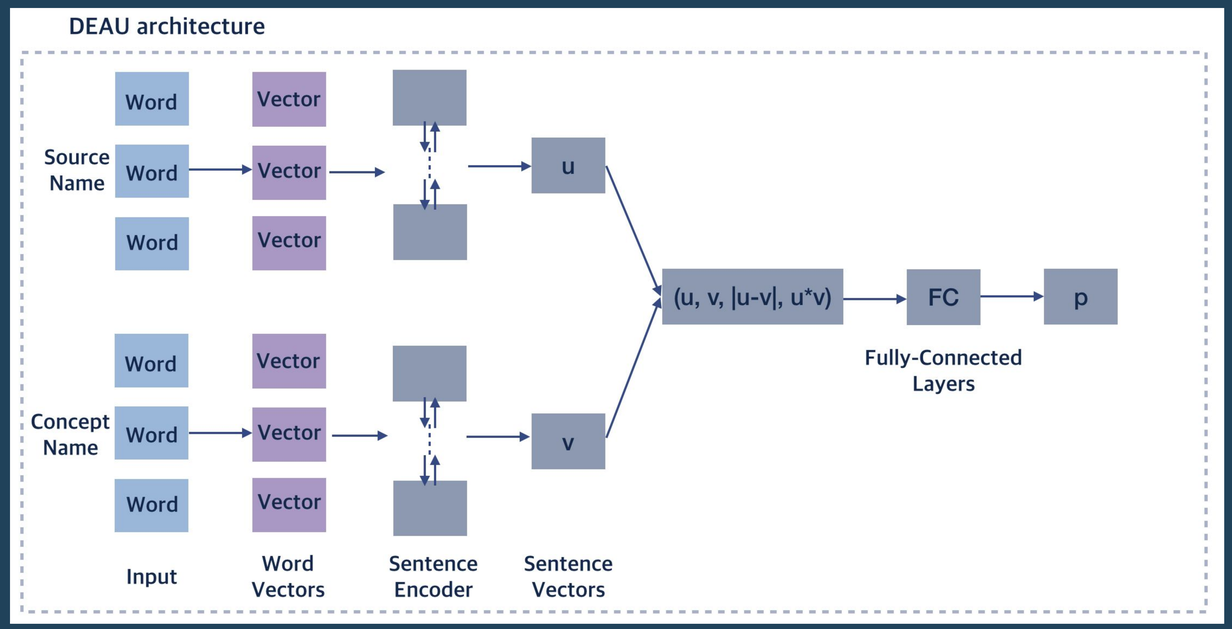
- By comparisons of the matching results between Usagi and DEAU, we confirmed that Usagi has a higher performance when the source name match with the words of the concept name.
- On the other hand, DEAU showed a high degree of similarity finding ability even the words of the concept name and the source name were formed in totally different forms.
- We expect that DEAU will provide better results for the current mapping performance.

DEAU is automated tool to aid users for map standard vocabulary of OMOP-CDM.

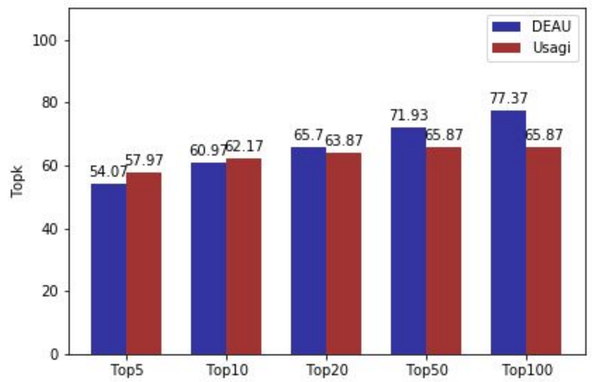
Take a picture to download the full paper

Contact: kam.hyejin@evidnet.co.kr

DEAU architecture

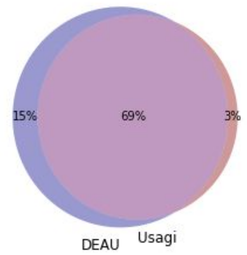


The percent of matching the real targets



- Top*k* evaluation metrics to compare the performance of Usagi and DEAU. This figure is the result of percent matching the real targets in top*k* between Usagi and DEAU tool.
- Usagi performed better until top10 but DEAU was better performance from top20.

The Venn diagram of matching the real targets



- The degree of overlap (purple part) of the set of the sources that Usagi and DEAU matched real targets in a Venn diagram.

Evaluation metrics of Usagi and DEAU

Tool	Presicion@100	recall@100
Usagi	0.0120	0.5663
DEAU	0.0097	0.7737

- To compare the performance of the 1:n mapping, we adopted precision at rank 100 and recall at rank 100 of Usagi and DEAU.

Yourim Lee. MS[1], Ba Rom Kang. BS[2], Soondong Kim. MS[1], Soo-Yeon Cho. RN, MPH[1], Aelan Park, RN, CMD, Ph.D[1], Ha Young Kim. Ph.D[3], Byungkon Kang. Ph.D[4], Hye Jin Kam. Ph.D[1]

[1]EvidNet, Seongnam, South Korea; [2]Dept.of Data Science Suwon South Korea; [3]Graduate School of Information, Yonsei University, Seoul, South Korea; [4]Dept. of Computer Science, SYNU Korea, Incheon, South Korea

Acknowledgement: This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea)

evidnet

OHDSI