Learning Embeddings from Free-Text Triage Notes Using Pretrained Transformer Models

Émilien Arnaud, Mahmoud Elbattah, Maxime Gignon, Gilles Dequen
Université de Picardie Jules Verne (UPJV), France
mahmoud.elbattah@u-picardie.fr

https://www.researchgate.net/publication/358867707_Learning_Embeddings_from_Free-text_Triage_Notes_using_Pretrained_Transformer_Models

UNIVERSITÉ de Picardie Jules Verne

Study Context

- Exploring pre-trained BERT models as a mechanism for learning embeddings from clinical notes.

- A use case of triage notes in the French language, from the University Hospital of Picardie Jules Verne in France.

UNIVERSITÉ de Picardie Jules Verne

Data Description

More than 260K ED records over the period of January 2015 to June 2019.

#	Field Name	Type
1	Arrival (Week Day /Hour)	Categorical
2	Gender	Categorical
3	Origin	Categorical
4	Arrival Modlaity	Categorical
5	Accompaniers	Categorical
6	Family Status	Categorical
7	Waiting Modality	Categorical
8	Reason for Encounter	Categorical
9	Circumstances	Categorical
10	Age	Numeric
11	Oxygen Flow	Numeric
12	Heart Rate	Numeric
13	Respiration Rate	Numeric
14	Systolic Blood Pressure	Numeric
15	Diastolic Blood Pressure	Numeric
16	Pain Scale	Numeric
17	Temperature	Numeric
18	Oxygen Saturation	Numeric
19	Capillary Blood Glucose	Numeric
20	Capillary Blood Hemoglobin	Numeric
21	Bladder volume	Numeric
22	Capillary Blood Ketones	Numeric
23	Breath Test of Alcohol	Numeric
24	Nurse Triage Scale	Numeric
25	Nurse Notes	Text
26	Psychiatric History	Text
27	Surgical History	Text
28	Medical History	Text

Data Description (cont'd)

Specialty / Label	Hospitalization %
Surgery / CHIR	19.7%
Short-Term Hospitalization Unit / UHCD	42.4%
Medical Specialty / MED	33%
Other	4.9%

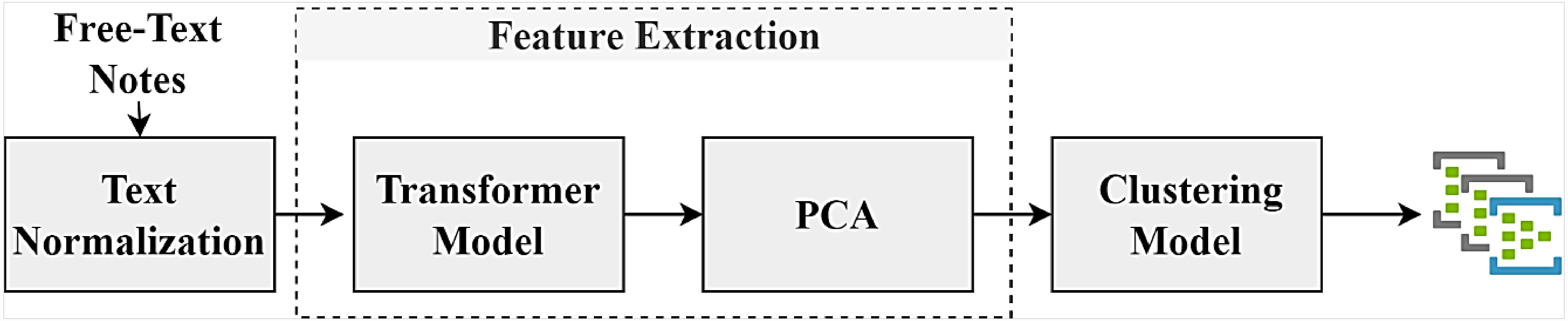
Our Earlier Work

- Early prediction of hospitalization[1]
- Prediction of medical specialties for patients hospitalized [2].

[1] Arnaud, E., Elbattah, M., Gignon, M., & Dequen, G. (2020). Deep learning to predict hospitalization at triage: Integration of structured data and unstructured text. *In Proceedings of the IEEE International Conference on Big Data*. IEEE.

[2] Arnaud, E., Elbattah, M., Gignon, G & Dequen, G. (2021). NLP-based prediction of medical specialties at hospital admission using triage notes. In Proceedings of IEEE International Conference on Healthcare Informatics (ICHI).

Approach Overview



UNIVERSITÉ de Picardie Jules Verne

Transformer Models

- CamemBERT (Martin et al. 2019)
- FlauBERT (Le et al. 2019)
- mBART (Liu et al. 2020)
- All models were accessed through the HuggingFace repository.

UNIVERSITÉ de Picardie Jules Verne

Feature Extraction Experiments

Model	Params	Embedding Dimension	Runtime
CamemBERT	110M	768	31 min
FlauBERT	137 M	768	32 min
MBART	610M	1024	64 min

Full Transfer-Learning was applied for the feature extraction process.
Single Nvidia V-100 GPU was used.

Clustering Experiments

Parameter	Value
Number of Clusters (K)	2–10
Centroid Initialisation	k-means++
Similarity Metric	Euclidian Distance
Number of Iterations	200

UNIVERSITÉ de Picardie Jules Verne

Evaluation of Clusters

- Silhouette Score:

$$S(i) = \frac{(b(i) - a(i))}{\max\,(a(i), b(i))}$$

Where a(i) is the average distance of point i to all other points in containing cluster, and b(i) is the smallest average distance of point i to points in another cluster.

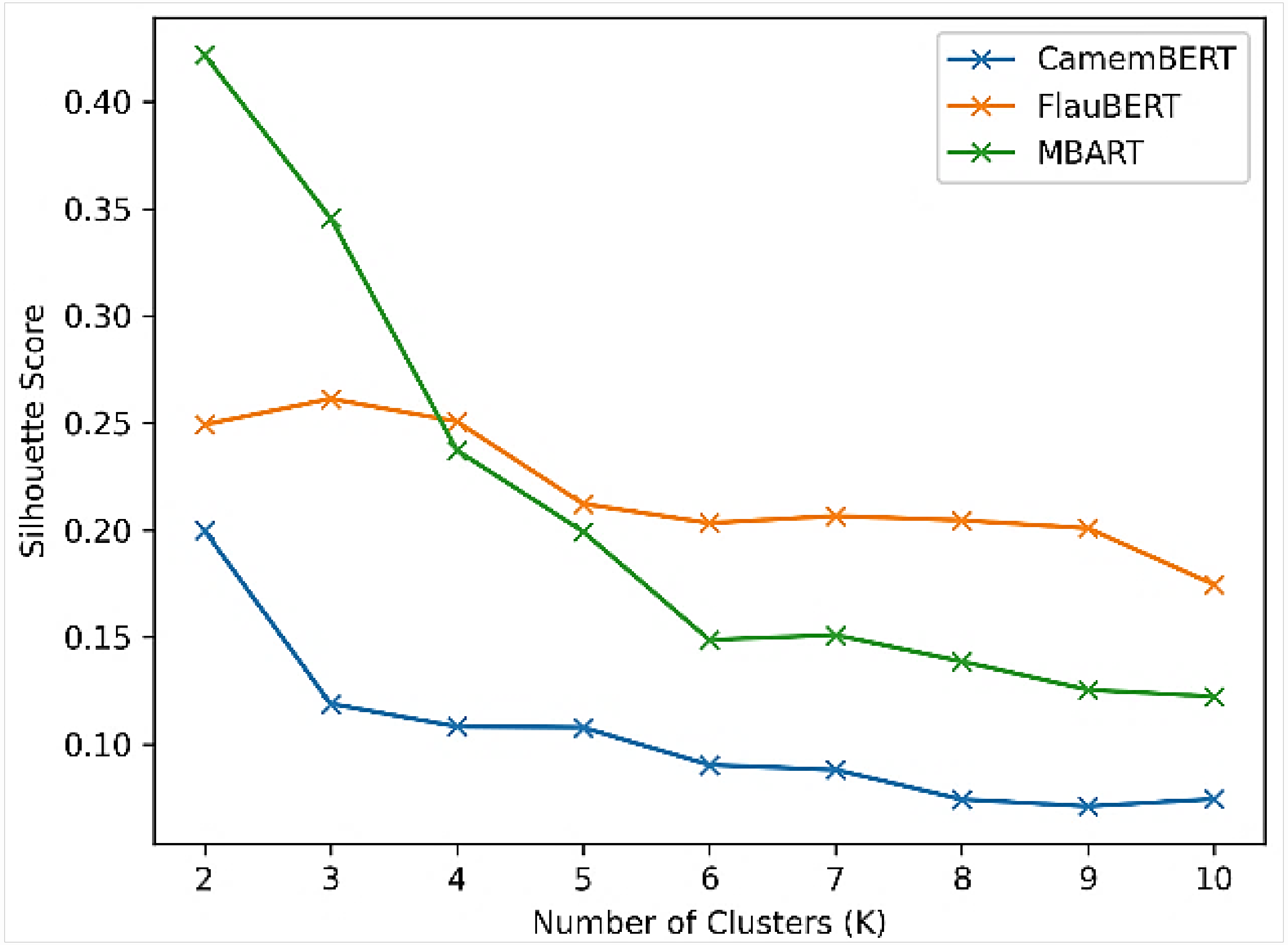
- Fowlkes-Mallows Score:

$$\text{FM SCORE} = \frac{\text{TP}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})}}$$

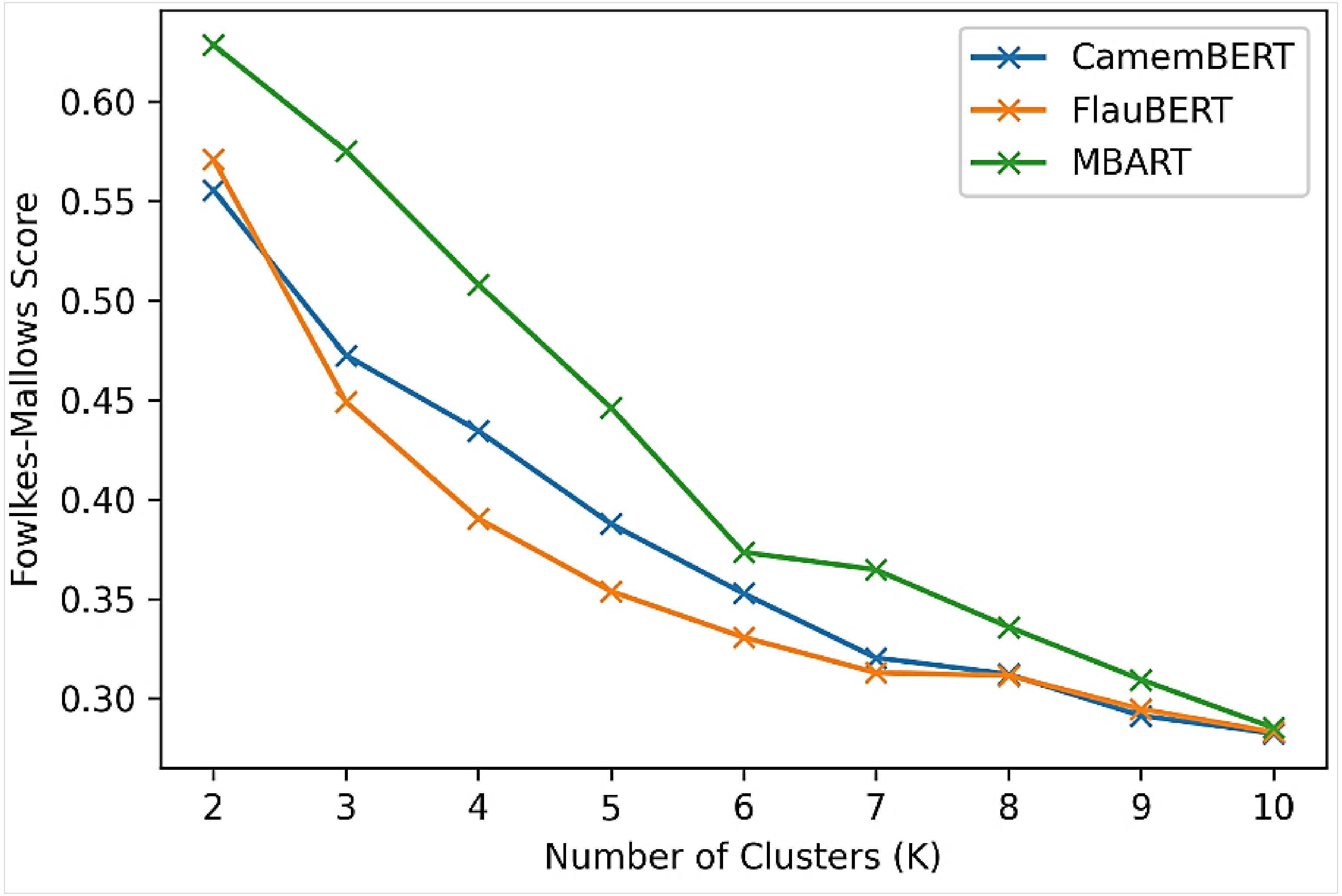
Where TP is the number of True Positives, FP is the number of False Positives, and FN is the number of False Negatives.

UNIVERSITÉ de Picardie Jules Verne

Results: Silhouette Score



Results: Fowlkes-Mallows Score



Conclusions

- BERT-based contextual embeddings could produce clusters of good coherence in general.

- Our experiments could largely validate the suitability of Transfer Leaning in this context.

- Pretrained transformers can serve as an effective mechanism for learning embeddings from free-text notes, ubiquitously in the healthcare environment.

UNIVERSITÉ de Picardie Jules Verne

Thank You!

mahmoud.elbattah@u-picardie.fr