

# Leveraging BERT for Extractive Text Summarization on Lectures

Tirer parti de Bart pour la synthèse extractive sur des cours

June 2019, Derek Miller, Georgia Institute of technology

# Objectif

synthèse automatique -> abstractive

-> extractive : action de ranger les phrases selon une certaine métrique avant de sélectionner les n premières

Tirer rapidement les informations les plus importantes d'un cours

# Données

Transcription de cours depuis Udacity



# Motivations

précédentes méthodes = “myopes”

pas encore pour cette tâche de méthode utilisant Deep Learning

pouvoir choisir la taille du résumé



# Pipeline

- 1) tokenizer le paragraphe d'entrée en phrases nettoyées
- 2) passer les phrases dans Bert pour obtenir un embedding
- 3) utiliser cet embedding dans l'algo de clustering k-means
- 4) garder les phrases les plus proches des centroids des clusters



# Choix de l'embedding pour K-Means

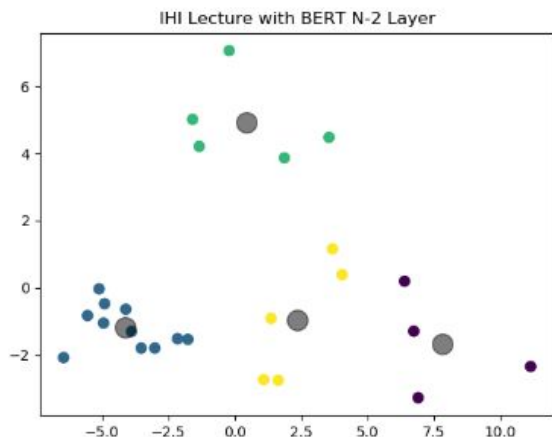


Figure 1 Introduction to Health Informatics lecture with BERT N-2 layer embeddings

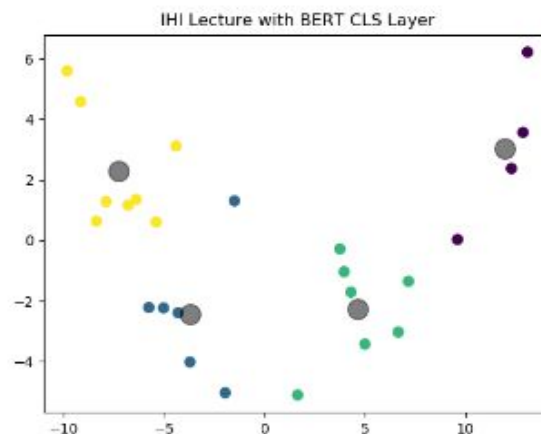


Figure 2 Introduction to Health Informatics lecture BERT [cls] layer embeddings.

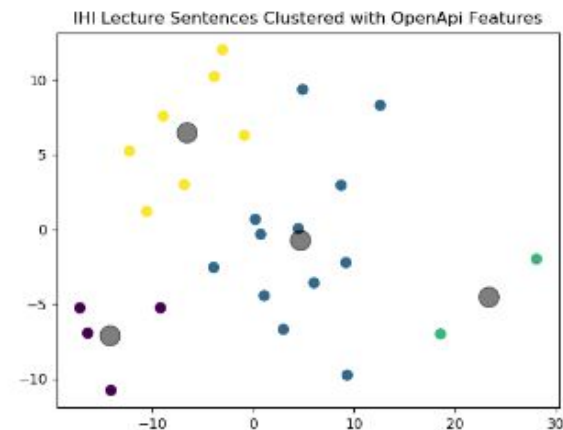


Figure 3 IHI GPT-2 embeddings

# Conclusion

- meilleure qualité de résumé, phrases plus variées
- amélioration possible : remplacer le contexte



Merci !

