DCU Team at the NTCIR-16 RCIR Task

Manh-Duy Nguyen, Thao-Nhu Nguyen, Binh Thanh Nguyen, Annalina Caputo, and Cathal Gurrin

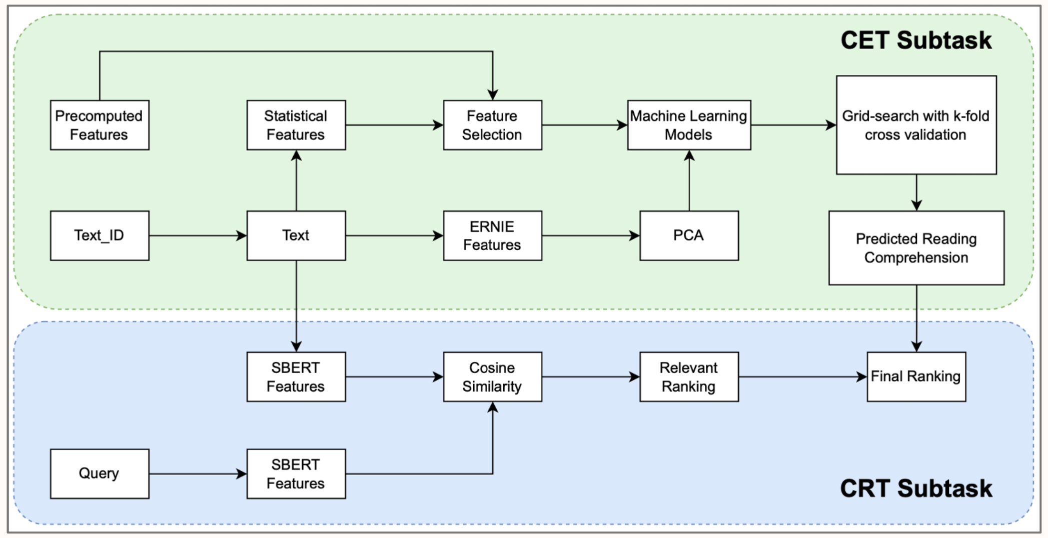
INTRODUCTION

- Reading plays a critical part in our daily lives.
- People tend to have different eye movements when they read texts with difficult concepts [1]
- Comprehension-evaluation task (**CET**): predict the level of comprehension of the reader based on their gaze behaviour
- Comprehension-based retrieval task (**CRT**): find a text relevant to a given topic using the comprehension level

DATASET

- The precomputed features from eye-tracking signals of 9 volunteers.
- Each volunteer read 24 pieces of text for each reading condition: reading, skimming, scanning, proofreading.
- 96 texts for 1 volunteer: 72 texts for train, 24 texts for test.
- In total: 648 training samples, 216 testing samples
- Each text has a topic: transportation, art, …
- Reading comprehension score: 0, 1, 2, 3
- 306 features in the dataset: text identifier, number of words, topic, reading time, 302 pre-calculated features of gaze behaviour.

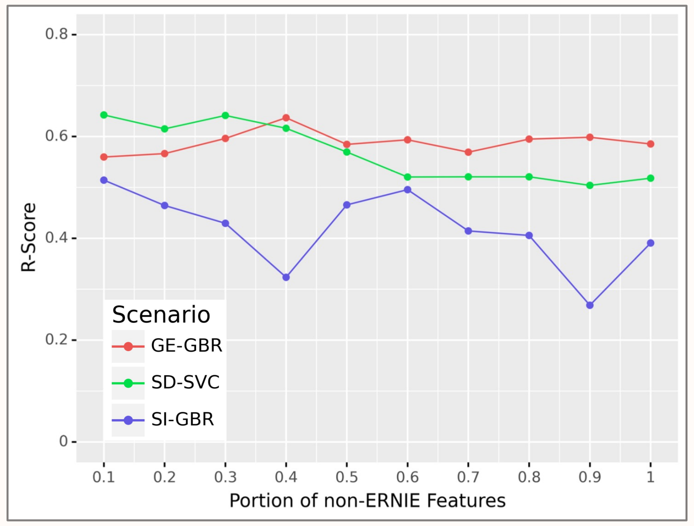
METHODOLOGY



Our Pipeline for CET and CRT

- 36 statistical features extracted from texts: number of nouns, verbs, adjectives, or entities, etc
- ERNIE [2] to encode a text into a 768-dimensional feature
- **NOT** use topic or the text identifier feature.
- 1108 features in total (304 + 36 + 768)
- Non-ERNIE Feature Selection: based on features importance
- PCA to reduce dimension for ERNIE features
- ML Models: conventional models for both regression and classification problem (Linear Regression, Random Forest, Gradient Boosting Tree, AdaBoost, Support Vector Machine)
- Classification: $p = \sum_{c=0}^{3} p_c * c = \sum_{c=1}^{3} p_c * c$
- Scenarios: **S**ubject-**I**ndependent, **S**ubject-**D**ependent, **G**eneral
- SBERT [3] to Encode text and topic into the same vector space
- Relevance: $sim(t, q) = (P(t) + 1) * R(t, q)$

CET Result



- Feature selection was better than using all features
- GBR worked best in SI, GE
- Used SVC in SD because of a few training samples
- ERNIE features not useful

- Model trained in SD gave best result on test set
- Combined models (0 + 4 + 1) trained in different scenarios gave similar result with model trained in SD.

RUN_ID	Scenario	#Features	#PCA	R-Score
0	SI-GBR	0.5	N/A	0.4038
8	SI-GBR	0.6	N/A	0.3389
3	SD-SVC	0.1	N/A	0.5119
4	SD-SVC	0.3	N/A	0.5992
2	SD-SVC	0.5	N/A	0.5600
5	GE-GBR	0.4	N/A	0.5165
1	GE-GBR	0.5	N/A	0.5529
6	GE-GBR	0.5	150	0.5232
7	Combine	-	-	0.6000

#PCA = *N/A* indicates not using ERNIE features

CRT Result

Keywords Type	Keywords
T1	animals with their life, habit, abilities, benefit and endangerment.
T2	animals and animals habit and endangerment.
T3	animals, animals habit.
T4	animals, elephants, wild, zoo.

- Tried different SBERT model structures with different types of inputs: from detail (T1) to general (T3), and most common words in a topic (T4).

RUN_ID	SBERT Type	top-*m*	Keywords Type	nDCG
1†	Fast-Mini	4	T1	0.5856
0	Fast-Mini	4	T1	0.6929
2	Fast-Mini	5	T1	0.7178
3	Fast-Mini	6	T1	0.7245
4	Fast-Mini	7	T1	0.7215
5	Fast-Mini	8	T1	0.7215
6	Mini	6	T1	0.7153
8	Base	6	T1	0.7149
9	Fast-Mini	6	T2	0.7271
10	Fast-Mini	6	T3	**0.7295**
11	Fast-Mini	6	T4	0.7164

† indicates using only SBERT similarity score

- Not using CET prediction still produced average result (RID = 1).
- General keywords (T3) obtained highest score.
- SBERT model with simplest version worked better than others.

References
[1] Johanna K Kaakinen and Jukka Hyona 2007. Perspective effects in repeated reading: An eye movement study. Memory & cognition 35, 6 (2007), 1323–1336.
[2] Yu Sun et. al. 2020. Ernie 2.0: A continual pre-training framework for language under- standing. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. 8968–8975.
[3] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019).