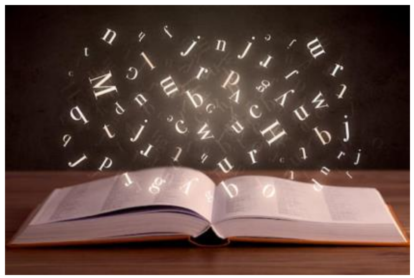
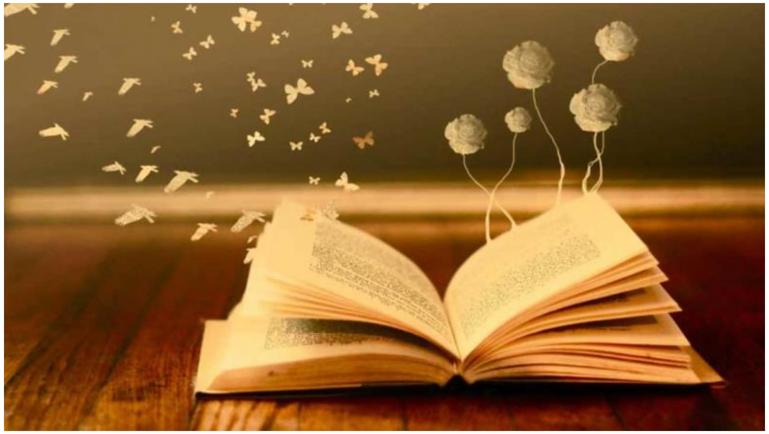


HCMUS at the NTCIR-16 RCIR Task

Kim-Nghia Liu, Vinh Dang, Thanh-Son Nguyen, Minh-Triet Tran
Faculty of Information Technology
University of Science
Vietnam National University Ho Chi Minh City

Content

- Introduction
- Related Work
- Methods
- Experiments
- Conclusion

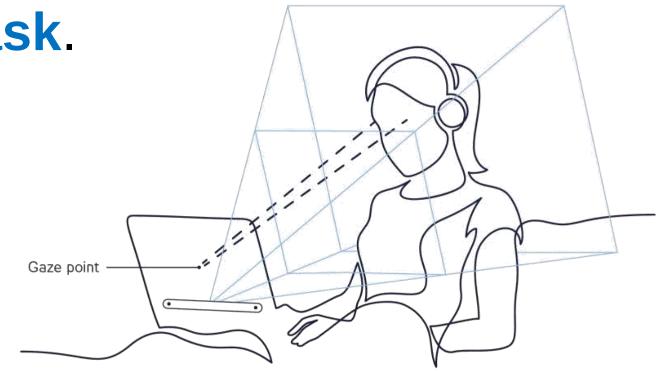


Introduction

- Reading comprehension of information in text form:
  - A new challenging and interesting field of research
  - Exciting problems related to reading comprehension, such as sorting texts based on their comprehension levels, ranking texts in various topics by integrating text comprehension-evidence into the IR process, etc.
- **Reading Comprehension in Information Retrieval** (RCIR [12]) in the NTCIR-16
  - focus on personalized retrieval techniques that can take advantage of useful information from eye-tracking to ranking text content
  - data captured from multi-modal sensors to monitor experimental participants in different types of reading behaviors, reading conditions: **sequential reading, skimming, scanning** or **proof-reading**.

Introduction

- **RCIR tasks:** to propose and develop solutions to utilize multi-modal signals (e.g. eye tracking, screenshots, etc) in the retrieval process in three sub-tasks:
  - **Comprehension-evaluation sub-task (CET)** aims to sort texts based on comprehension levels
  - **Comprehension-based Retrieval sub-task (CRT)** aims to rank texts by integrating text comprehension-evidence into the IR process.
  - The improvisation ideas to explore the RCIR dataset are encouraged for the **Insights (IT) sub-task**.

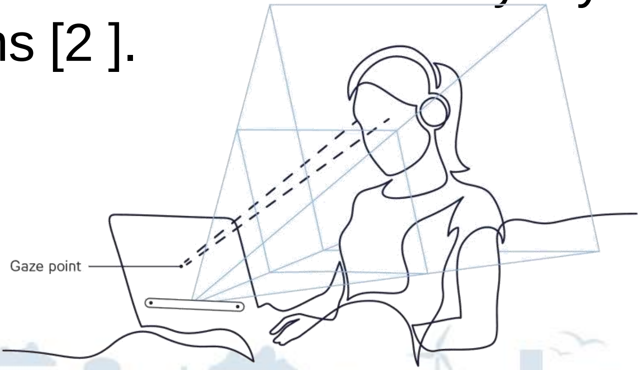


Our team

- **HCMUS team:** define the **common pipeline and strategy** for proposed solutions.
  - apply **data pre-processing techniques** to normalize the values of the attributes, or use PCA to reduce the dimensionality of data, as well as select meaningful attributes for information representation.
  - propose some **hand-crafted features** or use **BERT**[5] to encode information of text document in English texts, and propose several representations for the feature vectors.
  - use different machine learning techniques to **compute the final results**, namely Multilayer perceptron (MLP), Random Forest[4], AutoML[8]

Related Work

- Eye movements provide clues to analyze human behaviors and perception in reading, scene perception, and visual search [16].
- Rayner presents a study on 20 years of research on eye movements in reading and information processing [15].
- Noticeable difference in eye movement between skillful and novice readers➔ estimate the language skills of a reader by analyzing eye movement while reading English documents [20].
- Eye gaze ➔ predict a reader's understanding of the content of a document [8].
- A reader's understanding can be determined more accurately by using eye gaze than by answering questions [2].



Common Pipeline

- The first phase is to normalize data, reduce dimensionality, and select meaningful attributes for the data.
- The second phase is to propose different feature representations for data.
- The last phase is to predict the final result with an appropriate ML-based model for a given input feature vector.

Common Strategy as Guidance for Our Proposed Methods

- **Pre-processing:**
  - Normalize the field values to the range [−1, 1] or [0, 1]
  - Apply PCA[13] to reduce the dimensionality of data.
- **Feature Representation:**
  - combine both eye-tracking data (after the pre-processing phase) and text data, including text content and question-answer content.
  - use BERT[5] as a common utility for text content representation.
  - we also consider other potential attributes from text data, such as the number of words that are not common English words, or the total length of all questions (in character level), etc.

→ extra hand-crafted features may provide more valuable information for our solution in the CET sub-task.

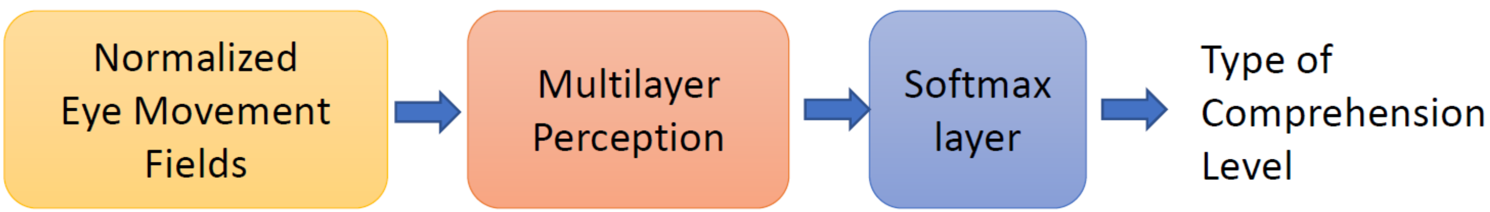
Common Strategy as Guidance for Our Proposed Methods

- **ML-based Evaluation:**
  - use different ML models, such as Random Forests[4], Multilayer perceptron (MLP), and AutoML[8].
  - employ different techniques for training and optimizing ML models, including Adam [14], SGD [17], Adagrad [7], etc., and several activation layers, such as ReLU [1], Leaky ReLU, Sigmoid, etc.
  - AutoML approaches seem to provide prominent results.

Method 1 - Using only Eye Movement Data and Multilayer Perceptron

Goal: to evaluate if we can estimate the level of reading comprehension only based on eye-tracking data.

- ❖ do not use text data in the text content and questions/answers
- ❖ exploit **only data captured from sensors** related to experimental participants' activities in reading, especially the eyes information of participants.



Classification with Multilayer perceptron

	layer_type(input_dim, output_dim)
layer1	Linear(276, 1024)
layer2	BatchNorm(1024, 1024)
layer3	Sigmoid(1024, 1024)
block4 x 3	Linear(1024, 1024)
	BatchNorm(1024, 1024)
	Sigmoid(1024, 1024)
layer5	Linear(1024, 4)
layer6	Softmax(4,4)

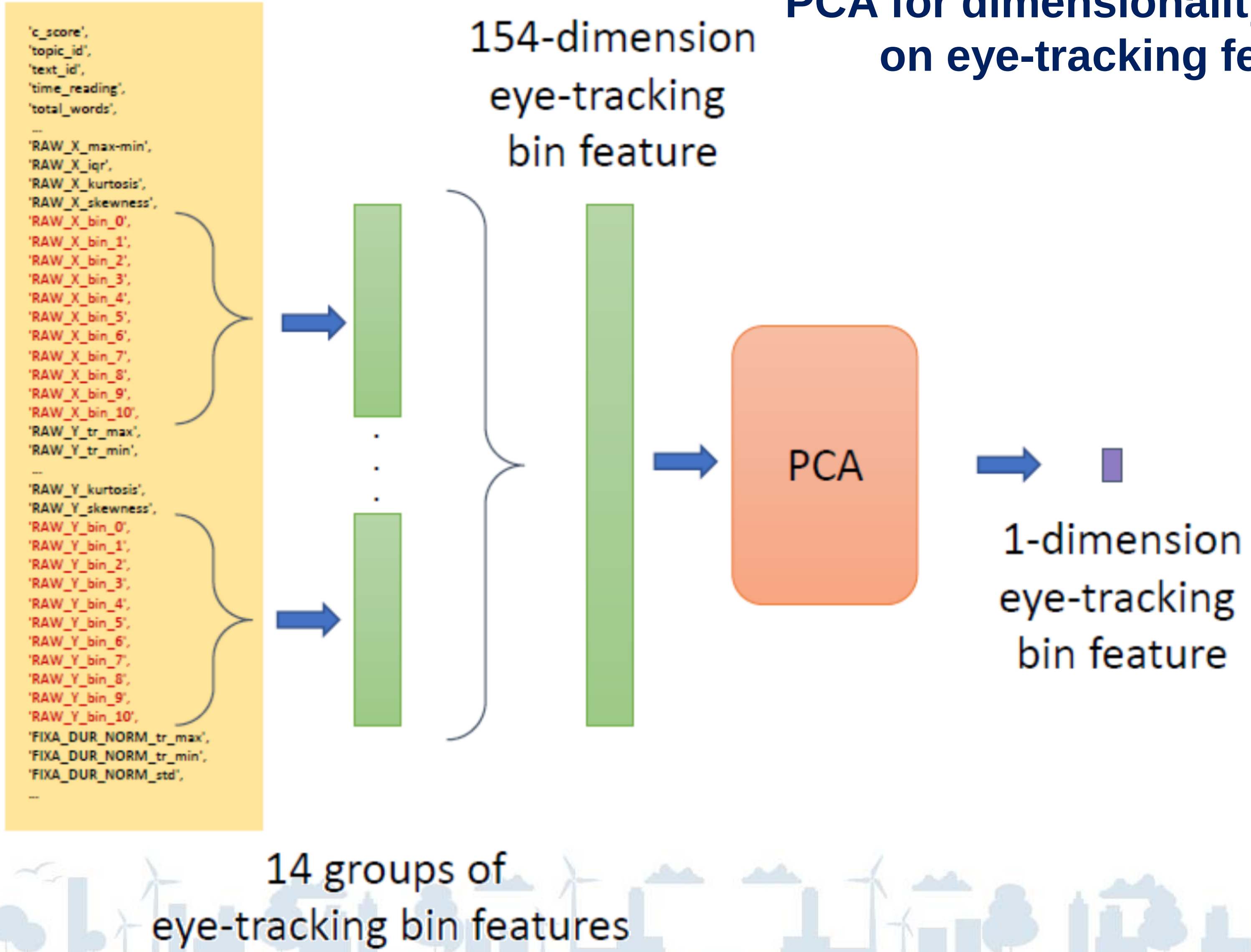
Method 2 - Combine Both Eye Movement and Text Data without Deep Feature

- Utilize **both eye movement and text data**, including the text content as well as questions and answers related to that text.
- Only use **traditional techniques**, **not deep learning methods**, to represent data feature and to predict output result.

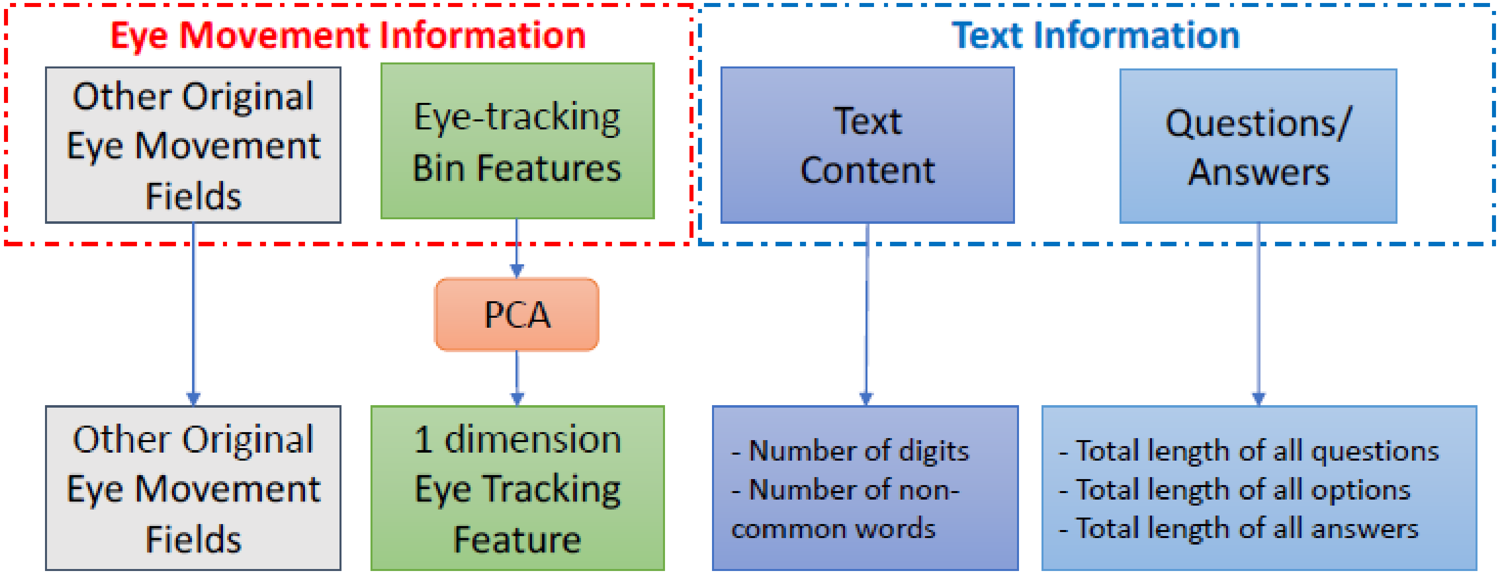
- **Proposed hand-crafted features for text encoding:**
  - The number of words that are actual digit numbers.
  - The number of words that are not common English words
  - The total length of all questions, in character level.
  - The total length of all options, for all questions, in character level.
  - The total length of all answers for all questions, in character level.

Method 2 - Combine Both Eye Movement and Text Data without Deep Feature

PCA for dimensionality reduction on eye-tracking features



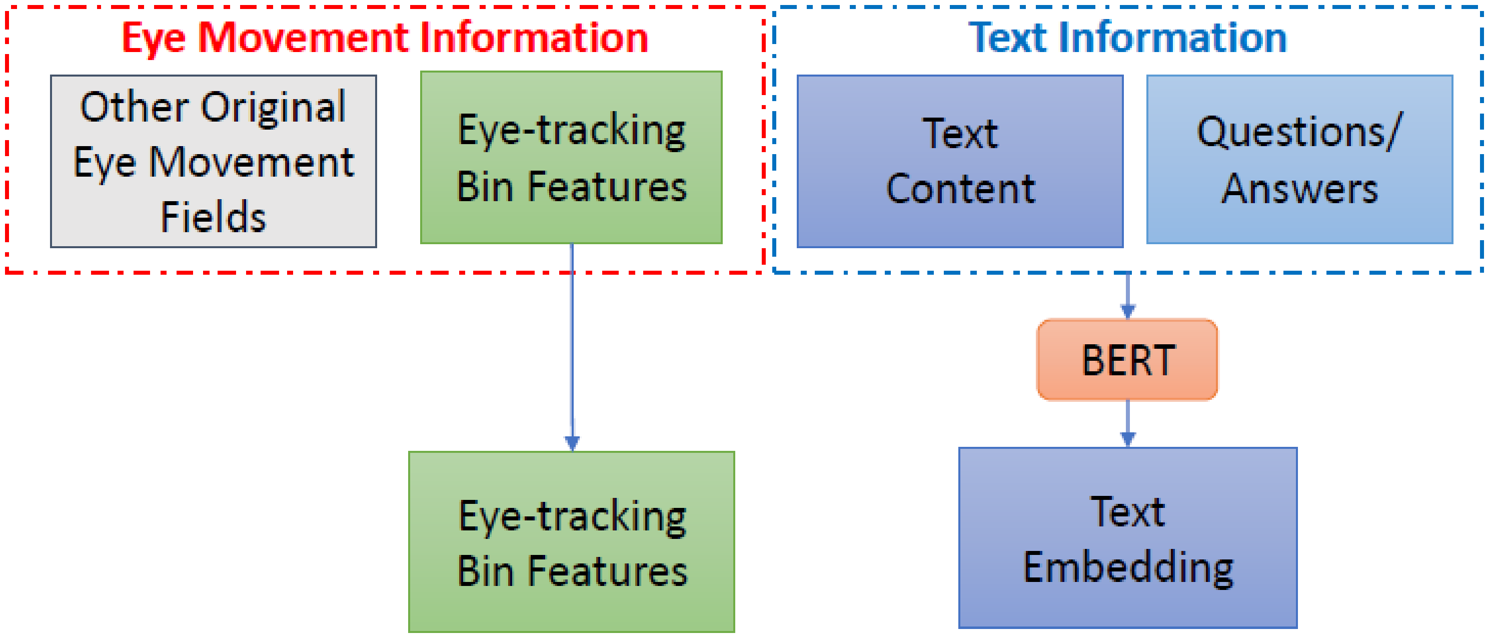
Method 2 - Combine Both Eye Movement and Text Data without Deep Feature



Feature representation in our second method

Method 3 - Deep Embedding and AutoML

- Use BERT to encode text content and combine this feature with the eye movement feature to form the feature vector.
- Employ AutoML[8] to search for the model configuration that performs the best accuracy on our data.



Feature representation in our third method

Official Experimental Results

- Using only eye movement information (as in Method 1) cannot provide the results as good as combining both eye movement and text information (as in Methods 2 and 3). When we utilize text information, the results can be boosted significantly from **0.4024** to **0.49182** and **0.50846**.
- In Method 2, we only extract some extra hand-crafted attributes from text information, and the result of Method 2 is slightly lower than that of Method 3, which employs BERT to encode the whole text into feature vectors.

Method	Result
Method 1	0.40242
Method 2	0.49182
Method 3	0.50846

Results on official test set
(Spearmans correlation coefficient)

Conclusion

- We propose three methods to solve the Comprehension evaluation Task (CET) in the **Reading Comprehension in Information Retrieval** (RCIR) challenge in **NTCIR-16**.
- Our best solution achieves the Spearmans correlation coefficient of **0.50846** in the official test set of the CET sub-task in RCIR challenge 2022.
- We aim to study different techniques further to boost the results for text comprehension evaluation by taking advantage of helpful information from eye-tracking systems.

Thank you for your attention!