KNUIR

KNUIR at the NTCIR-16 RCIR: Predicting Comprehension Level using Regression Models Based on Eye-Tracking Metadata

Kyungpook National University

Presented by Yumi Kim

Contents

0 KNUIR Team

Prof. Heesop, Kim

Members of KNUIR Team

Address:
Room 424-2
Department of Library & Information Science
KYUNGPOOK NATIONAL UNIVERSITY
80 Daehak-ro, Buk-gu, Daegu, Korea

0 Special Thanks to

Prof. Graham Healy and the RCIR organizers

Graham Healy
graham.healy@dcu.ie

NTCIR16-RCIR Task

HOME NEWS FAQ DATA/TASK RESOURCES

What is RCIR?

RCIR (Reading Comprehension in Information Retrieval) is an NTCIR-16 workshop data challenge task. The NTCIR-16 RCIR pilot task aims to motivate the development of a first generation of personalised retrieval techniques that integrate reading comprehension measures from eye tracking as a source of evidence when ranking text content. Participating researchers in the task will develop and benchmark approaches to integrate multi-modal signals (e.g. eye tracking, screenshots, etc) into the retrieval process for two sub-tasks: a) a comprehension-evaluation task (CET) that aims to sort texts in terms of comprehension levels, and b) a comprehension-based retrieval task (CRT) that aims to rank texts (for a variety of topics) by integrating text comprehension-evidence into the IR process. Both of these sub-tasks are exploratory in nature, but are designed to facilitate initial experimentation on the topic by the community.

Important Dates

31 Oct 2021 - Dataset release
3 Jan 2022 - Task registration due
14 Mar 2022 - Run submissions due
25 Mar 2022 - Participant paper due
1 May 2022 - Camera ready paper due
14-17 June 2022 - NTCIR-16 in NII, Tokyo

I INTRODUCTION

I-1 RCIR CET sub-task

- Participate in the NTCIR-16 reading comprehension information retrieval (RCIR) CET (Comprehension-evaluation Task) sub-task

I-2 Compare five regression model

-**Five regression models**:
Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, eXtreme Gradient Boosting (XGB) Regressor, and Voting Regressor.

I-3 Contribution

- Findings in our work may bring insight into the understanding of human reading and information seeking processes with the help of eye-tracking systems by applying various regression models.

II METHODS

- Overview

1

Implementation

- Python
- Scikit-learn packages
- Pandas
- Numpy and so on.

2

Five regression model

- Linear Regression
- Random Forest Regressor
- Gradient Boosting Regressor
- XGB Regressor
- Voting Regressor

3

Dataset

- Obtained in the NTCIR-16 RCIR Task

Compare regression models

Predicting Comprehension Level using five Regression Models and comparing the results

II METHODS

- Implementation

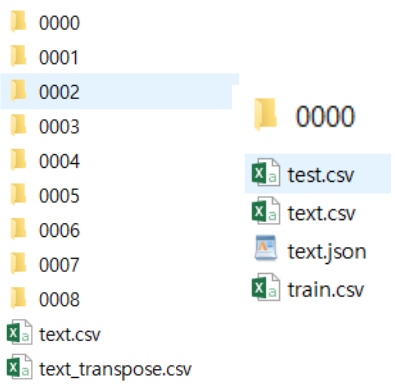
Building a regression model to **predict Comprehension Level**

- Python and Scikit-learn packages are mainly used for implementation of our CET prediction model.
- Scikit-learn is an open source python module integrating a wide range of machine learning algorithms.
- Pandas is a python library of rich data structures and tools for analysis and manipulation in statistics, finance, social sciences, and many other fields.
- NumPy is a python library for scientific computing.

II METHODS

- Dataset

❑ Structure of dataset



❑ Sample of dataset

Train.csv

	c_score	topic_id	text_id	time_reading	total_words	NUM_FIXA	RATE_FIXA	NUM_SACC	RATE_SACC	NUM_BLINK	...	RP_SIZE_bin_1	RP_SIZE_bin_2	RF
0	0	11	13217	45.228448	308	220.0	0.714286	131.0	0.425325	8.0	...	3.153885	3.412500	
1	1	41	8366	60.134963	494	293.0	0.593117	137.0	0.277328	15.0	...	2.860784	3.154468	
2	3	11	17628	60.135543	318	283.0	0.889937	158.0	0.496855	8.0	...	2.829330	3.114190	
3	0	16	7658	60.144139	315	278.0	0.882540	135.0	0.428571	12.0	...	3.430810	3.700230	
4	0	19	18047	34.546934	356	173.0	0.485955	66.0	0.185393	4.0	...	3.300285	3.594950	
5	0	11	3256	39.952412	242	180.0	0.743802	87.0	0.359504	11.0	...	2.435892	2.783734	
6	0	19	7993	60.161009	357	301.0	0.843137	153.0	0.428571	11.0	...	2.096030	2.462830	
7	0	41	10040	33.071858	446	161.0	0.360987	69.0	0.154709	7.0	...	3.271058	3.540126	
8	3	41	8477	56.362339	293	258.0	0.880546	147.0	0.501706	14.0	...	2.642291	2.958962	
9	0	16	16326	57.487725	435	264.0	0.606897	118.0	0.271264	14.0	...	2.694548	2.993596	

10 rows × 307 columns

Test.csv

	text_id	time_reading	total_words	NUM_FIXA	RATE_FIXA	NUM_SACC	RATE_SACC	NUM_BLINK	RATE_BLINK	RATE_X_BWD	...	RP_SIZE_bin_1	RP_SIZE
0	5181	37.032498	302	167.0	0.552980	94.0	0.311258	8.0	0.026490	0.343373	...	2.775414	3.
1	7394	59.151110	384	271.0	0.705729	143.0	0.372396	13.0	0.033854	0.274074	...	2.338765	2.
2	18017	52.926883	317	238.0	0.750789	133.0	0.419558	10.0	0.031546	0.253165	...	2.941839	3.
3	3418	57.368722	314	269.0	0.856688	115.0	0.366242	12.0	0.038217	0.298507	...	2.729999	3.
4	4564	32.588329	433	167.0	0.385681	58.0	0.133949	5.0	0.011547	0.475904	...	3.006721	3.
5	1631	49.295799	327	230.0	0.703364	137.0	0.418960	9.0	0.027523	0.423581	...	3.375280	3.
6	2360	55.408437	357	256.0	0.717087	129.0	0.361345	14.0	0.039216	0.301961	...	2.094938	2.
7	16637	47.562668	408	221.0	0.541667	92.0	0.225490	12.0	0.029412	0.286364	...	2.834623	3.
8	6099	25.711480	301	119.0	0.395349	64.0	0.212625	4.0	0.013289	0.449153	...	3.332863	3.
9	15656	60.144083	385	280.0	0.727273	127.0	0.329870	20.0	0.051948	0.311828	...	2.589615	2.

10 rows × 305 columns

- The dataset is structured into 9 directories (from 0000 to 0008).
- The reading data (5 features)
  : comprehension score, topic ids , text ids, the duration of reading, and the number of total words
- The pre-computed 302 features from the volunteer's eye-tracking data.

II METHODS

- Comparison of machine learning regression algorithms

❑ **We adopted five regression models in order to predict target results.**

- ❑ Linear Regression
- ❑ Random Forest Regressor
- ❑ Gradient Boosting Regressor
- ❑ XGB Regressor
- ❑ Voting Regressor

- The Gradient Boosting Regressor model, can handle non-linear correlation between input data and target result as well as correlation between features.
- Random Forest is an ensemble learning algorithms that constructs many decision trees during the training. It predicts the mode of the classes for classification tasks and mean prediction of trees for regression tasks.
- Xgboost (eXtreme Gradient Boosting) is a library that provides machine learning algorithms under the a gradient boosting framework. The reading data consists of comprehension score, topic and text ids, the duration of reading, and the number of total words.
- Voting Regressor combines the performances of the other four models to make predictions.

III RESULTS

- Each regression model of on the validation data

Table 2: Spearman's rank correlation coefficient (Spearman's ρ) and p-value of each model.

	Spearman's ρ (p-value) of Volunteer								
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Gradient Boost Regressor	0.71 (< 0.0002)	0.63 (< 0.0018)	0.27 (< 0.2330)	0.59 (< 0.0036)	0.22 (< 0.3310)	0.47 (< 0.0264)	0.15 (< 0.5144)	0.38 (< 0.0839)	0.86 (< 2.1227)
Random Forest Regressor	0.70 (< 0.0002)	0.76 (< 4.680)	0.22 (< 0.0334)	0.61 (< 0.0023)	0.12 (< 0.6082)	0.36 (< 0.0967)	0.06 (< 0.7980)	0.54 (< 0.0096)	0.77 (< 3.2613)
Linear Regression	0.27 (< 0.2079)	0.31 (< 0.1537)	0.27 (< 0.2188)	0.12 (< 0.5941)	0.06 (< 0.7854)	0.46 (< 0.0305)	0.07 (< 0.7681)	0.20 (< 0.3720)	0.41 (< 0.0588)
XGB Regressor	0.69 (< 0.0004)	0.64 (< 0.0012)	0.35 (< 0.1087)	0.59 (< 0.0041)	0.11 (< 0.6400)	0.46 (< 0.0324)	0.04 (< 0.8585)	0.47 (< 0.0276)	0.80 (< 8.4717)
Voting Regressor	0.71 (< 0.0001)	0.70 (< 0.0003)	0.35 (< 0.1114)	0.36 (< 0.1004)	0.11 (< 0.6123)	0.62 (< 0.0020)	0.11 (< 0.6354)	0.37 (< 0.0926)	0.81 (< 4.2279)

- Training data is automatically shuffled and divided into training data and validation data at a ratio of 7:3.
- Each volunteer's Spearman's ρ and p-value for each regression model of on the validation data.
- 1st, 2nd, 4th, and 9th volunteers are generally predicted well by the regression models.

III RESULTS

- Running results from NTCIR-16 RCIR

Table 3: Running Result from NTCIR-16 RCIR *

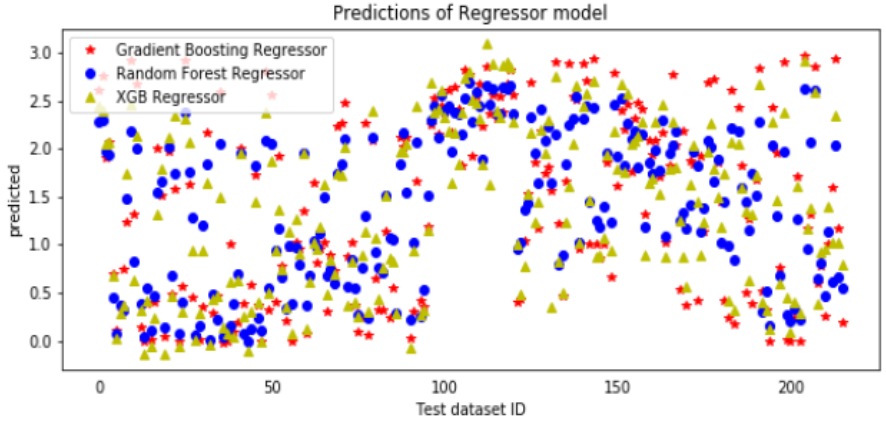
	ρ	p-value
Gradient Boost Regressor	0.53186	< 0.00000
Random Forest Regressor	0.57061	< 0.00000
Linear Regression	0.05021	< 0.46292
Voting Regressor	0.31124	< 0.00000

* The result of XGB Regressor is not included due to late submission.

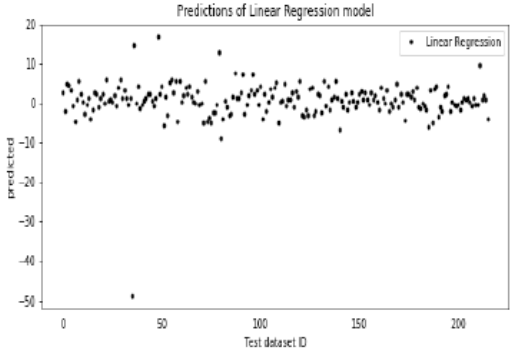
- The actual results of the submitted data from NTCIR-16
- Gradient Boosting and Random Forest Regressor show better performance.

III RESULTS

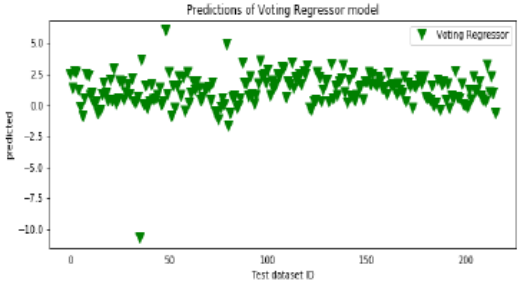
- Scatter plots for prediction with test data



(a) Gradient Boost Regressor, Random Forest Regressor, and XGB Regressor



(b) Linear Regression



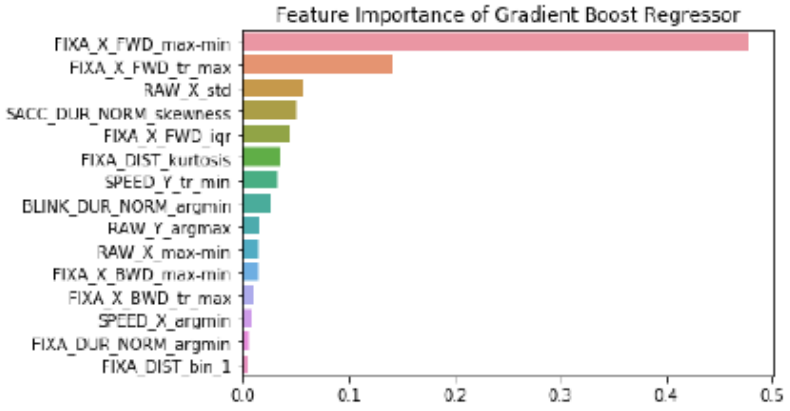
(c) Voting Regressor

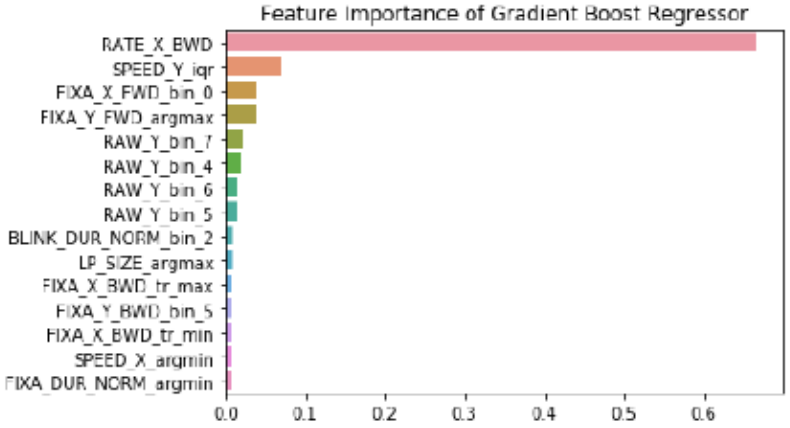
Figure 1: Scatter plots for prediction with test data.

- Scatter plot of the submitted prediction results
- Gradient Boosting Regressor, Random Forest Regressor, and XGB Regressor perform better with the prediction values reside from 0 to 3.
- Linear Regression and Voting Regressor perform worse with some predicted values that are out of the range.

III RESULTS

- Scatter plots for prediction with test data





- RATE_X_BWD
  - feature related to frequency of the eyes moving back and fixating on certain points
- RATE_BLINK
  - number of blinks divided by the total words in the text
- FIXA_X_FWD_tr_max, FIXA_X_FWD_max-min
  - features related to forward and backward movement distances.

IV CONCLUSION

IV-1 Concluding remarks

- We implemented and compared five regression models in order to predict the comprehen sion score based on reading data and eye-tracking metadata.
- Based on our findings, Gradient Boosting and Random Forest Regressor show better performance.
- The Spearman's ρ values of the two models benchmarked on the NTCIR-16 RCIR test set are **0.53** and **0.57**, respectively.
- We found that [RATE_X_BWD, RATE_BLINK, FIXA_X_FWD_tr_max, and FIXA_X_FWD_maxmin] affect the prediction result highly as important features.

IV-2 Future work

- Feature selection methods not only decrease complexity of the model and mitigate overfitting problems, but also select the most important features from the dataset.
- Our team is still working on feature selection for better prediction models.
- By using the low variance features, high correlation features, and univariate feature selection.

Thank you for listening!

NTCIR

감사합니다
Thank you!

- Heesop, Kim : heesop@knu.ac.kr
- Yumi, Kim : yumikim@knu.ac.kr