Engaging Content
Engaging People

DCU at the NTCIR-14 OpenLiveQ-2 Task

Piyush Arora & Gareth J.F. Jones
ADAPT Centre, School of Computing
Dublin City University, Ireland
{Piyush.Arora,Gareth.Jones}@dcu.ie

Date: 13th June 2019

Outline

- Task Overview
- Methodology
- Experiments
- Results
- Analysis
- Findings & Future Work

Task Overview

- **Task:** Rank a list of Japanese language questions matching a user's query

- **Dataset:** Yahoo queries and respective question-answers

- **Goal:** Effectively model information from the user click logs and relevance based metrics

- **Evaluation:**
  - Offline evaluation: metrics such as NDCG, ERR
  - Online evaluation: live Yahoo question answering platform

Snapshot

https://chiebukuro.yahoo.co.jp/search?p=喫煙&flg=3&class=1&ei=UTF-8&fr=common-navi

喫煙 ✕ **Search** official corner **Question and consultation**

+ Condition specification

Smoking 's passive smoking smoking seat smoking office searched

I am wondering if I should be a smoker . I am a university student man. Wh…

My surroundings smoke cigarettes anyway. There is a **smoking** area in the university, but even if my friend is in the **smoking** area and smokes, I hate the smell of cigarettes, and I don't like the sidestream smoke, so I don't get into the **smoking** area, everyone Wait outside until you finish smoking ...

Resolved 2018/02/24 06:16 18Views 262

Ways of life and love, troubles in relationships > Love consultation, troubles in relationships

I do not know the smoker 's feelings at all. If a human being is normal, I nee…

Smoking is a desire that does not require . Because it looks so cool, it looks so cool, so why not start it? As a result, too high money is paid, breath becomes stinking, aerobic exercise ability is also lost, and unnecessary image down is also caused, and smokers are unconditional ...

Resolved 2016/11/14 02 21 21Views 401

Manners, ceremonial occasions > manners > smoking manners

Search target

All (198, 221)

Answering received (752)

Voting Accepted (109)

Solved (197,360 cases)

order of display

Relevancy order

Notice

Wisdom bag search RSS function,

Original Japanese page translated using the Google translation

Challenges

- Queries are typically short and ambiguous in nature and might not capture the user's intention effectively
- For example for Japanese query: "喫煙", English translation: "smoking", can have multiple intentions:
  "dangers of smoking"
  "smoking health effects"
  "mechanism to quit smoking"
- Without understanding the user's intent and focus of the query, it becomes challenging to re-rank the questions
- **Aim:** Model textual based information and click logs based information to re-rank questions effectively

Learning To Rank Problem

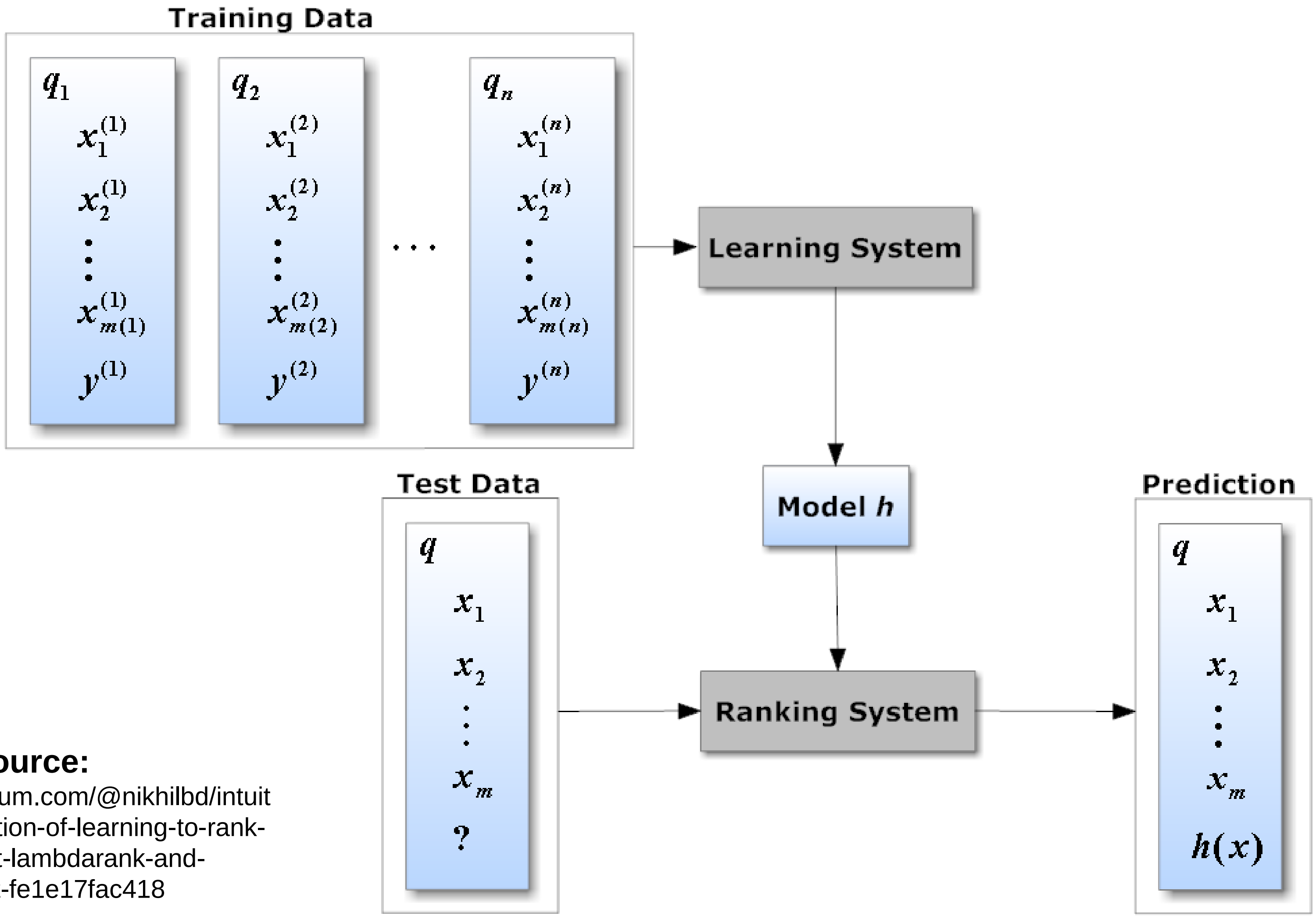


Image Source:
https://medium.com/@nikhilbd/intuitive-explanation-of-learning-to-rank-and-ranknet-lambdarank-and-lambdamart-fe1e17fac418

Resources and Tools

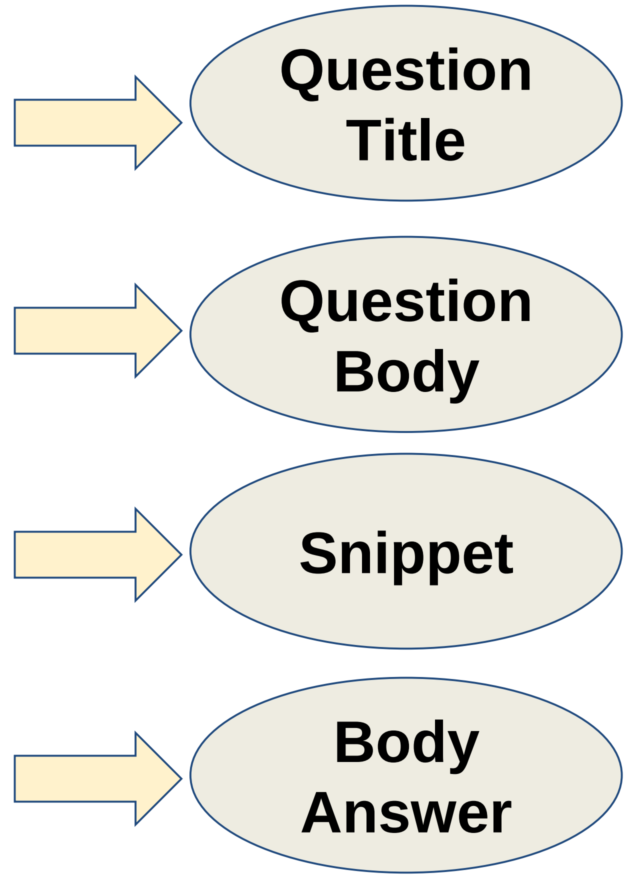
- Resources provided by the task organizers:
  - Pipeline for processing Japanese text
  - Pipeline for features extraction
  - Data set and click logs

- Used Lemur RankLib toolkit

- Total of 77 features

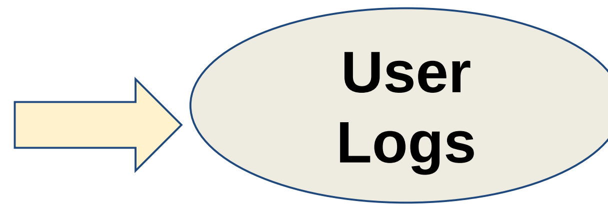
Content based features

Features	Features
tf_sum	tf_in_idf_sum &
log_tf_sum	bm25
norm_tf_sum	log_bm25
log_norm_tf_sum	lm_dir
idf_sum	lm_jm
log_idf_sum	lm_abs
icf_sum	dlen
log_tfidf_sum	log_dlen
tfidf_sum	



Click log based features

Features
answer_num
log_answer_num
view_num
log_view_num
is_open
is_vote
is_solved
rank
updated_at



Methodology

- **Learning to Rank (L2R) algorithms:**
  - Coordinate Ascent
  - MART
- **Feature Selection & Combination:**
  - Alternative combinations of the 5 feature set
- **Parameters optimization**
- **Scores Normalisation:**
  - Z-score normalization
  - Score average
  - Max based normalization

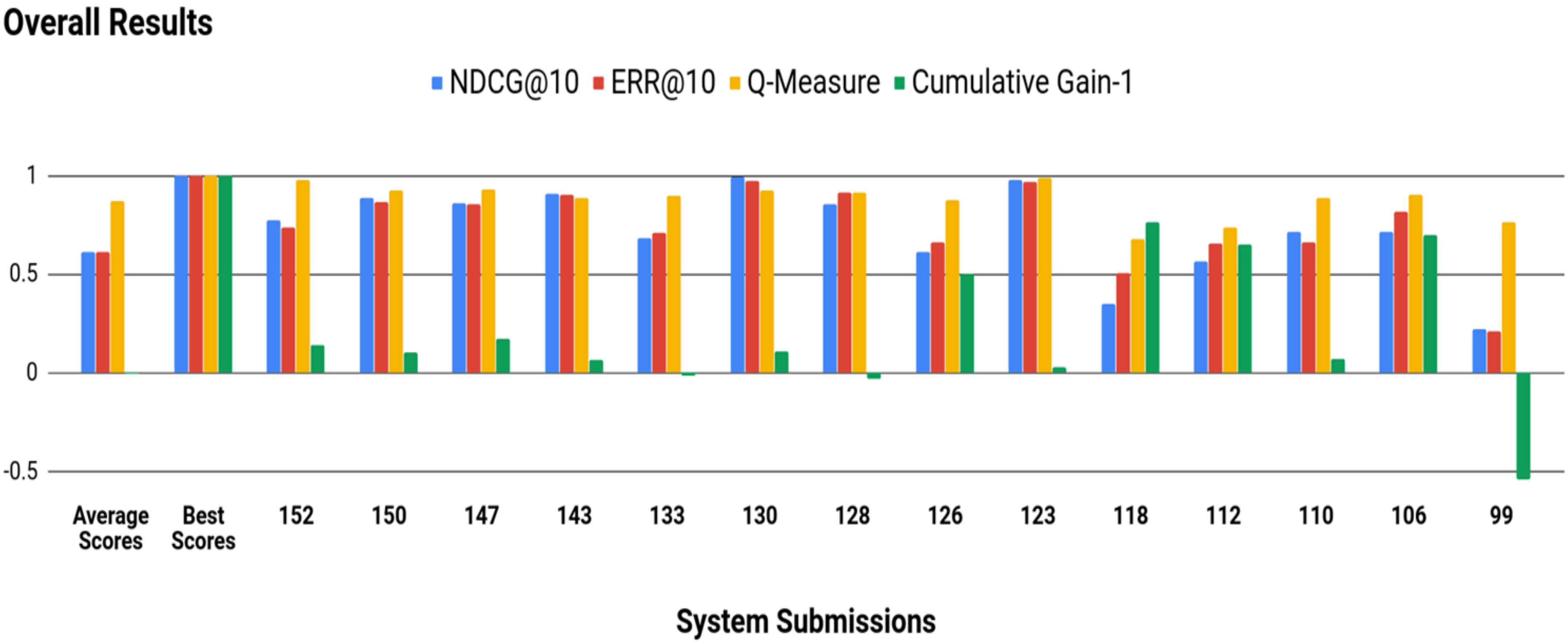
Dataset

	Training set	Test set
Number of Queries	1,000	1,000
Number of Questions	986,125	985,691
Number of click logs	288,502	148,388

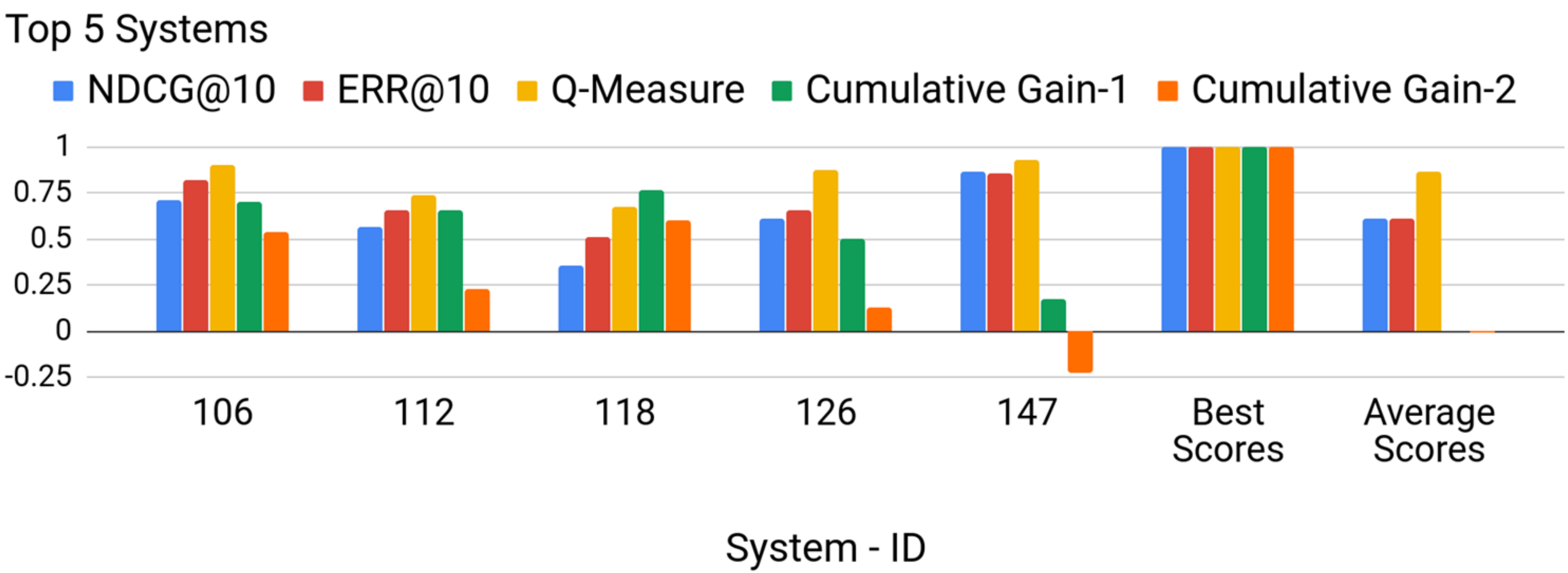
Our Submissions

- Total of 14 systems submitted
- Overall 65 participant submissions
- All 65 submissions evaluated & ranked using
  - NDCG@10, ERR@10, Q measure
  - Phase-1 online evaluation
- Top 30 systems selected for final online evaluation
- 5 of our systems selected in top 30 systems

Results


Overall Results
NDCG@10
ERR@10
Q-Measure
Cumulative Gain-1
1
0.5
0
-0.5
Average Scores
Best Scores
152
150
147
143
133
130
128
126
123
118
112
110
106
99
System Submissions

Best Models


Top 5 Systems
NDCG@10
ERR@10
Q-Measure
Cumulative Gain-1
Cumulative Gain-2
1
0.75
0.5
0.25
0
-0.25
106
112
118
126
147
Best Scores
Average Scores
System - ID

Systems Ranking

Systems	ID	NDCG@10	ERR@10	Q-Measure	Online Evaluation Phase-1	Final Online Evaluation
System-2	106	32	24	26	7	7**
System-4	112	36	35	64	8	10
System-5	**118**	**45**	**38**	**65**	**4**	**6****
System-7	126	34	34	32	14	12
System-12	147	21	23	20	29	23

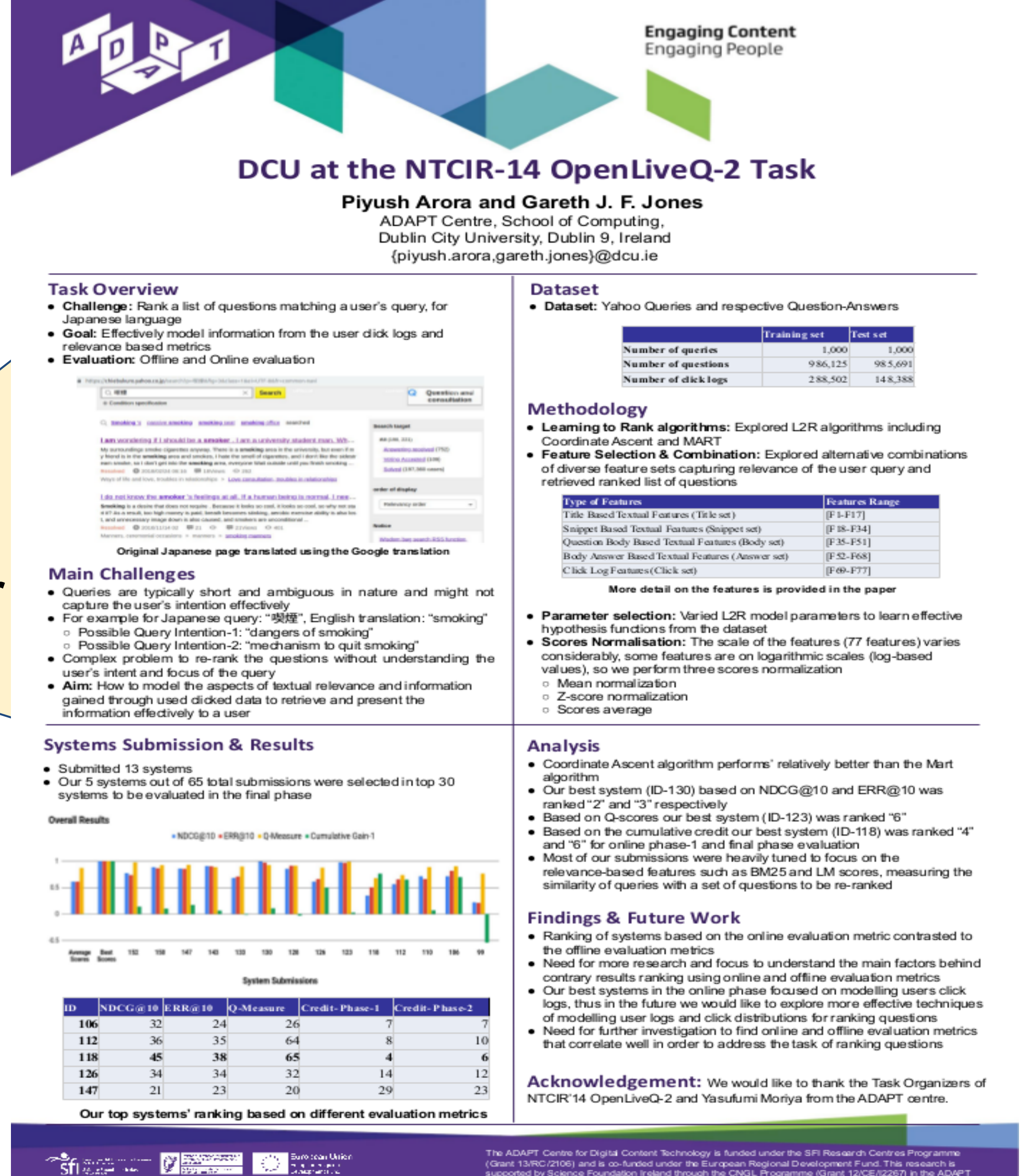
**** No significant differences between the top scored runs using Tukey's HSD tests**

Analysis

- Coordinate Ascent algorithm performs relatively better than the Mart algorithm
- Our best system (ID-130) based on NDCG@10 and ERR@10 was ranked “2” and “3” respectively
- Based on Q-scores our best system (ID-123) was ranked “6”
- Based on the cumulative credit our best system (ID-118) was ranked “4” and “6” for online phase-1 and final phase evaluation
- Most of our submissions were heavily tuned to focus on relevance-based features (for e.g BM25 and LM scores)

Findings & Future Work

- Ranking of systems based on the online evaluation metric differed from that for the offline evaluation metrics
- Need for more research to understand the factors behind contrary ranking results arising from the use of online and offline evaluation metrics
- Our best systems in the online phase focused on modelling users click logs
- **Future work:** explore more effective techniques for the exploitation of user logs and click distributions for ranking questions



Engaging Content
Engaging People

DCU at the NTCIR-14 OpenLiveQ-2 Task

Piyush Arora and Gareth J. F. Jones
ADAPT Centre, School of Computing,
Dublin City University, Dublin 9, Ireland
{piyush.arora,gareth.jones}@dcu.ie

Task Overview

- **Challenge:** Rank a list of questions matching a user's query, for Japanese language
- **Goal:** Effectively model information from the user click logs and relevance based metrics
- **Evaluation:** Offline and Online evaluation

Original Japanese page translated using the Google translation

Main Challenges

- Queries are typically short and ambiguous in nature and might not capture the user's intention effectively
- For example for Japanese query: "喫煙", English translation: "smoking"
  - Possible Query Intention-1: "dangers of smoking"
  - Possible Query Intention-2: "mechanism to quit smoking"
- Complex problem to re-rank the questions without understanding the user's intent and focus of the query
- **Aim:** How to model the aspects of textual relevance and information gained through used clicked data to retrieve and present the information effectively to a user

Systems Submission & Results

- Submitted 13 systems
- Our 5 systems out of 65 total submissions were selected in top 30 systems to be evaluated in the final phase

Overall Results

NDCG@10 ERR@10 Q-Measure Cumulative Gain-1

System Submissions

ID	NDCG@10	ERR@10	Q-Measure	Credit-Phase-1	Credit-Phase-2
106	32	24	26	7	7
112	36	35	64	8	10
118	**45**	**38**	**65**	**4**	**6**
126	34	34	32	14	12
147	21	23	20	29	23

Our top systems' ranking based on different evaluation metrics

Dataset

- **Dataset:** Yahoo Queries and respective Question-Answers

	Training set	Test set
Number of queries	1,000	1,000
Number of questions	986,125	985,691
Number of click logs	288,502	148,388

Methodology

- **Learning to Rank algorithms:** Explored L2R algorithms including Coordinate Ascent and MART
- **Feature Selection & Combination:** Explored alternative combinations of diverse feature sets capturing relevance of the user query and retrieved ranked list of questions

Type of Features	Features Range
Title Based Textual Features (Title set)	[F1-F17]
Snippet Based Textual Features (Snippet set)	[F18-F34]
Question Body Based Textual Features (Body set)	[F35-F51]
Body Answer Based Textual Features (Answer set)	[F52-F68]
Click Log Features (Click set)	[F69-F77]

More detail on the features is provided in the paper

- **Parameter selection:** Varied L2R model parameters to learn effective hypothesis functions from the dataset
- **Scores Normalisation:** The scale of the features (77 features) varies considerably, some features are on logarithmic scales (log-based values), so we perform three scores normalization
  - Mean normalization
  - Z-score normalization
  - Scores average

Analysis

- Coordinate Ascent algorithm performs' relatively better than the Mart algorithm
- Our best system (ID-130) based on NDCG@10 and ERR@10 was ranked "2" and "3" respectively
- Based on Q-scores our best system (ID-123) was ranked "6"
- Based on the cumulative credit our best system (ID-118) was ranked "4" and "6" for online phase-1 and final phase evaluation
- Most of our submissions were heavily tuned to focus on the relevance-based features such as BM25 and LM scores, measuring the similarity of queries with a set of questions to be re-ranked

Findings & Future Work

- Ranking of systems based on the online evaluation metric contrasted to the offline evaluation metrics
- Need for more research and focus to understand the main factors behind contrary results ranking using online and offline evaluation metrics
- Our best systems in the online phase focused on modelling users click logs, thus in the future we would like to explore more effective techniques of modelling user logs and click distributions for ranking questions
- Need for further investigation to find online and offline evaluation metrics that correlate well in order to address the task of ranking questions

Acknowledgement: We would like to thank the Task Organizers of NTCIR'14 OpenLiveQ-2 and Yasufumi Moriya from the ADAPT centre.

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. This research is supported by Science Foundation Ireland through the CNGL Programme (Grant 12/CE/I2267) in the ADAPT Centre for Digital Content Technology

More Details @ the Poster Session

Q/A

Acknowledgement:

- NTCIR’14 Organizers
- Task Organizers of NTCIR’14 OpenLiveQ-2
- Yasufumi Moriya from the ADAPT centre