RMIT at the NTCIR-12 MobileClick-2: iUnit Ranking and Summarization Subtasks

Kevin Ong, Ruey-Cheng Chen, and Falk Scholer
RMIT University
Melbourne, Australia

Introduction

- Learning-to-Rank (LTR) framework[1]
- Enhance iUnit ranking using feature-based approach
- Extend existing QSum[2] feature set

[1] Yang et al. (2016)
[2] Metzler & Kanungo. (2008)

Aim

- Maximise the score for ranking by exploring different novel features
  - 5 classes (BL, QSum, Semantic, QT, Ctx)
- Test our features on the summarisation subtask
- Check if LTR can deal with this level of homogeneity

Methods (1)

5 different classes

Class		Type
Baseline	BL	Language Model
Lexical / Synonymy	QSum	Metzler and Kanungo (2008)
Semantic Relatedness	Sem	Yang et al. (2016)
Question / Entity	QT	New
Document Context	Ctx	New

Features

Base (BL)

OddsRatio

QSum[1]

ExactMatch
TermOverlap
SynonymOverlap
LanguageModel
iUnitLength

Sem[2]

ESA
Word2Vec
Tagme

Question type and entities (QT)

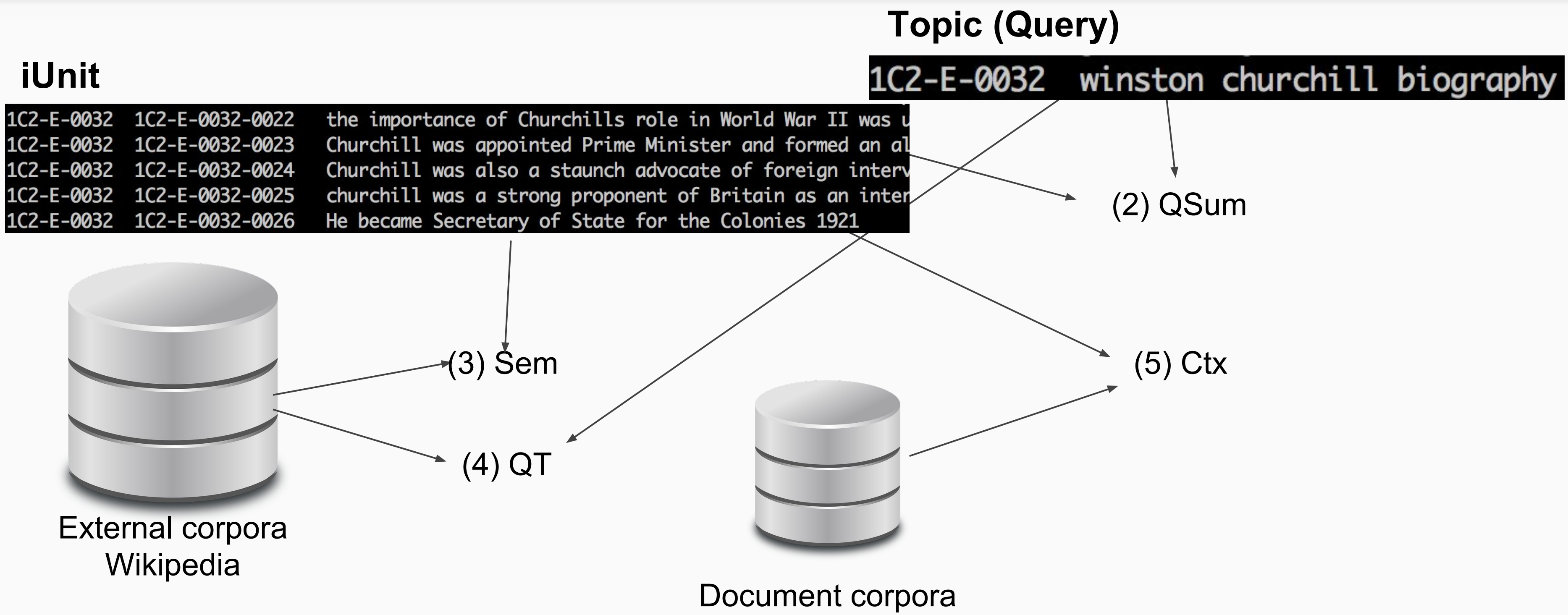
Leading5W1H
TopRankedWikiPage
WikipageRR
WikipagePassage

Contexts (Ctx)

CollectionFrequency
AverageSentencePos
AverageDocumentRR

[1] Metzler & Kanungo. (2008)
[2] Yang et al. (2016)

Methods (2)



Topic (Query)
1C2-E-0032 winston churchill biography
iUnit
1C2-E-0032 1C2-E-0032-0022 the importance of Churchills role in World War II was
1C2-E-0032 1C2-E-0032-0023 Churchill was appointed Prime Minister and formed an a
1C2-E-0032 1C2-E-0032-0024 Churchill was also a staunch advocate of foreign inter
1C2-E-0032 1C2-E-0032-0025 churchill was a strong proponent of Britain as an inte
1C2-E-0032 1C2-E-0032-0026 He became Secretary of State for the Colonies 1921
(2) QSum
(3) Sem
(4) QT
(5) Ctx
External corpora
Wikipedia
Document corpora

Results

English iUnit Ranking (Training)

Rank	Group ID	Score
1	RISAR	0.8878
2	UHYG	0.8788
3	ORG	0.8772
4	ALICA	0.8621

English iUnit Ranking (Test)

Rank	Group ID	Score
1	TITEC	0.9003
2	UHYG	0.8994
3	ORG	0.8975
4	RISAR	0.8972

English iUnit Summarization (Test)

Rank	Group ID	Score
1	TITEC	18.2596
2	ORG	16.8975
3	RISAR	16.047

Results -
English iUnit Ranking (1)

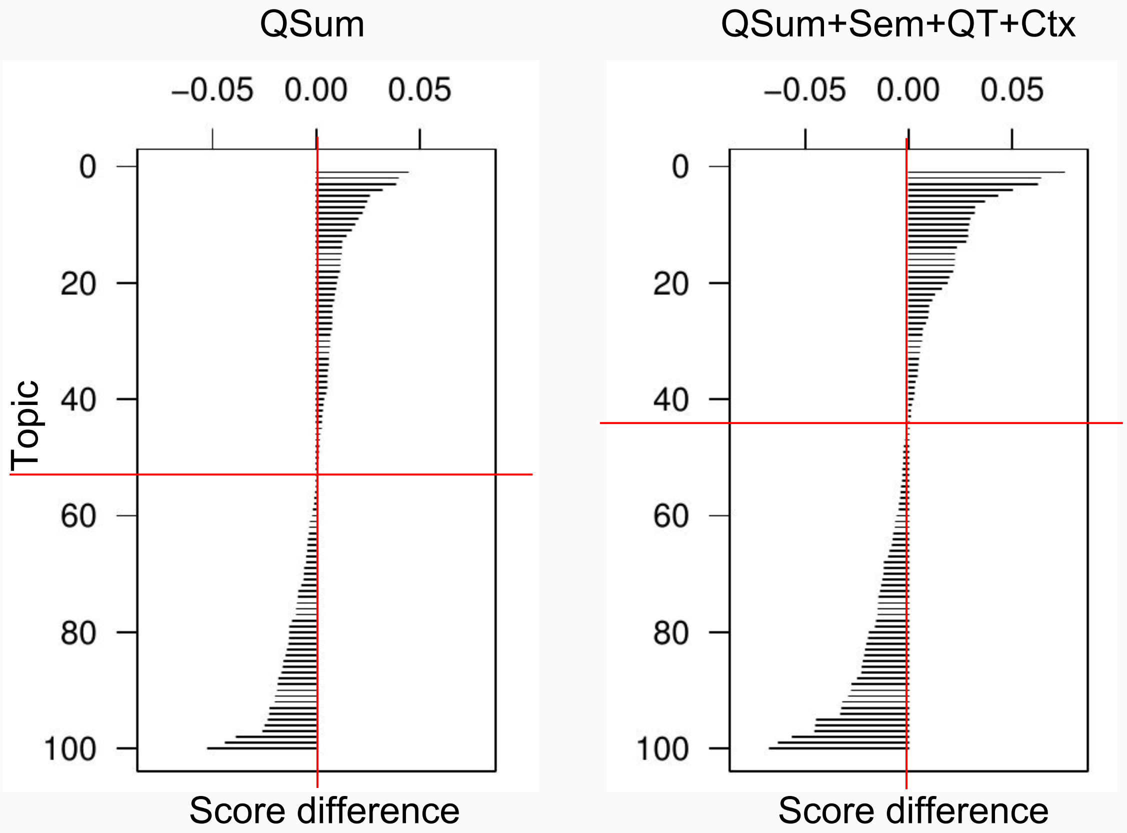
Method	Algorithm	nDCG@3	nDCG@5	nDCG@10	nDCG@20	Q
Base (**BL**)		0.7438	0.7643	0.7961	0.8325	0.8772
Base+QSum (**BQ**)	Coordinate Ascent	0.7669*	**0.7821**	**0.8137***	**0.8504****	**0.8882****
Base+QSum+Sem	LambdaMART	**0.7686***	0.7759	0.8085*	0.8422*	0.8848**
Base+QSum+Sem+QT	LambdaMART	0.7674*	0.7788	0.8070*	0.8453**	0.8875**
Base+QSum+Sem+QT+Ctx (**BQP**)	LambdaMART	0.7668*	0.7801*	0.8067	0.8435*	0.8819

Table 2: Performance results on the training set for English iUnit Ranking. Significant improvements are indicated by * and ** (Two-tailed t-test; $p < 0.05$ and $p < 0.01$ respectively).

Method	Algorithm	nDCG@3	nDCG@5	nDCG@10	nDCG@20	Q
Base (**BL**)		**0.7460**	**0.7596**	**0.8033**	0.8689	**0.8975**
Base+QSum (**BQ**)	Coordinate Ascent	0.7354	0.7557	0.8015	**0.8690**	0.8972
Base+QSum+Sem+QT+Ctx (**BQP**)	LambdaMART	0.7352	0.7532	0.8002	0.8666	0.8962

Table 3: Performance results on the test set for English iUnit Ranking.

Results -
English iUnit Ranking (2)



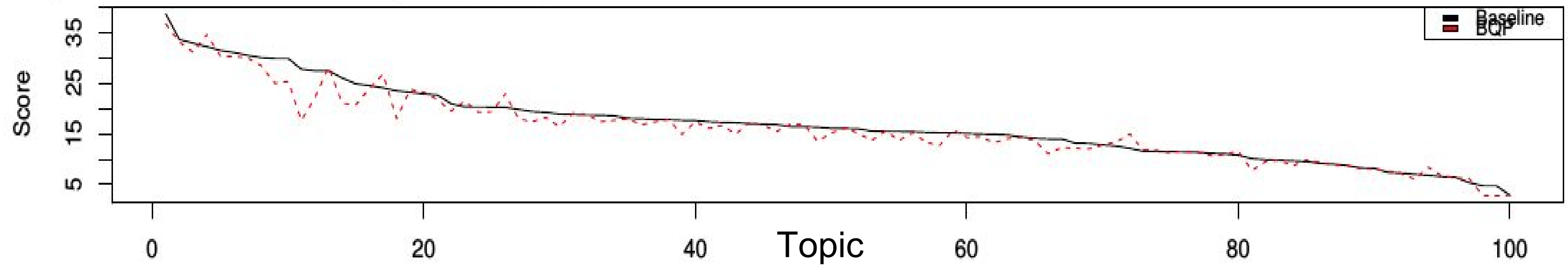
Results -

English iUnit Summarisation

Method	M
Base (**BL**), 2-layer	**16.8975**
Base+QSum+Sem+QT+Ctx (**BQP**), 2-layer	16.047

Table 4: Results for English iUnit Summarization.

Figure 2: Distribution of score for summarization subtask between Baseline and *BQP*. Baseline is observed to perform better than *BQP* features



Conclusion

For iUnit ranking

BL+QSum features outperformed newer features for training dataset

BL+QSum+Sem+QT+Ctx has a higher variance for gain/loss

For iUnit Summarisation

1 simple run submitted

Interesting to find LTR score better than BL occasionally

Generally

LTR may be useful to generate 2-layer summaries

終り