

UHYG at the NTCIR-12 MobileClick Task: Link-based Ranking on iUnit-Page Bipartite Graph

Sho Iizuka*
University of Hyogo, Japan
eo12w008@steng.u-hyogo.ac.jp

Manabu Nii
University of Hyogo, Japan
nii@eng.u-hyogo.ac.jp

Takayuki Yumoto
University of Hyogo, Japan
yumoto@eng.u-hyogo.ac.jp

Naotake Kamiura
University of Hyogo, Japan
kamiura@eng.u-hyogo.ac.jp

ABSTRACT

We participated in the iUnit ranking subtask and the iUnit summarization subtask of the NTCIR-12 MobileClick for the Japanese and English languages. Our strategy is based on link analysis on an iUnit-page bipartite graph. First, we constructed an iUnit-page bipartite graph considering the entailment relationship between the iUnits and the pages. Then, we ranked the iUnits by their scores based on link analysis. For the iUnit ranking subtask, we examined three types of entailment relationships and three types of link analysis, the degree of nodes, PageRank, and HITS. For the iUnit summarization subtask, we propose an intent-sensitive PageRank that is an extended version of the topic-sensitive PageRank based on the probability that users visit pages in a search result page.

Team Name

UHYG

Subtasks

iUnit Ranking Subtask(Japanese, English)
iUnit Summarization Subtask(Japanese, English)

Keywords

Bipartite graph, PageRank, HITS, topic-sensitive PageRank

1. INTRODUCTION

The NTCIR-12 MobileClick task consists of an iUnit ranking subtask and an iUnit summarization subtask[3]. We participated in both subtasks in the Japanese and English languages. In both subtasks, we constructed an iUnit-page bipartite graph and apply link analysis to it. We propose three methods to detect the entailment relationship between iUnits and pages. In the iUnit ranking subtask, we examine the degrees of the iUnit, PageRank[1], and HITS[4]. We compared combinations of three entailment detection methods and three link analysis methods and found the best combination. For the iUnit summarization subtask, we propose an intent-sensitive PageRank. It is based on the topic-sensitive PageRank[2], and it considers the probability that users visit pages in a search result page. We compared it with the original PageRank and examined its effectiveness.

*Present affiliation: Nara Institute of Science and Technology

2. IUNIT RANKING SUBTASK

2.1 Our Method for iUnit Ranking

Because iUnits are short texts, estimating their importance only from their texts is difficult. Therefore, we focused on the relationship between iUnits and Web pages that contain them. We constructed a bipartite graph between iUnit and Web pages based on their entailment relationship. Then, we computed the importance using link analysis-based approaches.

2.1.1 Constructing Bipartite Graph

In the bipartite graph, the nodes are iUnits and Web pages. If an entailment relationship existed between the iUnit and the Web page, we added an edge between them. We show an example of the bipartite graph in Figure 1. We used the Bag-of-Words model to judge the entailment relationships. We extracted nouns from iUnits and Web pages, and we compared the sets of nouns. To extract nouns, we used MeCab¹ for the Japanese language and Stanford POS Tagger² for the English language. We considered three types of entailment detection methods, ALL, ANY, and RATE. In the ALL method, we considered them to have an entailment relationship if all of the nouns in the iUnit appeared in the Web page. In the ANY method, we considered them to have the entailment relationship if any of the nouns in the iUnit appeared in the Web page. In the RATE method, we computed the rate of the nouns in the iUnits that appeared in the Web pages. Then, we used the rate as the weight of the edge in the weighted bipartite graph.

2.1.2 Link-based Ranking

Then, we applied link analysis methods into the bipartite graph. We considered three link analysis approaches, DEG, PR, and HITS.

The DEG approach is based on the idea that important iUnits appear in many Web pages. In this approach, we used the degrees of the iUnits as the importance.

In the PR approach, we applied the PageRank algorithm[1] to the bipartite graph. The PageRank is based on the idea that good pages are linked from other good pages. The

¹MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/>

²Stanford Log-linear Part-Of-Speech Tagger, <http://nlp.stanford.edu/software/tagger.shtml>

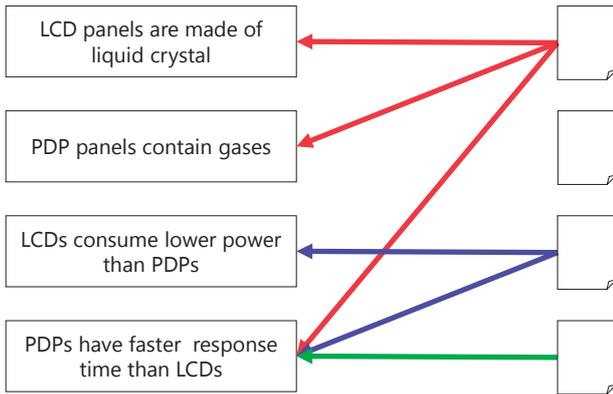


Figure 1: Bipartite graph between iUnits and Web pages

PageRank score is calculated using the following equation.

$$pr(v) \leftarrow (1 - \alpha) \frac{1}{|D|} + \alpha \sum_{s \in \text{in}(v)} \frac{pr(s)}{\text{outdeg}(s)} \quad (1)$$

$pr(v)$ is the PageRank score of page v and $|D|$ is the number of pages. $\text{in}(v)$ is the set of the pages that link to v , and $\text{outdeg}(v)$ is the number of the pages that are linked from v . α is a damping factor. Mihalcea and Tarau proposed the PageRank for weighted graphs[6, 5]. We define the iUnit version of the PageRank $pr'(v)$ based on PageRank for weight graphs as follows:

$$pr'(v) \leftarrow (1 - \alpha) \frac{1}{|U| + |D|} + \alpha \sum_{s \in \text{link}(v)} \frac{w_{s,v} pr'(s)}{\sum_{s' \in \text{link}(s)} w_{s,s'}} \quad (2)$$

$|U|$ is the number of the iUnits and $\text{link}(v)$ is the set of the nodes connecting to node v . $w_{s,s'}$ is the edge weight between s and s' . We set $\alpha = 0.85$. In the PR approach, we obtained not only the PageRank scores of iUnits but also the PageRank scores of Web pages. However, we only used the PageRank scores as the importance of iUnits.

In the HITS approach, we applied the HITS algorithm[4]. In the original HITS approach, each Web page has two types of importance, an authority score and a hub score. Good authority pages are linked from good hub pages, and good hub pages link to good authority pages. We applied this idea to the bipartite graph between iUnit and Web pages. We assumed that good iUnits connect to good Web pages and that good Web pages connect to good iUnits. In our HITS approach, each iUnit only had an authority score and each Web page only had a hub score. The authority scores of iUnits were used as importance. We also extended the HITS for weighted graphs in a similar way with the PageRank for weighted graphs. The iUnit version of HITS is defined in

equation (3).

$$\begin{aligned} h(d) &\leftarrow \sum_{u \in \text{link}(d)} \frac{w_{u,d} a(u)}{\sum_{d' \in \text{link}(u)} w_{u,d'}} \\ a(u) &\leftarrow \sum_{d \in \text{link}(u)} \frac{w_{u,d} h(d)}{\sum_{u' \in \text{link}(d)} w_{u',d}} \end{aligned} \quad (3)$$

2.2 Evaluation

2.2.1 Evaluation by Training Data

We combined three entailment approaches and three link analysis approaches, and we computed the importance of iUnits. We tried eight approaches except for RATE+DEG because the DEG method does not consider edge weight. The results for the Japanese language are shown in Table 1 and Figure 2. The results for the English language are shown in Table 2 and Figure 3. From these results, we obtained different tendencies between the two languages.

Table 1: Q-measures on iUnit ranking subtask (Japanese)

Link analysis \ Entailment	ALL	ANY	RATE
DEG	0.832	0.759	-
PR	<u>0.834</u>	0.761	0.802
HITS	0.823	0.758	0.803
Baseline (random)		0.773	
Baseline (LM-based)		0.790	

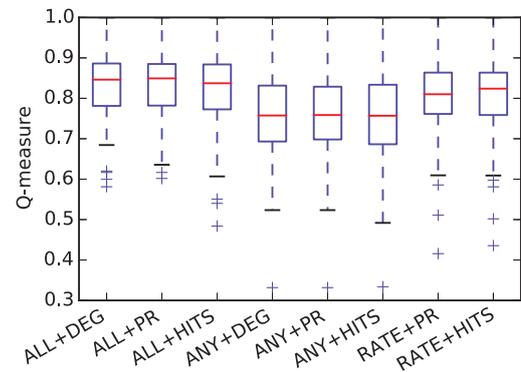


Figure 2: Distribution of Q-measures on iUnit ranking subtask (Japanese)

Table 2: Q-measures on iUnit ranking subtask (English)

Link analysis \ Entailment	ALL	ANY	RATE
DEG	0.825	0.879	-
PR	0.824	0.879	0.859
HITS	0.828	<u>0.881</u>	0.857
Baseline (random)		0.803	
Baseline (LM-based)		0.877	

First, we discuss about the results for the Japanese language. For Japanese, the best method was ALL+PR. The

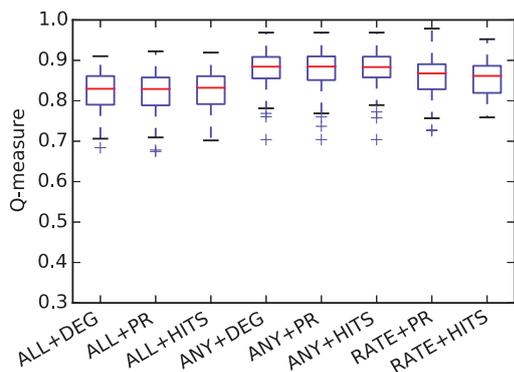


Figure 3: Distribution of Q-measures on iUnit ranking sub-task (English)

results using this method outperformed the results of the baselines. Among entailment detection methods, the ALL method was the best, and the RATE method and the ANY methods followed. We conducted paired two-tailed *t*-tests on pairs of Q-measures using arbitrary pairs of the eight methods. The results are shown in Table 3. In this table, the labels are sorted in descending order of the Q-measures. The Q-measures are shown in parentheses. The value in each cell is *p*-value between the corresponding methods. If the Q-measure using the method in the row was lower than the Q-measure using the method in the column, the value was omitted. If a statistically significant difference at the 1% level was found, the value was underlined. The results revealed statistically differences among the entailment detection methods, while no statistically differences were found among the link analysis methods except for two cases where ALL+PR was compared with ALL+HITS and where ALL+DEG was compared with ALL+HITS.

To analyze the performances in each query, we compared the Q-measures using ALL+PR with the Q-measures using a random baseline. We regarded the Q-measure using the random baseline as the easiness level for each query. If the Q-measure using ALL+PR was better than the Q-measure using the random baseline in the certain query, this meant that ALL+PR was good at the query. The results are shown in Figure 4. We found from this figure that ALL+PR was especially good at the celebrity category.

Next, we discuss the results for the English language. For English, the best method was ANY+HITS. Its Q-measure was better than the baselines. We compared the Q-measures in each query using ANY+HITS with the Q-measures using the random baseline. The results are shown in Figure 5. For the English language, we could not find obvious differences in the Q-measures using ANY+HITS among the categories.

We analyzed the differences using the entailment detection methods and the link analysis methods using a paired two-tailed *t*-test on pairs of Q-measures with arbitrary pairs of the eight methods. The results are shown in Table 4. The differences in the link analysis methods did not cause significant differences in the Q-measures at the 1% level. However, the differences in the entailment detection methods caused significant differences in the Q-measures at the 1% level. Surprisingly, the ANY method was the best, and the RATE method and the ALL methods followed among entailment

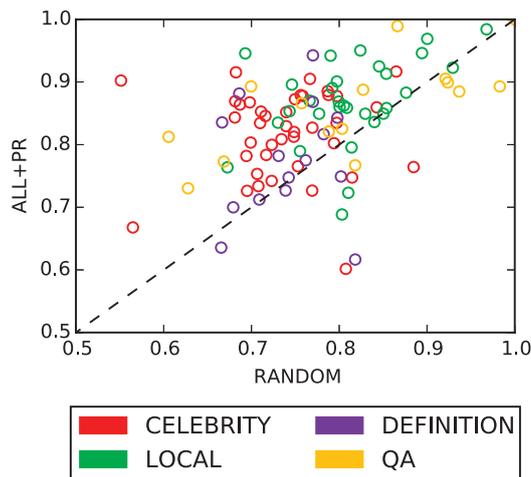


Figure 4: Differences in Q-measures between ALL+PR and Random (Japanese)

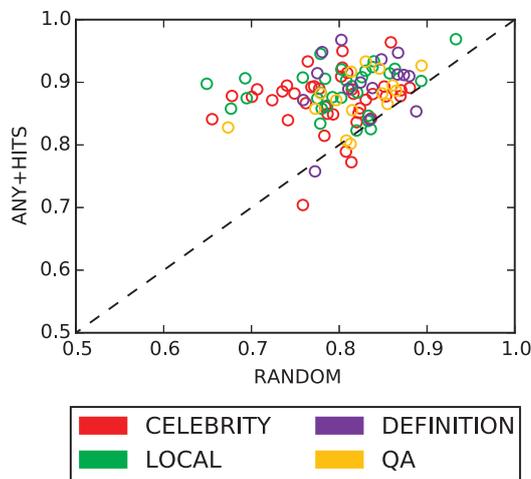


Figure 5: Differences in Q-measures between ALL+PR and Random (English)

detection methods. This tendency was the opposite to the tendency for the Japanese language.

To find the reason, we analyzed the relationship between the importance scores of the iUnits and their length. We calculated Kendall's τ between the importance scores of the iUnits and their length for each query. We show the results as histograms in Figure 6. We found that many queries having a positive τ in the Japanese training data and many queries having a negative τ in the English training data. This means that shorter iUnits tended to have higher importance in the Japanese training data while longer iUnits tended to have higher importance in the English training data. Although the ALL method and the ANY method were affected by such tendencies, the RATE method had no such effect. Therefore, the RATE method is more robust than other methods.

Table 3: p -values using our methods on iUnit ranking subtask (Japanese)
 paired two tailed t -test. underlined if $p < 0.01$

	ALL+DEG (0.832)	ALL+HITS (0.823)	RATE+HITS (0.803)	RATE+PR (0.802)	ANY+PR (0.761)	ANY+DEG (0.759)	ANY+HITS (0.758)
ALL+PR (0.834)	0.0813	<u>0.00222</u>	1.54×10^{-7}	7.38×10^{-7}	8.71×10^{-13}	2.27×10^{-13}	2.09×10^{-13}
ALL+DEG (0.832)	-	<u>0.00171</u>	3.60×10^{-7}	2.70×10^{-6}	1.50×10^{-12}	3.08×10^{-13}	2.62×10^{-13}
ALL+HITS (0.823)	-	-	<u>3.84×10^{-5}</u>	<u>0.000253</u>	<u>2.77×10^{-11}</u>	<u>4.13×10^{-12}</u>	<u>2.52×10^{-12}</u>
RATE+HITS (0.803)	-	-	-	0.629	<u>1.00×10^{-10}</u>	<u>2.24×10^{-11}</u>	<u>1.15×10^{-11}</u>
RATE+PR (0.802)	-	-	-	-	<u>1.44×10^{-11}</u>	<u>5.39×10^{-12}</u>	<u>7.61×10^{-12}</u>
ANY+PR (0.761)	-	-	-	-	-	0.431	0.177
ANY+DEG (0.759)	-	-	-	-	-	-	0.204

 Table 4: p -values using our methods on iUnit ranking subtask (English)
 paired two-tailed t -test. underlined if $p < 0.01$

	ANY+DEG (0.879)	ANY+PR (0.879)	RATE+PR (0.859)	RATE+HITS (0.857)	ALL+HITS (0.828)	ALL+DEG (0.825)	ALL+PR (0.824)
ANY+HITS (0.881)	0.146	0.180	<u>1.86×10^{-8}</u>	<u>1.18×10^{-9}</u>	<u>2.21×10^{-14}</u>	<u>4.35×10^{-15}</u>	<u>6.85×10^{-15}</u>
ANY+DEG (0.879)	-	0.392	<u>2.31×10^{-8}</u>	<u>3.40×10^{-8}</u>	<u>1.85×10^{-13}</u>	<u>1.11×10^{-14}</u>	<u>8.37×10^{-15}</u>
ANY+PR (0.879)	-	-	<u>3.12×10^{-8}</u>	<u>1.49×10^{-7}</u>	<u>3.67×10^{-13}</u>	<u>1.34×10^{-14}</u>	<u>6.95×10^{-15}</u>
RATE+PR (0.859)	-	-	-	0.167	<u>9.90×10^{-10}</u>	<u>1.09×10^{-11}</u>	<u>4.33×10^{-12}</u>
RATE+HITS (0.857)	-	-	-	-	<u>7.73×10^{-12}</u>	<u>8.88×10^{-13}</u>	<u>2.99×10^{-12}</u>
ALL+HITS (0.828)	-	-	-	-	-	0.0179	0.0378
ALL+DEG (0.825)	-	-	-	-	-	-	0.444

2.2.2 Formal Runs

On the basis of the aforementioned, we adopted the ALL method for the Japanese language and the ANY method for the English language. We submitted the results of ALL+DEG and ALL+PR for the Japanese test data and the results of ANY+DEG and ANY+PR for the English test data.

The average Q-measures are shown in Table 5 for the Japanese test data. We compared them using a paired two-tailed t -test, and we obtained $p = 2.14 \times 10^{-6}$. Therefore, ALL+DEG was significantly better than ALL+PR. Figure 7 shows the differences in the Q-measures using ALL+DEG and ALL+PR.

Table 5: Q-measures for test data (Japanese)

Method	Q-measure
ALL+DEG	0.839
ALL+PR	0.812

For the English test data, the average Q-measures are shown in Table 6. We compared them using a paired two-tailed t -test, and we obtained $p = 0.00366$. Therefore, ANY+DEG was significantly better than ANY+PR. Figure

7 shows the differences in the Q-measures using ANY+DEG and ANY+PR.

Table 6: Q-measures for test data (English)

Method	Q-measure
ANY+DEG	0.903
ANY+PR	0.899

3. IUNIT SUMMARIZATION SUBTASK

3.1 Our Method for iUnit Summarization

In the iUnit summarization subtask, a summary consists of two layers. On the first layer, we first ranked user intent keywords in a descending order of point-wise mutual information (PMI), and then we ranked iUnit in a descending order of PageRank defined in equation (2) up to the summary length limit. On the second layer, we ranked the iUnit in a descending order of intent-sensitive PageRank for each user intent.

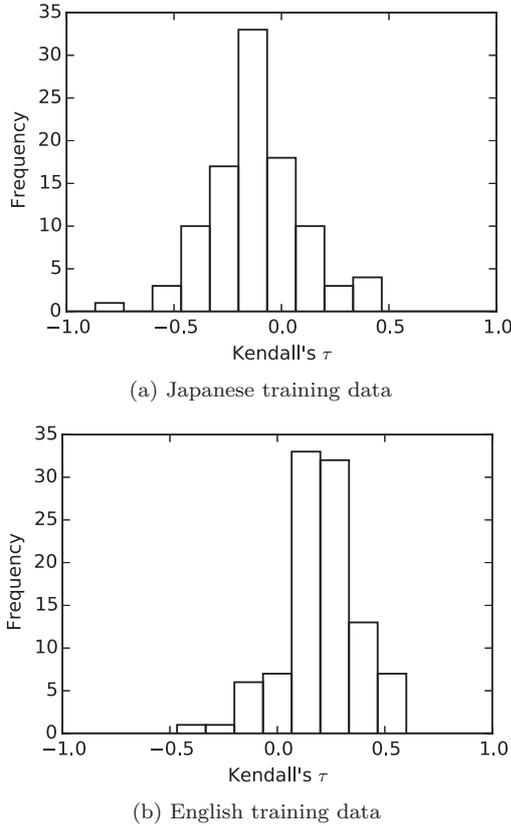


Figure 6: Histograms of Kendall's τ between length and importance of iUnit

3.1.1 PMI-based User Intent Ranking

In our approach for the iUnit summarization subtask, we put users' intent in the first layer of summarization.

For ranking in the first layer, we used PMI. We regard $PMI(i, D_q)$ as the importance of user intent i , where i is the intent keyword and D_q is the set of Web pages relevant to query q . PMI is defined as follows:

$$PMI(i, D_q) = \log \frac{P(i, D_q)}{P(i)P(D_q)} = \log \frac{P(i|D_q)}{P(i)} \quad (4)$$

In this equation, $P(i|D_q)$ is a probability that i appears in D_q . $P(i)$ is a probability that i appears in a general corpus. Instead of a general corpus, we used D_o , which is a set of Web pages removed D_q from the Web page dataset of MobileClick-2. $P(i)$ is calculated using the following unigram language model:

$$P(i) \simeq P(i|D_o) = \prod_{t \in i} \frac{TF(t, D_o)}{|T_{D_o}|} \simeq \prod_{t \in i} \frac{TF(t, D_o) + \beta}{|T_{D_o}| + \beta|T|}, \quad (5)$$

where $|T_D|$ is the number of words in document set D , and $TF(t, D)$ is the number of words t in D . $|T|$ is the number of the unique words in D_o . β is a smoothing parameter, and we set $\beta = 1.0$.

$P(i|D_q)$ is also calculated using an unigram language model.

$$P(i|D_q) = \prod_{t \in i} \frac{TF(t, D_q)}{|T_{D_q}|} \simeq \prod_{t \in i} \frac{TF(t, D_q) + \beta}{|T_{D_q}| + \beta|T|}, \quad (6)$$

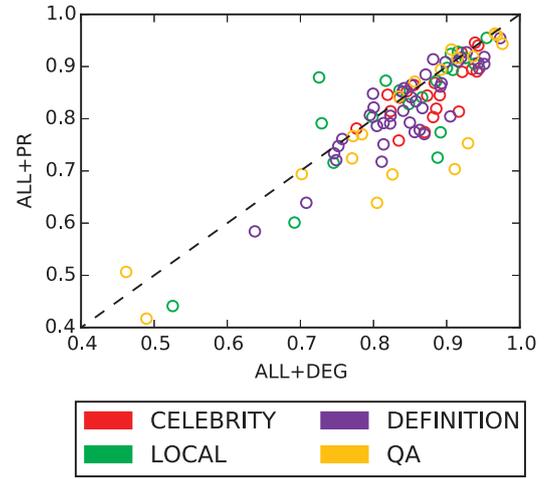


Figure 7: Differences in Q-measures between ALL+DEG and ALL+PR (Japanese)

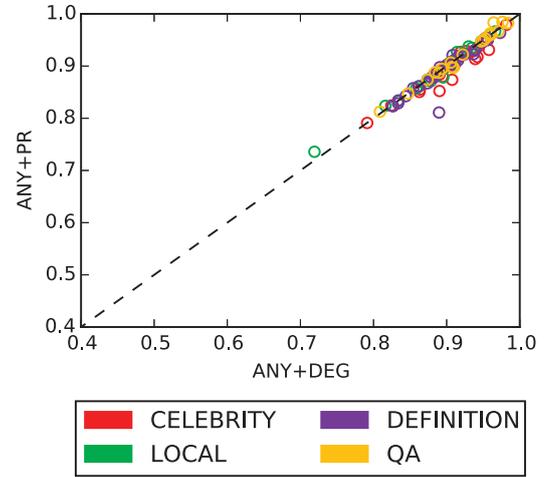


Figure 8: Differences in Q-measures using ANY+DEG and ANY+PR (English)

3.1.2 Intent-sensitive iUnit Ranking

In this method, we also constructed the iUnit-page bipartite graph. We applied the intent-sensitive PageRank to the bipartite graph.

The intent-sensitive PageRank is an extension of the topic-sensitive PageRank[2]. The topic-sensitive PageRank is also an extension of the PageRank to consider the topic of each page. It is defined as follows:

$$pr(v) \leftarrow (1 - \alpha)p_v + \alpha \sum_{s \in \text{in}(v)} \frac{pr(s)}{\text{outdeg}(s)} \quad (7)$$

p_v is a probability that a random surfer teleports to v , and that is defined with the relevancy between v and the topic. In the original topic-sensitive PageRank, the relevant pages have the same probability. In the intent-sensitive PageRank, we allocated higher probability to more relevant pages.

First, we ranked pages by BM25 defined in equation (8).

$$\text{BM25}(q, d) = \sum_{t \in q} \text{IDF}(t) \frac{(k_1 + 1) \text{TF}(t, d)}{\text{TF}(t, d) + k_1(1 - b + b \frac{L_d}{L_{ave}})} \quad (8)$$

$$\text{IDF}(t) = \log \frac{|D| - \text{DF}(t) + 0.5}{\text{DF}(t) + 0.5} \quad (9)$$

We used intent keywords as query q . $t \in q$ means that query q contains word t , and $t \in d$ means that document d contains word t . $\text{TF}(t, d)$ is the number of words t in document d , and $\text{DF}(t)$ is the number of words t in document set D . $|D|$ is the number of documents. L_d and L_{ave} are the number of words in document d and the average number of words in the documents, respectively. k_1 and b are parameters, and we set $k_1 = 2.0$ and $b = 0.75$.

We assume the probability for the k -th ranked page follows the geometric distribution. We define it as equation (10).

$$p_k = \gamma \lambda (1 - \lambda)^{k-1} \quad (10)$$

In teleportation vector \mathbf{p} , elements corresponding to iUnits are 0. γ is a coefficient normalizing p_k to satisfy condition $\sum_k p_k = 1$. To decide λ , we consider the probability that users visit the k -th ranked page in the search result page. Nakamura et al. conducted an online survey about user behavior on a Web search, and 1000 people answered [7]. According to the results of the survey, 53.2% of the respondents browse only the top five search results. Therefore, we set $\lambda = 0.15$ so that λ satisfies $\sum_{k=1}^5 p_k \simeq 0.532$ (Actually, $\sum_{k=1}^5 p_k = 0.556$).

We considered edge weights, and we defined the intent-sensitive PageRank in equation (11).

$$pr'(v) \leftarrow (1 - \alpha) p_v + \alpha \sum_{s \in \text{link}(v)} \frac{w_{s,v} pr'(s)}{\sum_{s' \in \text{link}(s)} w_{s,s'}} \quad (11)$$

α is a parameter to balance the effect of the link structure and page relevancy. $\alpha \simeq 1$ causes less effect of page relevancy and $\alpha \simeq 0$ causes more effect in the page relevancy. We set $\alpha = 0.1$. We set all of the edge weights to 1.0 in the experiments.

To obtain different intent-sensitive PageRank score for each intent, we changed teleportation vector \mathbf{p} for each intent, while we used the same bipartite graph for all intents.

3.2 Evaluation

We especially focused on the intent-sensitive iUnit ranking, and we compared it with the intent-insensitive ranking. In the intent-insensitive ranking, we used the PageRank defined in equation (2). We submitted the results of the intent-sensitive ranking and the intent-insensitive ranking for both languages. To detect entailment, we used the ALL method for the Japanese language and the ANY method for the English language.

The M-measures for the Japanese test data are shown in Table 7. This table shows that our proposed methods outperformed the baseline methods. We compared the intent-sensitive ranking and the intent-insensitive ranking using a paired two-tailed t -test, and we obtained $p = 2.87 \times 10^{-6}$. Therefore, the intent-sensitive PageRank was significantly better than the PageRank.

We show the distribution of M-measures in Figure 9. From this figure, we found that the intent-sensitive PageRank was better in some queries, but the PageRank was better in some queries. We show the queries and intent keywords where M-measures were improved using the intent-sensitive PageRank in Table 8. These queries have multiple meanings and their intent keywords specify the meanings. We show the highly ranked pages for the query “euro” in Table 9. For example, iUnit “UEFA EURO 2016” is highly ranked by the intent-sensitive PageRank for intent keyword “football competition,” although it is lowly ranked by PageRank.

However, PageRank works better than intent-sensitive PageRank for some queries. We show the examples in Table 10. In these queries, BM25 fails to rank relevant pages. For example, for query “word origin of Japan,” BM25 highly ranked the pages about the origin of other words in the QA sites. This caused unintended teleportation vectors in the intent-sensitive PageRank.

Table 7: M-measures in iUnit summarization subtask (Japanese)

Method	M-measure
Proposed (intent-sensitive)	22.834
Proposed (intent-insensitive)	21.111
Baseline (two-layered)	17.438
Baseline (random)	15.037
Baseline (LM-based)	12.799

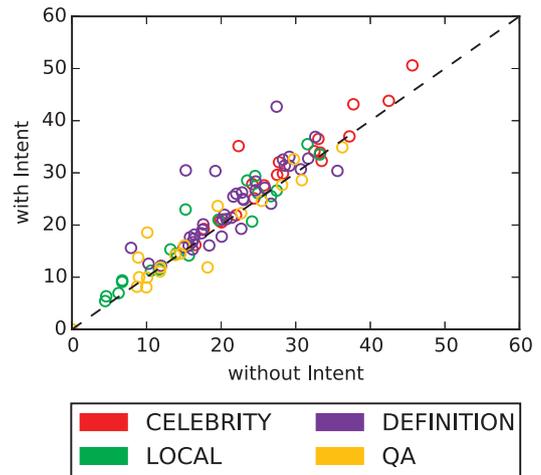


Figure 9: Differences in M-measures by intent-sensitive PageRank and intent-insensitive PageRank (Japanese)

The M-measures for the English test data are shown in Table 11. In contrast to the Japanese test data, the baseline methods were better than our methods. One of the reasons was that our methods did not consider the length of the iUnits. Because the total length of the summaries are limited in the summarization task, the scores of the iUnits per their length were more important than the scores of the iUnits themselves. However, we adopted the ANY method for the English test data. In this method, longer iUnits tended to have larger scores. To avoid this problem,

Table 8: Queries and intent keywords where M-measures were improved by intent-sensitive PageRank

Query	Intent-sensitive M-measure	Intent-insensitive M-measure	Intent keywords
euro	30.477	15.257	currency / overview Euro Co., Ltd. football competition
aladdin	15.614	7.918	band company name music drama character game cartoon
how to make macaron	18.557	10.100	cookware ganache dough final touch preparation
april fool	30.353	19.224	related work example origin name overview
Millet	35.135	22.337	MILLET JAPAN painter online greengrocery art work French sweets

we needed to normalize the scores by the length of the iUnits. Another way to avoid it was to use the ALL method instead of the ANY method. In the ALL method, shorter iUnits tended to have larger scores. When we unofficially submitted the result for English task using a topic-sensitive PageRank with the ALL method, M-measure was 16.567. It was better than topic-sensitive PageRank with the ANY method and it was comparable to the two-layered baseline.

We compared the intent-sensitive ranking and the intent-insensitive ranking using a paired two-tailed t -test, and we obtained $p = 0.0118$. We show the distribution of M-measures in Figure 10. From Figure 10, the Q-measures of the intent-sensitive ranking and the intent-insensitive rankings are the same in many of the queries. Therefore, small differences between the intent-sensitive ranking and the intent-insensitive ranking were found.

4. CONCLUSIONS

In this paper, we proposed a link-based ranking method on an iUnit-page graph for the NTCIR-12 MobileClick task. In the iUnit ranking subtask, we considered three entailment detection methods, ALL, ANY, and RATE and three link analysis methods, DEG, PR, and HITS. For the Japanese training data, the ANY method was the best while the ALL method was the best for the English training data. This was caused by the difference in the correlation between the length and importance score of the iUnits in the Japanese and English training data. However, no significant differences were found among the link analysis methods. For the iUnit summarization subtask, we proposed an intent-sensitive PageRank. It is based on a topic-sensitive PageRank and it considers the probability that users visit pages. We compared the intent-sensitive PageRank with the origi-

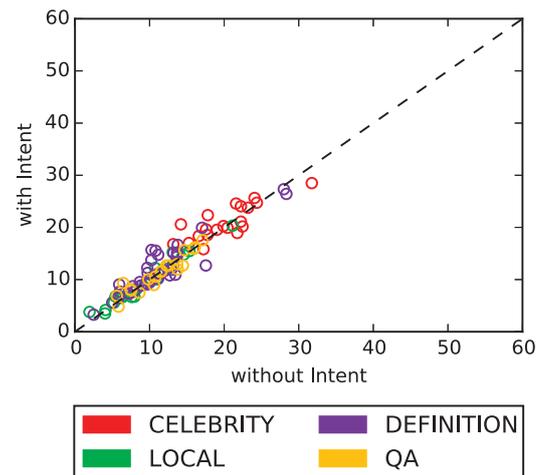


Figure 10: Difference in M-measures by intent-sensitive PageRank and intent-insensitive PageRank (English)

nal PageRank and demonstrated its effectiveness.

Acknowledgments

This work was supported by a Grant-in-Aid for Young Scientists (B) 24700097 from JSPS.

5. REFERENCES

- [1] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

Table 9: Highly ranked iUnits for query “euro” by intent-sensitive PageRank

Intent keywords	iUnit
-	euro one of world currencies such as US dollar and Japanese yen exchange rate mentioned in newspapers and news FAX. 03-xxxx-xxxx Foreign exchange [shop name]
currency / overview	euro exchange rate mentioned in newspapers and news UEFA EURO 2016 historical exchange rate of each pair of currencies one of world currencies such as US dollar and Japanese yen
Euro Co., Ltd.	euro Euro Co., Ltd. FAX. 03-xxxx-xxxx 24 teams x-y-z Higashi Kasai, Edogawa, Tokyo
football competition	euro UEFA EURO 2016 host country France 24 teams 15th UEFA Europe Championship

Table 10: Queries and intent keywords where M-measures were decreased by intent-sensitive PageRank

Query	Intent-sensitive M-measure	Intent-insensitive M-measure	Intent keywords
word origin of Japan	11.873	18.179	origin Zipangu
Reasons why white chocolate is white	8.060	9.972	reason
nlb	19.276	22.696	Network Load Balancing [product name]
baguette	30.379	35.596	shop name bread
junior high school Higashiyodogawa ward	20.670	24.146	Osaka City municipal Higashiyodogawa junior high school Osaka City municipal Zuiko junior high school

Table 11: M-measures in iUnit summarization subtask (English)

Method	M-measure
Baseline (two-layered)	16.898
Baseline (random)	14.105
Baseline (LM-based)	13.269
Proposed (intent-sensitive)	13.055
Proposed (intent-insensitive)	12.598

[2] T. Haveliwala. Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796, 2003.

[3] M. P. Kato, T. Sakai, T. Yamamoto, Matthew, V. Pavlu, H. Morita, and S. Fujita. Overview of the NTCIR-12 MobileClick Task. In *Proceedings of the 12th NTCIR Conference*, 2016.

[4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[5] R. Mihalcea. Graph-based ranking algorithms for

sentence extraction, applied to text summarization. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, number 4. Association for Computational Linguistics, 2004.

[6] R. Mihalcea and P. Tarau. TextRank: Bringing order into texts. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, 2004.

[7] S. Nakamura, S. Konishi, A. Jatowt, H. Ohshima, H. Kondo, T. Tezuka, S. Oyama, and K. Tanaka. Trustworthiness Analysis of Web Search Results. In *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries*, ECDL’07, pages 38–49, 2007.