YJST at the NTCIR-12 IMine-2 Task

Yasuaki Yoshida
Yahoo Japan Corporation
yayoshid@yahoo-corp.jp

Hideyuki Maeda
Yahoo Japan Corporation
hidmaeda@yahoo-corp.jp

Tatsuhiro Niwa
Yahoo Japan Corporation
tniwa@yahoo-corp.jp

Sumio Fujita
Yahoo Japan Corporation
sufujita@yahoo-corp.jp

ABSTRACT

Yahoo Japan Search Technology (YJST) team participated in the Query Understanding subtask of NTCIR-12 IMine-2. We explored various search log mining techniques to discover subtopics against the given original topics. For Vertical Identification, we trained a Gradient Boosted Decision Tree (GBDT) learner to identify a vertical label to each subtopic using several complex features including topical probabilities based on random walks on click graphs, and query distribution analyses through several commercial vertical search services and so on. Our best official submission run of subtopic mining achieves higher D#-nDCG@10 score than the average, but below the median of the best runs of all the participating team. In other measures such as QU-score or V-score, our best result performed as poorly as below the median. Although our task campaign was clearly not successful enough to confirm the adequacy of our service solutions, we try to analyze the results and the data as failure analyses are the only way to make a progress towards the future.

Team Name

Yahoo Japan Search Technology

Subtasks

Query Understanding subtask (Japanese)

Keywords

query intent, diversity, subtopics

1. INTRODUCTION

The YJST team participated in the Query Understanding subtask [12]. The goal of the subtask is to discover and rank the subtopics against given topics according to the relevance and importance and finally to assign a vertical label to discovered subtopics. Naturally the system consists of two modules namely, subtopic mining, which discovers and ranks subtopics and vertical identification, which assigns a vertical label to each subtopics.

For subtopic mining, we extended our base system, product query suggestion system introducing three additional procedures. First we added new subtopic candidates to the base system. We generate additional subtopics by using chronic candidates of the base system, best rank co-click queries and IMine-2 Co-topic queries which we provided as IMine-2 offical datasets [6]. In the second step, we rerank subtopic candidates using RankSVM [9] implementation. In the final step, we screen aforementioned candidate subtopics out by using co-click, co-session, and co-topic relations.

For Vertical Identification, we trained a GBDT learner to identify a vertical label to each subtopic using several complex features including topical probabilities based on random walks on click graphs, and query distribution analyses through several commercial vertical search services and so on.

In Section 2 and 3, we describe Subtopic Mining and Vertical Identification phases respectively. In Section 4, we explain our evaluation experiments and discuss the results. Section 5 concludes the work.

2. SUBTOPIC MINING

For subtopic mining, we extended our base system, i.e. operational query suggestion system of our web search service, in order to optimize it for the IMine-2 task. Due to intellectual property restrictions, we refrain from explaining the details of the operational system. The overview of our extension, integrated with existing system is shown in Figure 1.

2.1 Base System and Extensions

Our base system generates 10 subtopics for each given query by using query logs from our commercial web search service, which holds the biggest market share among Japanese web search services. The characteristic features of the system include:

- Subtopic queries expand input queries by adding one term to the input query.
- Subtopic queries reflect the recent trends since they are extracted from recent query logs, though they do not cover users' intents.
- Subtopics do not cover all the facets of original query; the system is optimized to maximize the CTR i.e. to satisfy maximum users with limited number of subtopics but not to maximize the number of facets it covers.

Our extended system consists of the three phases namely, subtopic candidate generation which assures better coverage, subtopic candidate re-ranking which optimize the ranking to maximize the click trough rate (CTR) of the links to

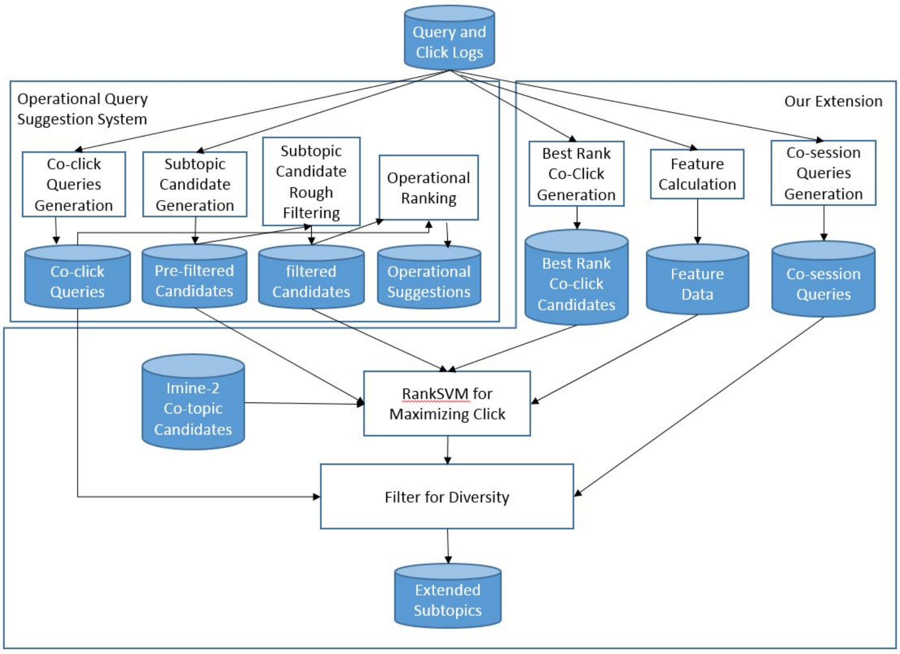

Figure 1: Overview of Our System

suggestions, and finally subtopic candidate filtering which diversifies the results by eliminating redundant subtopics.

2.2 Subtopic Candidate Generation

We used the data from three sources for subtopic mining as described in subsequent sections. We mergeed subtopic candidates from all the three sources, and create the candidate set.

2.2.1 *Candidates from the Base System*

The base system outputs 200 subtopic candidates. These candidates were generated by using the query logs for a period of 28 days, from July 21st 2015 to August 17th 2015.

2.2.2 *Pre-filtered Candidates*

The system roughly filters subtopics and reduce the number of candidates to 200 in it's early phase of the generating process. The pre-filtered candidates are lists of subtopics preceding the filtering, which we also used to improve the coverage.

2.2.3 *Co-topic Queries of IMine-2 Query logs*

We used co-topic queries, which we provided to IMine-2 official dataset as candidates [6]. As mentioned in [4], such query relations represent subtopics of original queries, in other words drilled down queries, rather than synonymous or parallel move queries [1]. These subtopics are based on query logs for the duration of about 4 years, covering data much longer than that of 28 days used in the operational system.

2.2.4 *Best Rank Co-Click Queries*

These queries represent not only subtopic relations but also synonymous relations. Let q_h and q_l be the best rank co-click queries. These queries have the following relations:

- A URL u is clicked in SERP of queries q_h and q_l.
- u is ranked higher by q_h than q_l.

The best ranking restriction strongly suggests that q_h specifies the intent of q_l. We also extract the best rank co-click queries by using query logs of 28 days, from July 21 2015 to August 17 2015. The co-click queries (q_h, q_l) should hold the following conditions:

- In the SERP of q_h, the clicked URL u should be in top three.
- The URL u should be clicked at least three times in SERP of both q_h and q_l.

2.3 Subtopic Re-ranking

In view of maximizing the CTR, we re-rank subtopic candidates by using the extended liblinear [7] library, incorporating a RankSVM [9] learner as learning algorithm options. We optimized the ranking by directly maximizing the mean reciprocal rank against our own gold standard data produced based on search user behavior logs. Features include frequencies of queries, CTR when presented to users as suggestion, the rank position when clicked and so on[1].

2.4 Subtopic Filtering

For improving the diversity of subtopic listing, we apply three kinds of filters; co-click filter, co-session filter, and co-topic filter. Each filter remove "similar" subtopics to the original topic or its higher ranked subtopics in order to reduce the redundant information.

Co-clcik filter The idea behind this is that a query q_i and another query q_j whose sets of clicked URLs are similar to each other may have the similar intent.

[1]For confidential reasons, we cannot reveal more details.

Co-session Filter This filter is based on the rational that the two query q_i, q_j are searched in a small time interval, q_i and q_j are likely to have the similar intent. We used the following procedure:

1. Extract user sessions by 5 minute interval.
2. Filter sessions which include over 200 queries.
3. For each session, create query combinations (q_i, q_j) where q_i is searched before q_j in the session.
4. Calculate difference of sequence number for each query combination.
5. Calculate combination frequency $f_{i,j}$ of each (q_i, q_j), average difference of sequence $s_{i,j}$.
6. Filter the combination whose $f_{i,j}$ is less than 30 or whose $s_{i,j}$ is more than 10.

We extract these query combinations by using query logs for a period of 28 days from July 21 2015 to August 17 2015.

Co-topic Filter This filter is based on the idea that if a query q_2 is expanded by adding some terms to another query q_1, q_1 and q_2 have similar intent.

2.5 Supplementary Subtopics for Short Lists

We obtain ranked subtopics by the procedures described above. If the number of subtopics is less than 10, we append additional subtopics by processing constituent sub strings of the original query.

3. VERTICAL IDENTIFICATION

In this section, we describe our approach of Vertical Identification which follows Subtopic Mining. The goal of the module is to identify the relevant verticals for each of the subtopic queries generated by the Subtopic Mining part. Since we focus on Japanese topics, we treat only the vertical labels, namely *Web, Image, News, QA, Encyclopedia,* and *Shopping.*

We treat the task as a Multi-label Classification problem. Hence, we learned the classifiers for each vertical using GBDT. Thus learned models corresponding to each topic (web, image, shopping, etc.) are represented by M1, M2, M3 etc.. in figure 2. The queries generated by the Subtopic Mining phase are passed to each model(M1, M2, M3, etc.) which computes a probability of the input query being classified into the modeled vertical by a logistic regression as shown in figure 2.

The GBDT method [2] [3], which is an additive regression model over an ensemble of shallow regression trees. It iteratively fits an additive model:

$$F_m(x) \quad = \quad F_{m-1}(x) + \beta_m T_m(x; \Theta_m) \ ,$$

where $T_m(x; \Theta_m)$ is a regression tree at iteration m, weighted by parameter β_m, with a finite number of parameter Θ_m, consisting of split regions and corresponding weights, which are optimized such that a certain loss function is minimized as follows:

$$(\beta_m, \Theta_m) \quad = \quad argmin_{\beta,\Theta} \sum_{i=1}^{N} L(y_i, F_{m-1}(x) + \beta T_m(x; \Theta)) \ .$$

At iteration m, tree $T_m(x; \Theta_m)$ is induced to fit the negative gradient by least squares:

$$\hat{\Theta} \quad = \quad argmin_{\Theta,\beta} \sum_{i=1}^{N} (-g_m(x_i) - \beta_m T_m(x; \Theta_m))^2 \ .$$

where $-g_m(x_i)$ is the gradient over current prediction function:

$$-g_m(x_i) \quad = \quad -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \ .$$

Each non-terminal node in the tree represents the condition of a split on a feature space and each terminal node represents a region. The improvement criterion to evaluate splits of a current terminal region R into two subregions (R_ℓ, R_r) is as follows:

$$i^2(R_\ell, R_r) \quad = \quad \frac{w_\ell w_r}{w_\ell + w_r}(y_\ell - y_r)^2 \ ,$$

where y_ℓ, y_r are the mean response of left and right subregions respectively, and w_ℓ, w_r are the corresponding sums of weights. We evaluate the relative importance of each feature by the normalized sum of $i^2(R_\ell, R_r)$ through all the nodes corresponding to the feature.

Let C be a set of topics, $score(c_i)$ is :

$$score(c_i) = GBDTscore(c_i) + b_i \tag{1}$$

$GBDTscore(c_i)$ is a classification decision score for each topic i computed by the model. We define b_i as a threshold for each vertical class. Then, for each query, we extract a vertical class which has the highest score as below.

$$c^* = \arg\max_{c_i \in C} score(c_i) \tag{2}$$

The set of features used in the task is listed in Table 1 is described in more detail in the following sub-sections.

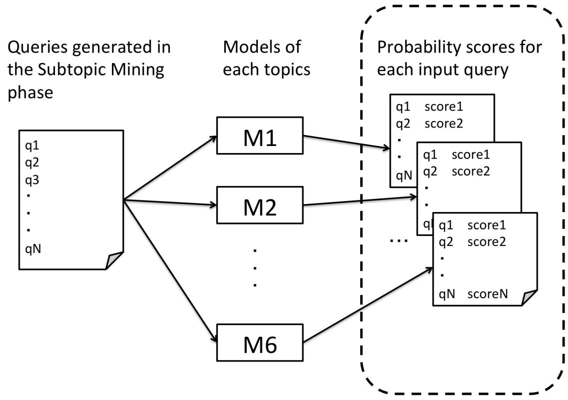

Figure 2: Vertical Identification Process

3.1 Vertical Search Features (VSF)

We used vertical search services (ex: Yahoo! Shopping, Yahoo! Image etc) in Yahoo! JAPAN, to calculate each vertical score based on the relative frequency in corresponding Yahoo! JAPAN service logs. The features F1 to F5 in table 1 represent them.

Table 1: Features Set

Feature	Description
F1	VSF(News)
F2	VSF(Image)
F3	VSF(QA)
F4	VSF(Encylopedia)
F5	VSF(Shopping)
F6	VSF(Web)
F7	SLMF(News)
F8	SLMF(Image)
F9	SLMF(QA)
F10	SLMF(Encylopedia)
F11	SLMF(Shopping)
F12	RWF(News)
F13	RWF(Image)
F14	RWF(QA)
F15	RWF(Encylopedia)
F16	RWF(Shopping)
F17	RMF(Web)

3.2 Statistical Language Model Features (SLMF)

For each vertical, we estimated a word n-gram language model based on Yahoo! JAPAN search logs. We used $P(q|\theta_i)$, probability of generating input query q given the language model of i_{th} vertical model as the features of the GBDT learner. The features F6 to F9 in table 1 represent these.

3.3 Random Walk Features (RWF)

We constructed a click bipartite graph using Web Search logs [5]. The click bipartite graph is a triplet $\mathcal{G} = (\mathcal{U}, \mathcal{Q}, \mathcal{E})$ where $\mathcal{U}$ is a set of URLs, $\mathcal{Q}$ is a set of queries and $\mathcal{E}$ is a set of edges representing a click event on a URL in $\mathcal{U}$ in response to a query in $\mathcal{Q}$. This graph is represented by the adjacency matrix A,where element $A_{i,j}$ is the count of click events on URL i in response to query j. We normalize the adjacency matrix by out degrees,i.e., the sum of each row,to obtain the transition matrix P as

$$P_{i,j} = \frac{A_{i,j}}{\sum_j A_{i,j}} \tag{3}$$

Let S be a subset of $\mathcal{V} = \mathcal{U} \cup \mathcal{Q}$, a set of seed nodes, representing examples of a vertical intent. We define vector s of dimension $|\mathcal{V}|$ as

$$s_i = \begin{cases} \frac{1}{|S|} & (\mathcal{V}_i \in S) \\ 0 & (otherwise) \end{cases} \tag{4}$$

The score of each vertex is computed iteratively according to

$$m_{t+1} = \alpha P m_t + (1 - \alpha)s \tag{5}$$

until a convergence is achieved. The random walk scores thus obtained for each query is used as features of GBDT. The features F6 to F9 in table 1 represent these features.

4. EXPERIMENTS

4.1 Subtopic Mining Experiments

We carried out experiments measuring the I-rec, D-nDCG and D#-nDCG [11] scores on the test set of NTCIR-10 INTENT-2 Task [10] as well as IMINE-2 test set, given that the INTENT-2 Subtopic Mining Subtask is similar to that of IMine-2.

The abbreviation list of examined extension methods is shown in Table 2.

Table 2: Names of Extension Methods

Name	Extension
PFC	Pre-filtered candidates of Base Suggestion System
IM2Q	Co-topic Queries of IMine-2 Query Logs
BRCC	Best Rank Co-click Queries
CC	Co-click filter
CS	Co-session filter
CT	Co-topic filter

We computed scores using NTCIREVAL [8] and the results are shown in Table 3. In the run description, T (True) means the extension is turned on and F (False), turned off.

4.2 Vertical Identification and Official Runs

In the vertical identification phase, we extracted the vertical intent of the query as described in Section 3. To evaluate the results, we used the test dataset of NTCIR-10 INTENT-2 Task [10] . For each query in the test dataset, we labeled a vertical intent manually. We calibrate the threshold parameter b by applying Algorithm 1.

Algorithm 1 Learning parameter b

Require: $iterations \Leftarrow 0$, $b \Leftarrow 0$
  while b is not converged **do**
    for $c \in C$ **do**
      $w(c) \Leftarrow 0$
    end for
    while $q \in Q$ **do**
      if $Topic(q) \neq CorrectTopic(q)$ **then**
        $w(Topic(q)) \Leftarrow w(Topic(q)) - \alpha$
        $w(CorrectTopic(q)) \Leftarrow w(CorrectTopic(q)) + \alpha$
      end if
    end while
    while $q \in Q$ **do**
      for $c \in C$ **do**
        $score(q, c) \Leftarrow score(q, c) + w(c)$
      end for
      $b \Leftarrow b + w(c)$
    end while
    $iterations \Leftarrow iterations + 1$
  end while
  $N \Leftarrow iterations$

We prepared the five submission runs using two Subtopic Mining outputs. In each run, we changed the iteration cut off when calibrating the threshold to generate different outputs. Finally, we submitted the five runs for Japanese Query Understanding subtask as shown in Table 4.

Table 3: Evaluation results of subtopic mining against INTENT-2 and IMINE-2 test set ; all measures are of @10.

Run Description							INTENT-2			IMINE-2		
Run#	PFC	IM2Q	BRCC	CC	CS	CT	I-rec	D-nDCG	D#-nDCG	I-rec	D-nDCG	D#-nDCG
SM01	F	F	F	F	F	F	**0.1962**	**0.1829**	**0.1896**	0.6046	**0.5214**	**0.563**
SM02	F	F	F	F	F	T	0.1919	0.1779	0.1849	0.614	0.4882	0.5511
SM03	F	F	F	F	T	F	0.1684	0.1672	0.1678	0.596	0.4678	0.5319
SM04	F	F	F	F	T	T	0.1826	0.17	0.1763	0.6013	0.457	0.5291
SM05	F	F	F	T	F	F	0.1789	0.173	0.176	0.604	0.506	0.555
SM06	F	F	F	T	F	T	0.192	0.1763	0.1841	0.6136	0.4863	0.5499
SM07	F	F	F	T	T	F	0.1673	0.1643	0.1658	0.5887	0.4657	0.5272
SM08	F	F	F	T	T	T	0.1804	0.1672	0.1738	0.5953	0.4556	0.5255
SM09	F	F	T	F	F	F	0.1962	0.1829	0.1896	0.6046	0.5214	0.563
SM10	F	F	T	F	F	T	0.1862	0.1727	0.1795	0.5653	0.4462	0.5057
SM11	F	F	T	F	T	F	0.1695	0.1668	0.1682	0.5879	0.4656	0.5267
SM12	F	F	T	F	T	T	0.1758	0.1622	0.169	0.538	0.4124	0.4752
SM13	F	F	T	T	F	F	0.1789	0.1737	0.1763	0.604	0.506	0.555
SM14	F	F	T	T	F	T	0.1863	0.1705	0.1784	0.5648	0.445	0.5049
SM15	F	F	T	T	T	F	0.1672	0.1635	0.1653	0.5823	0.4639	0.5231
SM16	F	F	T	T	T	T	0.1723	0.1589	0.1656	0.5337	0.4122	0.473
SM17	F	T	F	F	F	F	0.1633	0.1716	0.1675	0.5728	0.4731	0.5229
SM18	F	T	F	F	F	T	0.1551	0.1629	0.159	0.5732	0.4469	0.5101
SM19	F	T	F	F	T	F	0.1314	0.1392	0.1353	0.5218	0.3821	0.452
SM20	F	T	F	F	T	T	0.13	0.1347	0.1324	0.5139	0.3715	0.4427
SM21	F	T	F	T	F	F	0.1564	0.1628	0.1596	0.5683	0.4581	0.5132
SM22	F	T	F	T	F	T	0.154	0.1606	0.1573	0.5712	0.4415	0.5064
SM23	F	T	F	T	T	F	0.1301	0.1368	0.1335	0.5198	0.3807	0.4503
SM24	F	T	F	T	T	T	0.1287	0.1328	0.1308	0.5148	0.3704	0.4426
SM25	F	T	T	F	F	F	0.1551	0.1648	0.1599	0.5513	0.4355	0.4934
SM26	F	T	T	F	F	T	0.148	0.1563	0.1522	0.5322	0.4105	0.4714
SM27	F	T	T	F	T	F	0.1278	0.1344	0.1311	0.5029	0.3621	0.4325
SM28	F	T	T	F	T	T	0.123	0.1286	0.1258	0.466	0.3356	0.4008
SM29	F	T	T	T	F	F	0.1469	0.1549	0.1509	0.5496	0.4272	0.4884
SM30	F	T	T	T	F	T	0.1456	0.1536	0.1496	0.5302	0.4083	0.4692
SM31	F	T	T	T	T	F	0.1265	0.1326	0.1295	0.5026	0.3636	0.4331
SM32	F	T	T	T	T	T	0.1218	0.1267	0.1242	0.4653	0.3358	0.4005
SM33	T	F	F	F	F	F	0.1933	0.1773	0.1853	0.6007	0.5064	0.5536
SM34	T	F	F	F	F	T	0.1847	0.1709	0.1778	0.6237	0.4983	0.561
SM35	T	F	F	F	T	F	0.1612	0.1595	0.1603	0.5983	0.4728	0.5355
SM36	T	F	F	F	T	T	0.1734	0.1614	0.1674	**0.6247**	0.484	0.5544
SM37	T	F	F	T	F	F	0.1702	0.1652	0.1677	0.6022	0.5025	0.5523
SM38	T	F	F	T	F	T	0.1824	0.1683	0.1753	0.6179	0.4972	0.5576
SM39	T	F	F	T	T	F	0.1588	0.156	0.1574	0.5876	0.4731	0.5303
SM40	T	F	F	T	T	T	0.1711	0.1587	0.1649	0.6223	0.4863	0.5543
SM41	T	F	T	F	F	F	0.1845	0.168	0.1763	0.5673	0.4603	0.5138
SM42	T	F	T	F	F	T	0.1662	0.1593	0.1627	0.5611	0.4436	0.5024
SM43	T	F	T	F	T	F	0.1552	0.1506	0.1529	0.5706	0.4347	0.5026
SM44	T	F	T	F	T	T	0.1562	0.1503	0.1532	0.5568	0.4276	0.4922
SM45	T	F	T	T	F	F	0.1627	0.1571	0.1599	0.5679	0.4596	0.5138
SM46	T	F	T	T	F	T	0.1638	0.1573	0.1606	0.5599	0.4469	0.5034
SM47	T	F	T	T	T	F	0.1528	0.1476	0.1502	0.5603	0.4378	0.499
SM48	T	F	T	T	T	T	0.1527	0.147	0.1498	0.553	0.4318	0.4924
SM49	T	T	F	F	F	F	0.1606	0.1666	0.1636	0.5834	0.4698	0.5226
SM50	T	T	F	F	F	T	0.1536	0.1595	0.1566	0.5864	0.4499	0.5181
SM51	T	T	F	F	T	F	0.132	0.1349	0.1355	0.539	0.3968	0.4679
SM52	T	T	F	F	T	T	0.1294	0.1305	0.1299	0.5386	0.3935	0.4661
SM53	T	T	F	T	F	F	0.1523	0.1587	0.1555	0.5829	0.4595	0.5212
SM54	T	T	F	T	F	T	0.1513	0.1568	0.154	0.5844	0.4468	0.5156
SM55	T	T	F	T	T	F	0.1308	0.1324	0.1316	0.537	0.3953	0.4661
SM56	T	T	F	T	T	T	0.127	0.1282	0.1276	0.5395	0.3923	0.4659
SM57	T	T	T	F	F	F	0.1571	0.1624	0.1597	0.5547	0.4336	0.4941
SM58	T	T	T	F	F	T	0.1476	0.1531	0.1504	0.5435	0.4179	0.4807
SM59	T	T	T	F	T	F	0.1284	0.1304	0.1294	0.5199	0.3772	0.4485
SM60	T	T	T	F	T	T	0.1247	0.1262	0.1254	0.4936	0.3587	0.4262
SM61	T	T	T	T	F	F	0.1489	0.1544	0.1517	0.5642	0.4272	0.4957
SM62	T	T	T	T	F	T	0.1453	0.151	0.1482	0.5415	0.4158	0.4787
SM63	T	T	T	T	T	F	0.1272	0.1285	0.1278	0.5195	0.378	0.4488
SM64	T	T	T	T	T	T	0.1223	0.1239	0.1231	0.4928	0.3576	0.4252

Table 4: Official Runs

Name	# iter	SM #	QU-Score
YJST-Q-J-1Q	N	SM01	**0.5486**
YJST-Q-J-2Q	≈N/2	SM01	0.5253
YJST-Q-J-3Q	≈0	SM40	0.4436
YJST-Q-J-4Q	≈N/2	SM40	0.4914
YJST-Q-J-5Q	N	SM40	0.5192

4.3 Evaluation Results

4.3.1 Subtopic Mining

As shown in Table 3, all the extension methods we adopted negatively affected in pre-submission experiments using INTENT-2 test set.[2] At this point, we had had to seriously reconsider the strategy to the current task as well as precise re-investigation on the possible software implementation and experiment operation problems.

Supplementary added subtopics did not improve the results at all. This strongly suggests that we should have generated subtopic candidates from other sources than search logs. Although DCG measures did not improve through the all experiments, we found some improvements of I-rec@10 in IMINE-2 test set, when applying some methods as follows:

- *PFC* tends to improve I-rec@10.
- *CT* improves I-rec@10 when *BRCC* is turned off.

As *PFC* candidates includes many recent, rather major subtopics, they are somehow effective. When applying *IM2Q*, subtopics about recent events tend to descend in ranking. *CT* removes the duplicated intents by filtering subtopics similar to the higher ranked ones. By combining *PFC*, *CS* and *CT*, the SM36 run achieved the best I-rec@10 although the difference is not statistically significant against SM00.

Through all official measures, our best performing result is below the median of all best runs of each participating team. This is partly due to our insufficient preparation, inefficient operations and inappropriate planning, but also it suggests the limitations of purely search log based techniques to the query understanding task, although we did not carry out enough comparative work to reach any conclusive observations.

4.3.2 Vertical Identification

Table 5 shows the evaluation results of the Vertical Identification phase, where the calibration of the threshold parameter largely affects the effectiveness. The results obtained by applying N iterations where the parameters are converged, achieve the best effectiveness. This strongly suggests that the GBDT learners should be calibrated given the data distribution being largely uneven. In both INTENT-2 and IMINE-2 tasks, ”Web” and ”Shopping” intents are much popular than other intents.

In official Vertical Identification evaluation, alleged irrelevant subtopics are assigned no vertical labels, therefore the results of Vertical Identification are subject to the effectiveness of the preceding Subtopic Mining phase. It makes difficult for us to compare our Vertical Identification effectiveness directly with other participating teams.

[2]SM09 is identical to SM01 due to the implementation restriction since BRCC does not work at this combination.

Table 5: V-scores of Official Runs

Run Name	SM Run#	# of iter	V-score
YJST-Q-J-1Q	SM01	N	**0.5336**
YJST-Q-J-2Q	SM01	≈N/2	0.4869
YJST-Q-J-3Q	SM40	≈0	0.3318
YJST-Q-J-4Q	SM40	≈N/2	0.4275
YJST-Q-J-5Q	SM40	N	**0.4831**

5. CONCLUSIONS

We participated in the IMine-2 Query Understanding subtask, where we tacked the two technical issues namely, subtopic mining, and vertical identification. For the subtopic mining, we intended to improve our base system in terms of coverage, click maximization and diversity of subtopic. For the vertical identification, we learned a classifier for each vertical using search query logs, and identified the relevant verticals for each subtopic. The experimental results of the subtopic mining showed that our extensions to diversify subtopics have a measurable improvement in terms of I-rec@10.

6. REFERENCES

[1] P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. From ”dango” to ”japanese cakes”: Query reformulation models and patterns. In *WI-IAT ’09*, pages 183–190, 2009.

[2] J. H. Friedman. Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38:367–378, 1999.

[3] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[4] S. Fujita, G. Dupret, and R. A. Baeza-Yates. Semantics of query rewriting patterns in search logs. In *ESAIR*, pages 7–8, 2012.

[5] S. Fujita, K. Machinaga, and G. Dupret. Click-graph modeling for facet attribute estimation of web search queries. In *Proceedings of 9th international conference on Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO 2010), Paris, France*, Apr. 2010.

[6] S. Fujita and Y. Ozawa. Specifications of Related Queries Extracted from Yahoo Search Logs. included in data deliverables.

[7] Liblinear. `https://www.csie.ntu.edu.tw/~cjlin/liblinear/`.

[8] Ntcireval. `http://research.nii.ac.jp/ntcir/tools/ntcireval-en.html`.

[9] C. pei Lee and C.-J. Lin. Large-scale Linear RankSVM. *Neural Computation*, 26(4):781–817, 2014. `http://www.csie.ntu.edu.tw/~cjlin/papers/ranksvm/ranksvml2.pdf`.

[10] T. Sakai, Z. Dou, T. Yamamoto, Y. Liu, M. Zhang, and R. Song. Overview of the ntcir-10 intent-2 task, 2013.

[11] T. Sakai and R. Song. Evaluating Diversified Search Results Using Per-intent Graded Relevance. In *ACM SIGIR 2011*, pages 1043–1052, 2011.

[12] T. Yamamoto, Y. Liu, M. Zhang, Z. Dou, K. Zhou, I. Markov, M. P. Kato, H. Ohshima, and S. Fujita. Overview of the ntcir-12 imine-2 task. In *Proceedings of NTCIR-12*, 2016.