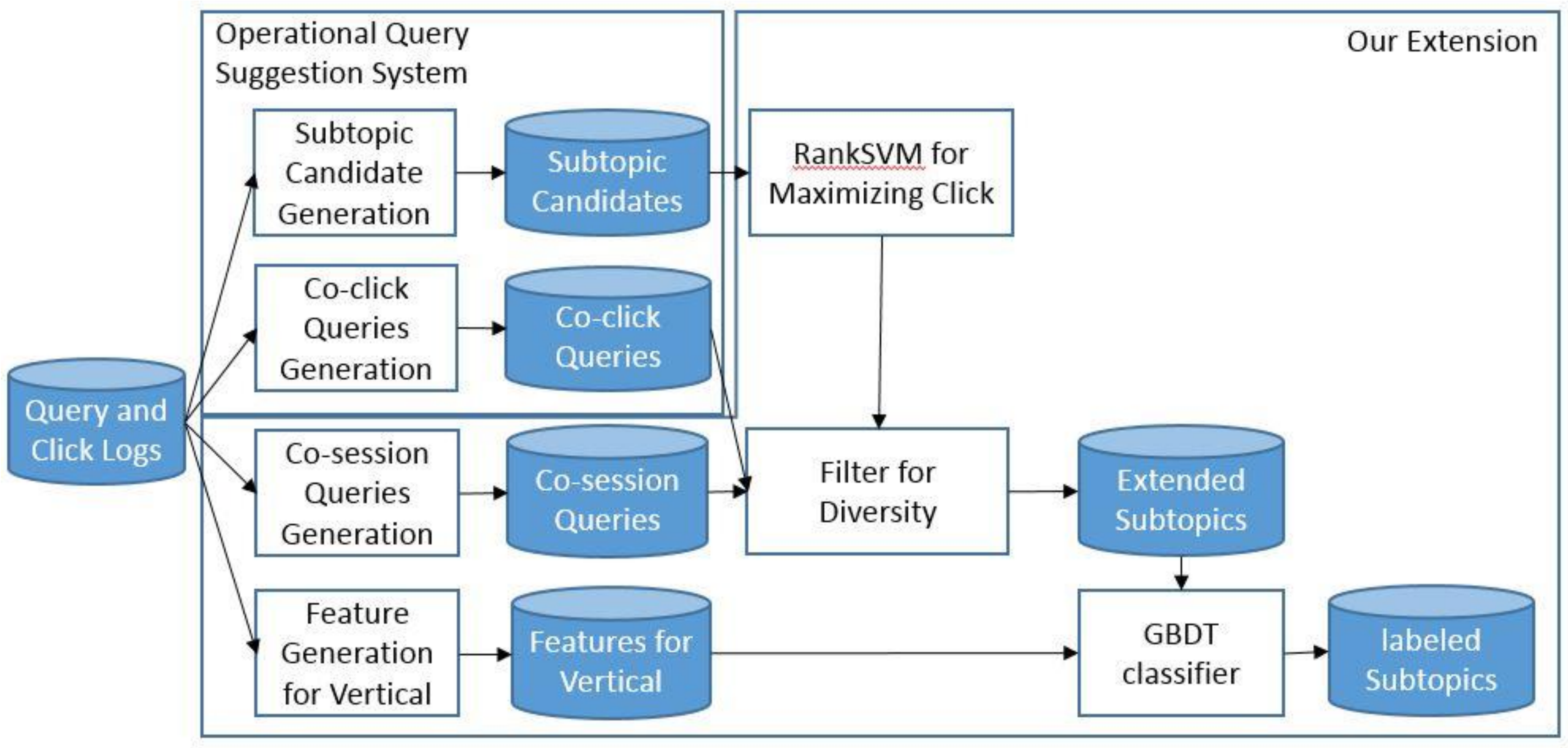
YJST at the NTCIR-12 IMine-2 Task

Yasuaki Yoshida, Hideyuki Maeda, Tatsuhiro Niwa, Sumio Fujita
(Yahoo Japan Corporation)
{yayoshid,hidmaeda,tniwa,sufujita}@yahoo-corp.jp

Basic Idea

◆To obtain diversified and ranked subtopic lists, we refined candidate subtopics from our base system by using co-click, co-session, and co-topic relations.

◆To identify vertical labels of ranked subtopics, we trained GBDT learner using several complex features.



Subtopic Mining

Method

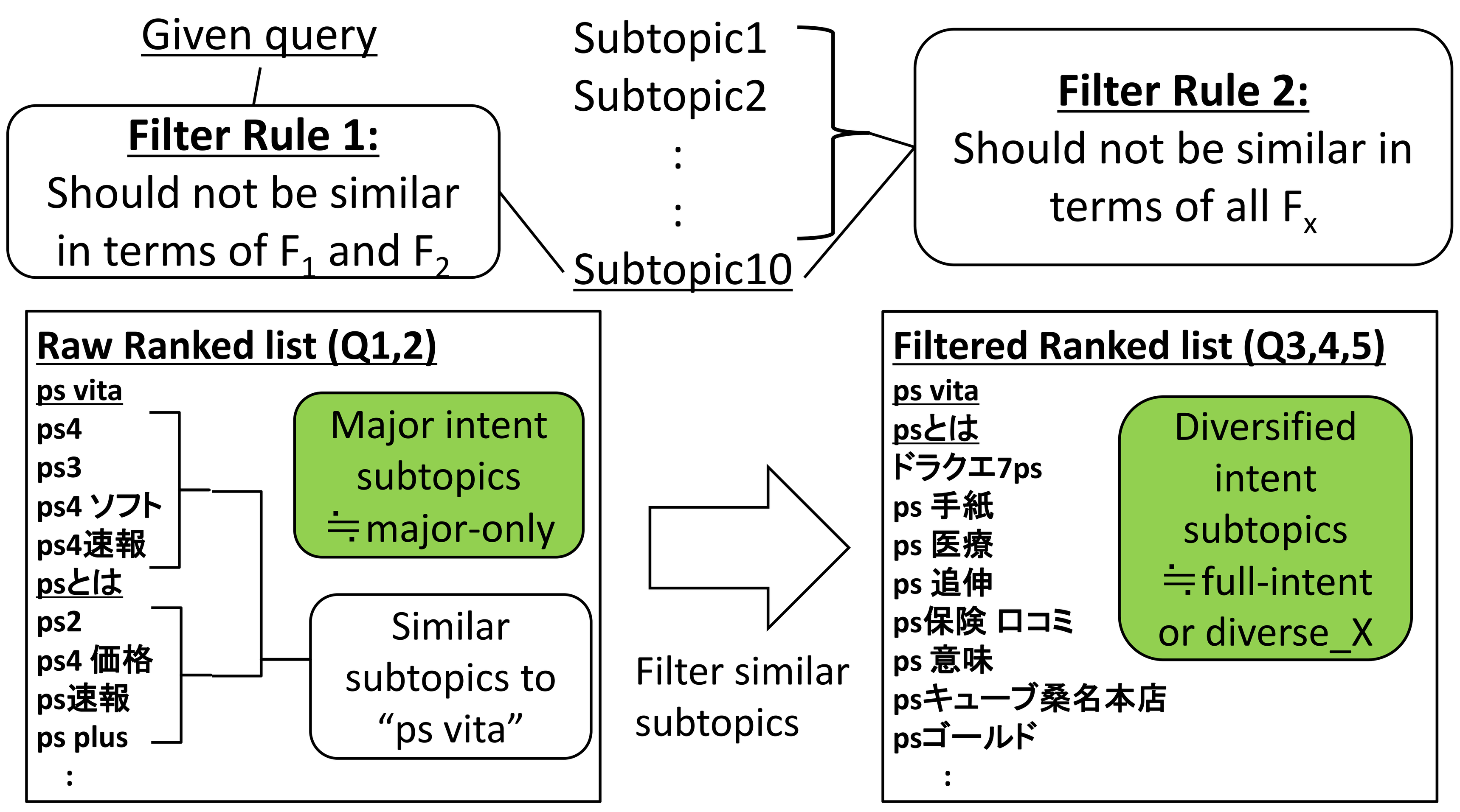
■ Subtopic filtering extensions

We use 3 relations described below to filter "similar" subtopics:

F_1. Co-Click Relations (from base system)

F_2. Co-Session Relations

F_3. Co-Topic Relations



■ Additional Extension

Blending Q1 and Q3 to create a run similar to "major-main" patterns. The major-main pattern always has highest score under these conditions. (※ experimentally confirmed, the definition of "major-main" and details of the experiments are shown in upper right.)

- if the # of "correct intents" is less than or equal to 10.
- if all of gains for correct subtopics are 1.
- if D#-nDCG = 0.5･D-nDCG + 0.5･I-rec

"Filter Rule 2" make Q3-5 similar to full-intent or diverse_X patterns, not to major-main_X patterns. So we create additional extension to imitate major-main_X form:

B_X : top *X* subtopics are the same as Q1 (**similar to major-only**), rest (*10–X*) subtopics are selected from Q3 (**similar to full-intent or diverse_X**).

Experimental Results (D#-nDCG)

■ Non-filter run was the best and applying filters harms D#-nDCG.

■ Additional extension runs, marked B_X , outperform our official runs significantly.

Run	Extension	D#-nDCG
Blend7	B_7 with Q1,Q3	0.5683
Blend8	B_8 with Q1,Q3	0.5682
Blend9	B_9 with Q1,Q3	0.5670
Blend5	B_5 with Q1,Q3	0.5663
Blend6	B_6 with Q1,Q3	0.5660

Run	Extension	D#-nDCG
Blend4	B_4 with Q1,Q3	0.5641
Q1(=Q2)	Nothing	0.5637
Blend3	B_3 with Q1,Q3	0.5599
Blend2	B_7 with Q1,Q3	0.5576
Q3(=Q4,5)	F_1,F_2,F_3	0.5554

※ D#-nDCG score simulation

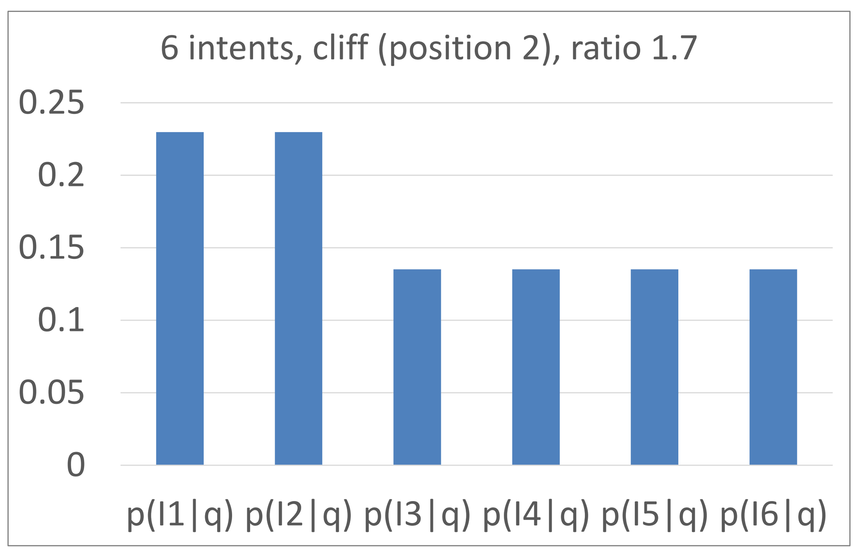
◆To reveal the characteristics of D#-nDCG, we examine how D#-nDCG with various Qrels behaves against the results by various strategies.

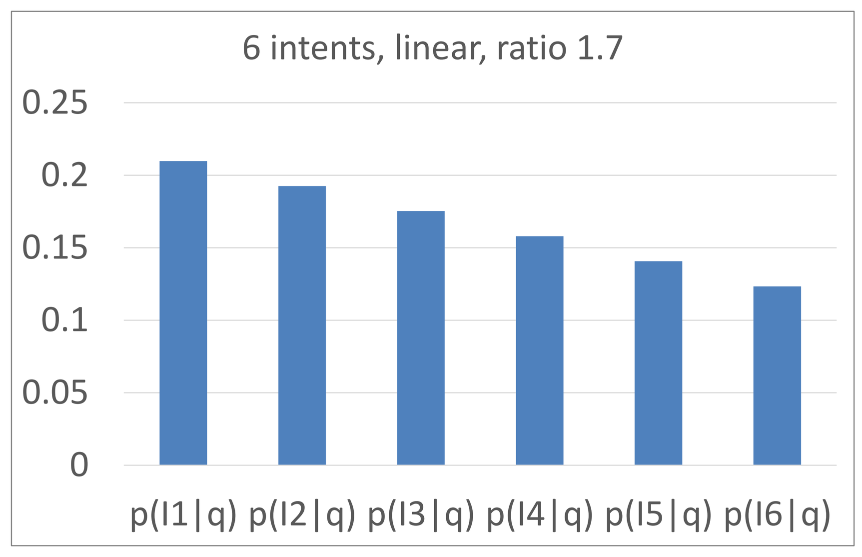
(Qrels include some intents for each given query as correct intents with probability p(i|q), and each intent includes some subtopics as correct data)

■ Generate Correct Qrels with various p(i|q) patterns

- The # of correct intents: 2, 3, 4, 5, 6, 7, 8, 9, 10
- The shape of wave when sorted by p(i|q) descent: linear, cliff (all p(i|q) are either minimum or maximum, we try all possible cliff positions)
- The ratio of maximum p(i|q) to minimum p(i|q): 1.0, 1.1, 1.2, … , 1.9, 2.0, 3.0, 4.0, … , 9.0, 10.0

EXAMPLES:

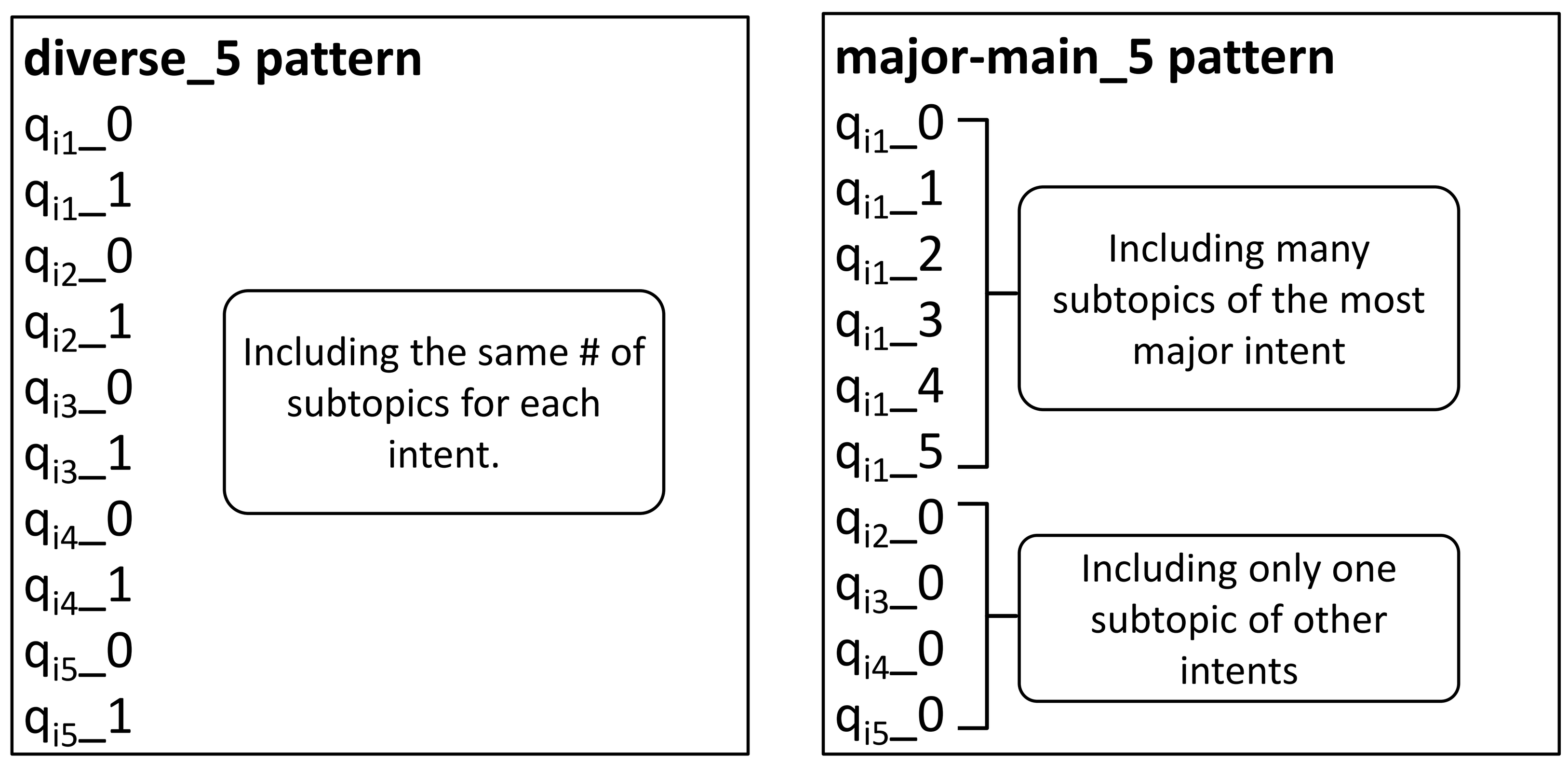




■ Generate Subtopics with different strategies

- major-only: Select 10 subtopics which include only most major intent.
- full-intent: Select 10 subtopics which include 10 intents.
- diverse_X: Select 10 subtopics which include X intents. Each intent has the same # of subtopics.
- major-main_X: Select 10 subtopics which include X intents. The most major intent has (11-X) subtopics and other intents have 1 subtopic.

EXAMPLES:



Experimental Result

■ major-main_X achieved the highest score with respect to the all thus generated Qrels, so major-main_X is the "ideal" pattern for D#-nDCG.

Vertical Identification

Method

■ GBDT (Multi-label)

- 3 kinds of features for each vertical
  1. Vertical Search Feature (VSF)
  2. Statistical Language Model Features (SLMS) (except Web)
  3. Random Walk Features (RWF)
- Total score is:

$$score(c_i) = GBDTscore(c_i) + b_i \quad b_i: threshold\ for\ each\ vertical$$

- b_i is also learned for each vertical
- Change the # of iterations to generate runs to submit

Experimental Result (V-score)

Run	Subtopic Mining Filter	# of iterations	V-score
Q1	Nothing	N (# to converge b_i)	0.5336
Q2	Nothing	N / 2	0.4869
Q3	F_1,F_2,F_3	0	0.3318
Q4	F_1,F_2,F_3	N / 2	0.4275
Q5	F_1,F_2,F_3	N	0.4831