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:
    - F1: Co-Click Relations (from base system)
    - F2: Co-Session Relations
    - F3: Co-Topic Relations

- Subtopic1: Should not be similar in terms of F1 and F2
- Subtopic2: Should not be similar in terms of all Fx

- 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^x - D-nDCG + 0.5^x-1-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:

Experimental Results (D#-nDCG)
- Non-filter run was the best and applying filters harms D#-nDCG.
- Additional extension runs, marked Bx, outperform our official runs significantly.

Experimental Result (V-score)

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:
    \( score(c_i) = GBDTscore(c_i) + b_i \) threshold for each vertical
    - \( b_i \) is also learned for each vertical
  - Change the # of iterations to generate runs to submit

Experimental Results (V-score)