ASURA: Best-Answer Estimation System for NTCIR-8 CQA Pilot Task
Daisuke Ishikawa
National Institute of Informatics

Outline of ASURA

ASURA-1 model:
- Simple best answer estimation system
- Based on only answers
- 5 features

ASURA-2 model:
- Enhance ASURA-1 model
- Based on questions and answers
- 13 features

Learning environment:
- Machine learning: SVM (TinySVM 0.09)
  - Solver Type: C-SVM (default)
  - Kernel: linear (default)
- Training data: 300 questions extracted from each category at random at the same rate as the test data
  - Questions: 300 items
  - Answers: 600 items (300 best answers, 300 normal answers)
- Test data: 1500 questions (official test collection)
  - Questions: 1500 items
  - Answers: 4113 items (1500 best answers, 2613 normal answers)
- Classification: Binary classification (Positive data is a best answer and negative data is a normal answer)

Computing environment:
- OS: CentOS 5.3 (x86, 64, 64-hi)
- CPU: Xeon 2.9 GHz Quad Core
- Memory: 16 GB
- Disk Array: 1 TB x 12 (RAID 6.0, 4 Gbps FC)

Official Results

In all evaluations, ASURA-2 performed better than ASURA-1.

In GA-nDCG and GA-Q, the performance of ASURA-2 was good.

In categories 5 (Internet) and 6 (sports), ASURA-2 performed better than BASELINE-2.

In categories 10 (news) and 11 (travel), BASELINE-2 performed better than ASURA-2.

For more detailed results, please refer to the text in the image.