

# 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: 7443 items (1500 best answers, 5943 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-bit)
- CPU: Xeon 2.0 GHz Quad Core
- Memory: 16 GB
- Disk Array: 1 TB x 12 (RAID 6.0, 4 Gbps FC)

Feature	ASURA-1: 5 features model	
1	Number of characters in answer.	Detailed
2	Number of appearances of 'desu' and 'masu'.	Polite
3	Existence of character string of 'http' or 'keiken' or 'taiken'. (exist = 1, not exist = 0)	Evidence
4	Average number of characters in answer group.	Detailed
5	Average number of appearances of 'desu' and 'masu' in answer group.	Polite

Feature	ASURA-2: 13 features model	
1 - 5	is the same as for ASURA-1.	
6	Kind number of keywords based on hiragana in answer.	Detailed
7	Kind number of keywords based on non-hiragana in answer.	Detailed
8	Kind number of keywords based on hiragana in question.	Detailed
9	Kind number of keywords based on non-hiragana in question.	Detailed
10	Number of characters in question.	Detailed
11	Number of agreements of keywords based on hiragana in question and answer.	Compatibility
12	Number of agreements of keywords based on non-hiragana in question and answer.	Compatibility
13	Category number (1-14) in question is acquired, 100 is added to category number, and number is set to 1 as feature. (For example, feature:value is 101:1 in case of category=yahoo, category number=1)	Other

## Official Results

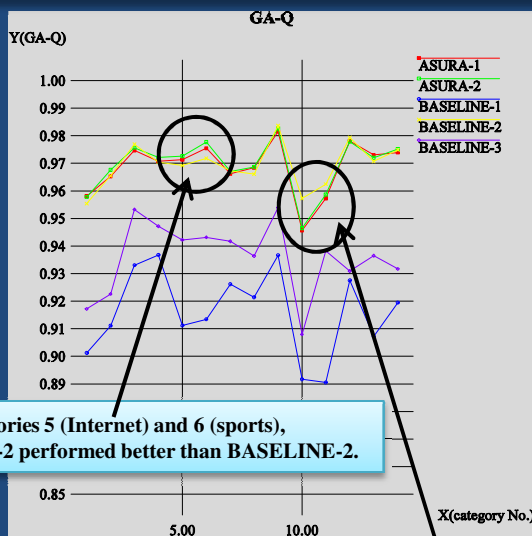
In all evaluations, ASURA-2 performed better than ASURA-1.

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

	BA-Hit@1	GA-Hit@1	GA-nG@1	GA-nDCG	GA-Q
ASURA-1 (5 features)	0.4813	0.9940	0.9140	0.9734	0.9680
ASURA-2 (13 features)	0.4840	<b>0.9953</b>	0.9166	<b>0.9742</b>	<b>0.9689</b>
BASELINE-1 (random)	0.2713	0.9920	0.7751	0.9311	0.9169
BASELINE-2 (answer length)	<b>0.4847</b>	<b>0.9953</b>	<b>0.9170</b>	0.9735	0.9680
BASELINE-3 (timestamp)	0.3820	0.9940	0.8213	0.9460	0.9359

In GA-Hit@1, the performance of ASURA-2 and BASELINE-2 was the same.

However, the performance of BASELINE-2 was better in BA-Hit@1 and GA-nG@1.



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.