ASURA: Best-Answer Estimation System for NTCIR-8 CQA Pilot Task

Daisuke Ishikawa
National Institute of Informatics
OUTLINE

1. Introduction
2. ASURA Overview
3. ASURA-1: 5-feature model
4. ASURA-2: 13-feature model
5. Official Results
6. Conclusions
1. Introduction

• An increase in activity surrounding social media research targeting community-type Q&A sites;
  – Yahoo! Chiebukuro (http://chiebukuro.yahoo.co.jp/)
  – Oshiete! Goo (http://oshiete.goo.ne.jp/)

• Want to find out if it is possible for a computer to select the Best Answer.

• Used ASURA, a best answer estimation system.
OUTLINE

1. Introduction
2. ASURA Overview
3. ASURA-1: 5-feature model
4. ASURA-2: 13-feature model
5. Official Results
6. Conclusions
2. ASURA Overview

• 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
2. ASURA Overview (con’t)

- **Learning environment:**
  - Machine learning: SVM (TinySVM 0.09)
  - Solver Type: C-SVM (default)
  - Kernel: linear (default)
  - Training data: 300 questions randomly extracted from each category at same rate as test data
    - Questions: 300
    - Answers: 600 (300 best answers, 300 normal answers)
  - Test data: 1500 questions (official test collection)
    - Answers: 7443 items (1500 best answers, 5943 normal answers)
  - Classification: Binary classification  (Positive: best answer, negative: normal answer.)
2. ASURA Overview (con’t)

- 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)
OUTLINE

1. Introduction
2. Outline of ASURA
3. ASURA-1: 5-feature model
4. ASURA-2: 13-feature model
5. Official Results
6. Conclusions
3. ASURA-1: 5-feature model

ASURA-1 is a simple best answer estimation system that consists of these factors.

– **Detailed**: Explanations were given in detail.
– **Evidence**: Answers with sources (URLs, etc.) included as evidence of information.
– **Polite**: Answers written in polite Japanese.
3. ASURA-1: 5-feature model (con’t)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Feature no.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed</td>
<td>1</td>
<td>Number of characters in answer.</td>
</tr>
<tr>
<td>Polite</td>
<td>2</td>
<td>Number of 'desu', です and 'masu', ます.</td>
</tr>
<tr>
<td>Evidence</td>
<td>3</td>
<td>Existence of character string of 'http' or 'keiken', 経験, or 'taiken', 体験. (exist = 1, not exist = 0)</td>
</tr>
<tr>
<td>Detailed</td>
<td>4</td>
<td>Average number of characters in answer group.</td>
</tr>
<tr>
<td>Polite</td>
<td>5</td>
<td>Average number of 'desu', です and 'masu', ます in answer group.</td>
</tr>
</tbody>
</table>
OUTLINE

1. Introduction
2. ASURA Overview
3. ASURA-1: five-feature model
4. ASURA-2: 13-feature model
5. Official Results
6. Conclusion
4. ASURA-2: 13-feature model

- ASURA-2 model
  - Based on questions and answers.
  - Calculates **Detailed** more strictly on the basis of the number of keywords.
  - Has added features based on the **compatibility** of question and answers.
  - Contains added category information for the features.
## 4. ASURA-2: 13 features model (con’t)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Feature no.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 - 5</td>
<td>The same as for ASURA-1.</td>
</tr>
<tr>
<td>Detailed</td>
<td>6</td>
<td>Number of keywords based on hiragana in answer.</td>
</tr>
<tr>
<td>Detailed</td>
<td>7</td>
<td>Number of keywords based on non-hiragana in answer.</td>
</tr>
<tr>
<td>Detailed</td>
<td>8</td>
<td>Number of keywords based on hiragana in question.</td>
</tr>
<tr>
<td>Detailed</td>
<td>9</td>
<td>Number of keywords based on non-hiragana in question.</td>
</tr>
<tr>
<td>Detailed</td>
<td>10</td>
<td>Number of characters in question.</td>
</tr>
<tr>
<td>Compatibility</td>
<td>11</td>
<td>Number of agreements of keywords based on hiragana in question and answer.</td>
</tr>
<tr>
<td>Compatibility</td>
<td>12</td>
<td>Number of agreements of keywords based on non-hiragana in question and answer.</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>Category numbers (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 for category=yahoo, category number=1)</td>
</tr>
</tbody>
</table>
OUTLINE

1. Introduction
2. ASURA Overview
3. ASURA-1: 5-feature model
4. ASURA-2: 13-feature model
5. Official Results
6. Conclusions
5. Official Results

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

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

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

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.
5. Official Results

GA-Q

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.
OUTLINE

1. Introduction
2. ASURA Overview
3. ASURA-1: 5-feature model
4. ASURA-2: 13-feature model
5. Official Results
6. Conclusions
6. Conclusions

- **ASURA model:**
  - ASURA-1 is a 5-feature model based on factors of 'best answers' selected by an individual.
  - ASURA-2 is a 13-feature model that has features based on the compatibility of questions and answers.

- **Official results:**
  - ASURA-2 Performed better than ASURA-1 in every case.
  - ASURA-2 Performed well in GA-nDCG and GA-Q.
  - In categories 5 (Internet) and 6 (sports), ASURA-2 performed well.

- **Future work:**
  - Analyze categories that did not perform well.
  - Verify effectiveness of each feature.