Overview of NTCIR-8 Community QA Pilot Task
(Part I): Test Collection and Task

Daisuke Ishikawa †, Tetsuya Sakai ‡ and Noriko Kando †
† National Institute of Informatics
‡ Microsoft Research Asia

with special thanks to
Yohei Seki and Kazuko Kuriyama
OUTLINE

1. Introduction
2. Yahoo! Chiebukuro Data
3. Specifications of CQA Pilot Task
4. Question Sampling
5. Assessment
6. Assessment results
1. Introduction

• 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/)
• Identify high-quality content on these sites (Agichtein 2008, etc).
• Propose a task in which a computer identifies best answers from Yahoo! Chiebukuro.
OUTLINE

1. Introduction
2. Yahoo! Chiebukuro Data
3. Specifications of CQA Pilot Task
4. Question Sampling
5. Assessment
6. Assessment results
2. Yahoo! Chiebukuro Data

- **Best Answer**:
  - (1) Most convincing and most satisfying answer as the “Best Answer.”
  - (2) The Best Answer can be selected by vote by other users.
- Only the Best Answers selected by the questioner (1) are recorded in the “Yahoo! Chiebukuro” data version 1.0.

### Details of Yahoo Chiebukuro data version 1.0:
- Data range: 2004/4/1 ~ 2005/10/31
- Questions resolved: 3116009 items (about 916 MB)
- Best answer: 3116008 items (about 935 MB)
- Other answer: 10361777 items (about 2.3 GB)
- Category: 285 categories
OUTLINE

1. Introduction
2. Yahoo! Chiebukuro Data
3. Specifications of CQA Pilot Task
4. Question Sampling
5. Assessment
6. Assessment results
3. Specifications of CQA Pilot Task

Rank all posted answers by answer quality (as estimated by system) for every question.

Yahoo Chiebukuro data version 1.0: 3 million questions

1500 questions selected at random

Test collection for CQA: 1500 questions

4 university students

Best Answer: Questioner selected

Test

Good Answers: 4 assessors individually assessed
OUTLINE

1. Introduction
2. Yahoo! Chiebukuro Data
3. Specifications of CQA Pilot Task
4. Question Sampling
5. Assessment
6. Assessment results
4. Question Sampling

1. **Target 15 categories on Yahoo! Chiebukuro website.**
2. Correspond 285 categories in Yahoo! Chiebukuro Data version 1.0 to top 15 categories.
3. Determine maximum number of answers to collect.
4. Determine number of distributions of each category based on ratio calculated from number of questions of each category to all questions.
5. Randomly extract 1500 questions from Yahoo! Chiebukuro data version 1.0 based on number of distributions.
6. Tag 1500 questions and distribute to participants.

1. yahoo
2. entertainment
3. health
4. lifeguide
5. internet
6. sports
7. love
8. education
9. school
10. news
11. travel
12. business
13. career
14. manners
15. computer
16. chat
17. others
4. Question Sampling

1. Target 15 categories on Yahoo! Chiebukuro website.
2. **Correspond 285 categories in Yahoo! Chiebukuro Data version 1.0 to top 15 categories.**
3. Determine maximum number of answers to collect.
4. Determine number of distributions of each category based on ratio calculated from number of questions of each category to all questions.
5. Randomly extract 1500 questions from Yahoo! Chiebukuro data version 1.0 based on number of distributions.
6. Tag 1500 questions and distribute to participants.
4. Question Sampling

1. Target 15 categories on Yahoo! Chiebukuro website.

2. Correspond 285 categories in Yahoo! Chiebukuro Data version 1.0 to top 15 categories.

3. **Determine maximum number of answers to collect.**

4. Determine number of distributions of each category based on ratio calculated from number of questions of each category to all questions.

5. Randomly extract 1500 questions from Yahoo! Chiebukuro data version 1.0 based on number of distributions.

6. Tag 1500 questions and distribute to participants.
4. Question Sampling

1. Target 15 categories on Yahoo! Chiebukuro website.

2. Correspond 285 categories in Yahoo! Chiebukuro Data version 1.0 to top 15 categories.

3. Determine maximum number of answers to collect.

4. **Determine number of distributions of each category based on ratio calculated from number of questions of each category to all questions.**

5. Randomly extract 1500 questions from Yahoo! Chiebukuro data version 1.0 based on number of distributions.

6. Tag 1500 questions and distribute to participants.

<table>
<thead>
<tr>
<th>Category No.</th>
<th>Top category (label)</th>
<th>All questions</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yahoo</td>
<td>368470</td>
<td>222</td>
</tr>
<tr>
<td>2</td>
<td>entertainment</td>
<td>278311</td>
<td>167</td>
</tr>
<tr>
<td>3</td>
<td>health</td>
<td>256246</td>
<td>154</td>
</tr>
<tr>
<td>4</td>
<td>lifeguide</td>
<td>226276</td>
<td>136</td>
</tr>
<tr>
<td>5</td>
<td>internet</td>
<td>223960</td>
<td>135</td>
</tr>
<tr>
<td>6</td>
<td>sports</td>
<td>199324</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>love</td>
<td>199937</td>
<td>120</td>
</tr>
<tr>
<td>8</td>
<td>education</td>
<td>198855</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>school</td>
<td>148258</td>
<td>89</td>
</tr>
<tr>
<td>10</td>
<td>news</td>
<td>104863</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>travel</td>
<td>96945</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>business</td>
<td>70701</td>
<td>42</td>
</tr>
<tr>
<td>13</td>
<td>career</td>
<td>63370</td>
<td>38</td>
</tr>
<tr>
<td>14</td>
<td>manners</td>
<td>59310</td>
<td>36</td>
</tr>
<tr>
<td>15</td>
<td>computer</td>
<td>222</td>
<td>0</td>
</tr>
</tbody>
</table>
4. Question Sampling

1. Target 15 categories on Yahoo! Chiebukuro website.

2. Correspond 285 categories in Yahoo! Chiebukuro Data version 1.0 to top 15 categories.

3. Determine maximum number of answers to collect.

4. Determine number of distributions of each category based on ratio calculated from number of questions of each category to all questions.

5. **Randomly extract 1500 questions from Yahoo! Chiebukuro data version 1.0 based on number of distributions.**

6. Tag 1500 questions and distribute to participants.

---

<table>
<thead>
<tr>
<th>Category No.</th>
<th>Top category (label)</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yahoo</td>
<td>222</td>
</tr>
<tr>
<td>2</td>
<td>entertainment</td>
<td>167</td>
</tr>
<tr>
<td>3</td>
<td>health</td>
<td>154</td>
</tr>
<tr>
<td>4</td>
<td>lifeguide</td>
<td>136</td>
</tr>
<tr>
<td>5</td>
<td>internet</td>
<td>135</td>
</tr>
<tr>
<td>6</td>
<td>sports</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>love</td>
<td>120</td>
</tr>
<tr>
<td>8</td>
<td>education</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>school</td>
<td>89</td>
</tr>
<tr>
<td>10</td>
<td>news</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>travel</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>business</td>
<td>42</td>
</tr>
<tr>
<td>13</td>
<td>career</td>
<td>38</td>
</tr>
<tr>
<td>14</td>
<td>manners</td>
<td>36</td>
</tr>
<tr>
<td>15</td>
<td>computer</td>
<td>0</td>
</tr>
</tbody>
</table>
4. Question Sampling

1. Target 15 categories on Yahoo! Chiebukuro website.
2. Correspond 285 categories in Yahoo! Chiebukuro Data version 1.0 to top 15 categories.
3. Determine maximum number of answers to collect.
4. Determine number of distributions of each category based on ratio calculated from number of questions of each category to all questions.
5. Randomly extract 1500 questions from Yahoo! Chiebukuro data version 1.0 based on number of distributions.
6. Tag 1500 questions and distribute to participants.

Test collection for CQA:
• Questions: 1500
• Answers: 7443
  ✓ Best answers: 1500
  ✓ Normal answers: 5943
• Data size: 4.0 MB (UTF-8 encode)
OUTLINE

1. Introduction
2. Yahoo! Chiebukuro Data
3. Specifications of CQA Pilot Task
4. Question Sampling
5. Assessment
6. Assessment results
5. Assessment
Assessors' Backgrounds

<table>
<thead>
<tr>
<th>Assessor ID</th>
<th>eval-1</th>
<th>eval-2</th>
<th>eval-3</th>
<th>eval-4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M / F</strong></td>
<td>male</td>
<td>male</td>
<td>female</td>
<td>female</td>
</tr>
<tr>
<td><strong>Ages</strong></td>
<td>Twenties</td>
<td>Twenties</td>
<td>Twenties</td>
<td>Thirties</td>
</tr>
<tr>
<td><strong>Undergraduate</strong></td>
<td>third-year</td>
<td>fourth-year</td>
<td>third-year</td>
<td>third-year</td>
</tr>
<tr>
<td><strong>Arts / Sciences</strong></td>
<td>Sciences</td>
<td>Arts</td>
<td>Sciences</td>
<td>Arts</td>
</tr>
<tr>
<td><strong>Major field of study</strong></td>
<td>Informatics (Simulation)</td>
<td>Literary thoughts</td>
<td>Biochemistry</td>
<td>Asian history</td>
</tr>
<tr>
<td><strong>Use frequency of Yahoo! Chiebukuro</strong></td>
<td>Once every two weeks</td>
<td>two or three times a week</td>
<td>two or three times a month</td>
<td>once a month</td>
</tr>
<tr>
<td><strong>Posting question / answer</strong></td>
<td>not used</td>
<td>not used</td>
<td>not used</td>
<td>not used</td>
</tr>
</tbody>
</table>
5. Assessment
CQA Judgment System Version 0.1

- Login management menu
- Category selection menu
- Question number selection menu
- Question assessment function
- Answer assessment function
- Log preservation function
- Download function of preservation logs
- Display function
- Short cut function
5. Assessment

Criteria

• Question criterion:
  – Grade A: question is a question.
  – Grade B: question is not actually a question

• Answer criterion:
  – Grade A: satisfactory answer to the question.
  – Grade B: partially relevant answer, or a partially irrelevant answer.
  – Grade C: answer unrelated to the question.
OUTLINE

1. Introduction
2. Yahoo! Chiebukuro Data
3. Specifications of CQA Pilot Task
4. Question Sampling
5. Assessment
6. Assessment results
6. Assessment results

Kappa Coefficient between assessors

<table>
<thead>
<tr>
<th></th>
<th>eval-1 eval-2</th>
<th>eval-3 eval-4</th>
<th>eval-1 eval-3</th>
<th>eval-2 eval-4</th>
<th>eval-1 eval-4</th>
<th>eval-2 eval-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement answer A</td>
<td>2072</td>
<td>2255</td>
<td>2473</td>
<td>1962</td>
<td>3574</td>
<td>1531</td>
</tr>
<tr>
<td>Agreement answer B</td>
<td>2082</td>
<td>2286</td>
<td>1938</td>
<td>2543</td>
<td>1616</td>
<td>3535</td>
</tr>
<tr>
<td>Agreement answer C</td>
<td>55</td>
<td>51</td>
<td>39</td>
<td>50</td>
<td>42</td>
<td>77</td>
</tr>
<tr>
<td>Po</td>
<td>0.565</td>
<td>0.616</td>
<td>0.597</td>
<td>0.611</td>
<td>0.702</td>
<td>0.690</td>
</tr>
<tr>
<td>Pc</td>
<td>0.420</td>
<td>0.453</td>
<td>0.435</td>
<td>0.449</td>
<td>0.512</td>
<td>0.510</td>
</tr>
<tr>
<td>κ</td>
<td><strong>0.249</strong></td>
<td><strong>0.298</strong></td>
<td><strong>0.287</strong></td>
<td><strong>0.295</strong></td>
<td><strong>0.390</strong></td>
<td><strong>0.368</strong></td>
</tr>
</tbody>
</table>

The lowest kappa coefficient: **0.249**

The highest kappa coefficient: **0.390**
6. Assessment results

Kappa Coefficients between Assessors for Each Category

- Difference in kappa coefficients arose due to difference among categories.
  - Approximately 0.30 ~ 0.45 for category no. 8 (education)
  - Approximately 0.03 ~ 0.33 for category no. 7 (love)
CONTINUE TO OVERVIEW PART II