CAO: A Fully Automatic Emoticon Analysis System

Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki

Graduate School of Information Science and Technology
Hokkaido University
Presentation Outline

• Emoticons - Definition
• Database Construction
• CAO – Emoticon Analysis System
• Evaluation of CAO
• Conclusions and Future Work
Emoticons - Definition

Our working definition of emoticons...
Emoticons - Definition

Our working definition of emoticons...
Emoticons - Definition

Emoticons:

– Emoticons are representations of body language in online communication (more-less).

Therefore...
Emoticons - Definition

Emoticons:
• Are an important in part of communication [1,2] in online communities (blogs, forums, BBS, e-mails, chat-rooms, etc.)

Emoticons - Definition

Emoticons:

• But sometimes are difficult to understand

( Hóa > Hóa) 7:^(~)

(。㍉ ิ‿㍉) (/~一一~)/...: * : ☆*

PC

what tha...
Emoticons - Definition

Emoticons:
- But sometimes are difficult to understand (☣ ฺゝ☣ ฺ)
- Need to analyze them effectively

what tha...

PC
Emoticons - Definition

Emoticons:

• Can be roughly divided into:
  – 1-line Western (text-base or pictures)
Emoticons - Definition

Emoticons:
• Can be roughly divided into:
  – 1-line Western (text-base or pictures)
Emoticons - Definition

Emoticons:

• Can be roughly divided into:
  – 1-line Western (text-base or pictures)
  – 1-line Eastern
Emoticons - Definition

Emoticons:
• Can be roughly divided into:
  – 1-line Western (text-base or pictures)
  – 1-line Eastern
Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western (text-base or pictures)
  - 1-line Eastern
  - Multiline Eastern
Emoticons - Definition

Emoticons:

- Can be roughly divided into:
  - 1-line Western
  - 1-line Eastern
  - Multiline Eastern
Emoticons - Definition

Emoticons:

• Can be roughly divided into:
  – 1-line Western (text-base or pictures)
  – 1-line Eastern
  – Multiline Eastern

We focused on these, because...
Emoticons - Definition

Emoticons:

• Can be roughly divided into:
  – 1-line Western
  – 1-line Eastern
  – Multiline Eastern

  ← There already is some research
  + we were a little more ambitious

  ← We are not that crazy
Emoticons - Definition

Emoticons:

• Can be roughly divided into:
  – 1-line Western
  – 1-line Eastern
  – Multiline Eastern

There already is some research + we were a little more ambitious

We are not that crazy

Only a little research done here
Emoticons - Definition

Some examples:
\((^\*^0^*^/)\)
·°·(ﾉД`;)·°·
(;;;;^_^A
(°.°)
(以上学历
(丿丿丿丿)
(丿丿丿丿)
(下下下)
(丿丿丿丿)
(丿丿丿丿)
Emoticons - Definition

Some examples:
\(\left(\ast^\circ\ast\right)\) / Suddenly came inspiration!
・゚・(ﾉД`;)・゚・
(;^_^A
(°. °)
(^_-)ゆ--~
(==┬==Д==┬==)
(・．・。)人(・．・。)
(＊___________
( －　)
Emoticons - Definition

Some examples:

\( (^*o^*) / \)

・・(ﾉД`;)・・・

(;^_^A

(°．．°)

(^_-)у--~

(==╦==Д==╦==)

(．・_・．)人(．・_・．)

(＊﹏ー)

Suddenly came inspiration!

Since emoticons are representations of body language...
Emoticons - Definition

Some examples:

\((^{*\_\_\_}\_\_\_})/

\(\cdot \cdot (/Д`;)\cdot\)

\( (;^\_\_\_А)

\(° . °)

\( (^\_\_\_y--~)

\( (==╦==Д==╦==)

\( ( . \cdot \cdot \cdot )人( . \cdot \cdot \cdot )

\( ( * \_\_\_) \)

Suddenly came inspiration!

A structural approach to body language could be applicable here as well!
Emoticons - Definition

• Theory of kinesics:
• Non-verbal behavior is used in everyday communication systematically and can be described structurally.
• A minimal part = a kineme, the smallest meaningful set of body movements, e.g. raising eyebrows, etc.

Birdwhistell (1952, 1970)

Emoticons - Definition

- **Theory of kinesics:**
- **Non-verbal behavior is used in everyday communication systematically and can be described structurally.**
- A minimal part = a kine = a minimal meaningful set of body movements, e.g., raising eyebrows, etc.

<table>
<thead>
<tr>
<th>Emoticon</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>🧐</td>
<td>Blank-faced</td>
</tr>
<tr>
<td>👀</td>
<td>Slitted eyes</td>
</tr>
<tr>
<td>🙁</td>
<td>Eyes upward</td>
</tr>
<tr>
<td>🙁</td>
<td>Shifty eyes</td>
</tr>
<tr>
<td>🙁</td>
<td>Glare</td>
</tr>
<tr>
<td>😊</td>
<td>Tongue in cheek</td>
</tr>
<tr>
<td>😕</td>
<td>Pout</td>
</tr>
<tr>
<td>😊</td>
<td>Clenched teeth</td>
</tr>
<tr>
<td>😊</td>
<td>Toothy smile</td>
</tr>
<tr>
<td>😊</td>
<td>Square smile</td>
</tr>
<tr>
<td>😊</td>
<td>Open mouth</td>
</tr>
<tr>
<td>😊</td>
<td>Slow lick—lips</td>
</tr>
<tr>
<td>😊</td>
<td>Quick lick—lips</td>
</tr>
<tr>
<td>😊</td>
<td>Moistening lips</td>
</tr>
<tr>
<td>😊</td>
<td>Lip biting</td>
</tr>
</tbody>
</table>


Emoticons - Definition

Some examples:
\( (*^o^* ) / \)
・・・(ﾉﾞ`؛)・・・
(;^_^A
(°.°)
(^_-)ュ--~
(===¬==ノ===)
(・.・.・)_人(・.・.・)
(*_*)
### Emoticons - Definition

**Some examples:**

\[\begin{align*}
\text{\(\text{*^o^*}\)/} & \quad \text{Blank-faced} \\
\cdot \cdot (\text{\(\text{/D}`\);}) \cdot \cdot & \quad \text{Slitted eyes} \\
(\text{;^_^A} & \quad \text{Single raised brow (\(\sim\) indicates brow raised)} \\
(\cdot \cdot \cdot & \quad \text{Eyes upward} \\
(\text{^_-)y--~} & \quad \text{Shifty eyes} \\
(\text{==_==Д==_==)} & \quad \text{Lowered brow} \\
(\text{. \cdot _· . \text{人（. \cdot _· .）}} & \quad \text{Glare} \\
(\text{* ̄ ̄ ̄ー ̄ ̄ ̄)} & \quad \text{Tongue in cheek} \\
\end{align*}\]

<table>
<thead>
<tr>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{<em>^o^</em>})/</td>
<td>Blank-faced</td>
</tr>
<tr>
<td>(\sim)</td>
<td>Slitted eyes</td>
</tr>
<tr>
<td>(\text{/D}`);)</td>
<td>Eyes upward</td>
</tr>
<tr>
<td>(\text{;^_^A})</td>
<td>Single raised brow ((\sim) indicates brow raised)</td>
</tr>
<tr>
<td>(\cdot \cdot \cdot)</td>
<td>Shifty eyes</td>
</tr>
<tr>
<td>(\text{^_-)y--~})</td>
<td>Lowered brow</td>
</tr>
<tr>
<td>(\text{==<em>==Д==</em>==)})</td>
<td>Glare</td>
</tr>
<tr>
<td>(\text{. \cdot _· . \text{人（. \cdot _· .）}})</td>
<td>Tongue in cheek</td>
</tr>
<tr>
<td>(\text{* ̄ ̄ ̄ー ̄ ̄ ̄)})</td>
<td>Pout</td>
</tr>
<tr>
<td>(\text{^_-)y--~})</td>
<td>Medial brow contraction</td>
</tr>
<tr>
<td>(\text{==<em>==Д==</em>==)})</td>
<td>Medial brow contraction</td>
</tr>
<tr>
<td>(\text{. \cdot _· . \text{人（. \cdot _· .）}})</td>
<td>Medial brow contraction</td>
</tr>
<tr>
<td>(\text{* ̄ ̄ ̄ー ̄ ̄ ̄)})</td>
<td>Medial brow contraction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{^_-)y--~})</td>
<td>Raised brows</td>
</tr>
<tr>
<td>(\text{==<em>==Д==</em>==)})</td>
<td>Toothy smile</td>
</tr>
<tr>
<td>(\text{. \cdot _· . \text{人（. \cdot _· .）}})</td>
<td>Wide eyed</td>
</tr>
<tr>
<td>(\text{* ̄ ̄ ̄ー ̄ ̄ ̄)})</td>
<td>Square smile</td>
</tr>
<tr>
<td>(\text{^_-)y--~})</td>
<td>Wink</td>
</tr>
<tr>
<td>(\text{==<em>==Д==</em>==)})</td>
<td>Open mouth</td>
</tr>
<tr>
<td>(\text{. \cdot _· . \text{人（. \cdot _· .）}})</td>
<td>Sideway look</td>
</tr>
<tr>
<td>(\text{* ̄ ̄ ̄ー ̄ ̄ ̄)})</td>
<td>Slow lick—lips</td>
</tr>
<tr>
<td>(\text{^_-)y--~})</td>
<td>Focus on auditor</td>
</tr>
<tr>
<td>(\text{==<em>==Д==</em>==)})</td>
<td>Quick lick—lips</td>
</tr>
<tr>
<td>(\text{. \cdot _· . \text{人（. \cdot _· .）}})</td>
<td>Stare</td>
</tr>
<tr>
<td>(\text{* ̄ ̄ ̄ー ̄ ̄ ̄)})</td>
<td>Moistening lips</td>
</tr>
<tr>
<td>(\text{^_-)y--~})</td>
<td>Rolled eyes</td>
</tr>
<tr>
<td>(\text{==<em>==Д==</em>==)})</td>
<td>Lip biting</td>
</tr>
</tbody>
</table>
Emoticons - Definition

Some examples:

\( (*^0^*)/ \)
・・・(／Д＼)・・
(;^_^A
(°・°)
(^_-)y--~
(==¼==Д==¼==)
(・_・.)人(・_・.)
( * __________ )

<table>
<thead>
<tr>
<th>Emoticon</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(^o^</em>)</td>
<td>Blank-faced</td>
</tr>
<tr>
<td>( 。・。)</td>
<td>Single raised brow (¬ indicates brow raised)</td>
</tr>
<tr>
<td>(°。°)</td>
<td>Lowered brow</td>
</tr>
<tr>
<td>（^_^）</td>
<td>Wink</td>
</tr>
<tr>
<td>(°°)</td>
<td>Sidewise look</td>
</tr>
<tr>
<td>(°。°)</td>
<td>Focus on auditor</td>
</tr>
<tr>
<td>（°。°）</td>
<td>Stare</td>
</tr>
<tr>
<td>（°。°）</td>
<td>Rolled eyes</td>
</tr>
<tr>
<td>((φ°))</td>
<td>Slitted eyes</td>
</tr>
<tr>
<td>(°°)</td>
<td>Eyes upward</td>
</tr>
<tr>
<td>(°°°°)</td>
<td>Shifty eyes</td>
</tr>
<tr>
<td>(°”°”°”°”)</td>
<td>Glare</td>
</tr>
<tr>
<td>( の )</td>
<td>Tongue in cheek</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Pout</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Clenched teeth</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Toothy smile</td>
</tr>
<tr>
<td>( 。・。)</td>
<td>Square smile</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Open mouth</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Slow lick—lips</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Quick lick—lips</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Moistening lips</td>
</tr>
<tr>
<td>(。・。)</td>
<td>Lip biting</td>
</tr>
</tbody>
</table>
Emoticons - Definition

\( \left( {^o^\smile} \right) \)

- Additional area: \( \backslash \)
- Bracket: ()
- Additional area: *
- Face: \( {^o^\smile} \) Eyes: ^ ^
- Additional area: *
- Bracket: )
- Additional area: /

Assumption: Emoticons could be analyzed by dividing them to areas (kinemes)!
Database Construction

- Visited 7 online emoticon dictionaries:
  5. *Kaomoji Paradise*, 6. *Kaomojisyoyo* and
  7. *Kaomoji Station.*
Database Construction

• Used an affect analysis system to select and categorize only emotion-related labels.

Emoticon Dictionary

Emoticon labels:
- Hobbies
- Greetings
- Angry
- Happy
- Running
- Screaming
- Joyful
- Crying
- Working
- Working...

Database Construction

• Used an affect analysis system to select and categorize only emotion-related labels.

• Extract emoticons only from labels related to emotions

\((^*^O^*_^*)\)
Database Construction

- Obtained 10,137 unique emoticons classified with emotion types.

<table>
<thead>
<tr>
<th></th>
<th>joy, delight</th>
<th>liking, fondness</th>
<th>anger</th>
<th>surprise, amazement</th>
<th>sadness, gloom</th>
<th>excitement</th>
<th>dislike</th>
<th>shame, shyness</th>
<th>fear</th>
<th>relief</th>
<th>Overall</th>
<th>Emoticons</th>
</tr>
</thead>
<tbody>
<tr>
<td>3128</td>
<td>1988</td>
<td>1238</td>
<td>1227</td>
<td>1203</td>
<td>1124</td>
<td>704</td>
<td>526</td>
<td>179</td>
<td>99</td>
<td>11416</td>
<td>All extracted</td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td>1972</td>
<td>1221</td>
<td>1196</td>
<td>1169</td>
<td>1120</td>
<td>698</td>
<td>511</td>
<td>179</td>
<td>99</td>
<td>10137</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>63%</td>
<td>99%</td>
<td>99%</td>
<td>97%</td>
<td>97%</td>
<td>99%</td>
<td>99%</td>
<td>97%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>Ratio</td>
<td></td>
</tr>
</tbody>
</table>

\((^{*^O^*})/\)
Database Construction

- Automatically divide emoticons into:
  - Eyes [E]: ^ ^
  - Mouths [M]: o
  - Additional areas (inside emoticon) [S]: * *
  - Additional areas (outside emoticon) [S]: \/

\((^o^*)/)
Database Construction

• We have a set of databases!
  – Raw emoticons
  – Triplets (E-M-E)
  – Eyes (E-E)
  – Mouths (M)
  – Additional (S)

\(\left\langle{\left(^\wedge o^\wedge\right)}\right\rangle\)

<table>
<thead>
<tr>
<th>Areas</th>
<th>$E_1 ME_R$</th>
<th>$S_1$</th>
<th>$B_1$</th>
<th>$S_2$</th>
<th>$E_1 E_R$</th>
<th>$M$</th>
<th>$S_3$</th>
<th>$B_2$</th>
<th>$S_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy, delight</td>
<td>1298</td>
<td>1469</td>
<td>–</td>
<td>653</td>
<td>349</td>
<td>336</td>
<td>671</td>
<td>–</td>
<td>2449</td>
</tr>
<tr>
<td>anger</td>
<td>741</td>
<td>525</td>
<td>–</td>
<td>321</td>
<td>188</td>
<td>239</td>
<td>330</td>
<td>–</td>
<td>1014</td>
</tr>
<tr>
<td>sadness</td>
<td>702</td>
<td>350</td>
<td>–</td>
<td>303</td>
<td>291</td>
<td>170</td>
<td>358</td>
<td>–</td>
<td>730</td>
</tr>
<tr>
<td>fear</td>
<td>124</td>
<td>72</td>
<td>–</td>
<td>67</td>
<td>52</td>
<td>62</td>
<td>74</td>
<td>–</td>
<td>133</td>
</tr>
<tr>
<td>shame, shyness</td>
<td>315</td>
<td>169</td>
<td>–</td>
<td>121</td>
<td>110</td>
<td>85</td>
<td>123</td>
<td>–</td>
<td>343</td>
</tr>
<tr>
<td>liking, fondness</td>
<td>1079</td>
<td>1092</td>
<td>–</td>
<td>802</td>
<td>305</td>
<td>239</td>
<td>805</td>
<td>–</td>
<td>1633</td>
</tr>
<tr>
<td>dislike</td>
<td>527</td>
<td>337</td>
<td>–</td>
<td>209</td>
<td>161</td>
<td>179</td>
<td>201</td>
<td>–</td>
<td>562</td>
</tr>
<tr>
<td>excitement</td>
<td>670</td>
<td>700</td>
<td>–</td>
<td>268</td>
<td>243</td>
<td>164</td>
<td>324</td>
<td>–</td>
<td>1049</td>
</tr>
<tr>
<td>relief</td>
<td>81</td>
<td>50</td>
<td>–</td>
<td>11</td>
<td>38</td>
<td>26</td>
<td>27</td>
<td>–</td>
<td>64</td>
</tr>
<tr>
<td>surprise, amazement</td>
<td>648</td>
<td>405</td>
<td>–</td>
<td>231</td>
<td>183</td>
<td>154</td>
<td>279</td>
<td>–</td>
<td>860</td>
</tr>
<tr>
<td>overall</td>
<td>6185</td>
<td>5169</td>
<td>–</td>
<td>2986</td>
<td>1920</td>
<td>1654</td>
<td>3192</td>
<td>–</td>
<td>8837</td>
</tr>
</tbody>
</table>
CAO – Emoticon Analysis System

• Constructed CAO system for emoticon analysis with these databases.
CAO – Emoticon Analysis System

- Emoticon detection in (any) input
  - Use 455 characters most frequently (>10 times) appearing in emoticons ($x_1, x_2, \ldots x_{455}$)
  - If (any three $x$ appear in a row) {
    there is an emoticon in input
  }
CAO – Emoticon Analysis System

• Emoticon extraction from input (+ affect analysis)
  – Three steps:
    1. Looking for a “raw” emoticon (+checking emotion labels)
CAO – Emoticon Analysis System

• Emoticon extraction from input (+ affect analysis)
  – Three steps:
    1. Looking for a “raw” emoticon (+checking emotion labels)
    2. Looking for a triplet (+checking emotion labels)
CAO – Emoticon Analysis System

• Emoticon extraction from input (+ affect analysis)
  – Three steps:
    1. Looking for a “raw” emoticon (+checking emotion labels)
    2. Looking for a triplet (+checking emotion labels)
    3. Checking all combinations of triplets (eyes x mouth*)
      (+checking emotion labels)

*)Eyes=1,920
Mouths=1,654
All combinations: ExM=3,175,680
CAO – Emoticon Analysis System

- Emoticon extraction from input
  - Finally:
    - Extract additional areas (+checking emotion labels)
CAO – Emoticon Analysis System

- Emoticon extraction from input
  - Finally:
    - Extract additional areas (+checking emotion labels)
    - Summarize scores (to determine emotion types statistically most probable for this emoticon)
Evaluation of CAO

• Test set
  – A large corpus of blogs from: Ameba Blog*
  – 354,288,529 Japanese sentences in
  – 12,938,606 downloaded and parsed web pages
  – written by 60,658 unique bloggers

*) www.ameblo.co.jp
Evaluation of CAO

- Randomly extracted 1000 middle-sized* sentences as the test set
  - 418 of those sentences included emoticons.
    - annotate the sentences with 42 people (10 sentences per 1 person)
      *Question: What emotion was expressed in the sentence?
    - annotate emoticons from the sentences (different samples than in sentences)
      *Question: What emotion could be expressed with this emoticon?
    - Answers (emotion type, random order): a) System’s;
      b) Similar**; c) Completely different; d) Other (from the seven remaining);

*) 20-50 characters in Japanese
**) From the same affect space in two-dimensional model of affect
Evaluation of CAO

• Estimation of:
  – Emotion types (10 types)
  – General emotive features (valence and activation)* adjusted to Japanese like in Ptaszynski et al.**

## Evaluation of CAO

### Results

<table>
<thead>
<tr>
<th>Users</th>
<th>Emoticon</th>
<th>No emoticon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emoticon</td>
<td>394</td>
<td>24</td>
</tr>
<tr>
<td>No emoticon</td>
<td>0</td>
<td>582</td>
</tr>
</tbody>
</table>

No. of agreements = 976 (97.6%), Kappa = 0.95

In 24/418 cases there were no 3 usual chars in a row.
Evaluation of CAO

- Results

<table>
<thead>
<tr>
<th>Extraction</th>
<th>R</th>
<th>P</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94.3%</td>
<td>100%</td>
<td><strong>97.1%</strong></td>
</tr>
</tbody>
</table>

\[
\text{F-score} = 2 \frac{P \times R}{P + R}
\]

Errors only for the undetected emoticons
Evaluation of CAO

• Results
• Emotion Estimation on Separate Emoticons
  – Emotion types: 93.54%
  – General emotive features: 97.39%

Accuracy in determining probable emotion types a certain emoticon could be used to express
Evaluation of CAO

• Results
• Emotion Estimation on Sentences
  – Emotion types: 80.2%
  – General emotive features: 94.63%

Accuracy in determining emotion types expressed in a sentence, only with the use of emoticon*

*) a sentence needs to contain at least one emoticon
Evaluation of CAO

• Results

• Emotion Estimation on Sentences
  – Emotion types: 80.2%
  – General emotive features: 94.63%

1. The results were worse because meaning in sentences is conveyed also through lexical channel; but,
2. Results for general features were high → People sometimes misinterpret specific emotion type, but rarely valence/activation;

*) a sentence needs to contain at least one emoticon
Conclusions

• Presented a prototype system for automatic affect analysis of Eastern type emoticons, CAO.

• Inspired by Theory of Kinesics

• Gathered database of +10,000 emoticons and (almost) automatically expanded it to +3 mln.
Conclusions

• CAO is capable of:
  – Detecting emoticons in any input
  – Extracting emoticons form input
  – Dividing emoticons into semantic areas (eyes, mouths, etc.)
  – Estimating potential emotion types expressed by emoticons.
  – Affect analysis of sentences including emoticons
• CAO got almost ideal results in all tasks.
Future Work

Possible applications:

• Affect analysis/annotation of corpora

• Emotion detecting in online communication
  – Support for Internet messengers, blog services, forums, etc.

• Sentiment analysis (when looking only at valence)

• Detecting irony*

*) Carvalho, P., Sarmento, L., Silva, M. J., and de Oliveira, E. 2009. Clues for detecting irony in user-generated contents: oh..!! it's "so easy" ;-. In Proceeding of the 1st international CIKM Workshop on Topic-Sentiment Analysis For Mass Opinion (Hong Kong, China, November 06 - 06, 2009)
Thank you for your attention!

Read more in: “A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics”
http://www.computer.org/portal/web/tac
Details on Extraction of Emoticon Areas

In: Database Construction
Database Construction

- Determined all possible emoticon borders:

  2bit chars: (, 【, [, ,, 〈, 〈, }; 1-bit chars: (, [, ,<, |; [none]

  2bit chars: ), ]), ], |>; 1-bit chars: ), ], |;

\(\text{\(\text{(*^O^*)/}\)}\)
Database Construction

• Extract eye-mouth-eye triplets
  – Get rid of what is behind brackets (inclusively with brackets)
  – Get rid of additional areas from within emoticons (the only detail done manually)
• Make a database of emoticon triplets

\( (^*^O^*^) / \)
Database Construction

• Extract eyes and mouths
  • If an eye has more than 1 character, both eyes are the same;
    if (n characters from left and right match) {n=eye};
    if else (take n-1, n-2, n-3, ...)

\(\text{(*^O^-*)}\)
Database Construction

- Extract eyes and mouths
  - If an eye has 1 character, eyes could be the same or different;
  - else(
    take 1 char. from left and right as eyes;
    mouth is what is left inside;
  )

\(\left(\ast ^{O}\ast \right)\)
Database Construction

- Extract additional areas
  - Localize and extract additional areas
  - Make database of additional areas

\(\left(\star^{0}\star\right)\)
Database Construction

- We have a database!

\((^*^o^*^*)/\)
Details on Affect Analysis of Emoticons

In: CAO – Emoticon Analysis System
CAO – Emoticon Analysis System

- Emoticon affect analysis (along with extraction)
  - Emotion list extraction
    - For [1.]: Check emotion types annotated on raw emoticons
CAO – Emoticon Analysis System

- Emoticon affect analysis
  - Emotion list extraction
    - For [1.]: Check emotion types annotated on raw emoticons
    - For [2.]: Check emotion types annotated on triplets
CAO – Emoticon Analysis System

- Emoticon affect analysis
  - Emotion list extraction
    - For [1.]: Check emotion types annotated on raw emoticons
    - For [2.]: Check emotion types annotated on triplets
    - For [3.]: Check emotion types annotated on separate ExM combinations
CAO – Emoticon Analysis System

• Emoticon affect analysis
  – Finally
    • Check emotion types annotated on additional areas
CAO – Emoticon Analysis System

• Emoticon affect analysis
  – Finally
    • Check emotion types annotated on additional areas
    • Summarize score
    • Give output: list of emotions most probably expressed with this emoticon
Summarizing scores

In: Evaluation of CAO
Evaluation of CAO

• Summarizing scores
  – Occurrence
    • Sum of all emotion types found for all elements
  – Frequency
    • Sum for each element divided by number of all elements in each database
  – Unique frequency
    • Sum for each element divided by number of unique elements in each database
Evaluation of CAO

Detailed Description
Evaluation of CAO

• Training Set*
  – Raw emoticon database (Tr.S. gold standard)
    • Take emoticon from a database (e.g. from “joy”)
    • Process
    • Check result with gold standard

*) In training set evaluation we matched only triplets and all possible; matching also raw would give all 100%
Evaluation of CAO

• Training Set*
  – Raw emoticon database (Tr.S. gold standard)
• Take emoticon from a
• Process
• Check result with gold standard
• Ranking:
  1. Occurrence
  2. Unique Frequency
  3. Frequency
  (differences not significant=all equally good)

<table>
<thead>
<tr>
<th>Emotion type</th>
<th>CAO: Occurrence</th>
<th>Frequency</th>
<th>Unique Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>anger</td>
<td>0.811</td>
<td>0.771</td>
<td>0.767</td>
</tr>
<tr>
<td>dislike</td>
<td>0.631</td>
<td>0.800</td>
<td>0.719</td>
</tr>
<tr>
<td>excitement</td>
<td>0.786</td>
<td>0.769</td>
<td>0.797</td>
</tr>
<tr>
<td>fear</td>
<td>0.451</td>
<td>0.936</td>
<td>0.858</td>
</tr>
<tr>
<td>fondness</td>
<td>0.915</td>
<td>0.778</td>
<td>0.783</td>
</tr>
<tr>
<td>joy</td>
<td>0.944</td>
<td>0.802</td>
<td>0.860</td>
</tr>
<tr>
<td>relief</td>
<td>0.600</td>
<td>0.990</td>
<td>0.985</td>
</tr>
<tr>
<td>shame</td>
<td>0.706</td>
<td>0.922</td>
<td>0.910</td>
</tr>
<tr>
<td>sorrow</td>
<td>0.814</td>
<td>0.809</td>
<td>0.791</td>
</tr>
<tr>
<td>surprise</td>
<td>0.862</td>
<td>0.866</td>
<td>0.874</td>
</tr>
<tr>
<td>All approx.</td>
<td>0.852</td>
<td>0.804</td>
<td>0.818</td>
</tr>
</tbody>
</table>
Evaluation of CAO

• Test set
  – A large corpus of blogs from: Ameba Blog*
  – 354,288,529 Japanese sentences in
  – 12,938,606 downloaded and parsed web pages
  – written by 60,658 unique bloggers
• Randomly extracted 1000 middle-sized** sentences as the test set
  – 418 of those sentences included emoticons.
  • annotate the sentences with 42 people (10 sentences per 1 person)
    Question: What emotion was expressed in the sentence?
  • annotate emoticons from the sentences (different samples than in sentences)
    Question: What emotion could be expressed with this emoticon?

*) www.ameblo.co.jp
**)20-50 characters in Japanese
Evaluation of CAO

• Estimation of:
  – Emotion types (specific)
  – General emotive features (valence and activation)* adjusted to Japanese like in Ptaszynski et al.**


Evaluation of CAO

• Results

<table>
<thead>
<tr>
<th>Detection</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emoticon</td>
<td>No emoticon</td>
</tr>
<tr>
<td>Users</td>
<td></td>
</tr>
<tr>
<td>Emoticon</td>
<td>394</td>
</tr>
<tr>
<td>No emoticon</td>
<td>0</td>
</tr>
<tr>
<td>No. of agreements=976 (97.6%), Kappa=0.95</td>
<td>24</td>
</tr>
<tr>
<td>No emoticon</td>
<td>582</td>
</tr>
</tbody>
</table>
Evaluation of CAO

• Results

<table>
<thead>
<tr>
<th>Extraction</th>
<th>R</th>
<th>P</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94.3%</td>
<td>100%</td>
<td>97.1%</td>
</tr>
</tbody>
</table>

\[
\frac{394}{418} \quad \frac{394}{394} \quad 2 \frac{P \times R}{P + R}
\]
Evaluation of CAO

• Results

<table>
<thead>
<tr>
<th>Emotion Estimation on Separate Emoticons</th>
<th>CAO</th>
<th>Unique Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence</td>
<td>Frequency</td>
<td>2D space</td>
</tr>
<tr>
<td>Types</td>
<td>0.891472</td>
<td>0.966778</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotion Estimation on Sentences</th>
<th>CAO</th>
<th>Unique Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence</td>
<td>Frequency</td>
<td>2D space</td>
</tr>
<tr>
<td>Types</td>
<td>0.755171</td>
<td>0.908911</td>
</tr>
</tbody>
</table>

1. Unique Frequency
2. Frequency
3. Occurrence
Comparing CAO to other systems

In: Evaluation of CAO
Evaluation of CAO

- Comparing CAO to other systems

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Detection whether input equals emoticon</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>2. Detection of emoticon in sentence input</td>
<td>O (included in 3.)</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>3. Extraction of emoticon from any string of characters</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>4. Division into semantic areas</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>5. Database coverage</td>
<td>1,075</td>
<td>693</td>
<td>31</td>
<td>10,137 (expanded automatically to over 3 million)</td>
</tr>
<tr>
<td>6. Classification of emotion types</td>
<td>6 types (BBS-based)</td>
<td>7 types (Subjective)</td>
<td>6 types (Subjective)</td>
<td>10 types (Language/Culture Based)</td>
</tr>
<tr>
<td>7. Emotion estimation of separate emoticons</td>
<td>O (included in 8.)</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>8. Affect Analysis of sentences with emoticons</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
</tbody>
</table>
Evaluation of CAO

• Comparing CAO to other systems
  – In Training set:
    Comparison with Yamada et al. (2007)

Upgraded with our database and emotion classification

Evaluation of CAO

• Comparing CAO to other systems
  – In Training set:
    Comparison with Yamada et al. (2007)
    (their best < our worst)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>anger</td>
<td>0.702</td>
<td>0.815</td>
<td>0.877</td>
<td>0.811</td>
<td>0.771</td>
<td>0.767</td>
</tr>
<tr>
<td>dislike</td>
<td>0.661</td>
<td>0.809</td>
<td>0.919</td>
<td>0.631</td>
<td>0.800</td>
<td>0.719</td>
</tr>
<tr>
<td>excitement</td>
<td>0.700</td>
<td>0.789</td>
<td>0.846</td>
<td>0.786</td>
<td>0.769</td>
<td>0.797</td>
</tr>
<tr>
<td>fear</td>
<td>0.564</td>
<td>0.409</td>
<td>0.397</td>
<td>0.451</td>
<td>0.936</td>
<td>0.858</td>
</tr>
<tr>
<td>fondness</td>
<td>0.452</td>
<td>0.436</td>
<td>0.448</td>
<td>0.915</td>
<td>0.778</td>
<td>0.783</td>
</tr>
<tr>
<td>joy</td>
<td>0.623</td>
<td>0.792</td>
<td>0.873</td>
<td>0.944</td>
<td>0.802</td>
<td>0.860</td>
</tr>
<tr>
<td>relief</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>0.600</td>
<td>0.990</td>
<td>0.985</td>
</tr>
<tr>
<td>shame</td>
<td>0.921</td>
<td>0.949</td>
<td>0.976</td>
<td>0.706</td>
<td>0.922</td>
<td>0.910</td>
</tr>
<tr>
<td>sorrow</td>
<td>0.720</td>
<td>0.861</td>
<td>0.920</td>
<td>0.814</td>
<td>0.809</td>
<td>0.791</td>
</tr>
<tr>
<td>surprise</td>
<td>0.805</td>
<td>0.904</td>
<td>0.940</td>
<td>0.862</td>
<td>0.866</td>
<td>0.874</td>
</tr>
<tr>
<td>All approx.</td>
<td>0.675</td>
<td>0.751</td>
<td>0.802</td>
<td>0.852</td>
<td>0.804</td>
<td>0.818</td>
</tr>
</tbody>
</table>

Evaluation of CAO

- Comparing CAO to other systems
  - In Test set: their best < our worst (or 2\textsuperscript{nd} worst)

Statistical significance of results
Statistical significance or differences in training set evaluation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-gram</td>
<td>2-gram</td>
<td>3-gram</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anger</td>
<td>0.702</td>
<td>0.815</td>
<td>0.877</td>
<td>0.811</td>
<td>0.771</td>
<td>0.767</td>
</tr>
<tr>
<td>dislike</td>
<td>0.661</td>
<td>0.809</td>
<td>0.919</td>
<td>0.631</td>
<td>0.8</td>
<td>0.719</td>
</tr>
<tr>
<td>excitement</td>
<td>0.7</td>
<td>0.789</td>
<td>0.846</td>
<td>0.786</td>
<td>0.769</td>
<td>0.797</td>
</tr>
<tr>
<td>fear</td>
<td>0.564</td>
<td>0.409</td>
<td>0.397</td>
<td>0.451</td>
<td>0.936</td>
<td>0.858</td>
</tr>
<tr>
<td>fondness</td>
<td>0.452</td>
<td>0.436</td>
<td>0.448</td>
<td>0.915</td>
<td>0.778</td>
<td>0.783</td>
</tr>
<tr>
<td>joy</td>
<td>0.623</td>
<td>0.792</td>
<td>0.873</td>
<td>0.944</td>
<td>0.802</td>
<td>0.86</td>
</tr>
<tr>
<td>relief</td>
<td>1</td>
<td>0.999</td>
<td>1</td>
<td>0.6</td>
<td>0.99</td>
<td>0.985</td>
</tr>
<tr>
<td>shame</td>
<td>0.921</td>
<td>0.949</td>
<td>0.976</td>
<td>0.706</td>
<td>0.922</td>
<td>0.91</td>
</tr>
<tr>
<td>sorrow</td>
<td>0.72</td>
<td>0.861</td>
<td>0.92</td>
<td>0.814</td>
<td>0.809</td>
<td>0.791</td>
</tr>
<tr>
<td>surprise</td>
<td>0.805</td>
<td>0.904</td>
<td>0.94</td>
<td>0.862</td>
<td>0.866</td>
<td>0.874</td>
</tr>
</tbody>
</table>

All approx. 0.675 0.751 0.802 0.852 0.804 0.818 0.517 0.469

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>N</td>
<td>Y (&lt;5%)</td>
<td>Y (&lt;5%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In most cases differences were not significant (results equally good).

This means that although in training set Occurrence scored higher, Freq and UniqFreq were more probable to achieve better results in test set set evaluation.