Ideas for Using Large-Scale Corpora to Improve Verification of Emotion Appropriateness in Japanese

I. OPTIMIZING DATABASE OF EMOTEMES

1. REFINING EMOTEMES: Extracting expressive words appearing most often in 1st and 5th reviews [4].

- Amazon.com
  - Reviews already annotated:
    - very bad
    - neutral
    - very good

2. DISAMBIGUATING EMOTEMES:
   1) Limiting the list of possible emotion types/word:
      - Use Amazon annotations to verify polarity tendency of a word ("J-shape" or "reversed J-shape").
   2) Statistical disambiguation:
      - For every emoteme and emotive expression take 100 sentences (or all applicable) from every corpus.

3. IMPLEMENT IN CONVERSATIONAL AGENT
   - Perform processing on a local hard drive
   - In processing older texts.

REFERENCES

I. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki

PROTOTYPE METHOD

The method [1] determines whether the emotion is appropriate for the context.

For recognition of emotions it uses Ptaszynski’s affect analysis and annotation system, ML-Ask [2].

To verify the contextual appropriateness of those states it uses Shi’s Web mining technique [3] for gathering emotive common sense from the Internet.

II. UPDATING DATABASE OF EMOTIVE EXPRESSIONS

1. REFINING EMOTIVE EXPRESSIONS:
   - Check the concordance (hit rate) of every emotive expression from the data-base in Amazon and all three corpora.

2. EXPANDING EMOTIVE EXPRESSION DATABASE:
   - Extract syntactical patterns of emotive expressions from emotive sentences:
     - Kyou wa nante kimochi ii hi da na !
     - Today: THEM [er] [x]:joy day:OBJ [er] [es]
     - Today is such a nice day!
     - Pattern: THEM [er] [OBJ [er] [ei] [x2 = @]
     - Sore wa nante osoroshii koto dana !
     - Sore = osoroshii [fear]
   - Questionnaire for sentences with each new emotive expression candidate:
     - "Do you agree that this sentence expresses [emotion type]?"

III. IMPROVING WEB MINING

1. FROM BROWSER WEB MINING TO A STAND ALONE SYSTEM:
   - Gather large number of pages from the Web to a hard drive and index with HyperEstraier [10].
   - Perform processing on a local hard drive
   - Predictable increase in speed:
     - At most 15 mins. → ~5 secs.

IV. EXTRACTING CONVERSATIONAL STRATEGIES

1. ANALYSE CONVERSATION CORPORA WITH THE METHOD
   - Perform processing on a local hard drive
   - Recognize emotions
   - Verify appropriateness

2. FIND CONVERSATIONAL STRATEGIES
   - Extract sentence patterns with valuable conversational strategies

3. IMPLEMENT IN CONVERSATIONAL AGENT

CONCLUSIONS

- We presented a set of ideas for improving a new method for:
  1) discovering what are the emotions conveyed by a user in an utterance and 2) whether they are appropriate for the context they are used in.
- The planned improvements include using of different, large-scale corpora, like Amazon reviews, Web as corpus, or two conversation corpora.
- We also plan to extract conversational strategies for further implementation of the system into a conversational agent.

ABSTRACT

We present a set of ideas for improving the verification of emotion appropriateness in Japanese language. Emotion appropriateness verification is a new method for discovering not only what are the emotions conveyed by a user in an utterance also whether they are appropriate for the context they are used in. We present ideas to improve this method with the use of several corpora. The corpora we plan to use are Amazon reviews, Web as corpus and two corpora of natural conversations to improve the method and provide conversational strategies for implementation of the method into a conversational agent.

I. Lacks in database of emotemes

- ML-ASK
  - input
  - recognition
  - emotive common sense
  - emotive expression
  - emotive DB
  - output

WEB MINING

- Emotive common sense
- Emotive expressions database
- Emotive expression
- Emotive expression x : "kandou" [Excitement]

Corpus A

Corpus B

Corpus C

VI. What conversational strategy for each case?

- Check the concordance (hit rate) of every emotive expression from the data-base in Amazon and all three corpora.
- Recheck on the whole Web.
- Compress the least frequent into “Out-of-date” folder (used in processing older texts).

- Perform processing on a local hard drive
- Predictable increase in speed:
  - At most 15 mins. → ~5 secs.

- Check the concordance (hit rate) of every emotive expression from the data-base in Amazon and all three corpora.
- Recheck on the whole Web.
- Compress the least frequent into “Out-of-date” folder (used in processing older texts).

- Perform processing on a local hard drive
- Predictable increase in speed:
  - At most 15 mins. → ~5 secs.

- Check the concordance (hit rate) of every emotive expression from the data-base in Amazon and all three corpora.
- Recheck on the whole Web.
- Compress the least frequent into “Out-of-date” folder (used in processing older texts).

- Perform processing on a local hard drive
- Predictable increase in speed:
  - At most 15 mins. → ~5 secs.