EMOTIVE OR NON-EMOTIVE: THAT IS THE QUESTION

Michal Ptaszynski
Fumito Masui
Rafal Rzepka
Kenji Araki

Kitami Institute of Technology
Hokkaido University
PRESENTATION OUTLINE

1. Problem definition
2. Language Combinatorics
3. Experiment setup
4. Results and discussion
5. Conclusions and Future work
PROBLEM DEFINITION

MAINSTREAM:

POSITIVE VS. NEGATIVE EMOTION TYPES

DISREGARDED OR AS SUBTASK:

IS THE SENTENCE EMOTIONAL / NEUTRAL (PRESELECTION)
<table>
<thead>
<tr>
<th>PROBLEM DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>emotional</td>
</tr>
</tbody>
</table>

“Was the speaker in an emotional state?”
**PROBLEM DEFINITION**

"Was the speaker in an emotional state?"

Easy to ask laypeople because everyone thinks they are specialists in their own emotions.

“Non-emotional” would be probably better.
PROBLEM DEFINITION

Was the speaker in an emotional state?

Easy to ask laypeople because everyone thinks they are specialists in their own emotions.

PROBLEM DEFINITION

“Was the speaker in an emotional state?”

“Did the speaker present their the contents from first-person centric perspective or no specified perspective?”
**PROBLEM DEFINITION**

Was the speaker in an emotional state?

Did the speaker present their the contents from first-person centric perspective or no specified perspective?

Doesn’t have much to do with emotions. Only expressions of emotions tend to be first person centric as well.
### Problem Definition

<table>
<thead>
<tr>
<th>emotional</th>
<th>neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>subjective</td>
<td>objective</td>
</tr>
</tbody>
</table>

“Was the speaker in an emotional state?”

“Did the speaker present their the contents from first person centric perspective or no specified perspective?”

Opinion could be subjective but neutral
- “I think it will rain tomorrow.”
- “In my opinion the government should have applied a different policy.”

**Not talking about positive/negative.**

Doesn’t have much to do with emotions
→ Only expressions of emotions tend to be first person centric as well.
**Problem Definition**

**emotional**

- subjective

**neutral**

- objective

Opinion could be subjective but neutral
- “I think it will rain tomorrow.”
- “In my opinion the government should have applied a different policy.”

**Not talking about positive/negative.**
“Was the speaker in an emotional state?”

“Did the speaker present their the contents from first-person centric perspective or no specified perspective?”
“Was the speaker in an emotional state?”

“Did the speaker present their contents from first-person centric perspective or no specified perspective?”

“Was the sentence expressed with emphasis (distinguishable linguistic emotive features)?”
## Problem Definition

<table>
<thead>
<tr>
<th>Emotional</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emotive</strong></td>
<td><strong>Non-emotive</strong></td>
</tr>
<tr>
<td><strong>Subjective</strong></td>
<td><strong>Objective</strong></td>
</tr>
</tbody>
</table>

- All emotive sentences are emotional.
- All neutral and objective sentences are non-emotive.
- Some emotional sentences could be non-emotive.
- Subjective sentences could be emotive/non-emotive as well as emotional/neutral.
**PROBLEM DEFINITION**

<table>
<thead>
<tr>
<th>Emotional</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emotive</strong></td>
<td><strong>Non-emotive</strong></td>
</tr>
<tr>
<td>Subjective</td>
<td>Objective</td>
</tr>
</tbody>
</table>

Emotive/non-emotive could help distinguishing between
- emotional/neutral
- subjective/objective
PROBLEM DEFINITION

In linguistics:

Karl Buhler in 1934: 3 functions of language:
  descriptive, impressive, emotive

Stevenson in 1937: emotiveness
  (with regards to morality as a concept influenced by emotions)

Roman Jakobson in 1960: 6 functions of language:
  poetic, phatic, metalingual

PROBLEM DEFINITION

In linguistics:

- Emotive elements: **interjections/exclamations** (aah, ooh, whoa, great!), **hypocoristics** (endearments, dog → doggy), **emotive punctuation** (!, ??, ..., ~), **emoticons** (:-), ^o^), **onomatopoeia/mimetic expressions** (gitaigo in Japanese)
PROBLEM DEFINITION

In linguistics:

- Emotive elements: **interjections/exclamations** (aah, ooh, whoa, great!), **hypocoristics** (endearments, dog → doggy), **emotive punctuation** (!, ??, ..., ~), **emoticons** (:-), ^o^), **onomatopoeia/mimetic expressions** (gitaigo in Japanese)

- Plus - combinations of elements

ああ、今日はなんて気持ちいい日なんだ！(^0^)/
Oh, what a pleasant day today, isn’t it？(^0^)/
PROBLEM DEFINITION

In linguistics:

◆ Emotive elements: **interjections/exclamations** (aah, ooh, whoa, great!), **hypocoristics** (endearments, dog → doggy), **emotive punctuation** (!, ??, …, ~), **emoticons** (:-), ^o^), **onomatopoeia/mimetic expressions** (gitaigo in Japanese)

◆ Plus - combinations of elements

ああ、今日はなんて気持ちいい日なんだ！(^o^)/
Oh, what a pleasant day today, isn’t it ? (^0^)/

◆ Are there non-emotive elements ??
PROBLEM DEFINITION

In linguistics:

◆ Emotive elements: **interjections/exclamations** (aah, ooh, whoa, great!), **hypocoristics** (endearments, dog → doggy), **emotive punctuation** (!, ??, …, ~), **emoticons** (:-), (^o^), **onomatopoeia/mimetic expressions** (*gitaigo* in Japanese)

◆ Plus - combinations of elements

ああ、今日はなんて気持ちいい日なんだ！(^0^)/
Oh, what a pleasant day today, isn’t it? (^0^)/

◆ Are there non-emotive elements ??

◆ How to extract them?
This sentence contains the pattern:
ああ * なんて * なんだ！ (Oh, what a * isn’t it?)

1. This pattern cannot be discovered with n-gram approach.
2. This pattern cannot be discovered if one doesn’t know what to look for.

Need to find a way to extract such frequent sophisticated patterns from corpora.

*) pattern = something that frequently appears in a corpus (more than once).
SPEC – Sentence Pattern Extraction arChitecture

Sentence pattern = ordered non-repeated combinations of sentence elements.

For $1 \leq k \leq n$, there is all possible $k$-long patterns, and

$$\sum_{k=1}^{n} \binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \ldots + \frac{n!}{n!(n-n)!} = 2^n - 1$$
**SPEC – Sentence Pattern Extraction arChitecture**

**Sentence pattern** = ordered non-repeated combinations of sentence elements.

For \(1 \leq k \leq n\), there is all possible \(k\)-long patterns, and

\[
\sum_{k=1}^{n} \binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \ldots + \frac{n!}{n!(n-n)!} = 2^n - 1
\]

Extract patterns from all sentences and calculate occurrence.
SPEC – Sentence Pattern Extraction architecture

Sentence pattern = ordered non-repeated combinations of sentence elements.

For $1 \leq k \leq n$, there is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ all possible $k$-long patterns, and

$$\sum_{k=1}^{n} \binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \ldots + \frac{n!}{n!(n-n)!} = 2^n - 1$$

And then classify/comparative sentences with non-emotive

Normalized pattern weight

$$w_j = \left(\frac{O_{pos}}{O_{pos} + O_{neg}} - 0.5\right) \times 2$$

Score for one sentence

$$\text{score} = \sum w_j, (1 \geq w_j \geq -1)$$
LANGUAGE COMBINATORICS

SPEC – Sentence Pattern Extraction arChitecture
Sentence pattern = ordered non-repeated combinations of sentence elements.

For $1 \leq k \leq n$, there is all possible $k$-long patterns, and

$$
\sum_{k=1}^{n} \binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \ldots + \frac{n!}{n!(n-n)!} = 2^n - 1
$$

Normalized pattern weight

$$
w_j = \left( \frac{O_{pos}}{O_{pos} + O_{neg}} - 0.5 \right) \times 2
$$

Score for one sentence

$$
score = \sum w_j, (1 \geq w_j \geq -1)
$$
91 sentences close in meaning, but different emotional load (50 emotive, 41 non-emotive) gathered in an anonymous survey on 30 people of different background (students, businessmen, housewives).

**Examples:**

**Emotive**
- Takasugiru kara ne
- 'Cause its just too expensive
- Sugoku kirei na umi da naa
- Oh, what a beautiful sea!
- Nanto ano hito, kekkon suru rashii yo
- Have you heard? She’s getting married!

**Non-emotive**
- Kougaku na tame desu.
- Due to high cost.
- Kirei na umi desu
- This is a beautiful sea
- Ano hito kekkon suru rashii desu
- They say she is getting married.
EXPERIMENT SETUP

DATASET

91 sentences close in meaning, but different emotional load (50 emotive, 41 non-emotive) gathered in an anonymous survey on 30 people of different background (students, businessmen, housewives).

Emotive       Non-emotive

高すぎるからね       高額なためです。
Takasugiru kara ne   Kougaku na tame desu.
'Cause its just too expensive      Due to high cost.

すごくきれいな海だなあ         きれいな海です
Sugoku kirei na umi da naa   Kirei na umi desu
Oh, what a beautiful sea!      This is a beautiful sea

なんとあの人の、結婚するらしいよ
Nanto ano hito, kekkon suru rashii yo
Have you heard? She’s getting married!

あの日と結婚するらしいです
Ano hito kekkon suru rashii desu
They say she is getting married.

Examples:
## EXPERIMENT SETUP

### Preprocessing

<table>
<thead>
<tr>
<th>Sentence:</th>
<th>今日はなんて気持ちいい日なんだ！</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transliteration:</td>
<td>Kyōwanantekimochiiihinanda!</td>
</tr>
<tr>
<td>Translation:</td>
<td>What a pleasant day it is today!</td>
</tr>
</tbody>
</table>

### Preprocessing examples

1. **Tokens:**
   
   Kyō wa nante kimochi ii hi nanda!

2. **POS:**
   
   N TOP ADV N ADJ N COP EXCL

3. **Tokens+POS:**
   
EXPERIMENT SETUP

Pattern List Modification
1. All patterns
2. Zero-patterns deleted
3. Ambiguous patterns deleted

Weight Calculation Modifications
1. Normalized
2. Award length
3. Award length and occurrence
EXPERIMENT SETUP

Pattern List Modification
1. All patterns
2. Zero-patterns deleted
3. Ambiguous patterns deleted

Weight Calculation Modifications
1. Normalized
2. Award length
3. Award length and occurrence

Is it worth the time?
EXPERIMENT SETUP

Pattern List Modification
1. All patterns
2. Zero-patterns deleted
3. Ambiguous patterns deleted

Weight Calculation Modifications
1. Normalized
2. Award length
3. Award length and occurrence

All patterns vs. only n-grams

Automatic threshold setting

Is it worth the time?

Data is never perfectly balanced.
**EXPERIMENT SETUP**

- **Pattern List Modification**
  1. All patterns
  2. Zero-patterns deleted
  3. Ambiguous patterns deleted

- **Weight Calculation Modifications**
  1. Normalized
  2. Award length
  3. Award length and occurrence

- **10-fold Cross Validation**
  - All patterns vs. only n-grams
  - Automatic threshold setting

- **Is it worth the time?**
  - One experiment = 280 runs

- **Data is never perfectly balanced.**
EXPERIMENT SETUP

Score calculated in:
- Precision
- Recall
- Balanced F-score
- Accuracy
- Specificity
- Phi-coefficient
RESULTS AND DISCUSSION

Tokenized

Tokens + POS

Weight normalized

Length awarded
RESULTS AND DISCUSSION

Tokenized

Weight normalized

Length awarded

Tokens + POS

specific elements are more effective
RESULTS AND DISCUSSION

Tokenized

Weight normalized

Length awarded

Tokenized + POS

patterns better than n-grams

specific elements are more effective
RESULTS AND DISCUSSION

Specific elements are more effective than n-grams. Awarding length yields higher results.
**RESULTS AND DISCUSSION**

Tokenized

- Specific elements are more effective

- Awarding length yields higher results

- F=0.76

Tokens + POS

- Patterns better than n-grams

- F=0.76
  - P=0.64
  - R=0.95
RESULTS AND DISCUSSION

Tokenized

- Specific elements are more effective

Tokens + POS

- Patterns better than n-grams

• SPEC slightly worse than ML-Ask (F = 0.79, P = 0.8, R = 0.78)

SPEC    >    ML-Ask

fully automatic    handcrafted

length yields higher results

F=0.76
P=0.64
R=0.95
RESULTS AND DISCUSSION

Tokenized

Tokens + POS

• SPEC slightly worse than ML-Ask (F = 0.79, P = 0.8, R = 0.78)
  SPEC > ML-Ask
  fully automatic > handcrafted

• More efficient (user does nothing)
• Applicable to other lang.
• Can point out non-emotive el.

specific elements are more effective

patterns better than n-grams

length yields higher results

F=0.76
P=0.64
R=0.95
RESULTS AND DISCUSSION

Examples of extracted Patterns (Tokenized)

<table>
<thead>
<tr>
<th></th>
<th>Emotive</th>
<th></th>
<th>Non-emotive</th>
</tr>
</thead>
<tbody>
<tr>
<td>freq.</td>
<td>example pattern</td>
<td>freq.</td>
<td>example pattern</td>
</tr>
<tr>
<td>14</td>
<td>にた</td>
<td>11</td>
<td>い*</td>
</tr>
<tr>
<td>12</td>
<td>で</td>
<td>8</td>
<td>し*</td>
</tr>
<tr>
<td>11</td>
<td>ん*。</td>
<td>7</td>
<td>です。</td>
</tr>
<tr>
<td>11</td>
<td>と</td>
<td>6</td>
<td>は*です</td>
</tr>
<tr>
<td>11</td>
<td>ー</td>
<td>6</td>
<td>まし*。</td>
</tr>
<tr>
<td>10</td>
<td>にた*。</td>
<td>5</td>
<td>でした。</td>
</tr>
<tr>
<td>9</td>
<td>*よ</td>
<td>5</td>
<td>ます</td>
</tr>
<tr>
<td>9</td>
<td>*ん</td>
<td>5</td>
<td>い</td>
</tr>
<tr>
<td>8</td>
<td>し</td>
<td>4</td>
<td>です*。</td>
</tr>
<tr>
<td>7</td>
<td>ない</td>
<td>3</td>
<td>この<em>は</em>。</td>
</tr>
<tr>
<td>7</td>
<td>!</td>
<td>3</td>
<td>は*です。</td>
</tr>
<tr>
<td>6</td>
<td>ん*よ</td>
<td>3</td>
<td>で*ます</td>
</tr>
<tr>
<td>6</td>
<td>、*だ</td>
<td>3</td>
<td>が*た。</td>
</tr>
<tr>
<td>6</td>
<td>ちゃ</td>
<td>3</td>
<td>美味しい</td>
</tr>
<tr>
<td>6</td>
<td>よ。</td>
<td>3</td>
<td>た。</td>
</tr>
<tr>
<td>5</td>
<td>だ*。</td>
<td>2</td>
<td>た*、*。</td>
</tr>
<tr>
<td>5</td>
<td>に*よ</td>
<td>2</td>
<td>せ</td>
</tr>
<tr>
<td>5</td>
<td>が*よ</td>
<td>2</td>
<td>か</td>
</tr>
<tr>
<td>5</td>
<td>ん</td>
<td>2</td>
<td>さ</td>
</tr>
</tbody>
</table>

Emotive
- Casual wording
- Prolongation marks
- SFPs
- Subject particles

Non-emotive
- Official forms (desu-masu)
- More “periods”
- No exclamation marks, etc.
### RESULTS AND DISCUSSION

#### EXAMPLE SENTENCES

<table>
<thead>
<tr>
<th>Example 1.</th>
<th>Example 4.</th>
</tr>
</thead>
<tbody>
<tr>
<td>メガネ、そこにあったんだよ。&lt;br&gt;<code>Megane, soko ni atta n da yo</code>.&lt;br&gt;(The glasses were over there!)</td>
<td>高額なためです。&lt;br&gt;<code>Kougaku na tame desu.</code>&lt;br&gt;Due to high cost.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 2.</th>
<th>Example 5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ううん、舞台が見えないよ。&lt;br&gt;<code>Uun, butai ga mienai yo</code>.&lt;br&gt;(Ooh, I cannot see the stage!)</td>
<td>きれいな海です&lt;br&gt;<code>Kirei na umi desu</code>&lt;br&gt;This is a beautiful sea</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 3.</th>
<th>Example 6.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ああ、おなかがすいたよ。&lt;br&gt;<code>Aa, onaka ga suita yo</code>.&lt;br&gt;(Ohh, I’m so hungry)</td>
<td>今日は雪が降っています。&lt;br&gt;<code>Kyou wa yuki ga futte imasu.</code>&lt;br&gt;It is snowing today.</td>
</tr>
</tbody>
</table>
Presented research on extracting emotive patterns.

Used SPEC - a method for automatic extraction of patterns from sentences.

Extracted the patterns from a set of emotive and non-emotive sentences.

Classified sentences (test data) with those patterns.

Compared different preprocessing techniques (tokenization, POS, token-POS).

The best results obtained patterns with both tokens and POS

(F-score = 76%, Precision = 64%, Recall 95%).

Results for only POS were the lowest. This means the algorithm works better on less abstracted data.

The results of SPEC were compared to ML-Ask affect analysis system. ML-Ask achieved better Precision, but lower Recall. However, SPEC is fully automatic and thus more efficient and language independent.

Many of the automatically extracted patterns appear in handcrafted databases of ML-Ask, which suggests it could be possible to improve ML-Ask performance by extracting additional patterns with SPEC.

In the future we’ll try to quantify the correlation between emotive-subjective-emotional
THANK YOU FOR YOUR ATTENTION!

Michal Ptaszynski  Kitami Institute of Technology
ptaszynski@ieee.org
http://orion.cs.kitami-it.ac.jp/tipwiki/michal