Automatic Extraction of Emotive and Non-emotive Sentence Patterns

Michal Ptaszynski Fumito Masui
Kitami Institute of Technology

Rafal Rzepka and Kenji Araki
Hokkaido University

ABSTRACT

In this research we focus on automatic extraction of patterns from emotive (emotionally loaded) sentences. We assume emotive sentences stand out both lexically and grammatically and verify this assumption experimentally by comparing two sets of such sentences. We use a novel pattern extraction method based on the idea of language combinatorics. Extracted patterns are applied in a text classification task of discriminating between emotive and non-emotive sentences. The method reached balanced F-score of 76% with Precision equal to 64% and Recall 93%.

PROBLEM DEFINITION

今日はなんて気持ちいい日なんだ！(What a pleasant day today, isn't it?)

This sentence contains a pattern:

なんて＊なんだ！ (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 frequent patterns from corpora.

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).

Examples:

Emotive	Non-emotve
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、結婚するらしいよ	あの日と結婚するらしいです
Nanto ano hito, kekkon suru rashii yo	*Ano hito kekkon suru rashii desu*
Have you heard? She's getting married!	They say she is gatting married.

LANGUAGE COMBINATORICS

SPEC – Sentence Pattern Extraction arChitecture [2]

Sentence patterns = 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**

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

Normalized pattern weight

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

Score for one sentence

$$score = \sum w_j, (1 \geq w_j \geq -1)$$

EXPERIMENT SETUP

Preprocessing

Sentence:	今日はなんて気持ちいい日なんだ！
Transliteration:	*Kyōwanantekimochiiihinanda!*
Translation:	What a pleasant day it is today!
	Preprocessing examples
1. Tokens:	*Kyō wa nante kimochi ii hi nanda !*
2. POS:	N TOP ADV N ADJ N COP EXCL
3. Tokens+POS:	*Kyō*[N] *wa*[TOP] *nante*[ADV] *kimochi*[N] *ii*[ADJ] *hi*[N] *nanda*[COP] *!*[EXCL]

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

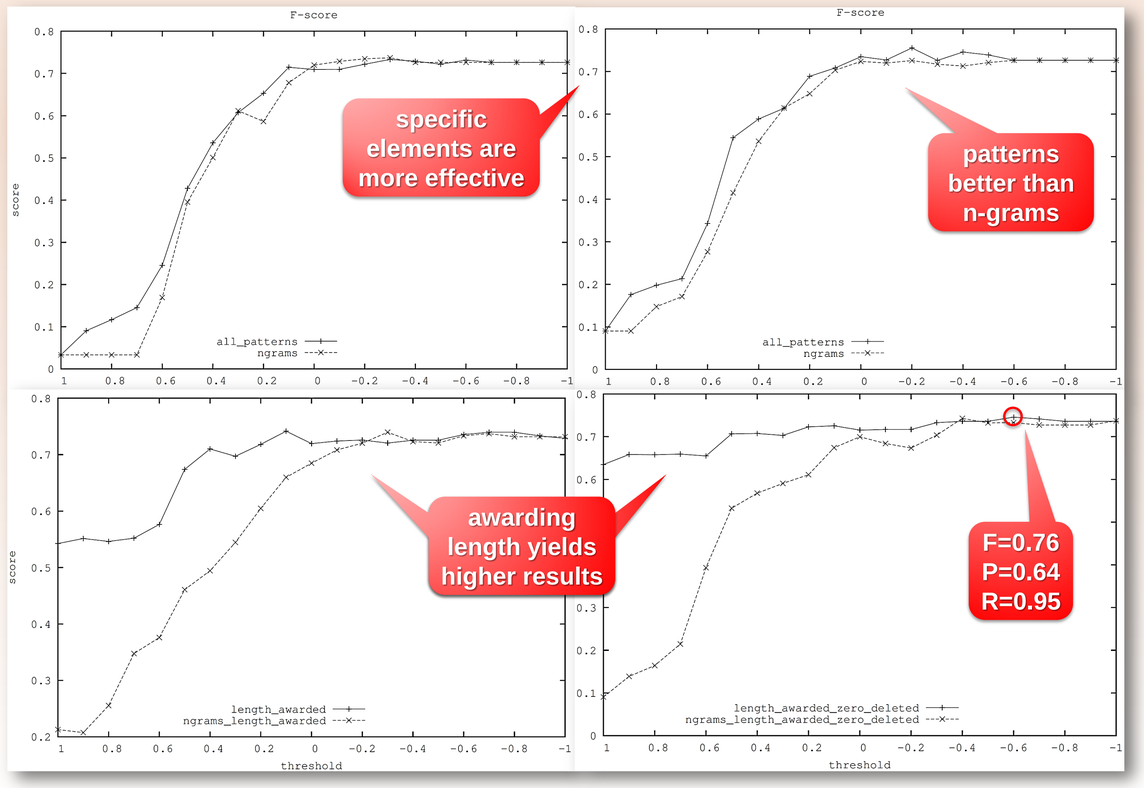
Automatic threshold setting

10-fold Cross Validation

EVALUATION EXPERIMENT

RESULTS

- **Token+POS > Tokenized > POS → algorithm works better on specific elements than more generalized**
- **Patterns > N-grams (sometimes n-grams get better Precision)**
- **Length awarded > normalized weight**
- **Highest results F-score = 0.76, Precision = 0.64, Recall = 0.95**
- **SPEC slightly worse than ML-Ask [3] (F = 0.79, P = 0.8, R = 0.78)**
- **SPEC – fully automatic > ML-Ask – handcrafted**



DETAILED ANALYSIS

Extracted patterns (Tokenized)

Emotive freq.	Emotive example pattern	Non-emotive freq.	Non-emotive example pattern
14	、＊た	11	い＊。
12	で	8	し＊。
11	ん＊。	7	です。
11	と	6	は＊です
11	ー	6	まし＊。
10	、＊た＊。	5	ました。
9	、＊よ	5	ます
9	、＊ん	5	い
8	し	4	です＊。
7	ない	3	この＊は＊。
7	！	3	は＊です。
6	ん＊よ	3	て＊ます
6	、＊だ	3	が＊た。
6	ちゃ	3	美味しい
6	よ。	3	た。
5	だ＊。	2	た＊、＊。
5	に＊よ	2	せ
5	が＊よ	2	か
5	ん	2	さ

EXAMPLE SENTENCES

Example 1.
メガネ、そこにあったんだよ。
Megane, soko ni atta n da yo .
(The glasses were over there!)

Example 2.
ううん、舞台が見えないよ。
Uun, butai ga mienai yo .
(Ooh, I cannot see the stage!)

Example 3.
ああ、おなかがすいたよ。
Aa, onaka ga suita yo .
(Ohh, I'm so hungry)

Example 4.
高額なためです。
Kougaku na tame desu.
Due to high cost.

Example 5.
きれいな海です
Kirei na umi desu
This is a beautiful sea

Example 6.
この本はとても怖いです
Kono hon wa totemo kowai desu .
This book is very scary.

Example 7.
今日は雪が降っています.
Kyou wa yuki ga futte imasu.
It is snowing today.

CONCLUSIONS

Presented **SPEC -** a **method for automatic extraction** of patterns from **emotive sentences**.
The patterns extracted from a set of emotive and non-emotive sentences were applied in text classification task.
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** state-of-the-art affect analysis system. ML-Ask achieved better Precision, but lower Recall. However, as fully automatic, **SPEC is 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**.

REFERENCES

[1] Kaori Sasai. 2006. The Structure of Modern Japanese Exclamatory Sentences: On the Structure of the *Nanto*-Type Sentence. *Studies in the Japanese Language*, Vol, 2, No. 1, pp. 16-31.
[2] Michal Ptaszynski, Rafal Rzepka, Kenji Araki and Yoshio Momouchi. 2011. Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion. *International Journal of Computational Linguistics (IJCL)*, Vol. 2, Issue 1, pp. 24-36.
[3] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki. 2009. Affecting Corpora: Experiments with Automatic Affect Annotation System -A Case Study of the *2channel* Forum-, In *Proceedings of The Conference of the Pacific Association for Computational Linguistics (PACLING-09)*, pp. 223-228.
[4] C. E. Shannon. 1948. A Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol. 27, pp. 379-423 (623-656), 1948.
[5] C. Potts and F. Schwarz. 2008. Exclamatives and heightened emotion: Extracting pragmatic generalizations from large corpora. Ms., UMass Amherst.