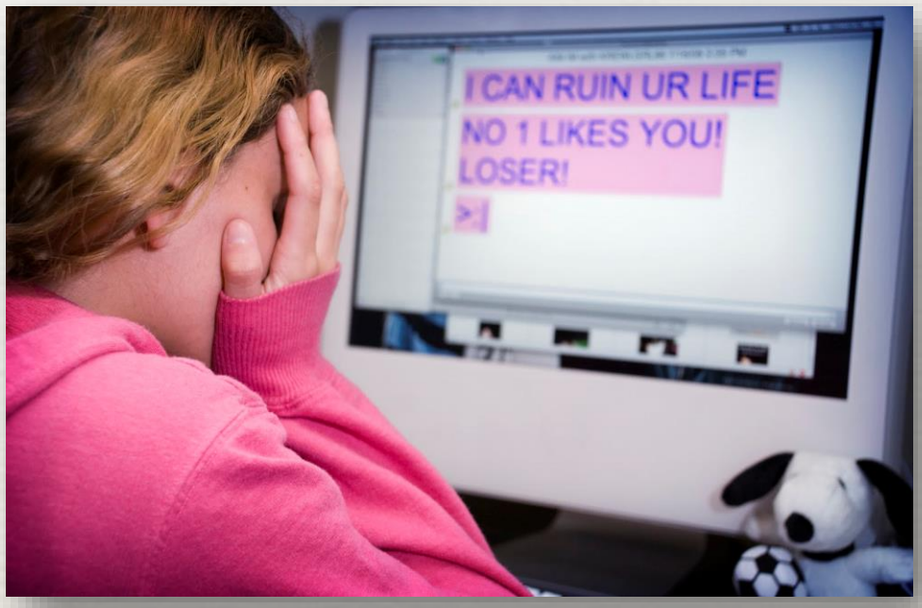
Brute Force Works Best Against Bullying

Michal Ptaszynski, Fumito Masui, Yasutomo Kimura,
Rafal Rzepka and Kenji Araki

Outline

1. Cyberbullying as social problem
2. Previous research
3. Proposed method
4. Experiments
5. Future work

Introduction



Cyberbullying

- Slandering and humiliating people on the Internet.
- Recently noticed social problem.

HELP by ICT

INTERNET PATROL

- Internet monitoring by PTA.
- Request site admin to remove harmful entries.
- High cost of time and fatigue for net-patrol members.

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

2013 PATENT

Language Combinatorics

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

SVM / optimization

Category Relevance Optimization

Automatic acquisition of harmful words

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

2013 PATENT

Language Combinatorics

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

SVM / optimization

Category Relevance Optimization

Automatic acquisition of harmful words

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

2013 PATENT

Language Combinatorics

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

SVM / optimization

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Ma
R. Rzepka and K. Araki. 2010. **Machine Learnin**
nd Affect Analysis Against Cyber-Bullying. In *A*
oceedings of AISB'10, 29th March – 1st April 2010.

Category Relevance Optimization

Automatic acquisition of harmful words

Affect analysis of cyberbullying data

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

SO-PMI-IR / phrases

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Ara
d Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cybe
lying with Machine Learning and Affect Analysis**. *International Journ
Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 201

SVM / optimization

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

2013 PATENT

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

Category Relevance Optimization

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

15 BUENOS AIRES

Language Combinatorics

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

Automatic acquisition of harmful words

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

Language Combinatorics

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

SVM / optimization

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

2013 PATENT

Category Relevance Optimization

Automatic acquisition of harmful words

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

2013 PATENT

Language Combinatorics

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

SVM / optimization

Category Relevance Optimization

Automatic acquisition of harmful words

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

2013 PATENT

Language Combinatorics

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

SVM / optimization

Category Relevance Optimization

Automatic acquisition of harmful words

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

2009 2010 2011 2012 2013 2014 2015

Previous Research

T. Matsuba, F. Masui, A. Kawai, N. Isu. 2011. **Study on the polarity classification model for the purpose of detecting harmful information on informal school sites** (in Japanese), In *Proceedings of NLP2011*, pp. 388-391.

Patent No. 2013-245813. Inventors: Fumito Masui, Michal Ptaszynski, Nitta Taisei. Patent name: *An Apparatus and Method for Detection of Harmful Entries on Internet*

Michal Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, K. Araki. 2015. **Brute Force Works Best Against Bullying**, *IJCAI 2015 Workshop on Intelligent Personalization (IP 2015)*, Buenos Aires, July 25-31, 2015.

SO-PMI-IR / phrases

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, K. Araki, and Y. Momouchi. 2010. **In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis**. *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

2013 PATENT

Language Combinatorics

S. Hatakeyama, F. Masui, M. Ptaszynski, K. Yamamoto. 2015. **Improving Performance of Cyberbullying Detection Method with Double Filtered Point-wise Mutual Information**. *ACM Symposium on Cloud Computing 2015 (SoCC'15)*, August 2015.

T. Nitta, F. Masui, M. Ptaszynski, Y. Kimura, R. Rzepka, K. Araki. 2013. **Detecting Cyberbullying Entries on Informal School Websites Based on Category Relevance Maximization.** In *Proceedings of IJCNLP 2013*, pp. 579-586.

SVM / optimization

Category Relevance Optimization

Automatic acquisition of harmful words

Michal Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka and K. Araki. 2010. **Machine Learning and Affect Analysis Against Cyber-Bullying**. In *Proceedings of AISB'10*, 29th March – 1st April 2010.

Affect analysis of cyberbullying data

2009 2010 2011 2012 2013 2014 2015

Dataset

- Actual data collected by Internet Patrol (annotated by experts)
- From unofficial school forums (BBS)
- Provided by Human Right Center in Japan (Mie Prefecture)
- According to the Definition by Japanese Ministry of Education (MEXT)
- 1,490 harmful and 1,508 non-harmful entries

Proposed Method

Language Combinatorics

SPEC – Sentence Pattern Extraction arChitecture

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

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

Extract patterns from all sentences and calculate occurrence.

Michal Ptaszynski, R. Rzepka, K. Araki and Y. Momouchi. 2011. **Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion.** *Int. J. of Computational Linguistics (IJCL)*, Vol. 2, Issue 1, pp. 24-36.

Language Combinatorics

Example: What a nice day !

That is why “brute force”

5-element pattern: What a nice day ! (1)

4-el. patterns:	3-el. patterns:	2-el. patterns:	1-el. patterns:
What a nice * !	a nice * !	What a	What
What a nice day	What a nice	What * !	a
What a * day !	What a * !	nice * !	nice
(5)	(10)	(10)	(5)
.	.	.	.
.	.	.	.
.	.	.	.

Language Combinatorics

SPEC – Sentence Pattern Extraction arChitecture

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

Classify new input with pattern list

M[illegible]K. Araki and Y. Momouchi. 2011. **Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion.** *Int. J. of Computational Linguistics (IJCL)*, Vol. 2, Issue 1, pp. 24-36.

Experiment setup

Preprocessing
1. Tokenization
2. POS
3. Tokens+POS

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

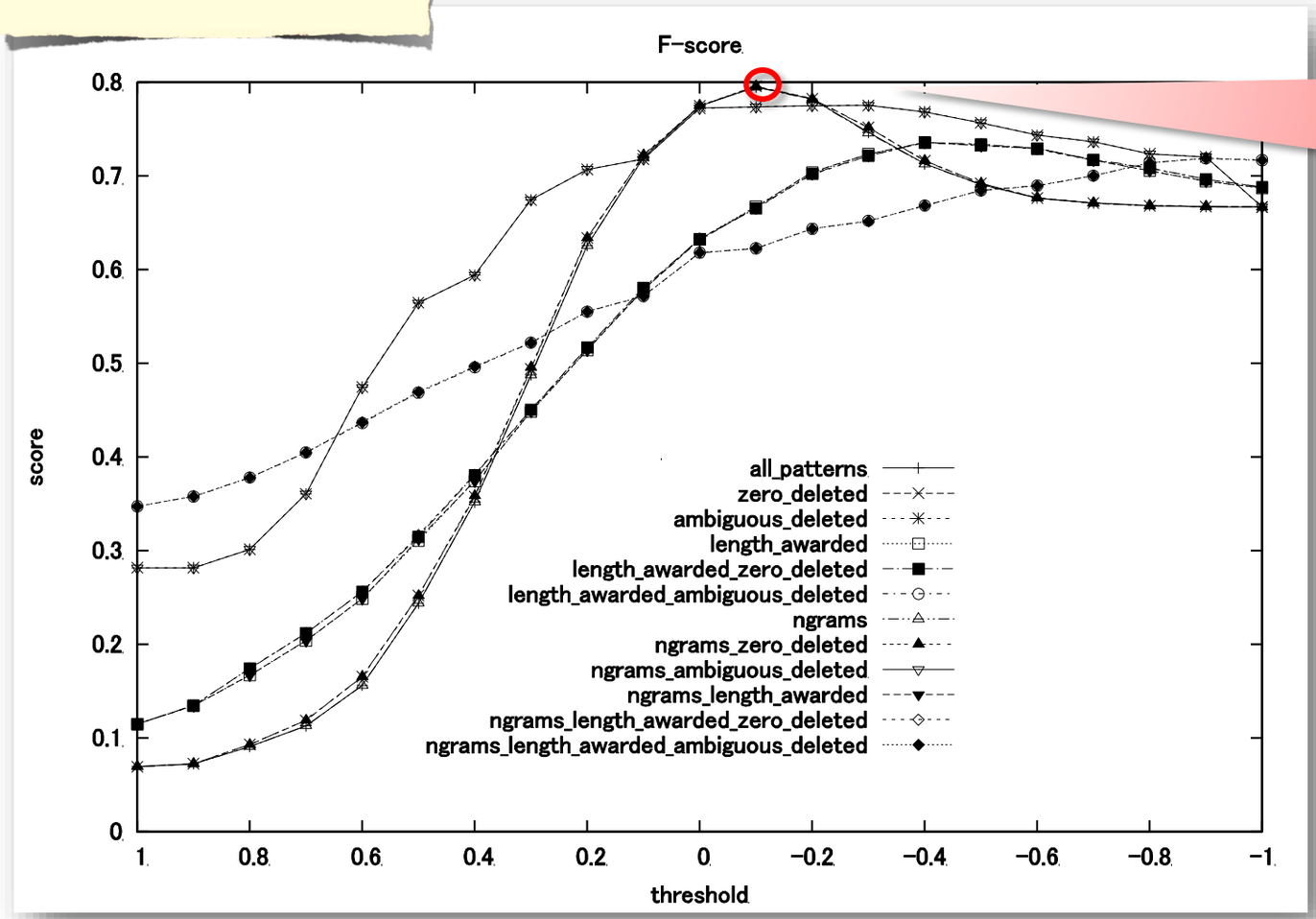
Is it worth the time?

One experiment = 420 runs

Data is never perfectly balanced.

Results

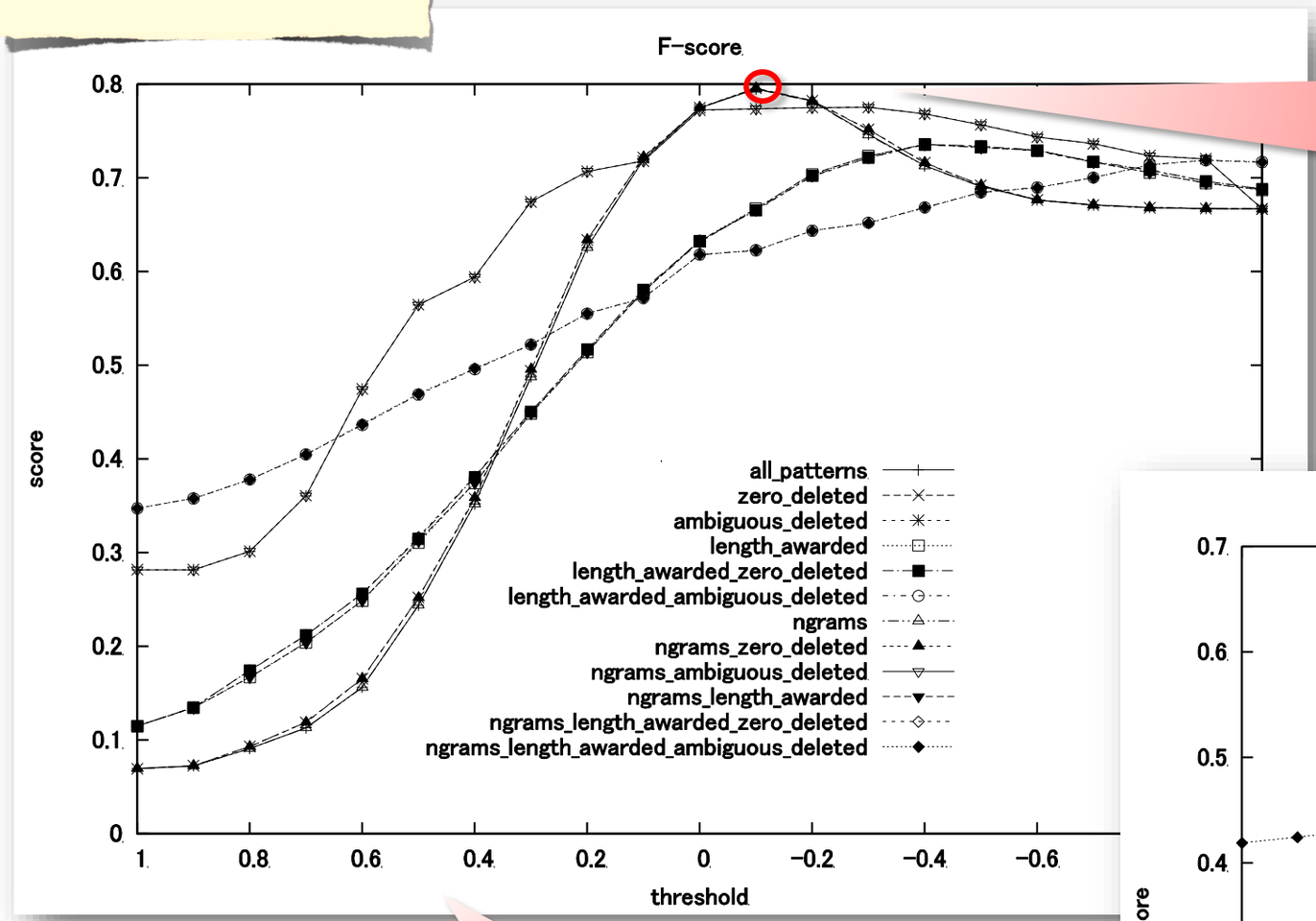
Tokens+POS



Best F-score
F=0.8
P=0.76
R=0.84

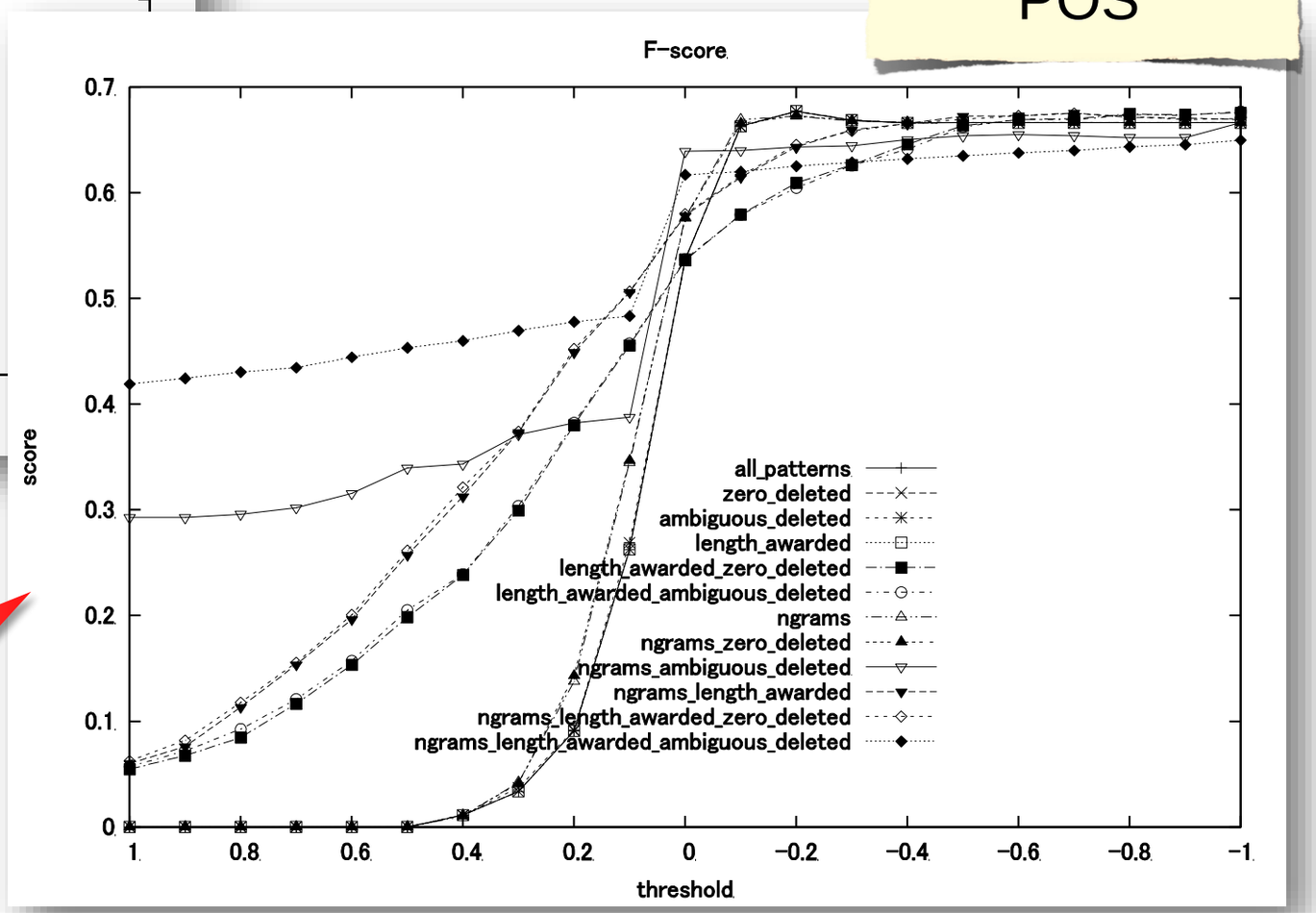
Results

Tokens+POS



Best F-score
F=0.8
P=0.76
R=0.84

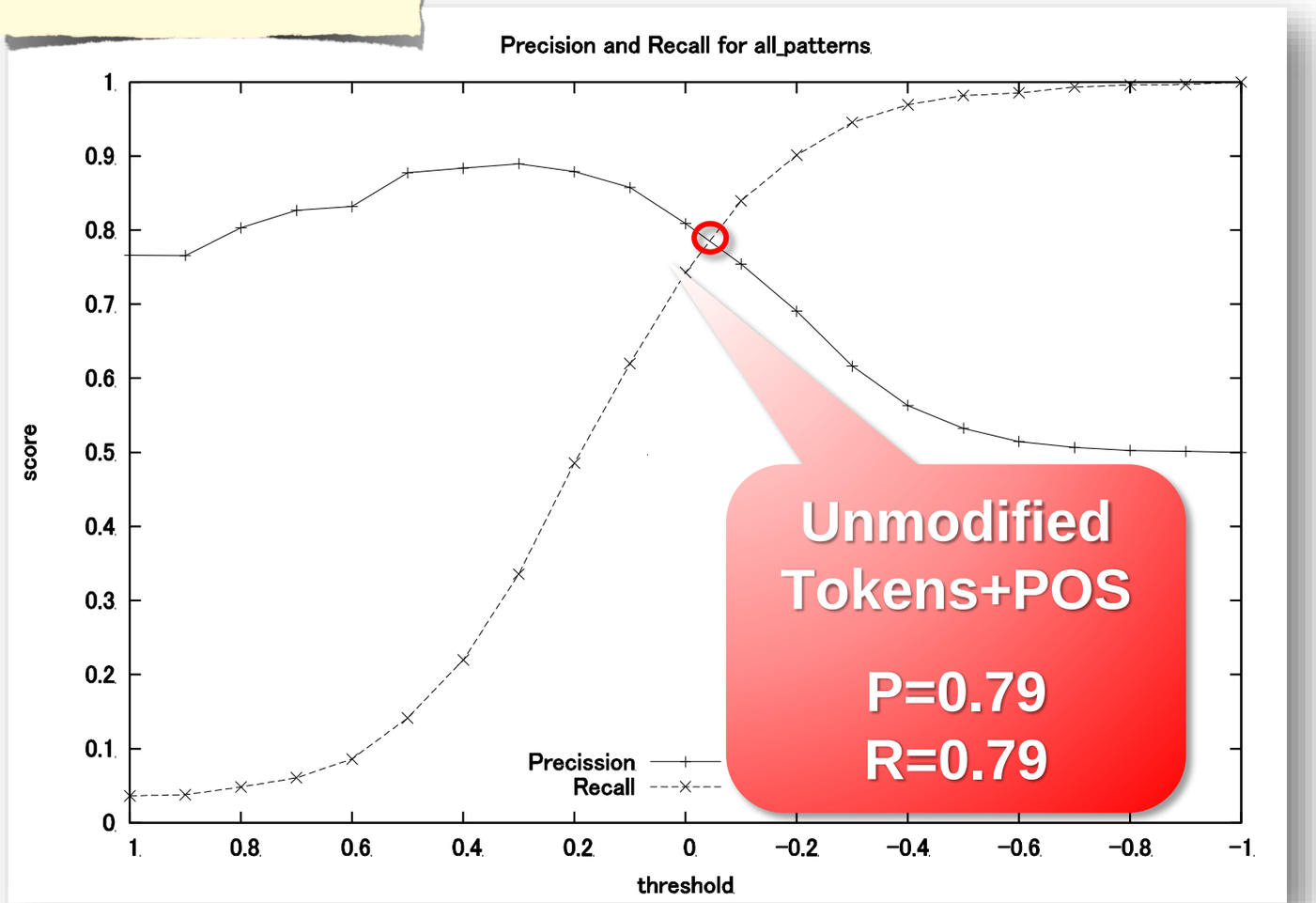
POS



specific elements are more effective than generalized ones

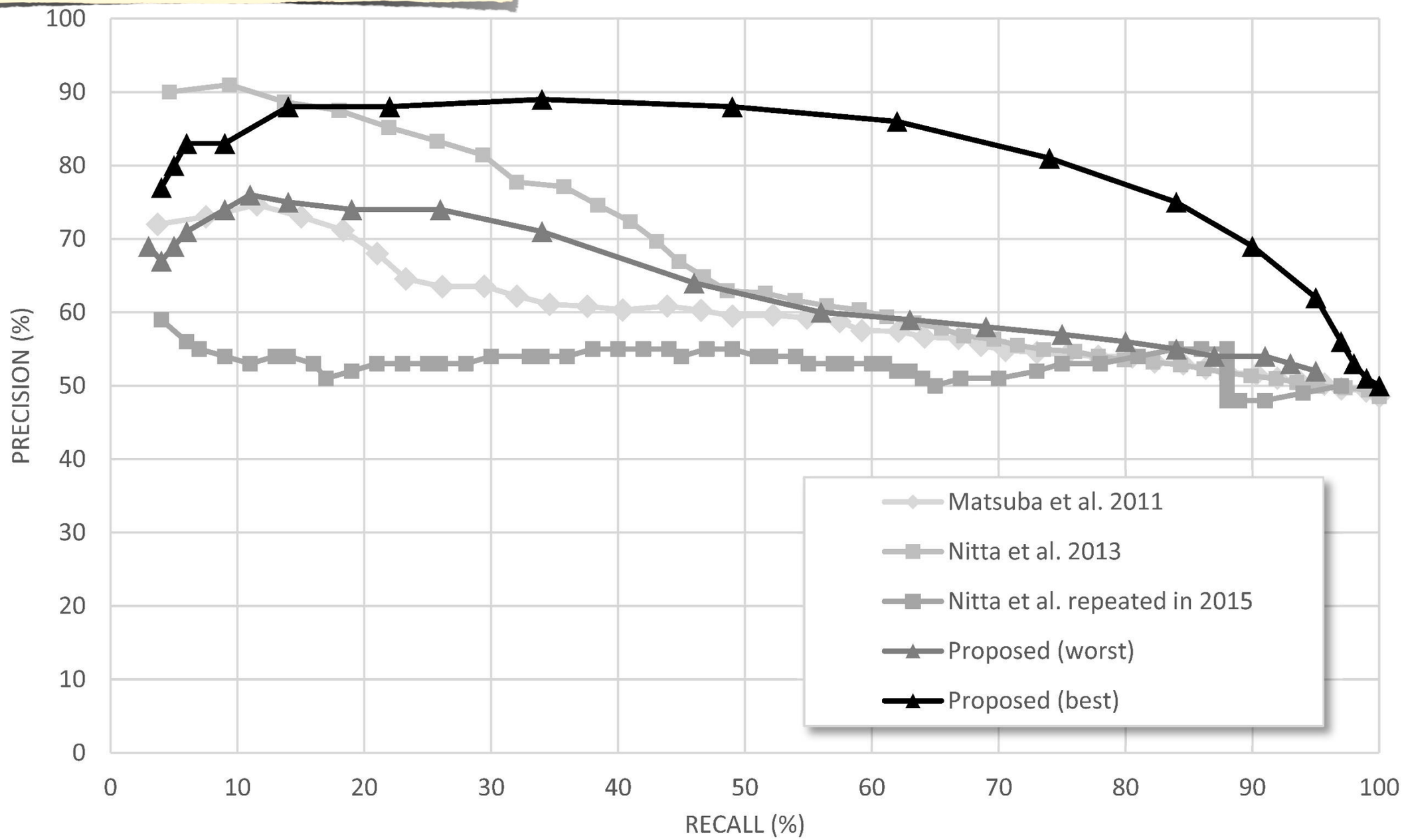
Results

Best BEP



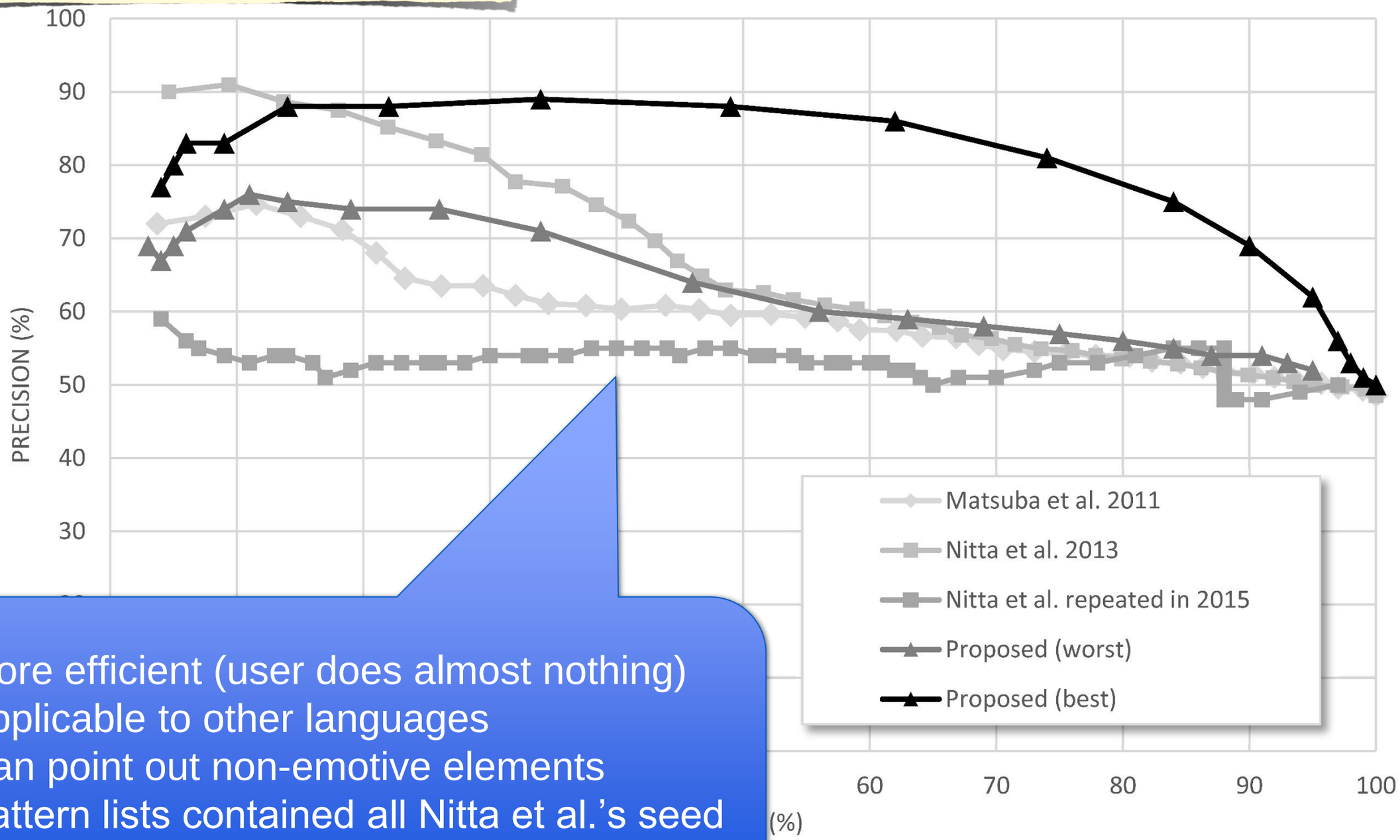
Results

Comparison with state-of-the-art



Results

Comparison with state-of-the-art



- More efficient (user does almost nothing)
- Applicable to other languages
- Can point out non-emotive elements
- Pattern lists contained all Nitta et al.'s seed words → could improve Nitta with patterns

Conclusions

- Presented research on cyberbullying detection.
- Proposed novel method.
  - Combinatorial algorithm applied in automatic extraction of sentence patterns.
- Used those patterns in classification of cyberbullying.
- Tested on actual data obtained by Internet patrol.
- Outperformed previous methods.
- Requires minimal human effort.

Future work

- Apply different preprocessing and classifiers for further improvement.
- Obtain new data by applying method in practice.
- Verify the actual amount of CB information on the Internet and reevaluate in more realistic conditions.

Thank you for your kind attention!

Michal Ptaszynski
paszynski@ieee.org