First Glance on Pattern-based Language Modeling

MICHAL PTASZYNSKI
FUMITO MASUI
RAFAL RZEPKA
KENJI ARAKI

Kitami Institute of Technology

Hokkaido University

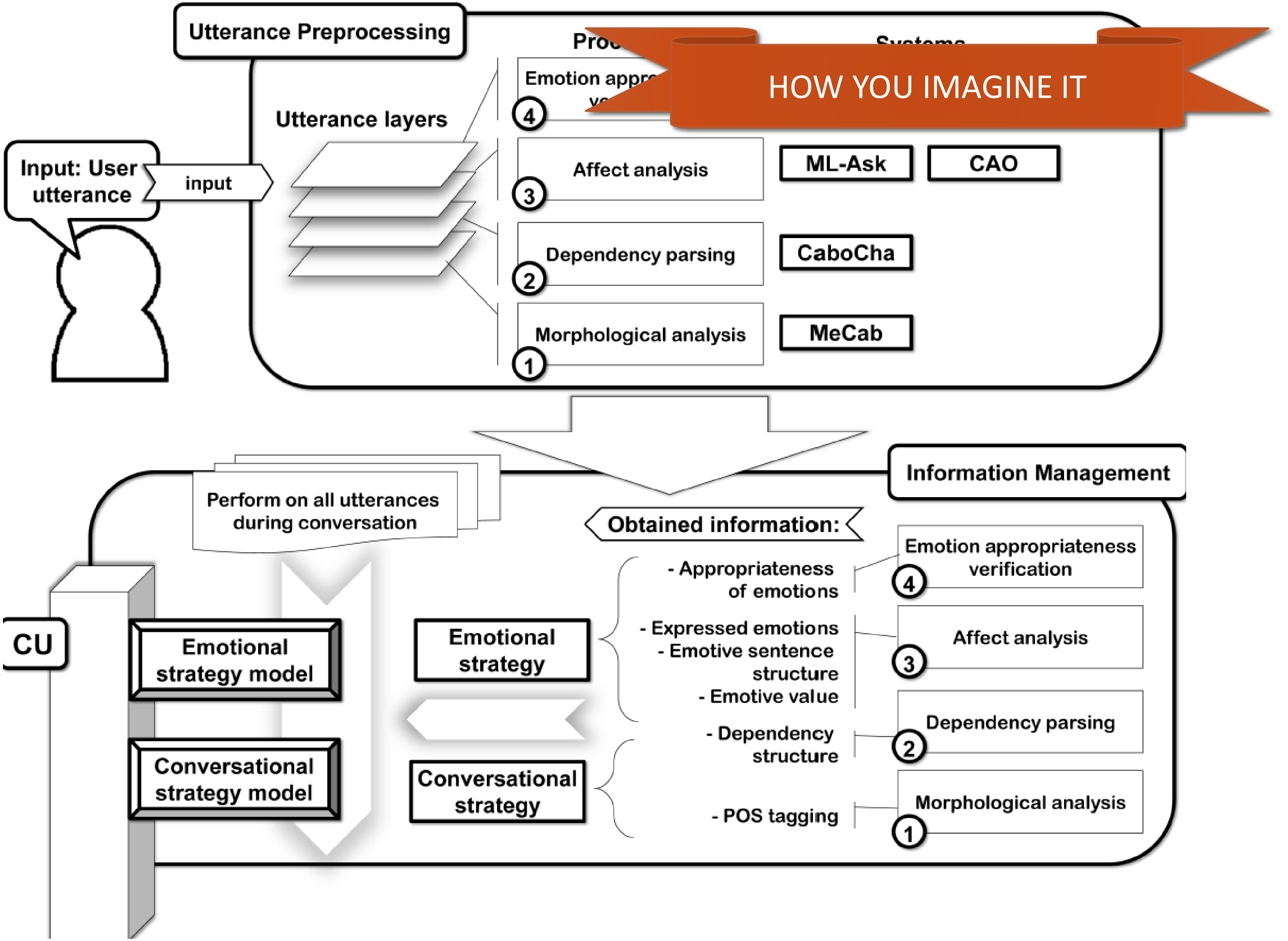
Presentation outline

1. Introduction
2. Language Models
3. Language Combinatorics
4. Applications
5. Conclusions and Future work

Introduction

Language modelling

HOW YOU IMAGINE IT



IN REALITY…

Introduction

Language modelling

John likes to watch movies. Mary likes movies too.
John also likes to watch football games.

```
{
    "John": 1,
    "likes": 2,
    "to": 3,
    "watch": 4,
    "movies": 5,
    "also": 6,
    "football": 7,
    "games": 8,
    "Mary": 9,
    "too": 10
}
```

```
[1, 2, 1, 1, 2, 0, 0, 0, 1, 1]
[1, 1, 1, 1, 0, 1, 1, 1, 0, 0]
```

Introduction

Language modelling

- Statistical representation of a piece of language data

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Unordered set of words

The dog bit the man = The man bit the dog

- No grammar
- No word order
- Just a bag of words…

Harris, Zellig. 1954. Distributional Structure. Word, 10 (2/3), pp. 146-162.

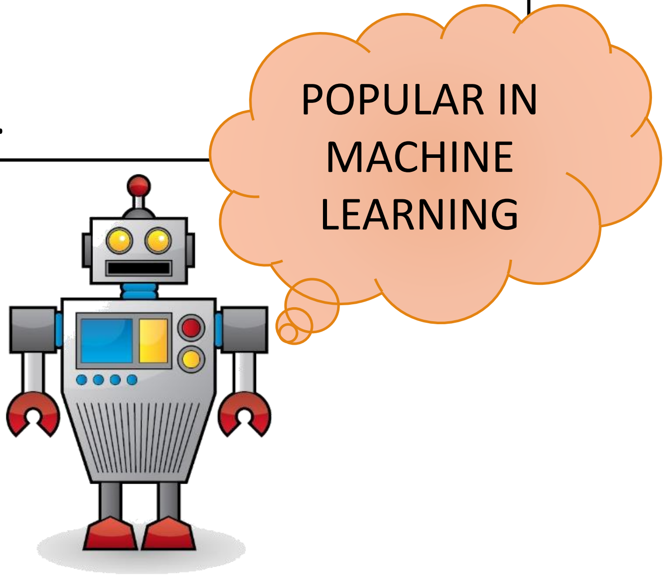
Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Unordered set of words

The dog bit the man = The man bit the dog

- No grammar
- No word order
- Just a bag of words…



Harris, Zellig. 1954. Distributional Structure. Word, 10 (2/3), pp. 146-162.

Language Models

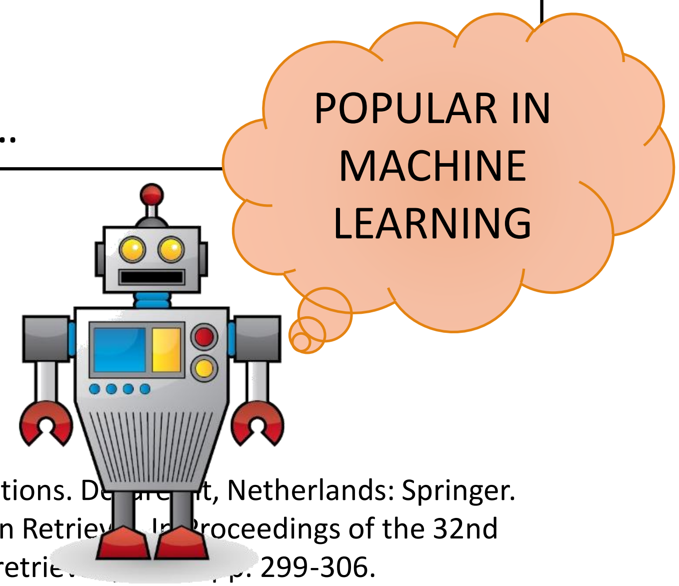
1. Bag-of-words
2. N-gram
3. Skip-gram

Unordered set of words

The dog bit the man = The man bit the dog

- No grammar
- No word order
- Just a bag of words...

Modifications:
- Positional Language Model
- Bag-of-concept



- Harris, Zellig. 1954. Distributional Structure. Word, 10 (2/3), pp. 146-162.
- E. Cambria and A. Hussain. 2012. Sentic Computing: Techniques, Tools, and Applications. D[illegible]t, Netherlands: Springer.
- Yuanhua Lv and ChengXiang Zhai. 2009. Positional Language Models for Information Retriev[illegible] In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrie[illegible] p. 299-306.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = set of n-long ordered sub-sequences of words.

The dog bit the man

2grams:
the dog | dog bit | bit the | the man

3grams:
the dog bit | dog bit the | bit the man

4grams:
the dog bit the | dog bit the man

- C. E. Shannon. 1948. A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379-423 (623-656), 1948.
- A.A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. Reprinted in Appendix B of: R. Howard. 1971. Dynamic Probabilistic Systems, Vol. 1: Markov Chains. John Wiley and Sons.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

POPULAR IN MACHINE TRANSLATION

Sentence = set of n-long ordered sub-sequences of words.

The dog bit the man

2grams:
the dog | dog bit | bit the | the man

3grams:
the dog bit | dog bit the | bit the man

4grams:
the dog bit the | dog bit the man

- C. E. Shannon. 1948. A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379-423 (623-656), 1948.
- A.A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. Reprinted in Appendix B of: R. Howard. 1971. Dynamic Probabilistic Systems, Vol. 1: Markov Chains. John Wiley and Sons.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = set of n-long ordered sub-sequences of words.

(1) John went to school today.

John went 👍
went to 👍
John * school 👎

- C. E. Shannon. 1948. A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379-423 (623-656), 1948.
- A.A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. Reprinted in Appendix B of: R. Howard. 1971. Dynamic Probabilistic Systems, Vol. 1: Markov Chains. John Wiley and Sons.

Problem definition

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

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).

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.

John went
went to
John * to

3gram: John went to
1skip2gram: John _ to

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.

John went
went to
John * to

To do this you need to…

3gram: John went to
1skip2gram: John _ to

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

$$\partial_1 w_{i-n+1}^{i} = w_{i-n+2}^{i}$$

$$\hat{P}_{\text{MKN}}(w_i | w_{i-n+2}^{i-1}) = \hat{P}_{\text{MKN}}(w_i | \partial_1 w_{i-n+1}^{i-1})$$

Skip-gram model with modified Kneser-Ney Smoothing

$$\hat{P}_{\text{MKN}}(w_i | (w_{i-n+1}^{i-1})) = \frac{\max\{N_{1+}(\bullet w_{i-n+1}^{i}) - D(c(w_{i-n+1}^{i})), 0\}}{N_{1+}(\bullet w_{i-n+1}^{i-1} \bullet)} + \gamma_{mid}(w_{i-n+1}^{i-1}) \hat{P}_{\text{MKN}}(w_i | w_{i-n+2}^{i-1})) \quad \ldots$$

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.
(2) John went to this awful place many people tend to generously call school today.

John went
went to
John * to
John * school
John * to * today

And still don't get the whole picture.

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Skip-grams cannot help extracting such patterns because…

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.
(2) John went to this awful place many people tend to generously call school today.

John went
went to
John * to
John * school
John * to * today

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.
(2) John went to this awful place many people tend to generously call school today.

John went
went to
John * to
John * school
John * to * today

1. The “skip” can appear only in one place.

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

1. **The "skip" can appear only in one place.**
2. **The same number of skips needs to be retained for each gap.**

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.
(2) John went to this awful place many people tend to generously call school today.

John went 👍
went to 👍
John * to 👍
John * school 👎
John * to * today 👎

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

1. The "skip" can appear only in one place.
2. The same number of skips needs to be retained for each gap.
3. Full control of the skip-length.

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.

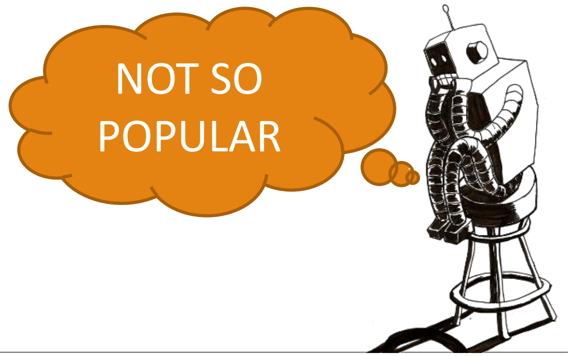
w s{1} w s{1} w

(2) John went to this awful place many people tend to generously call school today.

w s{1} w s{10} w

John went
went to
John * to
John * school
John * to * today

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.



Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.

w s{1} w s{1} w

≠

(2) John went to this awful place many people tend to generously call school today.

w s{1} w s{10} w

John went
went to
John * to
John * school
John * to * today

1. **The "skip" can appear only in one place.**
2. **The same number of skips needs to be retained for each gap.**
3. **Full control of the skip-length.**

- Xuedong Huang, Fileno Alleva, Hsiao-wuen Hon, Mei-yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II Speech Recognition System: An Overview, Computer, Speech and Language, Vol. 7, pp. 137–148.
- Guthrie, D., Allison, B., Liu,W., Guthrie, L., &Wilks, Y. (2006). A closer look at skip-gram modelling. In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006), pp. 1-4.
- Rene Pickhardt, Thomas Gottron, Martin Korner, Paul Georg Wagner, Till Speicher, Steffen Staab. 2014. A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1145-1154.

Language Models

1. Bag-of-words
2. N-gram
3. Skip-gram

Solution & simplification:

$$\binom{n}{k}$$

Language combinatorics

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

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

- 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.

Language combinatorics

<u>SPEC – Sentence Pattern Extraction arChitecture</u>

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

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

Extract patterns from all sentences and calculate occurrence.

- 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.

Language combinatorics

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

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

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

Score for one sentence $score = \sum w_j, (1 \geq w_j \geq -1)$

And then classify/ compare emotive sentences with non-emotive

i, Rafal Rzepka, Kenji Araki and Yoshio Momouchi. 2011. Language combinatorics: A sentence pattern extraction
sed on combinatorial explosion. International Journal of Computational Linguistics (IJCL), Vol. 2, Issue 1, pp. 24-36.

Language combinatorics

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

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

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

Score for one sentence $score = \sum w_j, (1 \geq w_j \geq -1)$

- 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.

Language combinatorics

1. Bag-of-words
2. N-gram
3. Skip-gram
4. Language combinatorics

Sentence = some words within an n-gram can be skipped over

(1) John went to school today.
(2) John went to this awful place many people tend to generously call school today.

John went
went to
John * to
John * school
John * to * today

Applications

Applications

1. Emotive / non-emotive [50 emotive and 41 non-emotive]
2. Future reference sentences [130 + 130 sentences]
3. Cyberbullying [1500 + 1500 sen.]
4. Conversations (male / female, social distance close / far, students / adults, …) [4000 sen., 6000 sen.]
5. Detection of depressive tendencies [10,000 sen.]
6. Determining specific emotions (joy, anger, fear, …) [~100 sen. x 10 classes (multiclass)]

- Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki. 2014. Automatic Extraction of Emotive and Non-emotive Sentence Patterns, In Proceedings of The Twentieth Annual Meeting of The Association for Natural Language Processing (NLP2014), pp. 868-871, Sapporo, Japan, March 17-21.
- Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki. 2014. Emotive or Non-emotive: That is The Question, In Proceedings of 5th Workhsop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2014), pp. 59-65, held in conjunction with The 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), Baltimore, USA, June 22-27.
- Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki. 2014. Detecting emotive sentences with pattern-based language modelling. In Proceedings of 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2014), Gdynia, Poland (to appear).
- Michal Ptaszynski, Dai Hasegawa, Fumito Masui, Hiroshi Sakuta, Eijiro Adachi. 2014. How Differently Do We Talk? A Study of Sentence Patterns in Groups of Different Age, Gender and Social Status. In Proceedings of The Twentieth Annual Meeting of The Association for Natural Language Processing (NLP2014), pp. 3-6, Sapporo, Japan, March 17-21.
- Yoko Nakajima, Michal Ptaszynski, Hirotoshi Honma, Fumito Masui. 2014. Investigation of Future Reference Expressions in Trend Information. In Proceedings of the 2014 AAAI Spring Symposium Series, "Big data becomes personal: knowledge into meaning – For better health, wellness and well-being –", pp. 31-38, Stanford, USA, March 24-26, 2014.

Conclusions and Future Work

- ✓ Little major development in language modelling
- ✓ None of the models catches the whole picture
- ✓ Presented a novel "pattern-based" language modelling method based on the idea of Language Combinatorics
- ✓ Applied the method to different datasets

In the near future:

- ➢ Apply to other data not limited to binary classification
- ➢ Analyze the behavior of different classifiers when trained on patterns

Thank you for your attention!

Michal Ptaszynski Kitami Institute of Technology

ptaszynski@ieee.org

http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/