SPASS: A Scientific Paper Writing Support System

Michal Ptaszynski Fumito Masui

Kitami Institute of Technology
Japan

Presentation outline

1. Motivation
2. Background
3. System Description
4. Evaluation
5. Conclusions
6. Future work

Motivation

- “All of the biggest technological inventions created by man - the airplane, the automobile, the computer - says little about his intelligence, but speaks volumes about his laziness. “ — Mark Kennedy

- “Efficiency is intelligent laziness.” — David Dunham

- “Progress isn't made by early risers. It's made by lazy men trying to find easier ways to do something.” — Robert A. Heinlein

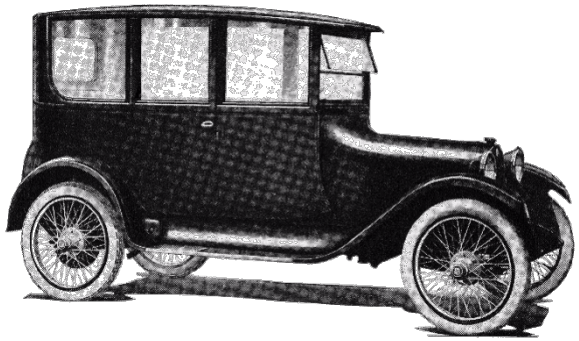
- “I don't think necessity is the mother of invention. Invention . . . arises directly from idleness, possibly also from laziness. To save oneself trouble.”

  — Agatha Christie, *An Autobiography*

- “Laziness is the first step towards efficiency.” — Patrick Bennett

Motivation

- History of mankind is based on laziness.
  "Where else can we save ourselves trouble?"



Motivation

- History of mankind is based on laziness.
  "Where else can we save ourselves trouble?"
- Automation – the keyword "What else can we automate?"

Motivation

- History of mankind is based on laziness.
  "Where else can we save ourselves trouble?"
- Automation – the keyword "What else can we automate?"

Motivation

Research process

- **creative part** (preparing descriptions of research background, literature review, discussion and detailed analysis of the results of experiments)
- **non-creative part** (laboriously preparing data for experiments, conducting the evaluation experiments, step-by-step manually changing feature sets to train and test classifiers, from the experiment results preparing tables, graphs, or descriptions of the results for the use in technical reports and scientific papers,

  Non-creative part of everyday research drill – laborious because requires the most of researcher's focus and precision

Motivation

People usually automate the non-creative part (laborious), and not the creative part (pleasant)

- Automated calculation process, but not what we use the calculations for
- Automated buying the drink but not drinking it
- Automated going to the theatre, but not watching the show

Non-creative - Laborious for us – easy for computers/ machines

Creative – pleasant for us difficult for computers

Background

- Weka (http://www.cs.waikato.ac.nz/ml/weka/)
- Wide range of machine learning algorithms; can be used as a stand-alone software, or can be called from a custom Java code to analyze data on the fly; allows data preprocessing, classification or clustering; provides simple visualizations of results; widely used.
- Needs especially prepared files with measurements in appropriate columns and cannot deal with plain unprocessed data (unprocessed collections of sentences, etc.); does not provide graphs in the format easily applicable in a research paper; does not provide natural language descriptions of results.

Background

- Nanba et al. (2000)
- Automatic generation of literature review
- Extracting from research papers short passages - summaries describing the essence of a paper.
- Dealt with automation of the creative part of research
- No further development.

Hidetsugu Nanba, Noriko Kando and Manabu Okumura. 2000. Classification of research papers using citation links and citation types: Towards automatic review article generation. In *Proceedings of 11th ASIS SIG/CR Classification Research Workshop*, pp. 117-134.

Background

- Shibata and Kurohashi (2005)
- Automatically generated summary slides from texts.
- Itemizing topic and non-topic parts extracted from syntactically analyzed text
- Similar to creating presentation slides from a scientific paper (one of our future tasks)
- In our method the parts created by the system are already grouped, which could help in the itemization.

Tomohide Shibata and Sadao Kurohashi. 2005. Automatic slide generation based on discourse structure analysis. In *Proceedings of IJCNLP 2005. Springer Berlin Heidelberg*, pp. 754-766.

Background

- Anonymous researchers involved in a campaign against dubious conferences - interesting, although not quite scientific (https://sites.google.com/site/dumpconf/)
- Generated pseudo-scientific papers by picking up random parts of actual papers
- Submitted fake-papers to specific conferences to verify review process
- Got accepted to the conferences
- Conclusion: review process of some conferences is not of the highest quality.
- Similar attempts in the future should consider the reputation and renown of the target conferences.

System Description

SPASS, or **S**cientific **Pa**per Writing **S**upport **S**ystem

1. Prepare data for experiment

2. Conduct the experiment under the conditions selected by user

3. Summarize results and prepare materials for a paper

System Description

- **User input**
- Text classification and analysis tasks.
- At present - up to two datasets (binary classification, “positive” and “negative”, but not limited to sentiment analysis)
- The user needs to prepare two separate files with sentences.
- The rest is done “with one click”.
- If the input consists of only one corpus the system will simply produce the most frequent patterns for the corpus.

System Description

- **Dataset Preprocessing**
- User can provide data on different level of abstraction

- **Tokenization:** All words, punctuation marks, etc. are separated by spaces.
- **Parts of speech (POS):** Words are replaced with their representative parts of speech.
- **Tokens with POS:** Both words and POS information is included in one element.
- **Lemmas**
- **Lemmas with POS**

System Description

Preprocessing example (Japanese)

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]*

System Description

- **Experiment Setup Preparation Module**
- Launch: one command

```
$ bash main.sh
```

- Options
- n-fold cross validation by adding parameter

`$ bash main.sh 5` (5-fold cross validation, default = 10)

- parameter: `1` (test data = training data)
- parameter: `-loo` (leave-one-out)

System Description

- **Pattern List Generation Module – all n-grams**

- **Weight calculations**
- Normalized
- Awarding length
- Awarding length and occurrence

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

$$w_l = w_j * k$$

$$w_{lo} = w_j * k * O$$

System Description

- **Pattern List Generation Module - all n-grams**

- **Pattern list modifications**
- Original pattern list
- Erasing all ambiguous patterns,
- Erasing zero-patterns (ambiguous patterns which appear in the same number on both sides).

...
1.0000:ー
1.0000:し
1.0000:〜
0.6923:よ
0.6667:だ
0.2000:に
0.2000:な
0.0182:。
0.0000:一緒 に
0.0000:ね 。
0.0000:から
0.0000:本
-0.0833:が
-0.1111:を
-0.2000:一緒
-0.6364:た 。
-1.0000:が すい た 。
-1.0000:い ます 。
-1.0000:が すい た
...

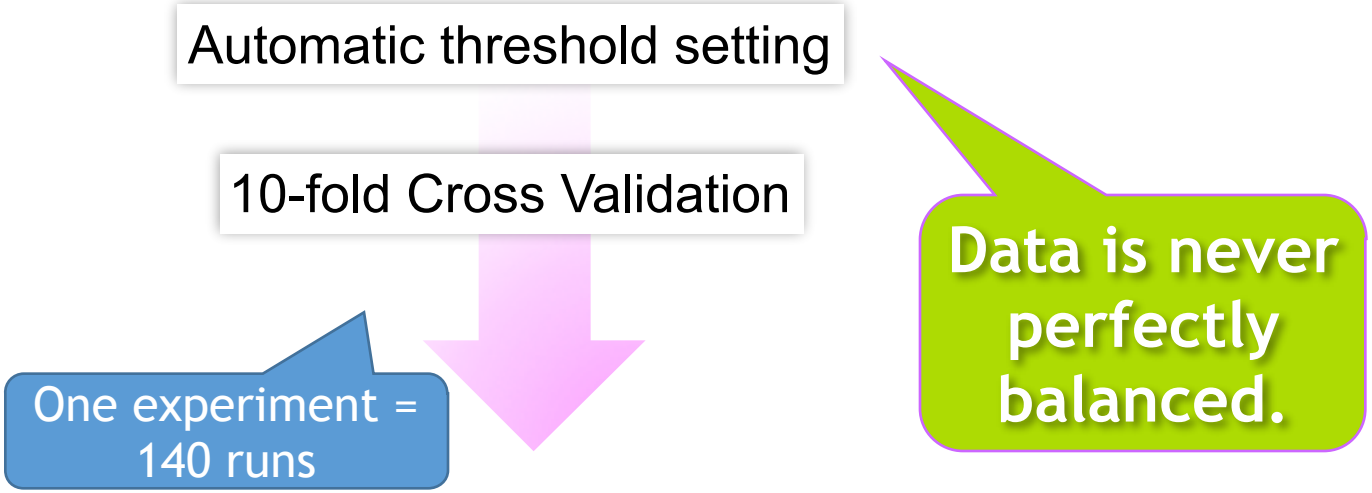
Experiment setup

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



System Description

- **Text Classification Module**

- Using weights of patterns calculate score for new input sentences
- Now – simple classifier

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

- In future add other classifiers (kNN, NN, SVM)

System Description

- **Contingency Table Generation Module**

Predicted class \ True class	Positive	Negative	Measures
Positive	True positive *TP*	False positive *FP*	Positive predictive value (PPV) $\frac{TP}{TP+FP}$
Negative	False negative *FN*	True negative *TN*	Negative predictive value (NPV) $\frac{TN}{FN+TN}$
Measures	Sensitivity $\frac{TP}{TP+FN}$	Specificity $\frac{TN}{FP+TN}$	Accuracy $\frac{TP+TN}{TP+FP+FN+TN}$

System Description

- **Contingency Table Generation Module**
- **LaTex Table Generation Module**

unmodified pattern list

Threshold	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00	-0.10	-0.20	-0.30	-0.40	-0.50	-0.60	-0.70	-0.80	-0.90	-1.00
Precision	0.00	0.00	0.10	0.40	0.35	0.50	0.61	0.60	0.57	0.57	0.58	0.57	0.55	0.54	0.52	0.52	0.51	0.50	0.50	0.50	0.50
Recall	0.00	0.00	0.02	0.10	0.10	0.20	0.30	0.44	0.56	0.82	0.92	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F-score	0.00	0.00	0.03	0.16	0.16	0.29	0.40	0.51	0.57	0.67	0.71	0.71	0.71	0.70	0.69	0.69	0.67	0.67	0.67	0.67	0.67
Accuracy	0.50	0.50	0.51	0.55	0.53	0.57	0.58	0.57	0.56	0.59	0.62	0.61	0.58	0.57	0.54	0.54	0.51	0.50	0.50	0.50	0.50
Specificity	1.00	1.00	1.00	1.00	0.96	0.94	0.86	0.70	0.56	0.36	0.32	0.26	0.16	0.14	0.08	0.08	0.02	0.00	0.00	0.00	0.00
phi-coefficient	0.00	0.00	0.03	0.15	0.08	0.18	0.20	0.17	0.14	0.19	0.30	0.29	0.23	0.20	0.12	0.12	0.03	0.00	0.00	0.00	0.00

zero deleted

Threshold	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00	-0.10	-0.20	-0.30	-0.40	-0.50	-0.60	-0.70	-0.80	-0.90	-1.00
Precision	0.00	0.00	0.10	0.40	0.40	0.50	0.59	0.60	0.58	0.56	0.58	0.57	0.55	0.54	0.53	0.52	0.51	0.51	0.50	0.50	0.50
Recall	0.00	0.00	0.02	0.10	0.12	0.24	0.30	0.46	0.62	0.82	0.92	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F-score	0.00	0.00	0.03	0.16	0.18	0.32	0.40	0.52	0.60	0.67	0.71	0.71	0.71	0.70	0.69	0.69	0.67	0.67	0.67	0.67	0.67
Accuracy	0.50	0.50	0.51	0.55	0.54	0.59	0.56	0.57	0.57	0.58	0.62	0.61	0.58	0.57	0.55	0.54	0.51	0.51	0.50	0.50	0.50
Specificity	1.00	1.00	1.00	1.00	0.96	0.94	0.82	0.68	0.52	0.34	0.32	0.26	0.16	0.14	0.10	0.08	0.02	0.02	0.00	0.00	0.00
phi-coefficient	0.00	0.00	0.03	0.15	0.12	0.22	0.15	0.17	0.16	0.18	0.30	0.29	0.23	0.20	0.13	0.12	0.03	0.03	0.00	0.00	0.00

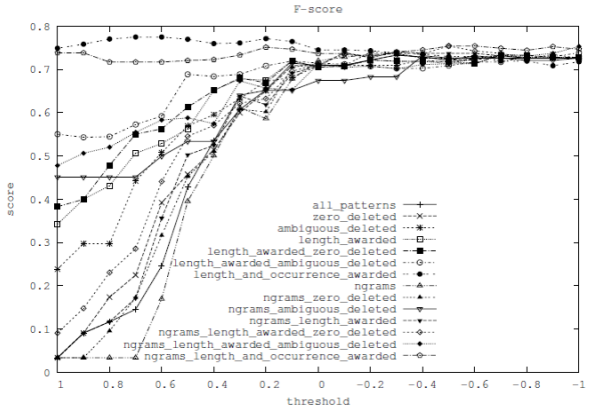
ambiguous deleted

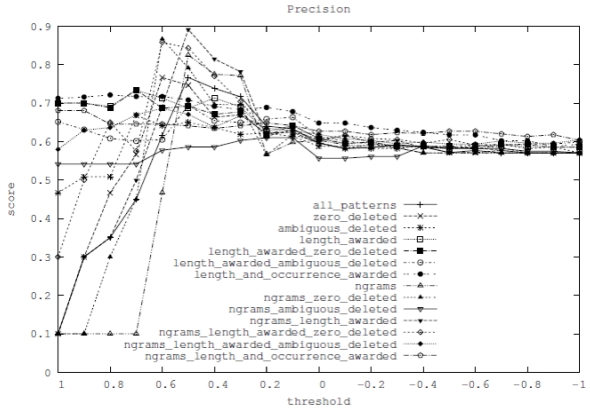
Threshold	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00	-0.10	-0.20	-0.30	-0.40	-0.50	-0.60	-0.70	-0.80	-0.90	-1.00
Precision	0.20	0.30	0.55	0.62	0.55	0.60	0.58	0.60	0.55	0.57	0.58	0.58	0.57	0.54	0.53	0.53	0.53	0.53	0.51	0.51	0.50
Recall	0.04	0.06	0.16	0.28	0.30	0.42	0.46	0.64	0.72	0.86	0.90	0.96	0.96	0.96	0.96	0.98	0.98	0.98	0.98	0.98	1.00
F-score	0.07	0.10	0.25	0.39	0.39	0.49	0.51	0.62	0.63	0.69	0.70	0.72	0.72	0.69	0.68	0.69	0.69	0.69	0.67	0.67	0.67
Accuracy	0.50	0.51	0.56	0.60	0.59	0.62	0.57	0.60	0.57	0.60	0.61	0.62	0.61	0.56	0.55	0.55	0.55	0.55	0.52	0.52	0.50
Specificity	0.96	0.96	0.96	0.92	0.88	0.82	0.68	0.56	0.42	0.34	0.32	0.28	0.26	0.16	0.14	0.12	0.12	0.12	0.06	0.06	0.00
phi-coefficient	0.00	0.03	0.17	0.24	0.20	0.27	0.15	0.21	0.15	0.23	0.27	0.31	0.29	0.16	0.13	0.13	0.13	0.13	0.05	0.05	0.00

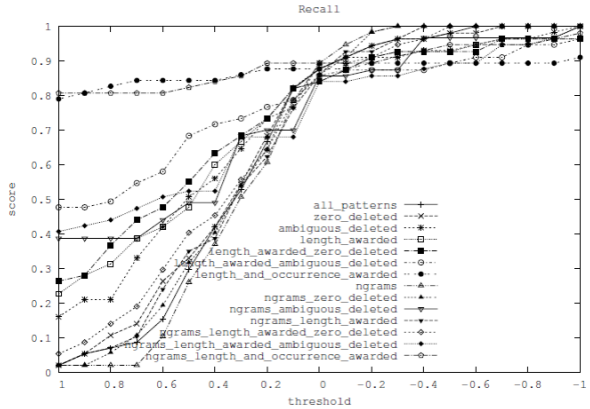
System Description

- **Contingency Table Generation Module**
- **LaTex Table Generation Module**
- **Graph Generation Module**

  Gnuplot (http://www.gnuplot.info/)







System Description

- **Contingency Table Generation Module**
- **LaTex Table Generation Module**
- **Graph Generation Module**
- **Result Analysis and Sentence Template Generation Module**

"When it comes to [weight calculations / pattern list modifications / ...], the highest [BEP / balanced F-score / Accuray / ...] was achieved by [zero_deleted / ambiguous _deleted / ...]".

System Description

- **Contingency Table Generation Modu**
- **LaTex Table Generation Module**
- **Graph Generation Module**
- **Result Analysis and Sentence Templ**
- **Most Useful Pattern Extraction Module**

  Useful in corpus linguistics

Emotive		Non-emotive	
freq.	example	freq	example
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

Evaluation

- No GUI
- One command -> everything done automatically = Nothing to ask users about usability features
- How to evaluate??

Evaluation

- No GUI
- One command -> everything done automatically = Cannot ask users about usability features
- How to evaluate??

1. Talk to users, ask their opinions
2. Use the system to perform research and get paper accepted (practical evaluation)

Evaluation

- User opinions
- “Add features/options”
  - Statistical significance calculation between all results
  - E-mail notification
  - Partial generation of presentation slides

Evaluation

- Practical evaluation

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 Nonemotive: 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 the 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014*, Gdynia, Poland, 15-17 September, 2014.

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.

Michal Ptaszynski, Dai Hasegawa, Fumito Masui. 2014. Women Like Backchannel, But Men Finish Earlier: Pattern Based Language Modeling of Conversations Reveals Gender and Social Distance Differences, In *9th International Conference on Natural Language Processing (PolTAL 2014)*, Samsung HLT Young Researchers Symposium, 2014.09.17-19, Warsaw, Poland.

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.

Analyze emotional and non-emotional sentences. Finding: completely automatic approach to extraction of emotional patterns can yield similarly good results to tools developed manually.

Conversation analysis task to find similarities between interlocutors of different age, gender, social distance and status. The system extracted several linguistic rules (confirmed statistically) which were previously unknown.

analysis of future related expressions for the task of future prediction from trend information.

Conclusions

- Research is a process requiring consisting of
  - creative tasks (literature review, writing a convincing discussion)
  - laborious non-creative tasks (data preparation, performing the experiments, preparing tables or graphs)
- Computers, poor at creative tasks, but good at laborious non-creative tasks should help researchers focus on the creative part of research
- We developed SPASS - a system which helps performing the laborious part of research.
- SPASS :
  - prepares the data for the experiments
  - automatically performs the experiments
  - from the results calculates the scores according to different measures (Precision, Recall, etc.)
  - creates tables in LaTex template containing all results
  - draws graphs informatively comparing each groups of results
  - generates descriptions of those results using sentence templates
- Does all that with a single command.

Future Work

- In the near future we plan to upgrade the system and implement additional functions:
- add various classification algorithms for more thorough evaluation.
- include automatic calculation of statistical significance of results.
- perform n-fold cross validation multiple times to further improve the objectivity
- add e-mail notification about the finalization of the experiment
- perform automatic summarization of sentence templates to increase the readability and informativeness of the descriptions.
- add generation of presentation slides in LaTex template from the results description.
- handle multi-label data

THANK YOU FOR YOUR ATTENTION!

Michal Ptaszynski Kitami Institute of Technology
ptaszynski@ieee.org

http://orion.cs.kitami-it.ac.jp/tipwiki/michal