Multi-label text classification of German language medical documents

Stephan Spat[a], Bruno Cadonna[a], Ivo Rakovac[a], Christian Gütl[b], Hubert Leitner[c], Günther Stark[c], Peter Beck[a]

[a] *Institute of Medical Technologies and Health Management, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria*

[b] *Institute for Information Systems and Computer Media, Graz University of Technology, Graz, Austria*

[c] *Steiermärkische Krankenanstaltenges. m.b.H., Graz, Austria*

Objective and Abstract

Generating metadata to improve quality of information retrieval in **medical free text documents**

⇒ **multi-label classification system** to classify medical free text documents written in German into predefined classes

- ***random sample of 1,500*** *medical free text* ***documents****, assigned manually to* ***1 to 8 categories*** *by a* ***domain expert*** *(physician)*

- *the sample was used to train and evaluate the* ***performance of 4 classification schemes: Naïve Bayes, k-NN, SVM*** *and* ***J48****,*

  additional tests of the effect of ***text preprocessing*** *were done*

- ***preprocessing improved performance****,*
  best results *were obtained by* ***J48 classification***

Introduction

- Increasing amount of **medical free text documents**
- Need for effective and timely retrieval
- **Additional metadata** to improve quality of information retrieval in medical free text documents

  ⇒ Development of a **multi-label text classification system** for medical free text documents originating from a **general hospital** in Austria

Methods

- random **sample of 1,500** free text **documents** written in German from an electronic medical record of a **general hospital** in Austria

- **domain expert** (physician) **manually classified** the retrieved documents into **one or more** of the following classes:

  - surgery,
  - vascular surgery,
  - casualty surgery,
  - internal medicine,
  - neurology,
  - anesthesia and intensive care,
  - radiology and
  - physiotherapy

Methods

- in average **1.47 labels** were assigned to a document
- automated **multi-label** text classification system was built in Java based on **Weka** [1], an **open source machine learning framework**
- **four** different kinds of **classification schemes** were compared:
  - **Naïve Bayes,**
  - **k-NN,**
  - **SVM and**
  - **J48**

Methods

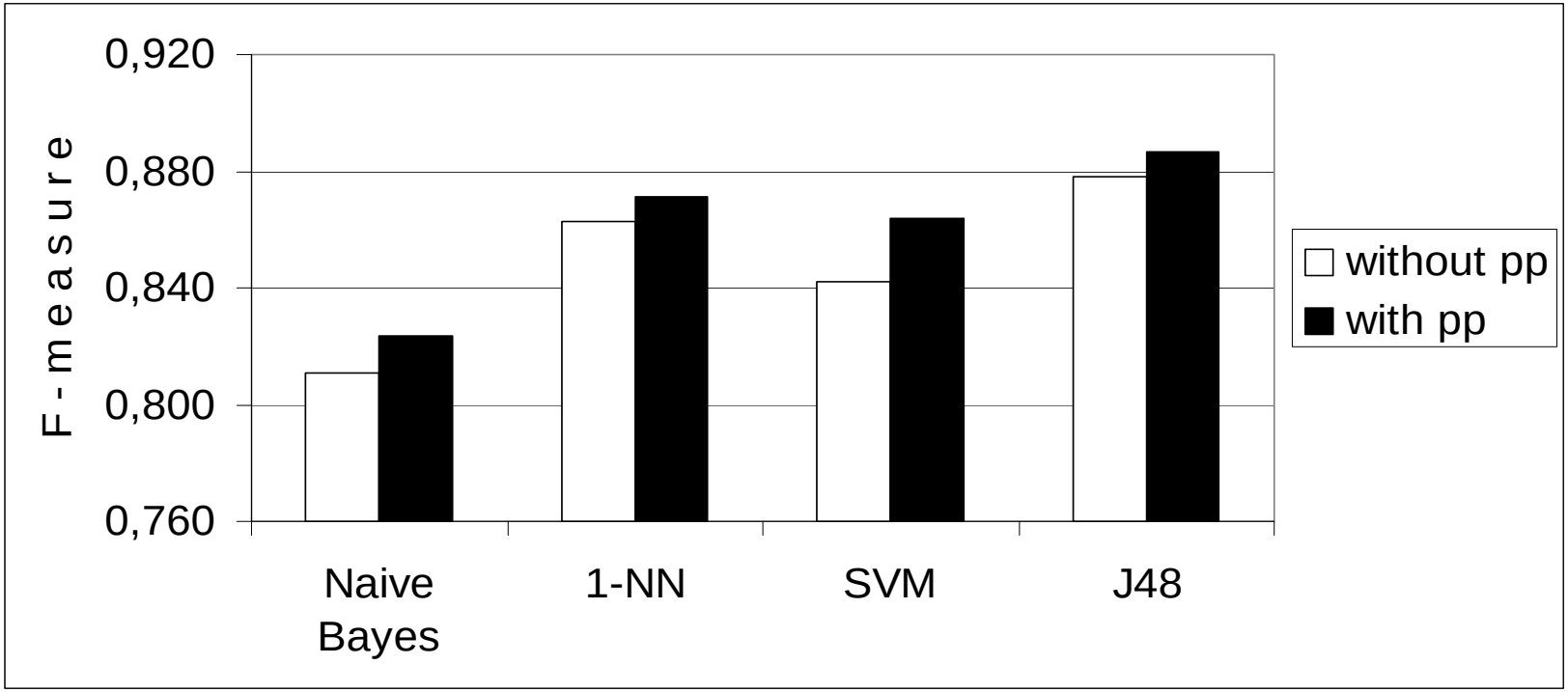
- **10-fold cross validation** was used for evaluation
- the influence of **text preprocessing** (e.g. stop-word-removal, stemming, lowercasing) was studied

Results

- Evaluation measure: **F-measure** [2]
- **J48 performed best**, followed by k-NN, SVM and Naïve Bayes
- J48 with text preprocessing achieved an **F-measure of 0.886**
- **text preprocessing improved the results**

Results

Comparison of the F-measure **with and without text preprocessing** (pp) of four different classification models:



Putting Knowledge to Work

Results

Evaluation results of document classification **with and without text preprocessing** and the improvement in percentage:

F-measure	Naïve Bayes	1-NN	SVM	J48
without text preprocessing	0.811	0.864	0.850	0.877
with text preprocessing	0.824	0.871	0.864	0.886
improvement	**+ 1.53** %	**+ 0.83** %	**+ 1.61** %	**+ 1.05** %

Results

Comparison of the F-measure achieved with document pre-processing **by twos** evaluated **classification models** in percentage:

	Naïve Bayes	1-NN	SVM	J48
Naïve Bayes	0.00 %	**- 5.45** %	**- 4.66** %	**- 7.07** %
1-NN	**+ 5.76** %	0.00 %	**+ 0.84** %	**- 1.72** %
SVM	**+ 4.88** %	**- 0.83** %	0.00 %	**- 2.53** %
J48	**+ 7.61** %	**+ 1.75** %	**+ 2.60** %	0.00 %

Conclusion

- **Classification of free text medical documents** written in German with machine learning classification schemes achieves **promising results,** comparable to [3]
- text preprocessing improves classification results
- **generated metadata** from classification is used in a prototype of an **information retrieval system** for score-calculation, thus influencing the display order of search results

References, Acknowledgement, contact details

References

- [1] Witten IH, Frank E: Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005
- [2] Hripcsak G, Rothschild AS: Agreement, the F-Measure, and Reliability in Information Retrieval. J Am Med Inform Assoc. 2005; 12(3): 296-298
- [3] Wilcox A, Hripcsak G: Classification algorithms applied to narrative reports. Proc AMIA Symp. 1999;:455-9

Contact details

- **Stephan Spat**
- Institute of Medical Technologies and Health Management,
  JOANNEUM RESEARCH Forschungsgesellschaft mbH, Elisabethstraße 11a, 8010 Graz, Austria
- Phone: +43 (0) 316 876 2157
- Email: stephan.spat@joanneum.at
- Homepage: http://www.joanneum.at/