Towards Bankruptcy Prediction:

Deep Sentiment Mining to Detect Financial Distress from Business Management Reports

Zahra Ahmadi[1], Peter Martens[1], Christopher Koch[2], Thomas Gottron[3], and Stefan Kramer[1]

[1] *Institut für Informatik*, *Johannes Gutenberg-Universität*, Mainz, Germany
[2] *FB Rechts- und Wirtschaftswissenschaften*, *Johannes Gutenberg-Universität*, Mainz, Germany
[3] *Innovation Lab*, *SCHUFA Holding AG*, Wiesbaden, Germany

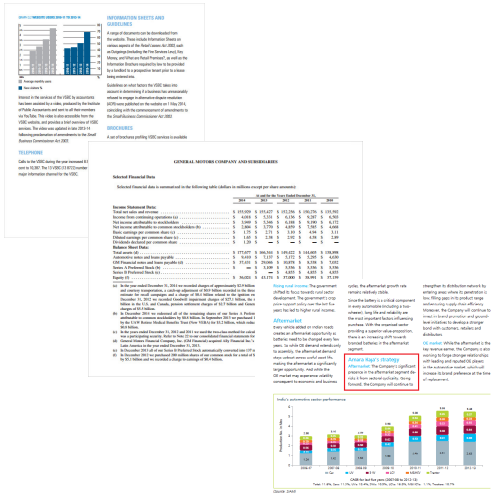
JG|U
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Introduction

Financial distress and bankruptcy is associated with very large costs

- Predict the bankruptcy likelihood of a company based on accounting-based indicates
- Limited predictive power of models on the backward-looking quantitative indicators
- Use another untapped unstructured source of information from texts published by or about companies

Challenges of the problem



$f(X)$?

suitable

target

variable

$L(X, f(X))$?

decent

prediction

performance

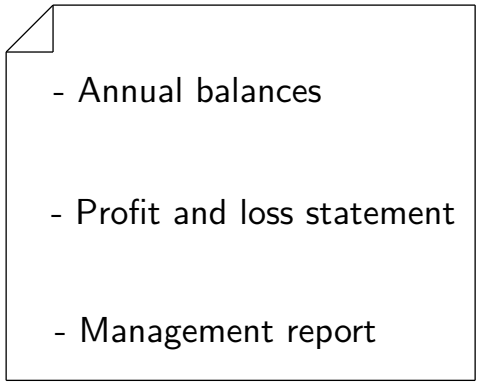
experts can

interpret

Data collection (1): management reports

Annual business reports from *Bundesanzeiger*

- 400,000 annual reports of 70,000 distinct mid-/big-sized companies between 2007 – 2015



- Not as standardized as *10-K forms* in the USA, we develop a manual strategy to excerpt the management reports:
  - Identify the headings of the attachments and remove the whole paragraph
  - Remove all tables to exclude profit and loss statements
  - Remove headings as they are typically neutral

Data collection (2): target class

Directly predicting insolvency as a target variable

- 20,000 applications in the German Insolvency Register platform
- 13,000 applications for small-sized companies and 1,000 for mid-/big-sized companies
- A company is insolvent once it applies for insolvency proceedings
- Prediction performance is not good

Data collection (2): Altman Z-Score

A target variable based on quantitative indicators

- A linear combination of 5 popular business ratios
- The coefficients were estimated by multiple discriminant analysis

$$Z'' = 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4,$$

$$X_1 = \frac{working\ capital}{total\ assets}, \qquad X_3 = \frac{earnings\ before\ interest\ and\ taxes}{total\ assets}$$

$$X_2 = \frac{retained\ earnings}{total\ assets}, \qquad X_4 = \frac{market\ value\ equity}{book\ value\ of\ total\ liabilities}$$

Data collection (3): *Amadeus* database

Calculate the Z"-Score of a company in a specific year with *Amadeus* information

- 2,586,023 records of 342,921 distinct German companies between 2004 – 2014 with 130 distinct attributes
- Match *Amadeus* and *Bundesanzeiger* using a plain-vanilla name by name matching

  $Amadeus : \{[\underline{amadeusID}, companyName, closeDate, attributes, ...]\}$
  $Companies : \{[\underline{companyID}, name, city, registryID]\}$

  $ManagementReports : \{[\underline{companyID}, content, \underline{publishDate}]\}$
- to choose the management report, we match

  $Amadeus.closeDate = ManagementReports.publishDate - 1$

Final dataset

Predict the Z"-Score class of a company for the next year:

$$\begin{cases} Z'' < 1.1 & \text{distressed zone} \\ 1.1 \leq Z'' < 2.6 & \text{grey zone} \\ Z'' \geq 2.6 & \text{safe zone} \end{cases}$$

	Reports	Distinct Companies
Distress Zone	3,094	783
Grey Zone	4,190	937
Safe Zone	5,374	1,576
Total	12,663	3,296

Proposed framework (1)

1. Filter sentences with *Multi-Class Correlated Pattern Mining* algorithm

   - NNs perform poorly on long texts, with increasing time and memory
   - Find the sentences with the most discriminative patterns for each class

   i. Generate n-grams of varied size for each sentence
   ii. Pick threshold θ from Pearson's χ^2 test and transform into min support for each class:

$$\theta_i = \frac{\theta N}{N^2 - n_i N + \theta n_i}$$

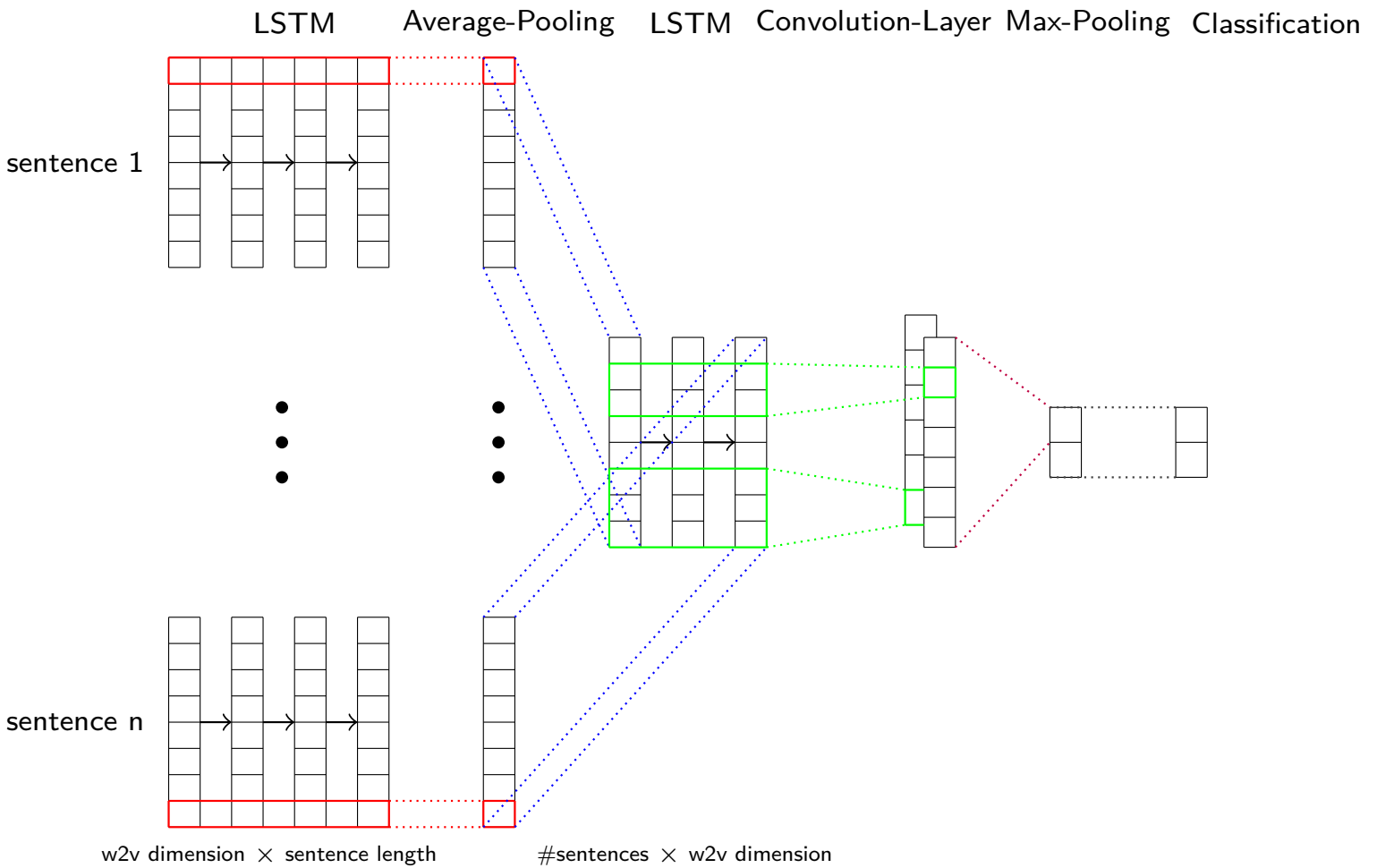
   iii. Find all frequent patterns and prune all patterns p for which

$$\sum_{i=1}^{d} \left(\frac{(a_i^{(p)} n_i - \sum_{j=1}^{d} \frac{(a_j^{(p)} n_j) n_i}{N})^2}{\sum_{j=1}^{d} \frac{(a_j^{(p)} n_j) n_i}{N}} + \frac{((1-a_i^{(p)}) n_i - \sum_{j=1}^{d} \frac{((1-a_j^{(p)}) n_j) n_i}{N})^2}{\sum_{j=1}^{d} \frac{((1-a_j^{(p)}) n_j) n_i}{N}} \right) < \theta$$

   iv. Merge the pruned pattern sets $\mathcal{F}_i$

Proposed framework (2)

2. Use Dependency Sensitive Convolutional Neural Network (DSCNN)
   - A multi-layer deep network with a convolutional layer on top of two LSTM networks

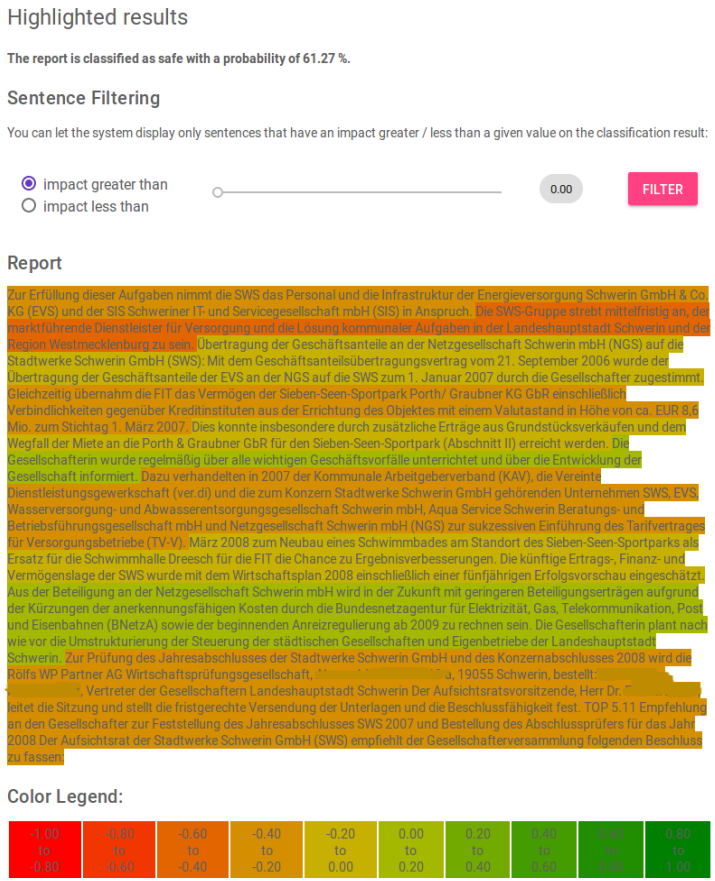


Visualization

Highlight those sentences with the largest impact on the classification

- Calculate the negative log-likelihood values from softmax function:

$$I_e(w) = \frac{S(e,c) - S(e,c,\neg w)}{S(e,c)}$$



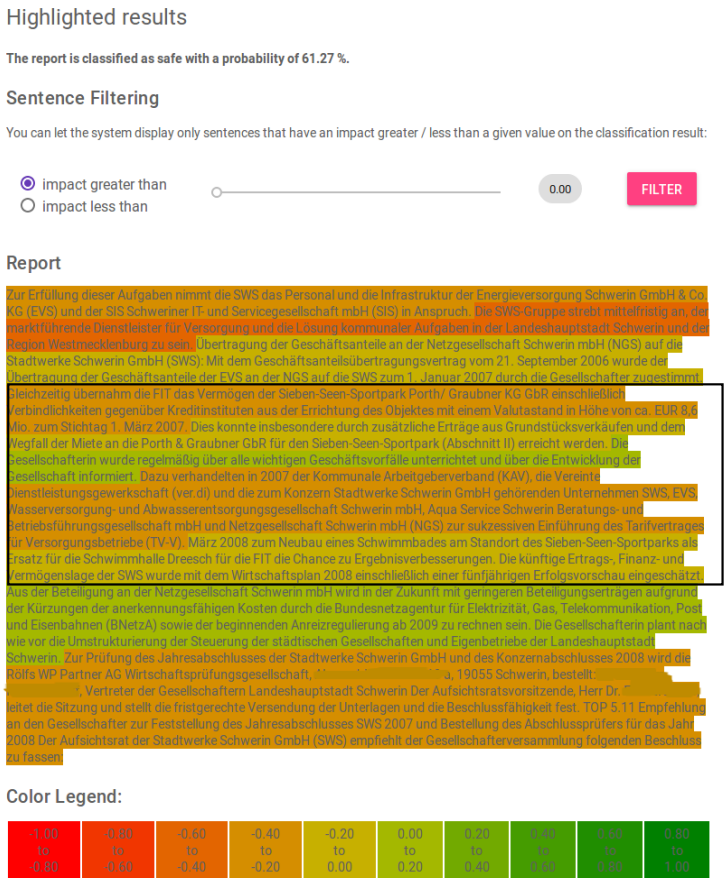
Visualization

Highlight those sentences with the largest impact on the classification

- Calculate the negative log-likelihood values from softmax function:

$$I_e(w) = \frac{S(e,c) - S(e,c,\neg w)}{S(e,c)}$$

Gleichzeitig übernahm die FIT das Vermögen der Sieben-Seen-Sportpark Porth/ Graubner KG GbR einschließlich Verbindlichkeiten gegenüber Kreditinstituten aus der Errichtung des Objektes mit einem Valutastand in Höhe von ca. EUR 8,6 Mio. zum Stichtag 1. März 2007. Dies konnte insbesondere durch zusätzliche Erträge aus Grundstücksverkäufen und dem Wegfall der Miete an die Porth & Graubner GbR für den Sieben-Seen-Sportpark (Abschnitt II) erreicht werden. Die Gesellschafterin wurde regelmäßig über alle wichtigen Geschäftsvorfälle unterrichtet und über die Entwicklung der Gesellschaft informiert. Dazu verhandelten in 2007 der Kommunale Arbeitgeberverband (KAV), die Vereinte Dienstleistungsgewerkschaft (ver.di) und die zum Konzern Stadtwerke Schwerin GmbH gehörenden Unternehmen SWS, EVS, Wasserversorgung- und Abwasserentsorgungsgesellschaft Schwerin mbH, Aqua Service Schwerin Beratungs- und Betriebsführungsgesellschaft mbH und Netzgesellschaft Schwerin mbH (NGS) zur sukzessiven Einführung des Tarifvertrages für Versorgungsbetriebe (TV-V). März 2008 zum Neubau eines Schwimmbades am Standort des Sieben-Seen-Sportparks als Ersatz für die Schwimmhalle Dreesch für die FIT die Chance zu Ergebnisverbesserungen. Die künftige Ertrags-, Finanz- und Vermögenslage der SWS wurde mit dem Wirtschaftsplan 2008 einschließlich einer fünfjährigen Erfolgsvorschau eingeschätzt.



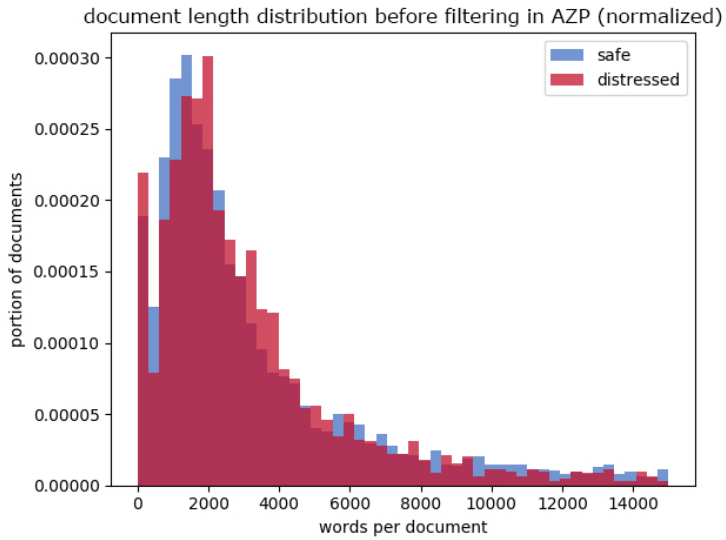
Data preparation

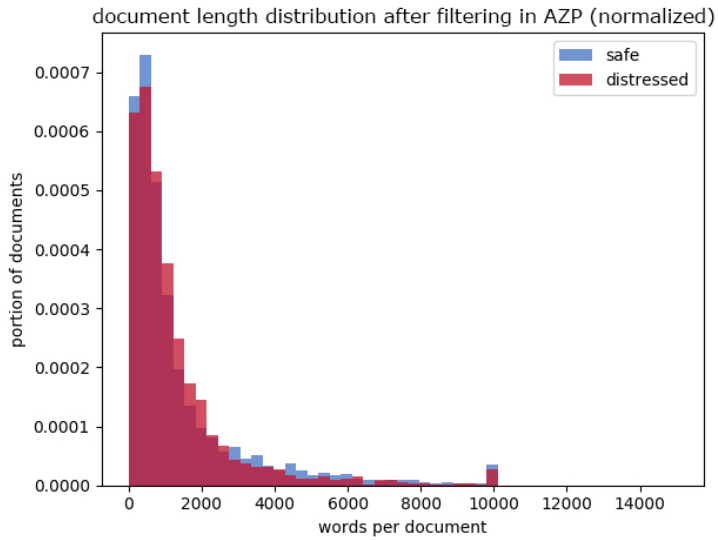
- Training data: reports with 2007 – 2012 reference date
- Test data: reports with 2013 – 2014 reference date
- To make a better separation between class labels, we only consider 2 classes: distressed and safe companies
- Apply some text preprocessing steps:
  - Remove all special characters except %, &, § ., , , (,)
  - Transform umlauts to become digraph
  - All tokens are lowercased
  - Numbers are replaced with a string that represents their category

Experiment settings

- Generate n-gram patterns for $n = 1, 2, 3, 4, 5$
- Choose θ_i so that the average text length ranges between 1000 – 1500 words

	distressed companies		safe companies	
	before filtering	after filtering	before filtering	after filtering
# words / report	4,210 ± 5,877	1,275 ± 1,567	4,494 ± 6,274	1,404 ± 1,809
# sentences / report	225 ± 307	57 ± 67	243 ± 331	64 ± 79
# words / sentence	18 ± 10	22 ± 10	18 ± 10	22 ± 10





Experimental results

- Apply a 2000-word cutoff on filtered documents for all deep networks
- In DSCNN: $Max_{sentencelength} = 32$, and $Max_{numberofsentences} = 63$
- Convolution layer contains 100 filters of size 3, 4, and 5

Method	Cohen's Kappa	Accuracy	Precision	Recall	F1-Score
SVM (2000-word cutoff)	0.4355	0.7635	0.7333	**0.9752**	0.8372
SVM (with filtering)	0.4528	0.7687	0.7409	0.9670	0.8390
SVM (complete text)	0.4475	0.7679	0.7373	**0.9752**	0.8397
CNN	0.0869	0.5712	0.4308	0.4308	0.4308
LSTM	0.2477	0.6549	0.5474	0.4834	0.5134
DSCNN (Wikipedia WV)	0.5139	0.7849	0.8022	0.5692	0.6659
DSCNN (Business WV)	0.5728	0.8069	0.8084	0.9045	0.8538
DSCNN (Business WV - 2000-word cutoff)	0.4045	0.7115	0.6006	0.6978	0.6456
DSCNN (Business WV - fine tuning)	**0.6544**	0.8385	**0.8821**	0.8551	0.8684
DSCNN (Wikipedia WV - fine tuning)	0.6523	**0.8414**	0.8544	0.8987	**0.8760**

Implementation challenges

✓ No business related pretrained word vectors
- Pretrained word vectors on German Wikipedia only contained about 30% of the relevant words
- Fine-tuning the word vectors, while the network learns, is the best solution

✓ Finding a viable minimum support threshold

× Long hours of training on web interface
- Django does not provide any interface to show the recent *loading-status* of the current process
- Switch to Ajax

Conclusion and future work

+ A novel framework for the prediction of insolvency for management annual reports based on the state-of-the-art deep networks
+ An effective filtering method for text reduction
+ A novel highlighting technique to scan a long report quickly

- Integrate the outputs of the framework with the *quantitative measures* of companies
- Validate the overall integrated predictive model over a longer period of time

Thank You!