Poster Presentations
Machine Translation for Film and TV Subtitles

✍️ Subtitles as an excellent basis for machine translation

✍️ Applied language technology: Neural Machine Translation

✍️ Translation industry: post-editing savings up to 30%
Aspect-based sentiment analysis to extract organoleptic wine profile

Ontology (aspects and opinions extraction) + Stacked RNN (relation extraction) = Wine profile & Buying intent
Hierarchical Classification for Economic News Articles

Fernando Benites  Mark Cieliebak

News Article
Twint erobert die Region Basler Kantonalbank und Bank Coop bieten Bezahl-App an
Weighted word overlap and word embeddings: A practical ensemble approach to Question Matching in a Dialogue Simulator

Don Tuggener, Institute of Computational Linguistics UZH / SUPSI / LifeLike

- Dialogue simulator for medical interviews (LifeLike)
- Replace mouse interaction with voice input
- Match user input to available questions using machine learning
- Handle non-matching inputs, speech-to-text errors

“What’s your work?”
Medical classifications include between 10 and 70 thousands single codes (or even more in case of SNOMED)

Stochastic systems have limitations to differentiate a big number of possible outcomes due to exponential growth of the training corpus needed

Semantic systems are advantageous since analyzing the inputs from inside and applying domain knowledge

The Concept Molecules facilitate development and maintenance of our interpreting system

The interpretation result, which contains implicit and explicit information from the input, allows for versatile data handling and analysis
We have answered this before: A study of the characteristics and solutions of the question retrieval and equivalence detection problem

Alireza Ghasemi, Silvia Quarteroni
ELCA Informatique SA
Making chatbots for the Swiss industry: Lessons learned

Lancelot Caron, Silvia Quarteroni
ELCA Informatique SA
BBDia: Diachronic Visualization of Semantically Related N-grams Using Word Embeddings

**Similar Words**
- Krieg
- Abstimmung, Republik
- Kanton Appenzell

**Similar Spellings**
- ähnliche Schreibweise:
  - Kriese
  - Republik
  - Kanton Appenzell

**Temporal Trends**

**Bigram Search**
- ähnliche Bedeutung:
  - kanton uri
  - kanton basel-landschaft
  - kanton schwyz
  - kanton thurgau
  - kanton basel-stadt
  - kanton oberwilien
  - kanton appenzell
  - kanton baselland
  - kanton st.Gallen
  - kanton baselstadt
A Twitter Corpus and Benchmark Resources for German Sentiment Analysis

Zurich University of Applied Sciences (ZHAW) - Winterthur, Switzerland
SpinningBytes AG, Küsnacht, Switzerland

Mark Cieliebak  Jan Deriu  Dominik Egger  Fatih Uzdilli

New Corpus: SB-10k

- 9738 German tweets
- Labels: "positive", "neutral", "negative" and "mixed"
- Each tweet annotated by 3 annotators
- Designed to cover a wide variety of unigrams and topics

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
<th>Mixed</th>
<th>Unknown</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1682</td>
<td>1077</td>
<td>5266</td>
<td>330</td>
<td>1428</td>
<td>9738</td>
</tr>
</tbody>
</table>

Benchmark for Sentiment Analysis in German

**SVM System**
- Features
  - n-grams, n = 1..4, POS-n-grams, n = 3..5, non-contiguous n-grams, n = 3..5
  - Character n-grams, n = 3..5
  - # upper-cased tokens, # of hashtags, # of POS tags
  - # continuous punctuations (max), last token punctuation (7, 1)
  - # mentioned words, # of negated tokens
- Lexicons: NRC-emotion, Bing/Lu, MGA, NRC-Hashtag/Sentiment, Sentiment140, Sentiment140-5-class, Rotterdam/Sentiment-3-class

**CNN System**

<table>
<thead>
<tr>
<th>Features</th>
<th>Training Corpus</th>
<th>Test Corpus</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>18/04</td>
<td>18/04</td>
<td>54.89</td>
<td>45.89</td>
<td>50.89</td>
</tr>
<tr>
<td>CNN</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>SVM</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>CNN</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>SVM</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>CNN</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>SVM</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>CNN</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>SVM</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
<tr>
<td>CNN</td>
<td>18/04</td>
<td>18/04</td>
<td>45.89</td>
<td>45.89</td>
<td>45.89</td>
</tr>
</tbody>
</table>

**Results**
- CNN outperforms SVM in all but one case (red)
- SB-10k generalizes better than MGS to unseen data
- Resulting F1-Scores match state-of-the-art

Corpus and source code are publicly available at www.spinningbytes.com/resources
Designing Cognitive Computing Architectures for Domain-Specific Decision Support Systems

Understanding
- Natural Language Processing
- Image Recognition

Reasoning
- Provide user-specific recommendations

Learning
- Machine Learning to generate or verify hypotheses

Johannes Forster (M.Sc.)
jforster@hm.edu

Nora Hollenstein (M.Sc.)
n.hollenstein@de.ibm.com
Taxonomy Induction using Hypernym Subsequences

Amit Gupta, Remi Lebret, Hamza Harkous, Karl Aberer
EPFL
An end-to-end pipeline for detecting and categorising customer complaints

Thomas Bögel, Henrik Matzen
IBM
Privacy in the Time of Bots: Answering Free-form Questions about Privacy Policies with Deep Learning

Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, Karl Aberer
EPFL, University of Michigan
KuBu - a chatbot for public transport

- Engage users
- Unique conversations
- Parse input via wit.ai

→ beta available: kubu.chat
Four Different Ways to Build a Chatbot About Movies

Erland Xhojaj
Yusuf Koc
Sandro Panighetti
Matteo Togni

Dirk Von Grünigen
Martin Weilemann
Hans Daniel Graf
Daniel Zürrer

Fernando Benites
Jan Deriu
Nico Neureiter
Pius von Däniken

Mark Cieliebak
Walter Eich
Stephan Neuhaus
Kurt Stockinger

Rule Based Question Answering

We developed a bot that is able to answer basic factual questions on movies, like “Who directed Inspector Clouseau?”. It is trained on a manually constructed database of movies to automatically infer the information and answer questions. The system was trained by hand to recognize how to answer factual questions, but other than a few hundred manually labelled examples, there is no need for additional training.

System Description

- (First) we trained the bot.
- The training data was crawled from Wikipedia articles on movies.
- We also included some questions that the bot could answer by itself.

Learning Dialogues End-to-End

We developed a neural network that learns how to respond to a dialogue partner. In contrast to the two chatbots above, the bot works out its own strategy for handling dialogues and is able to learn from its experiences. It is able to improve its performance over time by automatically adjusting its strategy and improving its understanding of the dialogue.

System Description

- (First) we trained the bot.
- The training data was crawled from Twitter articles on movies.
- We also included some questions that the bot could answer by itself.

Modelling Conversation Context

It is a second project where the bot is learning from a dialogue partner. In contrast to the two chatbots above, the bot works out its own strategy for handling dialogues and is able to learn from its experiences. It is able to improve its performance over time by automatically adjusting its strategy and improving its understanding of the dialogue.

System Description

- (First) we trained the bot.
- The training data was crawled from Twitter articles on movies.
- We also included some questions that the bot could answer by itself.

Microservice Architectures

The main development direction is cloud computing. The bot is able to interact with multiple services on demand. We used Kubernetes to deploy the bot on multiple cloud services. The bot is able to interact with multiple cloud services on demand. We used Kubernetes to deploy the bot on multiple cloud services.
Improving tourism marketing strategies by predicting the behavior of travelers using social media networks

- +200'000 comments and captions
- 3 years of pictures meta-information

Text Analysis – Semantic Concepts – Marketing Message