Sentiment Classification of Short Texts Using Artificial Neural Networks

Master's Qualifying Exam

Tiago Martinho de Barros
Prof. Dr. Hélio Pedrini
Prof. Dr. Zanoni Dias

University of Campinas
Institute of Computing

October 3, 2019

Table of Contents

1. Introduction
   - Problem Description
   - Objectives
2. Background
   - Concepts and Techniques
   - Related Work
3. Material and Methods
   - Methodology
   - Evaluation Metrics
   - Datasets
   - Baseline Results
   - Computational Resources
4. Work Plan and Schedule

Table of Contents

1. Introduction
   - Problem Description
   - Objectives
2. Background
   - Concepts and Techniques
   - Related Work
3. Material and Methods
   - Methodology
   - Evaluation Metrics
   - Datasets
   - Baseline Results
   - Computational Resources
4. Work Plan and Schedule

- Widespread Internet access.
- Opinion of end users and customers.
- Every day, 2.5 quintillion bytes of data are generated[1].
- Shifts in sentiment on social media have been shown to correlate with shifts in the stock market[2].
- Sentiment Analysis to gauge public opinion was an important factor in the 2012 presidential election in the USA[3].

[1] Domo, "Data Never Sleeps 5.0", 2017
[2] Ingle et al., "Sentiment Analysis of Twitter Data Using Hadoop", 2015
[3] Contently, "Social Media Sentiment Becomes Factor in Presidential Campaigns", 2012


Mark Duffy
@LordduFey
Replying to @F1
Schumacher was such a dirty driver.
24 7:01 PM - Jul 28, 2018

Introduction

Examples

Mark Duffy
@LordduFey
Replying to @F1
Schumacher was such a dirty driver.
24 7:01 PM - Jul 28, 2018

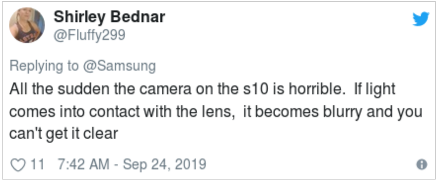
tami.
@Vetteleclerc
Ferrari got a 1-2 in one of their worst tracks.
Already love the 2nd half of the season #SingaporeGP
688 11:22 AM - Sep 22, 2019


Mark Duffy
@LordduFey
Replying to @F1
Schumacher was such a dirty driver.
24 7:01 PM - Jul 28, 2018

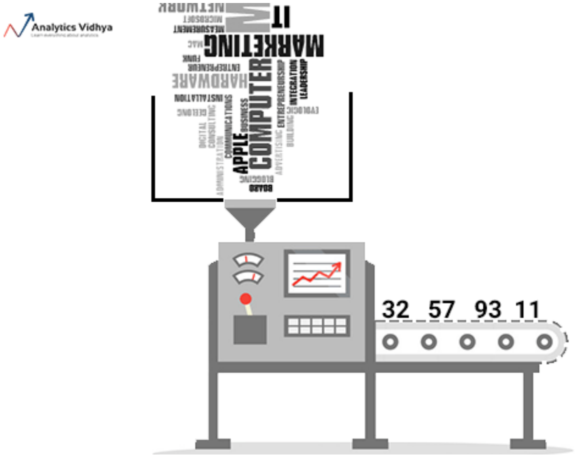

tami.
@Vetteleclerc
Ferrari got a 1-2 in one of their worst tracks.
Already love the 2nd half of the season #SingaporeGP
688 11:22 AM - Sep 22, 2019

Introduction

Problem Description



- Rule-based approach
- Lexicon-based approach
- Machine Learning approach



Embeddings from Language Models (ELMo): an example of unsupervised language representation learning method[1]

[1] Analytics Vidhya, "Learn ELMo for Extracting Features from Text", 2019

- **General objective**: Research and study state-of-the-art Sentiment Analysis techniques, and contribute to the field.
- **Contribution**: Propose and implement a competitive methodology to address the problem.

- **Specific objectives**:
  - Gain a deeper knowledge about pre-training unsupervised language representation learning methods.
  - Evaluate different options of further unsupervised pre-training and supplementary supervised training.
  - Experiment with and assess the different possibilities and combinations of preprocessing of the input text.
  - Propose a fine-tuning methodology that produces the competitive results in the considered scenarios.

- **Specific objectives**:
  - Gain a deeper knowledge about pre-training unsupervised language representation learning methods.
  - Evaluate different options of further unsupervised pre-training and supplementary supervised training.
  - Experiment with and assess the different possibilities and combinations of preprocessing of the input text.
  - Propose a fine-tuning methodology that produces the competitive results in the considered scenarios.

- **Specific objectives**:
  - Gain a deeper knowledge about pre-training unsupervised language representation learning methods.
  - Evaluate different options of further unsupervised pre-training and supplementary supervised training.
  - Experiment with and assess the different possibilities and combinations of preprocessing of the input text.
  - Propose a fine-tuning methodology that produces the competitive results in the considered scenarios.

- **Specific objectives**:
  - Gain a deeper knowledge about pre-training unsupervised language representation learning methods.
  - Evaluate different options of further unsupervised pre-training and supplementary supervised training.
  - Experiment with and assess the different possibilities and combinations of preprocessing of the input text.
  - Propose a fine-tuning methodology that produces the competitive results in the considered scenarios.

- **Specific objectives**:
  - Gain a deeper knowledge about pre-training unsupervised language representation learning methods.
  - Evaluate different options of further unsupervised pre-training and supplementary supervised training.
  - Experiment with and assess the different possibilities and combinations of preprocessing of the input text.
  - Propose a fine-tuning methodology that produces the competitive results in the considered scenarios.

Table of Contents

1 Introduction
- Problem Description
- Objectives

2 Background
- Concepts and Techniques
- Related Work

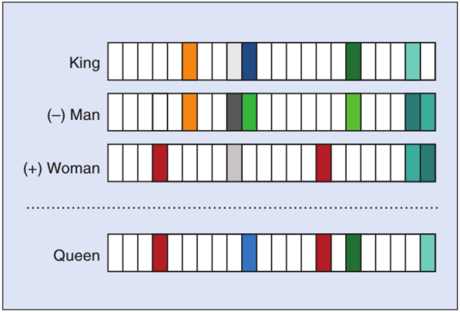
3 Material and Methods
- Methodology
- Evaluation Metrics
- Datasets
- Baseline Results
- Computational Resources

4 Work Plan and Schedule

- Document-level sentiment classification
- Sentence-level sentiment classification
- Aspect-level sentiment classification
    - *"The art direction of 'Star Wars: The Force Awakens' was amazing, but the plot was uninteresting, to say the least"*

- Document-level sentiment classification
- Sentence-level sentiment classification
- Aspect-level sentiment classification
    - *"The art direction of 'Star Wars: The Force Awakens' was amazing, but the plot was uninteresting, to say the least"*

- Language Modeling technique
- Transforms words into vectors of continuous real numbers
- Words with similar meanings tend to occur in similar contexts



An example of word embedding[1]

[1] Young et al., "Recent Trends in Deep Learning Based Natural Language Processing", 2018

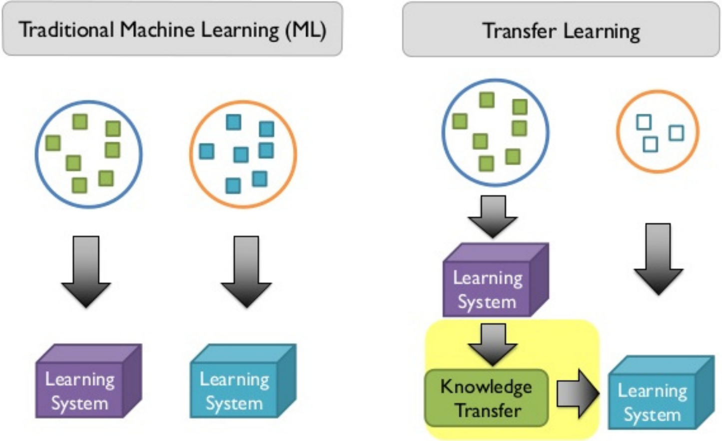
- Class of Machine Learning algorithms used in the learning of multiple levels of representation and abstraction, making it possible to model complex data relationships[1].
- They make use of several layers of non-linear processing units to extract and transform features.
- Each sucessive layer uses the output of the previous one as its input.
- Examples: LSTM[2], CNN[3], and Transformer[4].

[1] Goodfellow, Bengio, and Courville, "Deep Learning", 2016

[2] Hochreiter and Schmidhuber, "Long Short-Term Memory", 1997

[3] Kim, "Convolutional Neural Networks for Sentence Classification", 2014

[4] Vaswani et al., "Attention is All You Need", 2017



An illustration of Transfer Learning[1]

[1] Anchit Jain, "Improve Your Model Accuracy by Transfer Learning", 2018

- Dai and Le[1] proposed the supervised fine-tuning step after the unsupervised pre-training. Parameters obtained from the pre-training as a starting point for the supervised training model.
- **Embeddings from Language Models (ELMo)**[2]: contextualized word embedding.
- **Universal Language Model Fine-Tuning (ULMFiT)**[3]: addressing issues of over-fitting and catastrophic forgetting.

[1] Dai and Le, "Semi-Supervised Sequence Learning", 2015
[2] Peters et al., "Deep Contextualized Word Representations", 2018
[3] Howard and Ruder, "Universal Language Model Fine-Tuning for Text Classification", 2018

- **Generative Pre-trained Transformer (GPT)**[1]: combines unsupervised pre-training with Transformers, as opposed to LSTMs.
- **Bidirectional Encoder Representations from Transformers (BERT)**[2]: different training objective (*masked language modeling*).
- **XLNet**[3]: generalized auto-regressive pre-training method.
- **Supplementary Training on Intermediate Labeled-data Tasks (STILTs)**[4]: supplementary supervised training step between the unsupervised pre-training and the fine-tuning on the target task.

[1] Radford et al., "Improving Language Understanding by Generative Pre-Training", 2018

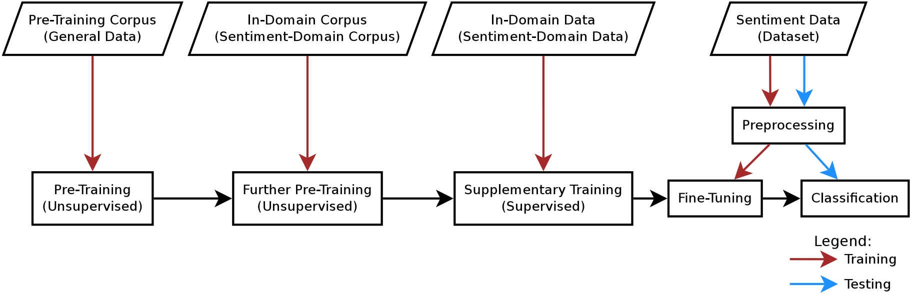
[2] Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 2019

[3] Yang et al., "XLNet: Generalized Autoregressive Pretraining for Language Understanding", 2019

[4] Phang, Févry, and Bowman, "Sentence Encoders on STILTs: Supplementary Training on Intermediate Labeled-data Tasks", 2018

Table of Contents

1. Introduction
   - Problem Description
   - Objectives
2. Background
   - Concepts and Techniques
   - Related Work
3. Material and Methods
   - Methodology
   - Evaluation Metrics
   - Datasets
   - Baseline Results
   - Computational Resources
4. Work Plan and Schedule



Proposed architecture for Sentiment Analysis

- Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

- Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- F1 score:

$$\text{F}_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Movie Reviews (MR)**[1]: short movie reviews dataset built using data from the review-aggregation website Rotten Tomatoes.
- **Customer Reviews (CR)**[2]: reviews for 14 products from Amazon.com and from CNET.
- **Multi-Perspective Question Answering (MPQA) Opinion Corpus**[3]: opinion polarity detection subtask of this question answering dataset.
- **Yelp Reviews Polarity (Yelp)**[4]: subset of the 2015 version of the Yelp dataset; 1 or 2 stars are negative, and 4 or 5 stars are positive.

[1] `http://www.cs.cornell.edu/people/pabo/movie-review-data`
[2] `http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html`
[3] `http://mpqa.cs.pitt.edu/corpora/mpqa_corpus`
[4] `http://www.yelp.com/dataset`

Dataset	Year	N	N^+	N^-	w	V
MR	2005	10662	5331	5331	21.01	18324
CR	2008	3746	2385	1361	18.38	5476
MPQA	2005	10514	3177	7337	3.04	5924
Yelp	2015	598000	299000	299000	134.04	214908

Comparative summary of datasets for sentiment classification

Movie Reviews (MR)					
Method	Accuracy (%)				
	A	B	C	D	E
Logistic Regression	78.4	78.5	78.9	78.0	78.6
Linear SVM	78.0	78.1	78.4	77.6	78.3
SVM with RBF kernel	78.7	79.2	78.6	78.9	79.6
Bernoulli Naïve Bayes	77.6	78.6	77.8	78.3	77.6
Multinomial Naïve Bayes	77.0	78.4	77.7	77.3	77.7
Random Forest	78.7	77.1	78.2	76.1	78.5

A	No preprocessing
B	Removing stop words (full list)
C	Removing stop words (minimal list)
D	Expanding contractions and removing stop words (full list)
E	Expanding contractions and removing stop words (minimal list)

Customer Reviews (CR)					
Method	Accuracy (%)				
	A	B	C	D	E
Logistic Regression	80.5	77.3	81.1	78.1	81.3
Linear SVM	78.7	79.5	77.1	77.1	79.2
SVM with RBF kernel	78.9	77.1	77.6	76.0	78.1
Bernoulli Naïve Bayes	77.6	78.1	77.9	77.6	77.1
Multinomial Naïve Bayes	79.5	75.7	78.7	75.5	78.7
Random Forest	79.2	77.6	78.1	76.8	75.7

A	No preprocessing
B	Removing stop words (full list)
C	Removing stop words (minimal list)
D	Expanding contractions and removing stop words (full list)
E	Expanding contractions and removing stop words (minimal list)

Multi-Perspective Question Answering (MPQA)					
Method	Accuracy (%)				
	A	B	C	D	E
Logistic Regression	90.0	89.5	89.4	89.4	89.4
Linear SVM	89.4	88.4	89.6	88.7	89.5
SVM with RBF kernel	91.2	89.8	90.6	89.9	90.5
Bernoulli Naïve Bayes	87.2	85.6	86.7	85.5	86.7
Multinomial Naïve Bayes	86.8	86.2	86.7	86.2	86.4
Random Forest	88.4	89.6	89.2	89.5	89.4

A	No preprocessing
B	Removing stop words (full list)
C	Removing stop words (minimal list)
D	Expanding contractions and removing stop words (full list)
E	Expanding contractions and removing stop words (minimal list)

Hardware:

- Processor: 3.5 GHz Intel i7-3770
- Main Memory: 32 GB of RAM
- Graphics Card: NVidia GeForce GTX TITAN Black
  - 2880 CUDA cores
  - Default memory DDR5 of 6 GB and 7 Gbps clock

Software:

- Python
- NumPy
- SciPy
- scikit-learn
- PyTorch
- TensorFlow
- Keras
- Sonnet

Table of Contents

1. Introduction
   - Problem Description
   - Objectives
2. Background
   - Concepts and Techniques
   - Related Work
3. Material and Methods
   - Methodology
   - Evaluation Metrics
   - Datasets
   - Baseline Results
   - Computational Resources
4. Work Plan and Schedule

Work Plan and Schedule

Activities	1st year				2nd year			
	1	2	3	4	1	2	3	4
Stage 1								
Literature review	•	•	•	•	•	•		
Data preparation		•	•	•				
Stage 2								
Selection of baseline neural network			•	•	•			
Development of the new model				•	•	•		
Experiments on the proposed methodology				•	•	•		
Stage 3								
Architecture refinement					•	•		
Result analysis					•	•	•	
Stage 4								
Result publication						•	•	•
Dissertation writing						•	•	•

Activity list for two-year Master's degree divided into trimesters