Seminarios
"Automatic detection of persuasion attempts on social networks"

Ruben Teimas

m47753@alunos.uevora.pt

Departamento de Informática
Escola de Ciências e Tecnologia

May 25, 2022

UNIVERSIDADE DE ÉVORA

Introduction

- ▶ Detection of persuasion attempts and hate speech on social networks has been in increasing demand as people spend more time on those platforms.
- ▶ Most of the content published is based on text, thus making it a text classification problem.
- ▶ Since we're trying to identify the techniques the problem is a multi-label problem.
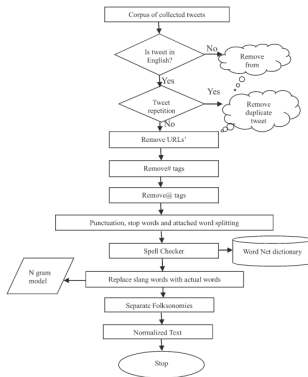
Figure: Text normalization flow.

Word representations and embeddings

▶ Bag of Words (*BoW*);
▶ Term frequency - Inverse document frequency (*Tf-Idf*);
▶ word2vec;

word2vec
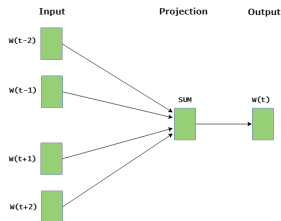
Figure: CBoW.
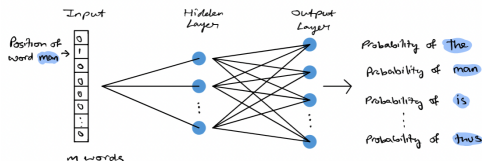
Figure: Skip-gram.

Binary Relevance

$L = \{l_1, l_2, l_3, l_4\}$

EX#	Label set
1	$\{l_1, l_3\}$
2	$\{l_2, l_4\}$
3	$\{l_1, l_3, l_4\}$
4	$\{l_4\}$

EX#	l_1
1	true
2	false
3	true
4	false

EX#	l_2
1	true
2	true
3	false
4	false

EX#	l_3
1	false
2	false
3	true
4	false

EX#	l_4
1	false
2	true
3	true
4	true
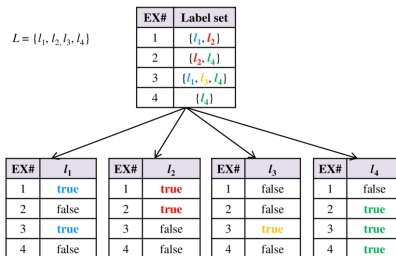
Figure: Binary relevance illustration.

Label Powerset

Example	x_1	...	x_7	Adventure	Drama	Comedy	Class
Game of Thrones	x_{11}	...	x_{17}	1	1	0	C110
The Big Bang Theory	x_{21}	...	x_{27}	0	0	1	C001
Rick and Morty	x_{31}	...	x_{37}	1	0	1	C101
College Romance	x_{41}	...	x_{47}	0	1	1	C011

Figure: Label Powerset illustration.

Real world usage

- ▶ Multi-label problem with 3 classes.
- ▶ Tried Random Forests, SVMs and Naive Bayes using Binary Relevance and Label Powerset.
- ▶ Best result came out of word unigram (feature extraction) Random Forests with Label Powerset, and it was an accuracy of 77.36%.
- ▶ In the next experiment also used part of speech (grammatical tagging) and got a little upgrade with an accuracy of 79.85%.
- ▶ word2vec was used instead of word unigram but got worst results mainly due to the small size of the corpus.

UNIVERSIDADE
DE ÉVORA
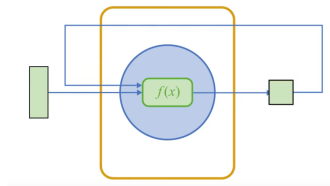
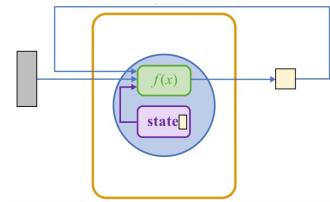Figure: RNN architecure.
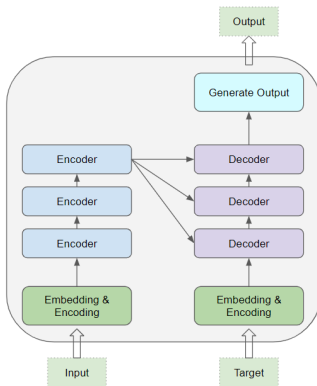

Figure: LSTM architecure.

Transformers

Figure: Over the top Transformer architecture.

Transformers

Lets have a look at the following sentences:

▶ The cat drank the milk because **it** was hungry.
▶ The cat drank the milk because **it** was sweet.

Figure: **Darker colors represent higher attention.**

Real world usage

- ▶ Multi-label problem with 20 classes.
- ▶ Used PTMs (pre-trained models) like BERT, RoBERTa and AlBERT.
- ▶ Use F1-micro score as its measure because it is good for multi-label classification.
- ▶ The final results for the winning group were 0.593.
- ▶ The corpus was small so the PTMs were just fine tuned with the training set available.
- ▶ The classes were imbalanced and Focal Loss (assigns higher weights to sparse samples) was used in order to balance it.

Thank you!

Thank you for your attention,
Ruben Teimas

GitHub github.com/TeimasTeimoso
LinkedId linkedin.com/in/ruben-teimas