

Summarization & Sentiment classification

Afonso Mendes, Shashi Narayan,
Shay B. Cohen, Sebastião Miranda, David Nogueira

amm@priberam.com, shashi.narayan@ed.ac.uk,
scohen@inf.ed.ac.uk, ssm@priberam.com, dan@priberam.com



Summarization : Goals

- Automatically synthesize a coherent summary of the most relevant events in a particular news article (single-document summarization)
- Enable media monitors to efficiently get the global picture and evolution of a story cluster (multi-document summarization)

Types of summaries

Extractive Summary:

- Washington (CNN)President Donald Trump insisted Thursday that he has remained consistent in his plans for a border wall, a day after his chief of staff, John Kelly, told Fox News he has changed his attitude on it.
- Trump was fuming after Kelly's Fox News interview, a source familiar with the matter told CNN, adding that the President hated the comments.
- Kelly on immigration: Trump has changed the way he's looked at a number of things
- Including services, the trade deficit with Mexico was \$68.3 billion from October 2016 through September 2017, the most recent figures available

Compressive Summary:

- ~~Washington (CNN)President Donald Trump insisted Thursday that he has remained consistent in his plans for a border wall, a day after his chief of staff, John Kelly, told Fox News he has changed his attitude on it.~~
- ~~Trump was fuming after Kelly's Fox News interview, a source familiar with the matter told CNN, adding that the President hated the comments.~~
- ~~Kelly on immigration: Trump has changed the way he's looked at a number of things~~
- ~~Including services, the trade deficit with Mexico was \$68.3 billion from October 2016 through September 2017, the most recent figures available~~

Abstractive Summary:

- Donald Trump said he has remained consistent in his plans for a border wall
- Trump was fuming after Kelly's Fox News interview
- Trump has changed the way he's looked at a number of things
- The trade deficit with Mexico was \$68.3 billion from October 2016, the most recent figures available

Summarization : Research within SUMMA

- ILP: Classical coverage-based optimization problem for multi-document summarization
- SideNet: Use side-information (e.g., image captions) to help generate the summary
- REFRESH: Reinforcement Learning based extractive summarization
- AMR: Parse text into a semantic graph, summarize this graph and generate text back again.
- STACKSUMM: Use stack-LSTMs to maintain an extractive and compressive summary representation

Summarization : Results

Models	CNN+DailyMail			
	R1	R2	RL	c _r
LEAD	39.6	17.7	36.2	.50
SUMMARUNNER Nallapati et al. (2017)	39.6	16.2	35.3	—
Refresh Narayan et al. (2018)	40.0	18.2	36.6	.48
StackSumm Sentence Extractive	41.7	18.6	37.8	.63
StackSumm BoW compressive	41.3	17.3	37.7	.77
StackSumm Compressive	40.9	18.0	37.4	.58
Pointer+Coverage See et al. (2017)	39.5	17.3	36.4	.87
LATENT Zhang et al. (2018)	41.1	18.8	37.4	—
Bottom-Up Gehrmann et al. (2018)	41.2	18.7	38.3	—

Table: Results on the combined CNN and DailyMail test set.

Sentiment classification: Goals

- Automatically attach to each storyline its corresponding sentiment polarity.
- Allow data analyst/media monitors to better understand the aggregate sentiment on social media.

Sentiment classification: System

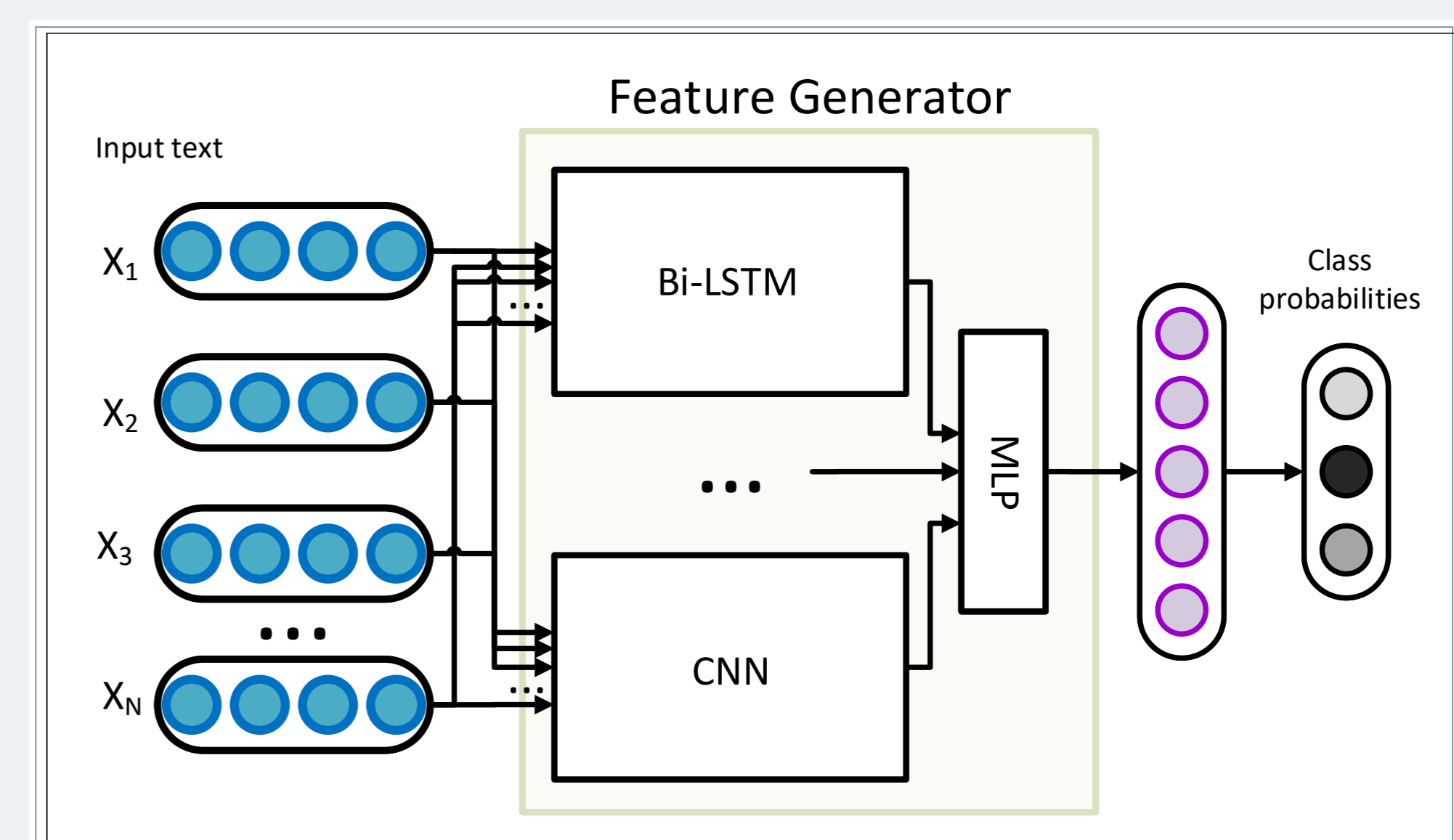
- **State-of-the-art industrial implementation**
- Able to leverage recent improvements in **language modelling representations** to boost the results even further.
- Pipeline is composed by:
 - 1 Twitter collection based on specific Twitter channels,
 - 2 alignment of social media text (tweets) to the corresponding articles (stories),
 - 3 sentiment polarity classification.

The screenshot shows a web interface for sentiment classification. At the top, there's a dropdown menu set to 'EN300Twitter(EN)'. Below it, a text input field contains a tweet: "That Mexico vs USA commercial with trump gets your blood boiling, Race war October 10th. Imagine that parking lot. Gadda". To the right of the input is an 'Analyse!' button. Below the input, a red circle with a sad face icon and the word 'negative' indicates the classification. Underneath, there are two columns: 'History' and 'Examples'. The 'History' column lists several tweets with their corresponding sentiment classifications (red for negative, green for positive, yellow for neutral). The 'Examples' column lists various tweets with their sentiment classifications.

Sentiment classification: System

The *Sentiment classification architecture* is largely based on the two best ranked systems from SemEval 2017 Task 4:

- Baziotis et al., 2017. "DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis.";
- Cliche. 2017. "BB twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs".



Sentiment classification architecture

Sentiment classification: Performance

English SemEval17 Task 4	macro-F1	Avg. recall	Accuracy
this system	0.666	0.668	0.650
Baziotis et al. 2017	0.677	0.681	0.651
Cliche. 2017	0.685	0.681	0.658

We achieve a very similar F1-score with a much lower model complexity, both in terms of the number of feature generator's submodules and the number of different pre-trained embeddings; and without relying on any external software for hyper-parameter optimization.

Portuguese dataset	macro-F1
this system	0.781
previous Priberam system	0.72