Deep learning for NLP in practice

Samuel Rönnqvist

Turku.ai, November 2, 2017

Deep learning for NLP in practice

Part 1: Background

Part 2: Practical, *NLP from scratch*

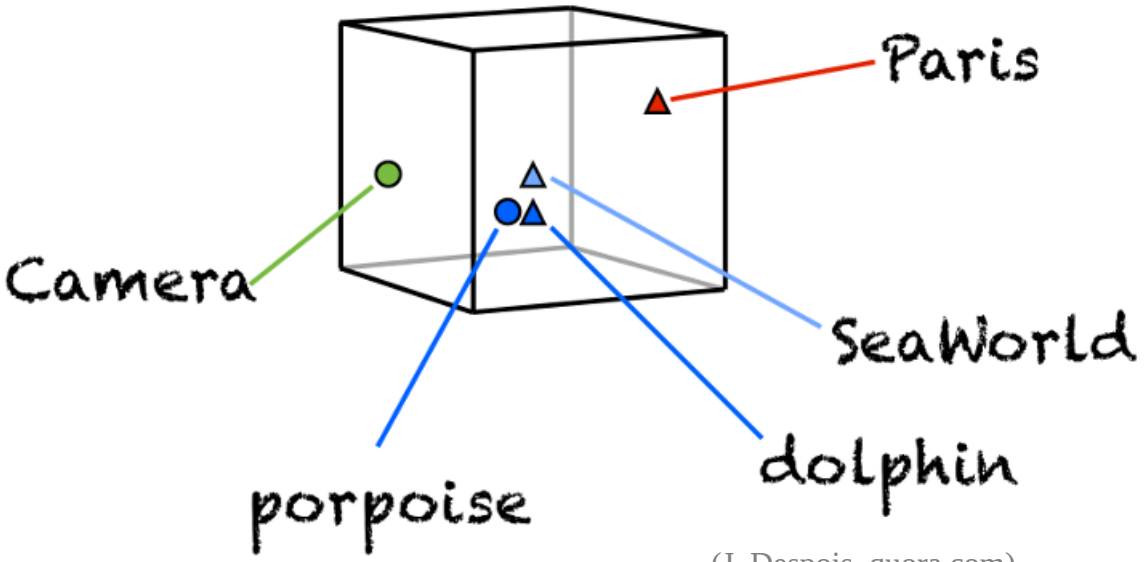
Deep learning for NLP

Deep learning: *representation learning* with neural networks

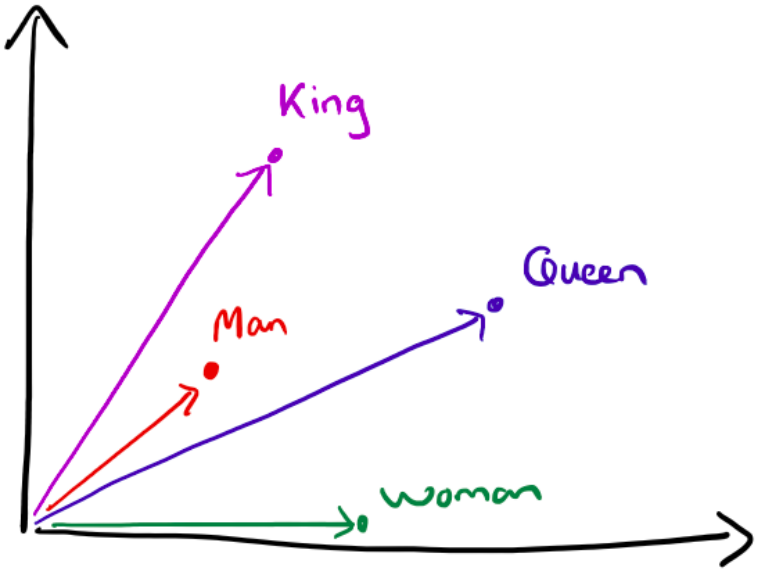
- Word vectors: learned semantic representations
- Recurrent neural networks: prediction with sequences

Word vectors

Vectors embed words in a semantic space



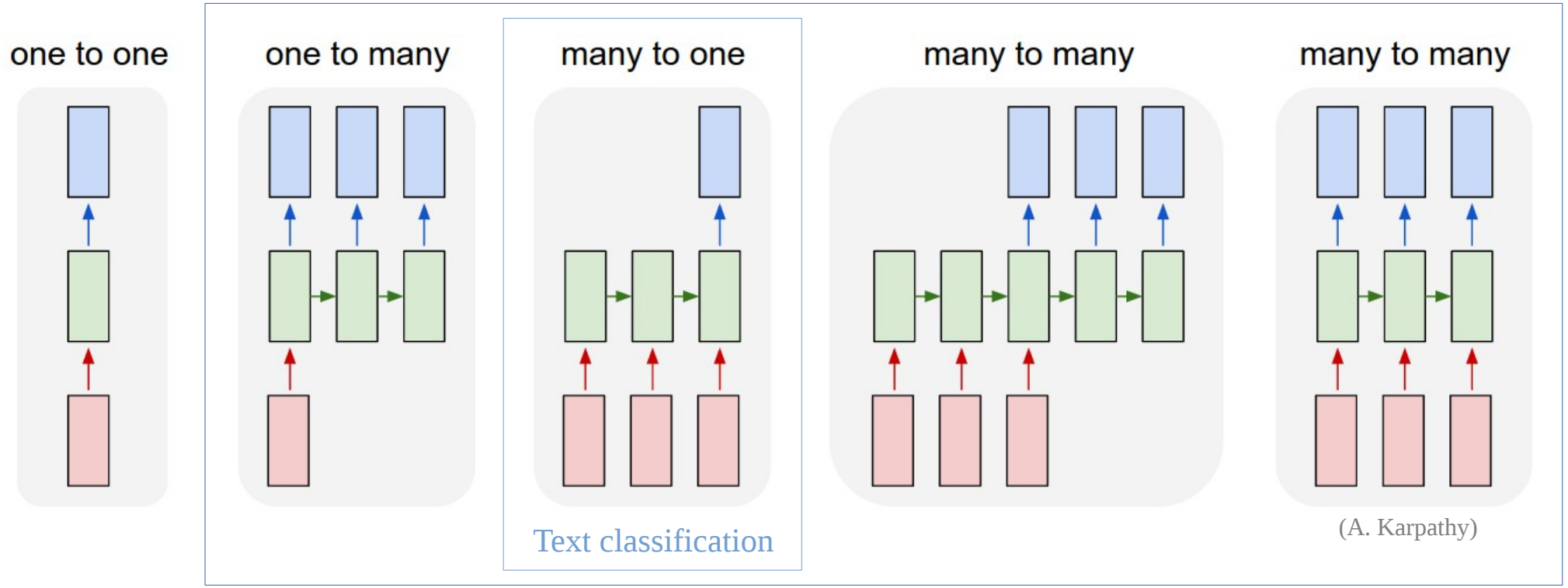
(J. Despois, quora.com)



(A. Colyer)

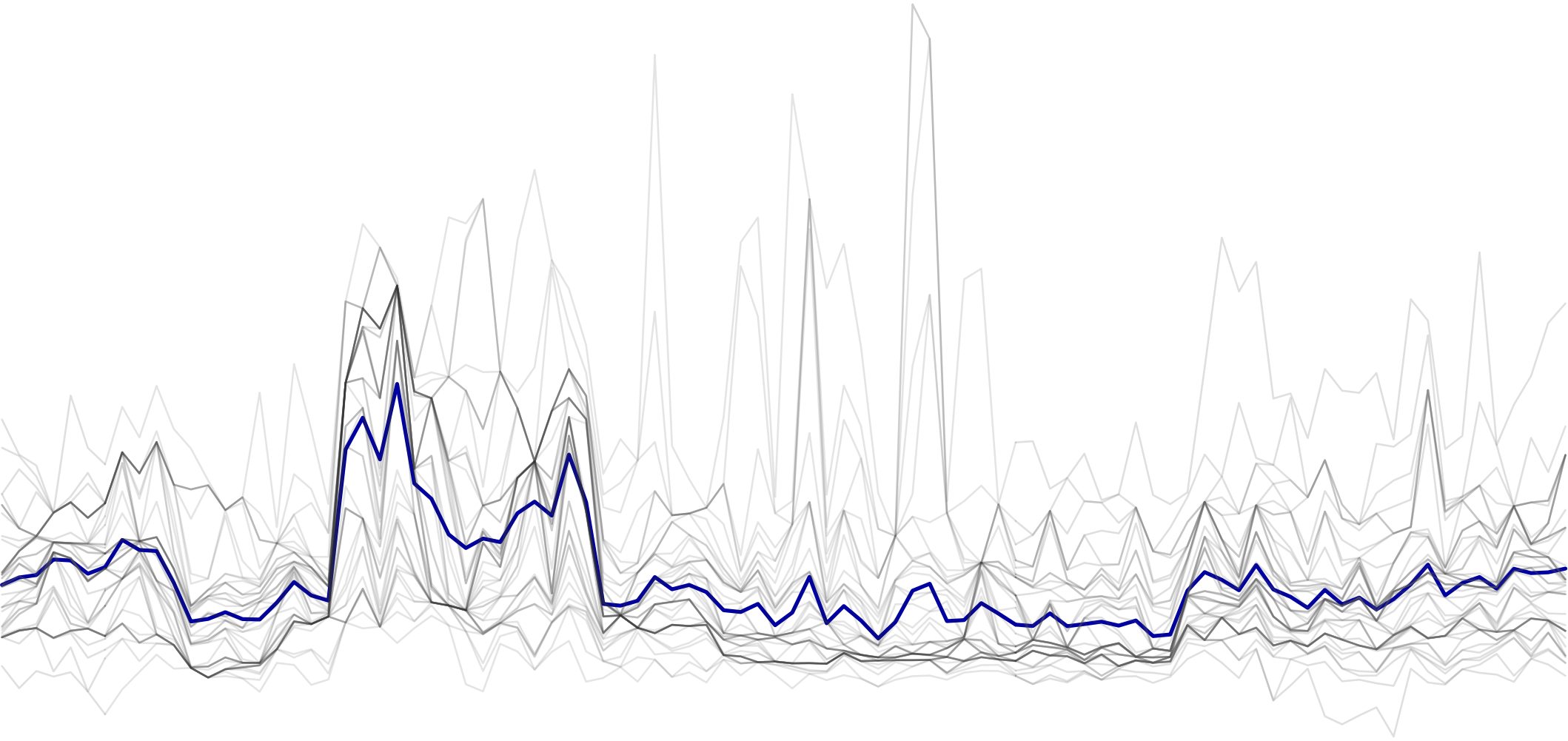
Recurrent modeling

NLP deals with sequences



Text classification examples

Case 1: Predicting bank distress from news (6.6M articles)

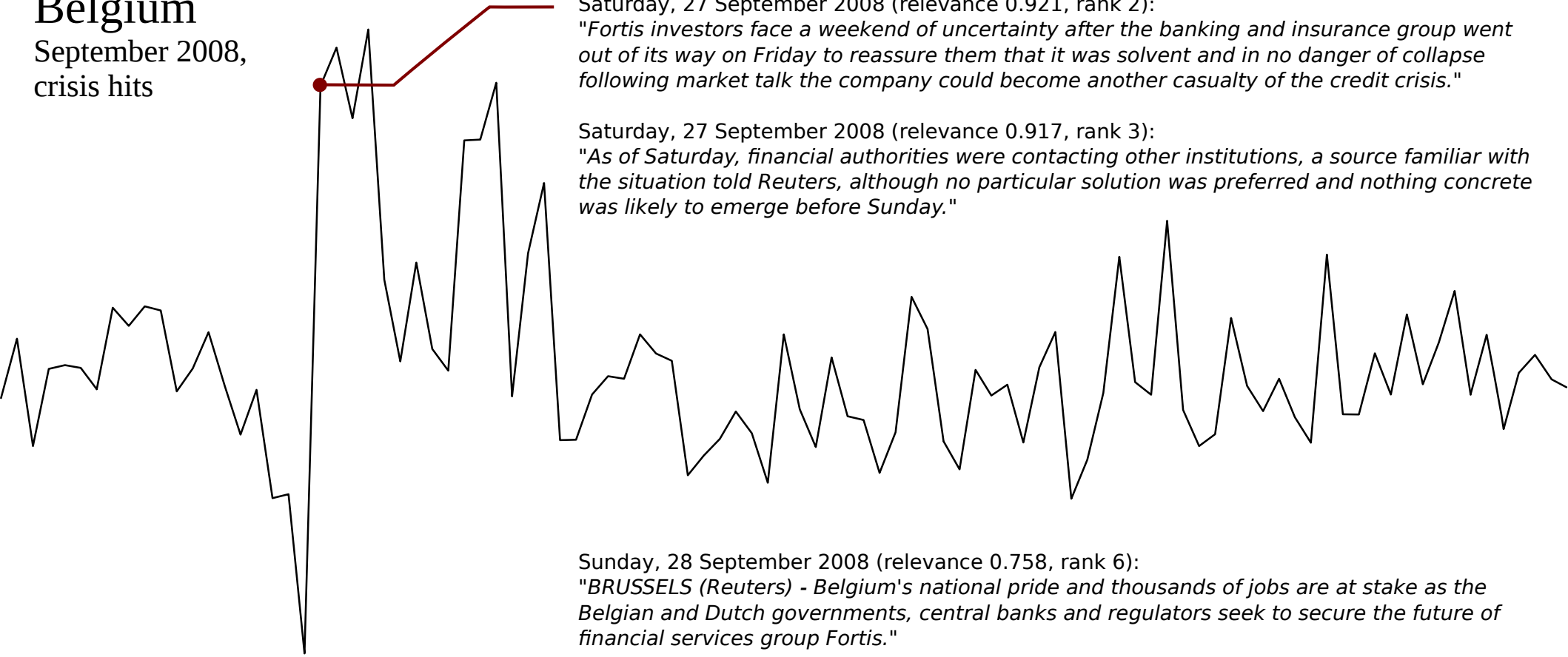


https://doi.org/10.1016/j.neucom.2016.12.110
http://arxiv.org/abs/1603.05670

Text classification examples

Belgium

September 2008, crisis hits



Saturday, 27 September 2008 (relevance 0.921, rank 2):
"*Fortis investors face a weekend of uncertainty after the banking and insurance group went out of its way on Friday to reassure them that it was solvent and in no danger of collapse following market talk the company could become another casualty of the credit crisis.*"

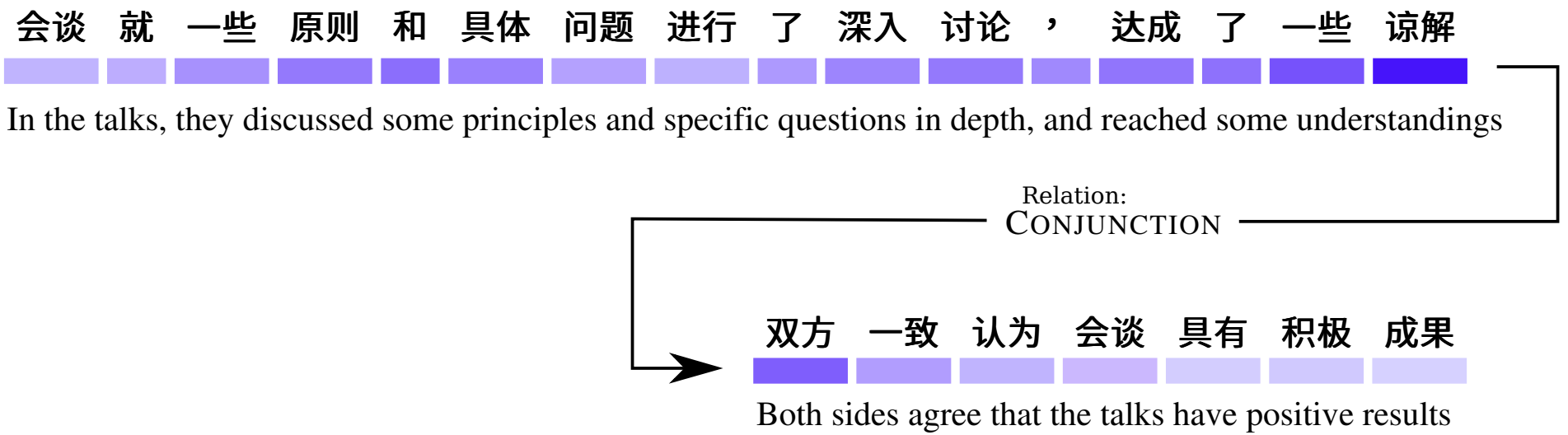
Saturday, 27 September 2008 (relevance 0.917, rank 3):
"*As of Saturday, financial authorities were contacting other institutions, a source familiar with the situation told Reuters, although no particular solution was preferred and nothing concrete was likely to emerge before Sunday.*"

Sunday, 28 September 2008 (relevance 0.758, rank 6):
"*BRUSSELS (Reuters) - Belgium's national pride and thousands of jobs are at stake as the Belgian and Dutch governments, central banks and regulators seek to secure the future of financial services group Fortis.*"

Monday, 29 September 2008 (relevance 0.889, rank 5):
"*Belgian, Dutch and Luxembourg governments rescued Fortis over the weekend to prevent a domino-like spread of failure by buying its shares for 11.2 billion euros.*"

Text classification examples

Case 2: Classifying implicit relations between sentences (in Chinese)



Text classification practical

Case: Sentiment analysis

IMDb movie review dataset:

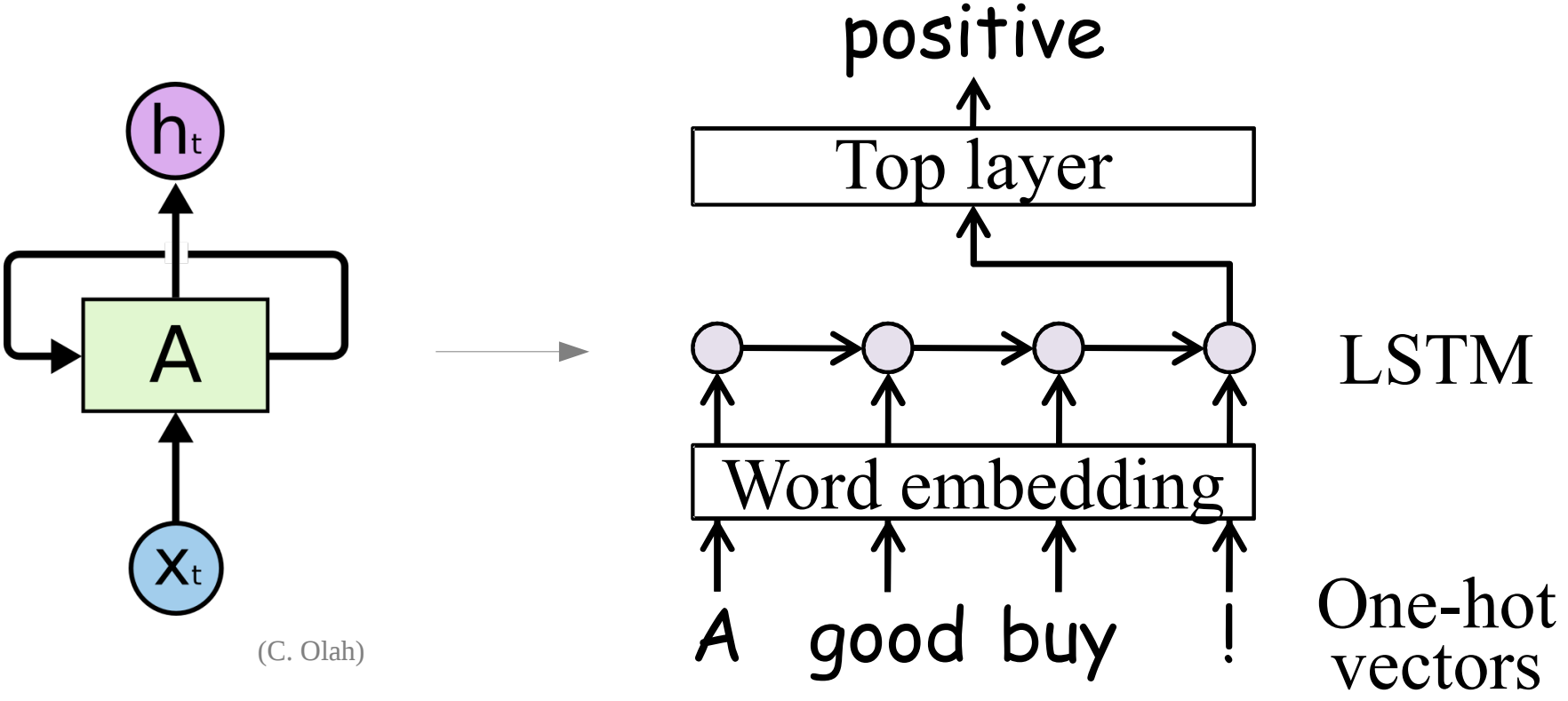
- 25k labeled comments: positive/negative
- 50k unlabeled comments

Modeling:

1. Train your first word vectors with *gensim*
2. Train your first RNN with *keras*

Text classification practical

Long Short-Term Memory network for classification



(C. Olah)

(Johnson & Zhang, 2016)

Thanks! Questions?

Code here:

`https://github.com/sronnqvist/deepNLPtutorial`

TBC: Doctoral defense, Agora, December 8, 2017