Deep Learning for Natural Language Processing

Continuous bag-of-words representations
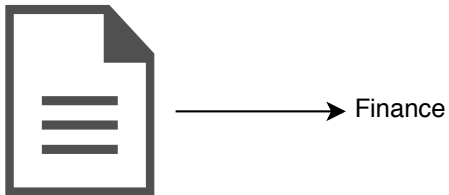
UNIVERSITY OF
GOTHENBURG

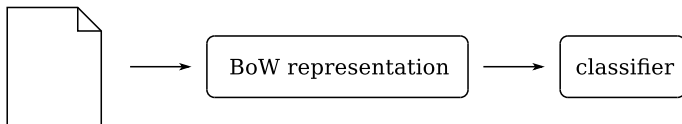**CHALMERS**

WASP | WALLENBERG AI
AUTONOMOUS SYSTEMS
AND SOFTWARE PROGRAM

Richard Johansson

`richard.johansson@gu.se`

Finance

recap: document classification with BoW representation


BoW representation → classifier

the continuous bag-of-words representation

▶ we return to the question of **representation**

▶ the **continuous bag-of-words** (CBoW) or **neural bag-of-words** representation computes a sum or mean of word embeddings over a document

$$\text{CBoW}(D) = \frac{1}{L} \sum_{i=1}^{L} \text{embedding}(w_i)$$

building a document categorizer with CBoW

▶ we can combine the CBoW representation with a linear or nonlinear (neural) classifier

```
[document]  ──→  [ CBoW representation ]  ──→  [ classifier ]
```
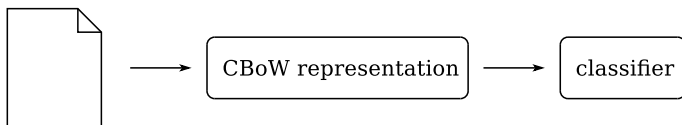
▶ when we train the complete system, the word embeddings are **learned automatically** for this task

▶ this is our first example of a **learned representation**: a key selling point for deep learning in NLP

famous examples of CBoW architecture

- DAN (Iyyer et al., 2015)
- fastText (Joulin et al., 2017)
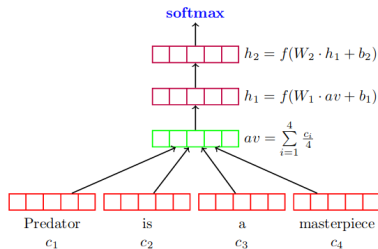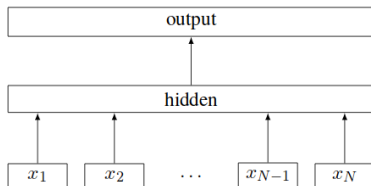
discussion

what do the word embeddings represent?

discussion

what do the word embeddings represent?

▶ intuitively, a word embedding represents the "meaning content" of this word **with respect to the target task**

limitations of the CBoW approach

- still more or less a word-spotting approach
- we can't handle **interactions** between words: negations, . . .
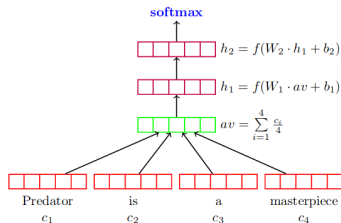- in some applications, we need a more generalized notion of "patterns"
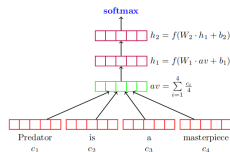
addressing the limitations of CBoW

▶ next block: **pre-training** embeddings
▶ and then: **convolutional** and **recurrent** models

case study: BoW and CBoW classification

▶ in the notebook for this lecture, we will implement document classifiers based on BoW and CBoW

recommended reading

- **Goldberg, chapter 8**: introduction to bag-of-words, embeddings, and continuous bag-of-words
- **Joulin et al. (2017)**: the fastText paper about text categorization

references I

M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé III. 2015. Deep unordered composition rivals synt. methods for text classification. In *ACL*.

A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov. 2017. Bag of tricks for efficient text classification. In *EACL*.