Language Technology: R&D

Word Embeddings

Ali Basirat

Department of Linguistics and Philology
Uppsala University

September, 2020

The Word

- Linguistics: the minimal syntactic unit of language
- Philosophy: the reflection of meaning in the mind
- Theology: the nature of God
- Cognitive science: the clusters of perceptual signals
- Artificial Intelligence: a symbol, a vector, a distribution, or a complex algebraic system

The Word

The Journey in AI/CL

- The importance: why word is important to the AI/CL communities?
- The use cases: which tasks would benefit from the study of words?
- Which models are examined by the community?
- What are the active lines of research?

Importance

Intelligent Machines

- Artificial intelligence: to design machines that simulate human intelligence, and think and behave like humans
- Turing test: an intelligent machine should behave equivalent to that of a human
- Communication system: a natural language is used to communicate with an intelligent machine

Importance

Language and Intelligence

- Humans use natural languages to communicate their intelligence
- Natural languages are brain products that have evolved gradually in centuries
- Natural languages can model almost whole the world
- Language is the jewel in the crown of cognition

Importance

Words of Language

- Words are fundamental elements of languages
- Syntax is the study of structures
- The word is the atomic element of syntax

Use Cases

Example

- Information retrieval, search engines, question answering, information extraction
- Machine translation
- Text analysis and language study
- Dialogue systems, and chat-bots
- Text summarization, story tellers, computational narrators
- Speech recognition
- Optical character recognition
- Many other use cases that deal with human languages

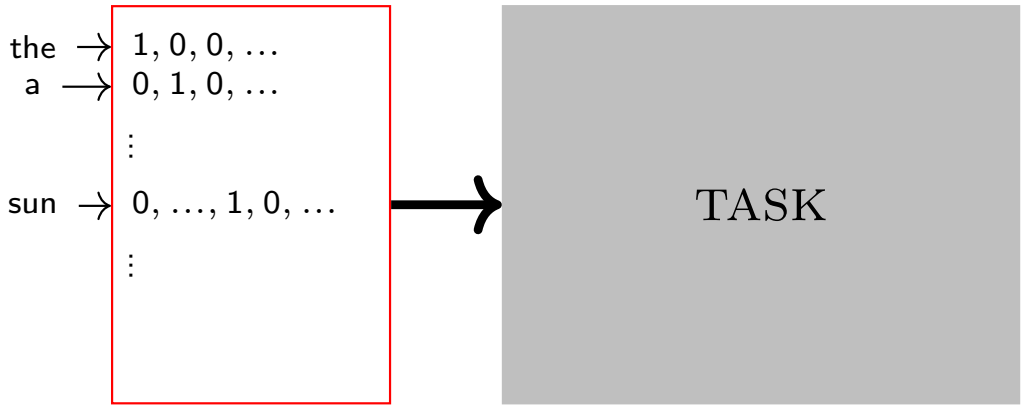
The Community

- Association for computational linguistics (ACL):
  - Journals: Computational Linguistics, Transactions of ACL
  - Conferences: ACL, EACL, NAACL, EMNLP, IJCNLP
- Association for the Advancement of Artificial Intelligence (AAAI)
- Other conferences on AI, Linguistics, Machine Learning, and Learning Representation (e.g., COLING, NIPS, ICLR, and ICML)

Which models are examined?

One-hot encoding

- Words are symbols independent of each other
- The relationships between words are modelled in separate tasks



Which models are examined?

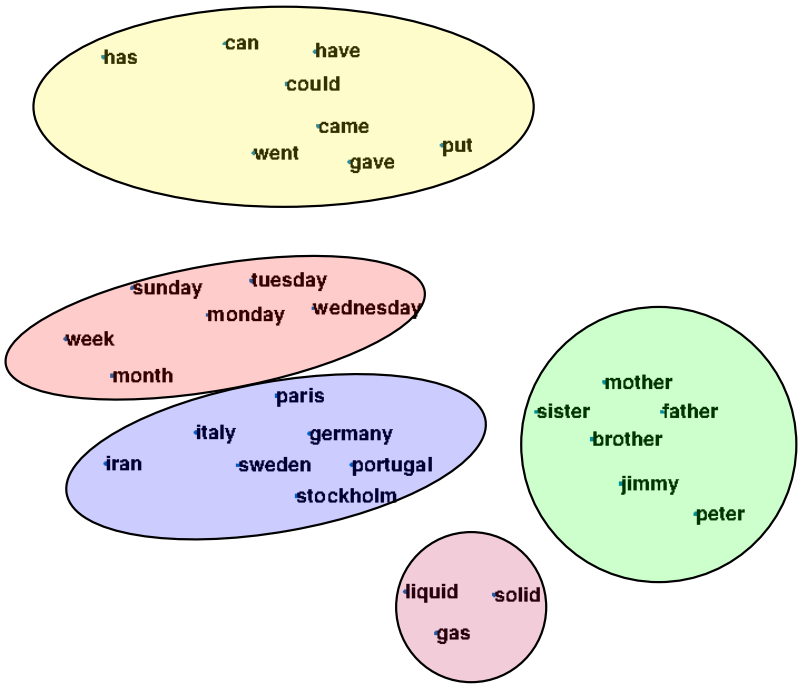
One-hot encoding

- Advantage: easy to implement - sparse vectors
- Disadvantages:
  - It does not model the interrelationships between words
  - A complex feature engineering should be performed by the target tasks
  - It does not tell us anything about the word properties (not good for linguistic studies)
  - No mechanism to handle out of vocabulary words

Which models are examined?

Word vectors

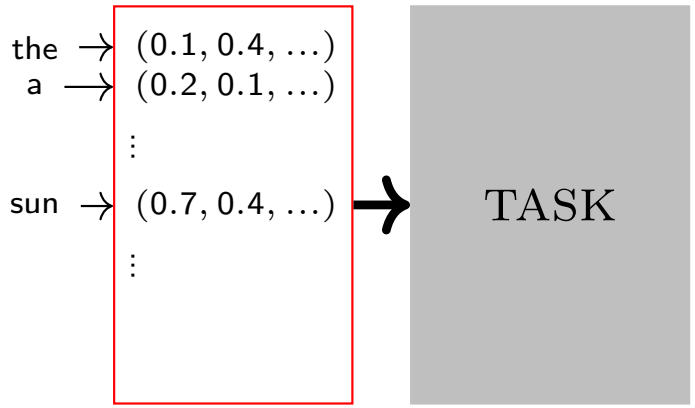
- Each words is represented as a vector (a list of real numbers)
- Vector similarity represent word similarity



Which models are examined?

Word vectors

- More complex word embedding learner
- Simpler feature engineering in the target task



Which models are examined?

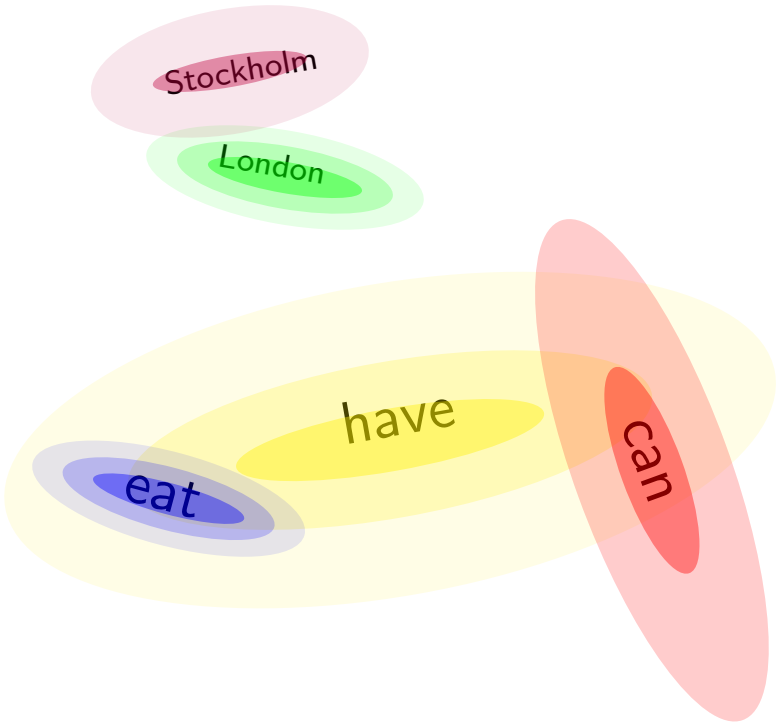
Word vectors

- Advantages:
  - No data annotation
  - Easy to train
  - Linguistically rich: very little feature engineering is needed
- Disadvantages
  - Does not encode polysemy and dynamics of word's meaning
  - Does not encode certain semantic aspects of words (e.g., is a noun countable or not?)

Which models are examined?

Random Word vectors

- Words are associated with *random* vectors
- Each word takes an area in a high-dimensional space
- Word similarities are measured by the distribution distances



Which models are examined?

Random Word vectors

- Advantages:
  - All advantages of word vectors
  - Encode multiple senses of words and models polysemy
  - Provide for modelling the complex semantic relations
- Disadvantages
  - Limited to a fixed number of senses for each word
  - Not studied enough in the literature

Which models are examined?

Contextualized Word vectors

- Each word in a context is associated with a vector
- Word vectors are generated according to the context of words
- The word similarities are measured according the contextual occurrence of words

Which models are examined?

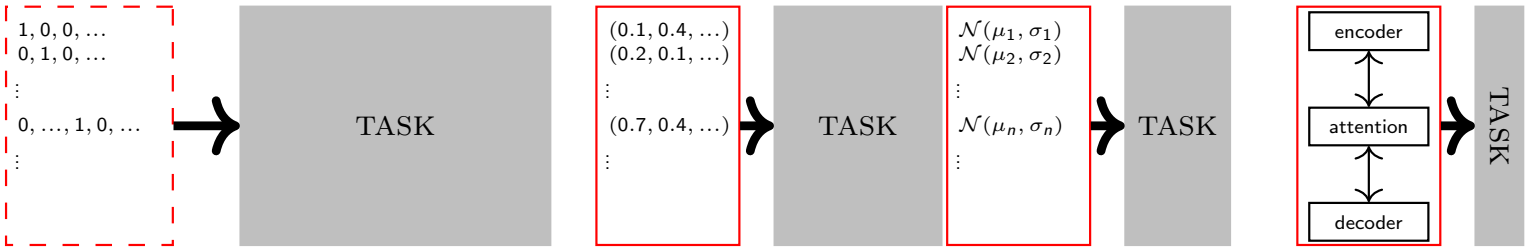
Contextual Word vectors

- Advantages:
  - No data annotation: word vectors are often trained on large raw corpora
  - Linguistically rich: almost no feature engineering is needed on the target tasks
  - Encode multiple senses of words and models polysemy
- Disadvantages
  - The training procedure is computationally heavy
  - Not suitable for modeling the static properties of words (e.g., grammatical gender)

Which models are examined?

Summary

- Word representation is becoming more and more important in natural language processing
- The target tasks become smaller and smaller as we have better representation of words



Research Lines

- New models and architectures of word embeddings
- Interpret the current models
- The application of words embeddings in new tasks
- Linguistic study of words - e.g., typology, nominal classification, etc.
- Compositional Semantics
- Survey of use cases, and architectures

Thank You

Questions?