Quantitative Methods and Machine Learning

Distributional Semantics

Mathias Müller

Problems with inferring the meaning of words

– Categorical nature

– Cannot be manipulated mathematically

– Can you define a degree of similarity between words?

Distributional Semantics

– What does semantics mean?

– What does it mean for semantics to be distributional?

"A word shall be known by the company it keeps"

– Firth, a long time ago

– How can a word be described by its context?

What shall "context" mean?

Projecting words into a vector space

– Key idea: represent words as vectors of numbers

Measures of distance or similarity

Dealing with high dimensionality

– Dimension equals vocabulary size ("curse of dimensionality")

– Densify: factorization (SVD, PCA)

– Dense vectors from word embeddings

Applications for distributed meaning