Black-Box Features for Quality Estimation

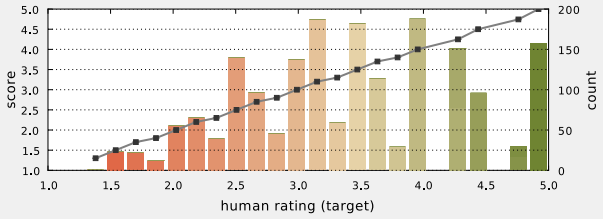
Christian Buck (christian.buck@ed.ac.uk)

THE UNIVERSITY of EDINBURGH

Task

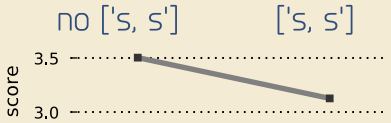
Task:

- predict quality [1=unusable, .., 5=perfect]
- given:
  - MT system + its training data
  - human judgment on 1832 translations
  - baseline feature set



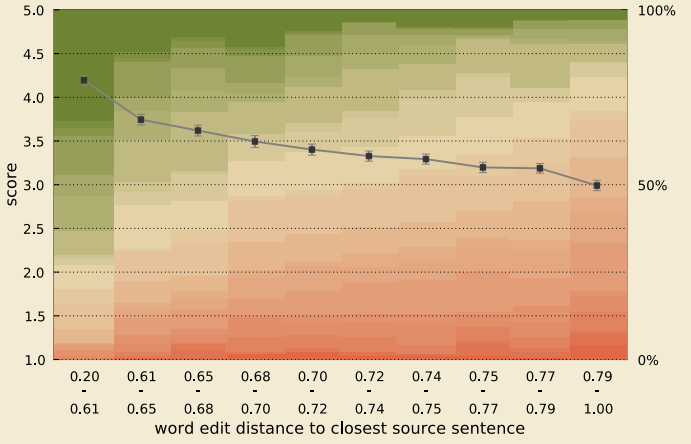
Indicators & Named Entities

- binary features that indicate:
  - numbers
  - uppercase words
  - punctuation, e.g. apostrophe
- Named Entities using Stanford NER Tagger
- best feature:
  - indicator of genitive case:



Edit Distance

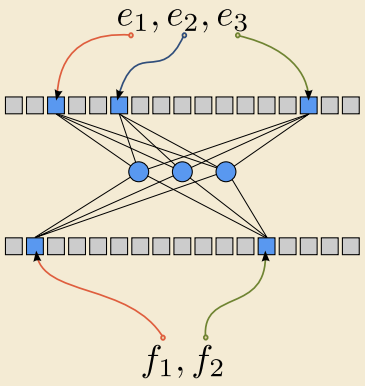
- edit distance to closest entry in corpus
- normalized by sentence length
- on source and target side
- on character and word level



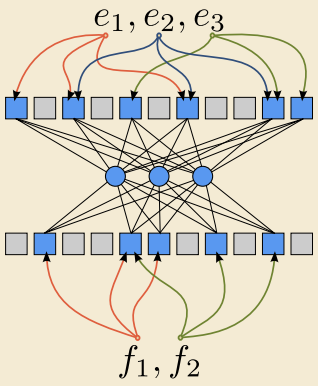
Neural Networks

- predict target words using vector space model
- to reduce dimensionality either:
  - filter to words in QE corpus
  - hash words, resembling bloom filter

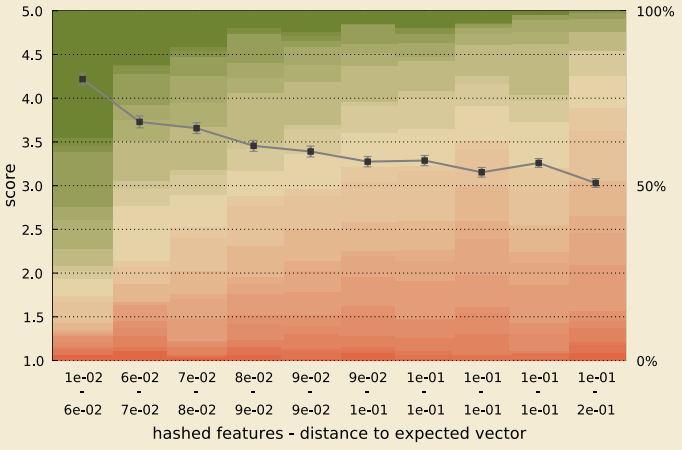
standard vector model



hashed source and target



- trained on 1.7M sentences
- features:
  - distance to expected vector (plot below)
  - geometric mean of expected word probabilities



Results (5-fold CV)

- ensemble of 500 neural networks
- improvements over baseline (72.7 / 57.7)

	RMSE	MAE
Indicators	+0.7	+0.5
Named Entities	+0.4	+0.5
Edit Distance	+0.0	+0.0
Neural Networks	+0.0	-0.1

Results (Shared Task)

indistinguishable from baseline :-(