

Barbara Plank, Anders Johannsen and Anders Søgaard
Center for Language Technology
University of Copenhagen, Denmark

bplank@cst.dk, ajohannsen@hum.ku.dk, soegaard@hum.ku.dk

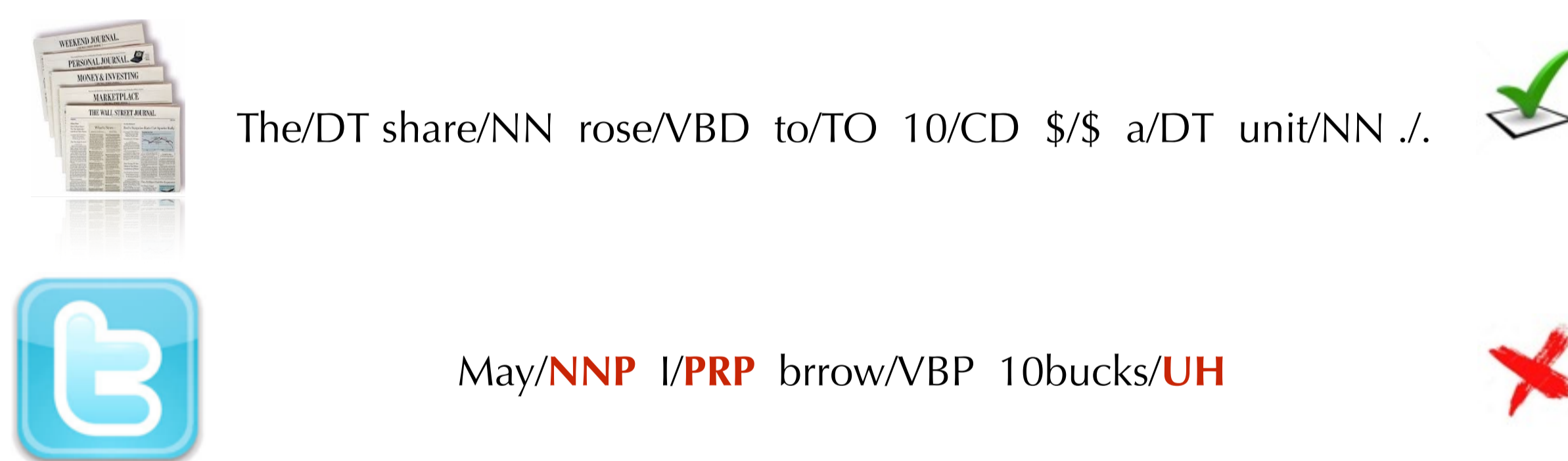
Motivation

- Labeled training in NLP is heavily biased
- Importance Weighting is one way to address data bias
- However, only few applications in NLP and mixed results (Jiang & Zhai, 2007; Foster et al. 2010; Søgaard & Haulrich, 2011; Plank & Moschitti, 2013)

Sample Selection Bias



Off-the-shelf POS tagger



Experimental Setup

- unsupervised Domain Adaptation (DA)



- English Web Treebank:
- Source: Ontonotes VWS
- Target domains (5): answers, emails, reviews, weblogs, newsgroups

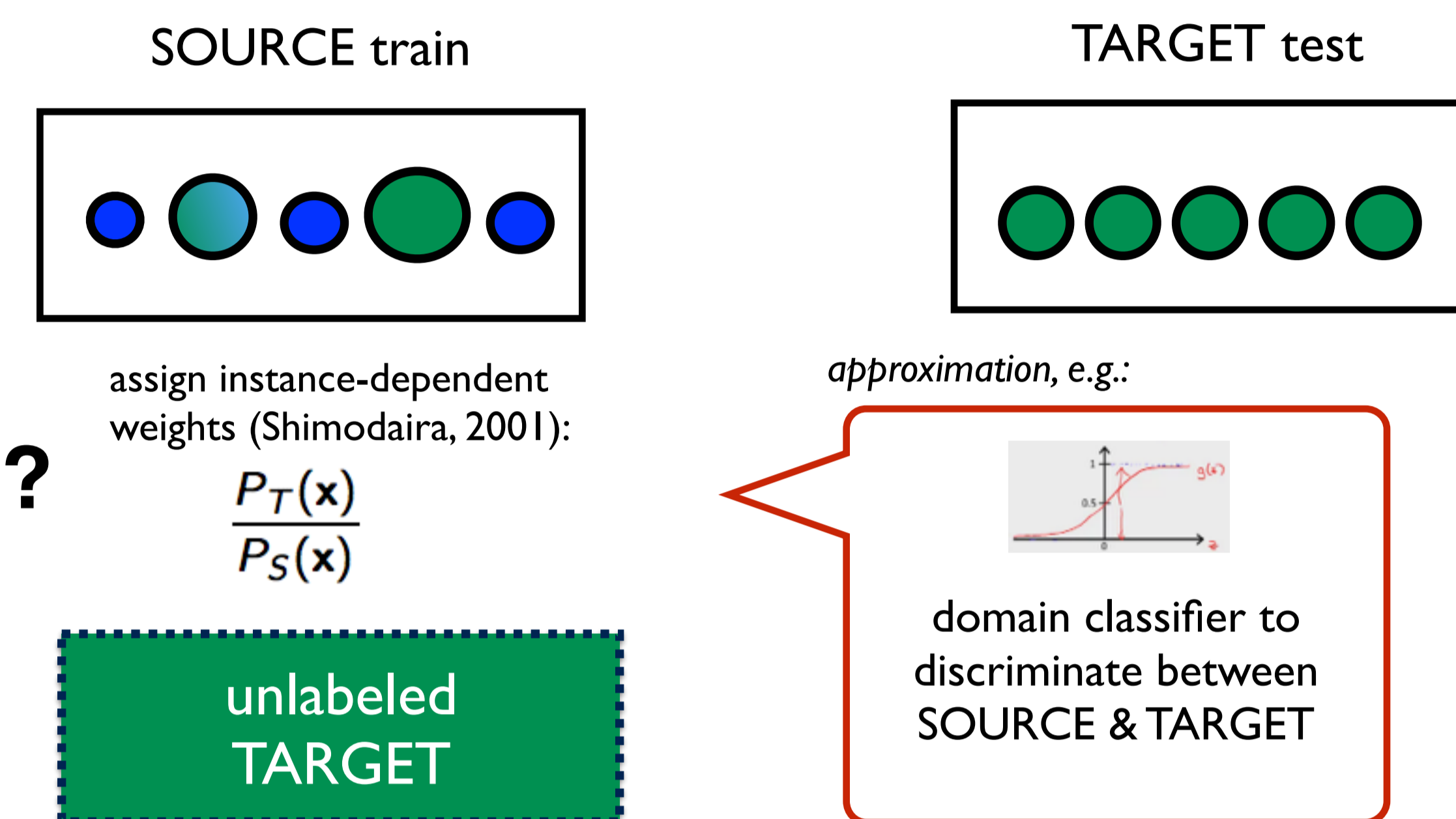
- Weighted structured perceptron

- Code available:
<https://bitbucket.org/bplank/importance-weighting-exp>
<https://github.com/coastalcph/rungsted>

Research Question

- Does importance weighting work for unsupervised domain adaptation of POS taggers?

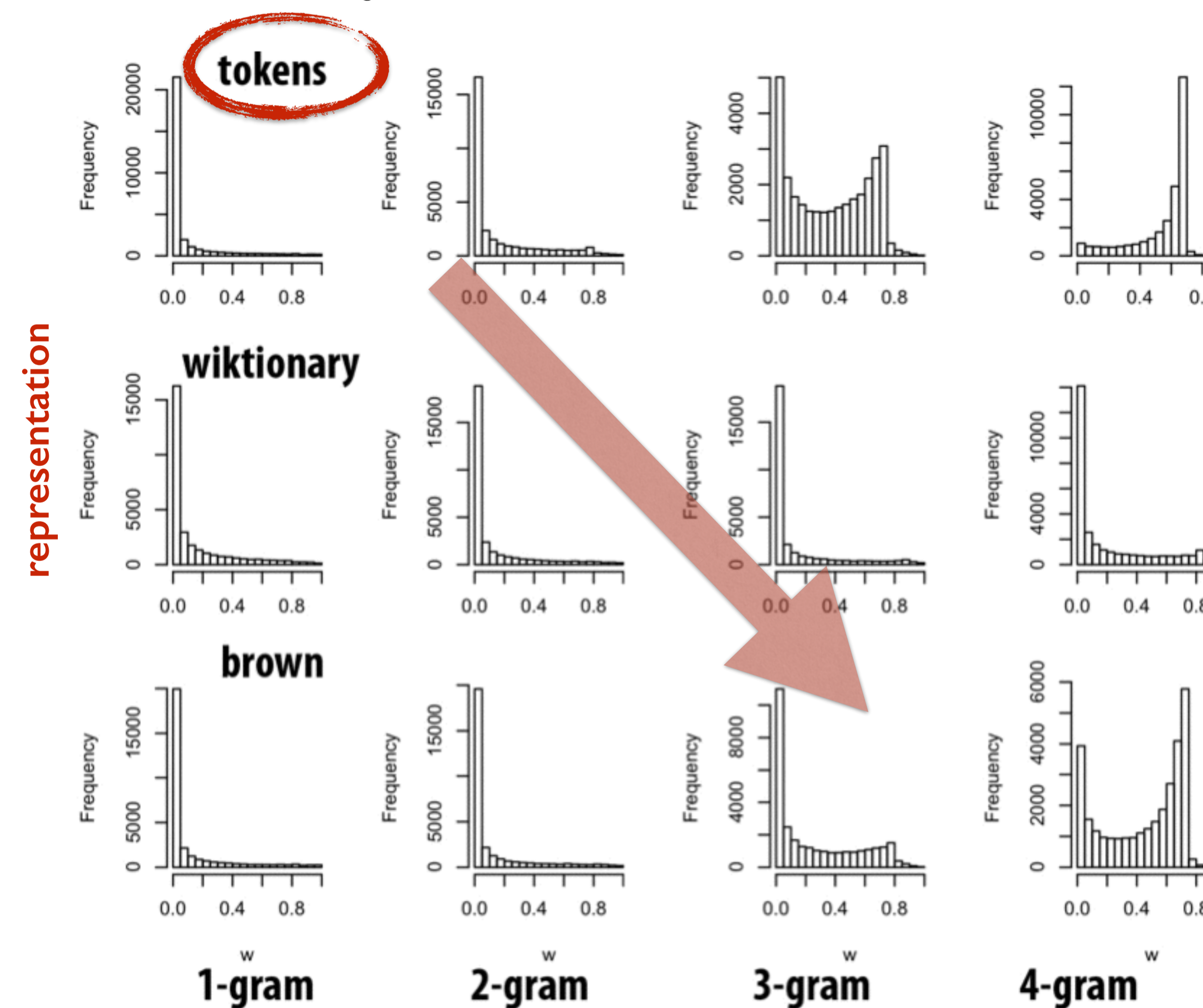
Importance weighting (IW)



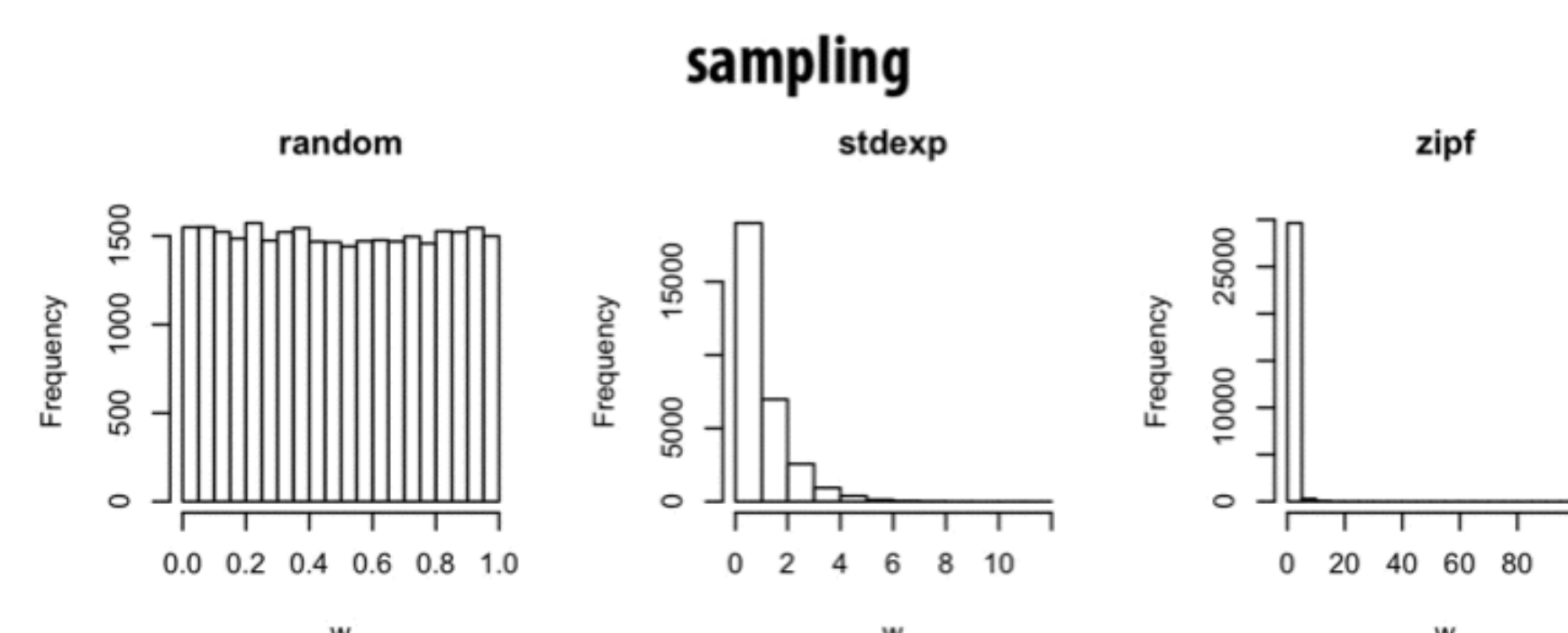
Weight functions

a) domain classifier

(Søgaard & Haulrich, 2011)

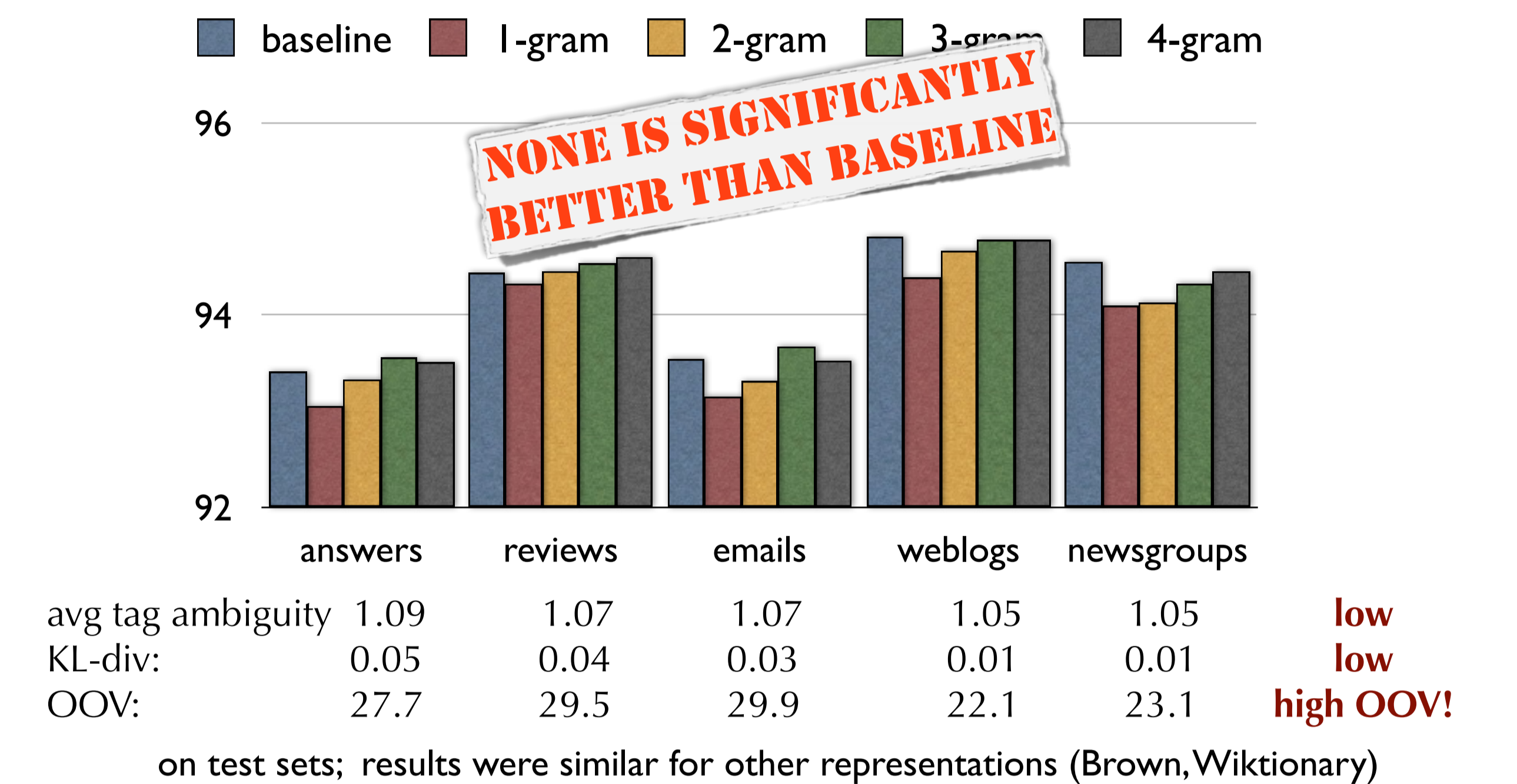


b) randomly sampled weights

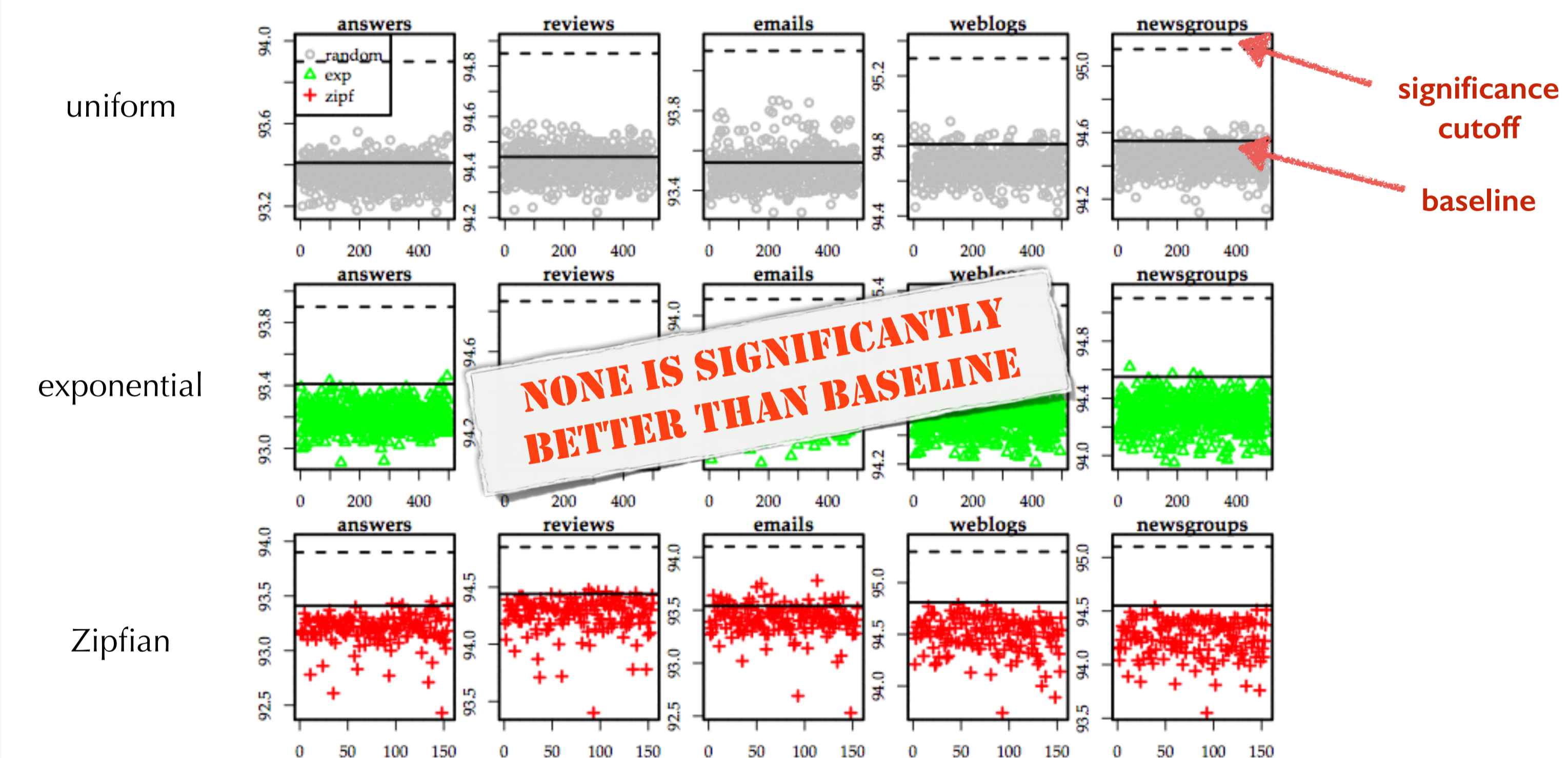


Results

Token-based domain classifier



Random weightings



(500 runs in each plot)

Conclusions & Future Work

- A negative result about importance weighting for unsupervised domain adaptation of POS taggers.
- None of the examined weightings lead to significant improvements
- Most errors due to unseen words (high OOV rate):
 - average word form ambiguity and POS bigram KL divergence low; thus, little space here for improvement for IW
 - Instead, robust improvements by using Wiktionary-based type constraints (Täckström et al., 2013)
- Future work: further weight functions, data sets, NLP tasks.

References:
Giovanni Cavallanti, Nicolò Cesa-Bianchi, and Claudio Gentile. 2006. Tracking the best hyperplane with a simple budget perceptron. In *COLT*.
Michael Collins. 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In *EMNLP*.
Corinna Cortes, Yishay Mansour, and Mehryar Mohri. 2010. Learning bounds for importance weighting. In *NIPS*.
George Foster, Cyril Goutte, and Roland Kuhn. 2010. Discriminative instance weighting for domain adaptation in statistical machine translation. In *EMNLP*.
Jing Jiang and ChengXiang Zhai. 2007. Instance weighting for domain adaptation in NLP. In *ACL*.
Oscar Täckström, Dipanjan Das, Slav Petrov, Ryan McDonald, and Joakim Nivre. 2013. Token and type constraints for cross-lingual part-of-speech tagging. *TACL*.