

# SUPERVISED RANKING OF TRIPLES FOR TYPE-LIKE RELATIONS

The Cross Triple Scorer at WSDM Cup 2017

**Faegheh Hasibi**

Norwegian University of Science and Technology

**Darío Garigliotti**

University of Stavanger

**Shuo Zhang**

University of Stavanger

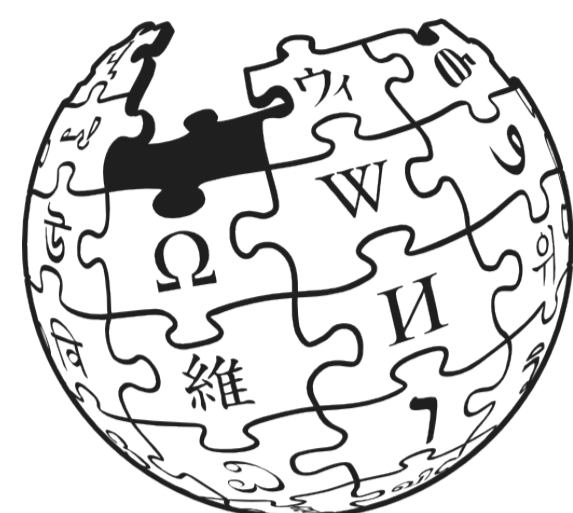
**Krisztian Balog**

University of Stavanger

## GOAL

“Given a list of triples from two type-like relations (profession and nationality), for each triple compute an integer score from 0..7 that measures the degree to which the subject belongs to the respective type.”

## DATA SOURCES



- ▶ Set of 33M annotated Wikipedia sentences
- ▶ First Wikipedia sentence/paragraph of a person

William Shakespeare	has-profession	Poet	7
William Shakespeare	has-profession	Lyricist	3
Bob Dylan	has-profession	Singer-songwriter	7
Bob Dylan	has-profession	Guitarist	5

Bob Dylan (*/ˈdɪlən/*; born Robert Allen Zimmerman, May 24, 1941) is an **American songwriter, singer, artist, and writer**. He has been influential in popular music and culture for more than five decades ...

## FEATURES

### Profession features

sumProfTerms{k}	The sum of top-k (TFIDF-weighted) profession terms from all Wikipedia sentences of the person
simCos{k}	Cosine similarity b/w the profession and person term vectors of top-k profession terms
simCosW2V	Cosine similarity b/w the profession and person embedding vectors (centroids of Word2Vec vectors)
simCosW2VPar	Cosine similarity b/w the profession and person embedding vectors based on 1st Wikipedia paragraph
isProfWPSent	Whether the profession occurs in the first Wikipedia sentence of the person
isProfWPPar	Whether the profession occurs in the first Wikipedia paragraph of the person
isFirstProfWPSent	Whether the profession is the 1st of the professions occurring in the 1st Wikipedia sentence of the person
isFirstProfWPPar	Whether the profession is the 1st of the professions occurring in the 1st Wikipedia paragraph of the person

### Nationality features

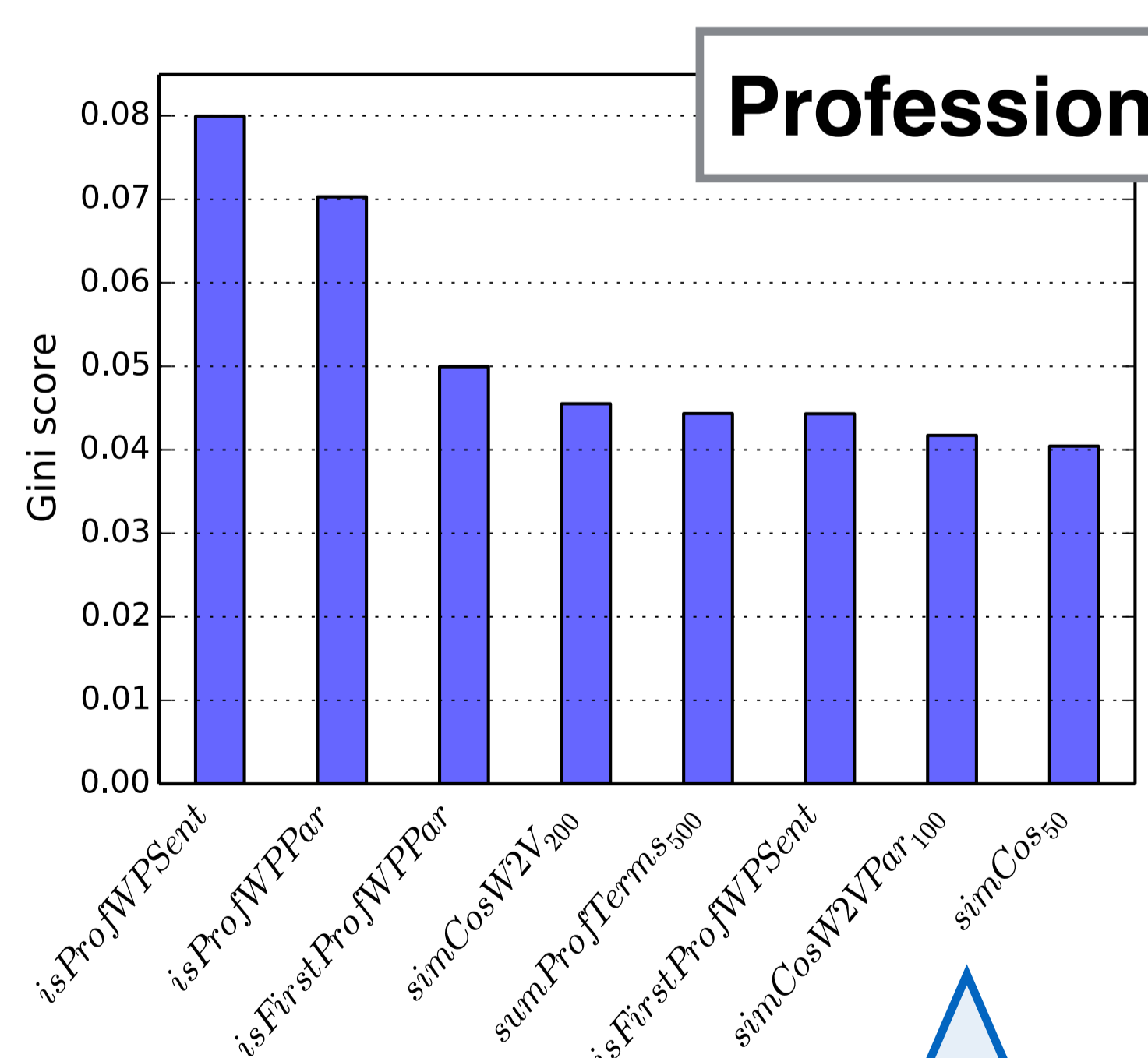
freqPerNat	The normalized frequency of a person co-occurring with a nationality in Wikipedia sentences
isNatWPSent	Whether the nationality occurs in the first Wikipedia sentence of the person
isNatWPPar	Whether the nationality occurs in the first Wikipedia paragraph of the person
isFirstNatWPSent	Whether the nationality is the 1st of the nationalities occurring in the 1st Wikipedia sentence of the person
isFirstNatWPPar	Whether the nationality is the 1st of the nationalities occurring in the 1st Wikipedia paragraph of the person

## RESULTS

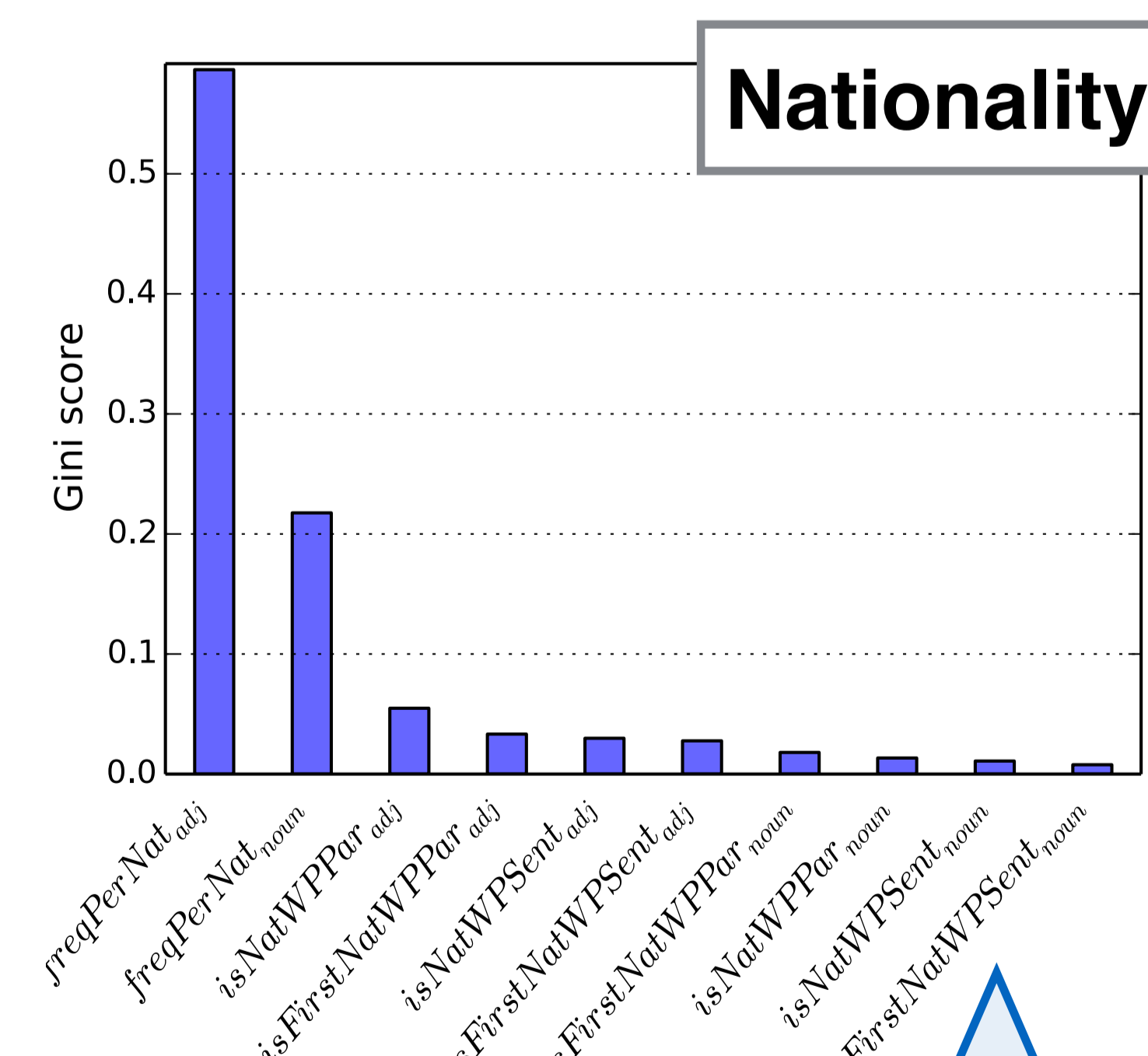
Model	Accuracy	ASD	Kendall
Profession (train)	0.77	1.61	0.34
Profession (test)	0.78	1.61	0.27
Nationality (train)	0.76	1.62	0.32
Nationality (test)	0.77	1.62	0.40

	cross [our]	bokchoy	goosefoot
Accuracy	0.78	<b>0.87</b>	0.75
rank	6	1	>7
ASD	<b>1.61</b>	1.63	1.78
rank	1	2	7
Kendall	0.32	0.33	<b>0.31</b>
rank	2	3	1

## FEATURE ANALYSIS



Features extracted from the 1st Wikipedia sentence/paragraph are of the highest importance.



Feature *freqPerNat*, extracted from Wikipedia sentences, plays the most important role.