

David Alfter

Språkbanken

University of Gothenburg

Clustering and visualizing the data



Språk-
BANKEN

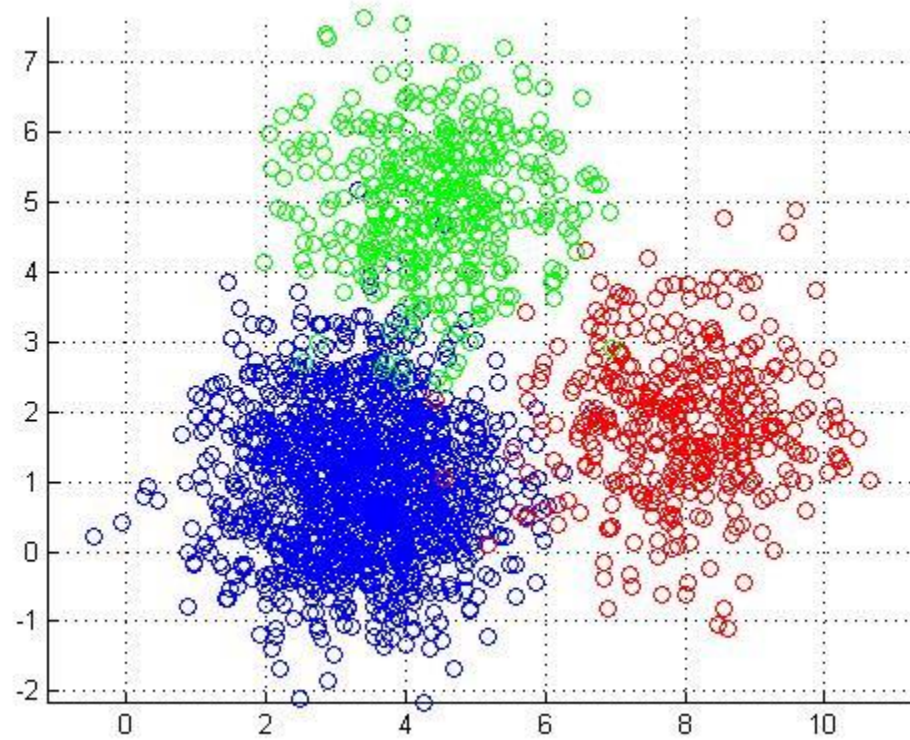


UNIVERSITY OF
GOTHENBURG

Data

MWE	Excuse me	Guess what?	be called sth	beat about/around the bush	become available/rich/a writer, etc.	break the ice	break the law	bring a lump to your th
Excuse me	0.0	1.0	1.2857142857142858	1.7142857142857142	0.8571428571428571	1.0	1.0	1.4285714285714286
Guess what?	1.0	0.0	1.0	1.2857142857142858	0.5714285714285714	1.2857142857142858	0.8571428571428571	1.4285714285714286
be called sth	1.2857142857142858	1.0	0.0	1.1428571428571428	0.8571428571428571	0.5833333333333334	0.5714285714285714	1.1428571428571428
beat about/around the bush	1.7142857142857142	1.2857142857142858	1.1428571428571428	0.0	1.4285714285714286	1.3333333333333333	0.42857142857142855	1.1666666666666667
become available/rich/a writer, etc.	0.8571428571428571	0.5714285714285714	0.8571428571428571	1.4285714285714286	0.0	1.0	0.8571428571428571	1.0
break the ice	1.0	1.2857142857142858	0.5833333333333334	1.3333333333333333	1.0	0.0	0.8571428571428571	0.85
break the law	1.0	0.8571428571428571	0.5714285714285714	0.42857142857142855	0.8571428571428571	0.8571428571428571	0.0	1.0
bring a lump to your throat	1.4285714285714286	1.4285714285714286	1.1428571428571428	1.1666666666666667	1.0	0.85	1.0	0.0
burn the midnight oil	2.0	2.0	1.2	1.0	1.5	1.0526315789473684	1.4	0.7142857142857143
can't/couldn't help doing sth	1.7142857142857142	1.5714285714285714	1.0	1.0	1.1666666666666667	0.42857142857142855	0.5714285714285714	1.2857142857142858
catch fire	1.0	1.0	0.7142857142857143	1.5714285714285714	1.1428571428571428	0.7142857142857143	0.8	1.0
change your mind	1.1428571428571428	0.14285714285714285	1.0	1.0	0.6428571428571429	1.2857142857142858	0.4	1.1666666666666667
come true	1.0	1.0	1.0	1.4285714285714286	0.5	1.0714285714285714	0.7894736842105263	0.8333333333333334
crack a joke	1.8	1.1428571428571428	1.1428571428571428	0.5714285714285714	1.4285714285714286	1.0	0.42857142857142855	0.8571428571428571
cross your mind	1.0	1.0	1.5	1.0	0.7142857142857143	0.5714285714285714	0.7142857142857143	0.5714285714285714
do the cleaning/cooking, etc.	0.8571428571428571	1.0	1.0	0.5714285714285714	1.0	0.7142857142857143	1.3333333333333333	1.0
draw a conclusion	1.037037037037037	0.5714285714285714	1.0	0.7142857142857143	1.0	1.1428571428571428	0.42857142857142855	1.0
drive sb mad/crazy, etc.	1.0	0.5714285714285714	0.14285714285714285	0.7142857142857143	1.0	0.2857142857142857	1.0	1.0
face the music	1.8571428571428572	1.5714285714285714	1.0	0.6428571428571429	1.1428571428571428	1.0	0.7142857142857143	1.0
fall flat	1.2857142857142858	0.8571428571428571	0.42857142857142855	1.0	1.7142857142857142	0.7142857142857143	0.4	1.0
fall in love	0.8	1.0	0.8571428571428571	1.1428571428571428	0.8571428571428571	1.1428571428571428	1.0	1.4285714285714286
feel at home	0.2	1.1428571428571428	0.14285714285714285	0.7142857142857143	0.6666666666666666	0.8571428571428571	0.7142857142857143	1.3571428571428572
follow suit	1.6428571428571428	1.5714285714285714	1.5714285714285714	1.0	1.2857142857142858	0.5714285714285714	0.8571428571428571	1.0
get a bus/train/taxi, etc.	0.7142857142857143	0.5714285714285714	1.4285714285714286	0.5714285714285714	0.4	1.0	0.5714285714285714	2.0

Clustering



Clustering

k-means with $n = 6$

Adverbial MWEs: 41.7% accuracy

Verbal MWEs: 50.0% accuracy

Clustering – Adverbial MWEs

Assigned ↓ True ->	A1	A2	B1	B2	C1	C2
A1	5	2	3	0	0	0
A2	0	5	2	2	0	1
B1	0	4	2	2	1	1
B2	0	1	0	6	1	2
C1	0	2	0	3	3	2
C2	0	0	0	2	4	4

Clustering – Verbal MWEs

Assigned ↓ True ->	A1	A2	B1	B2	C1	C2
A1	6	1	2	1	0	0
A2	2	5	3	0	0	0
B1	0	1	3	5	1	0
B2	0	1	1	5	3	0
C1	0	0	0	3	2	5
C2	0	0	0	1	0	9

Embedding projections

- <https://tinyurl.com/enetCollectVerbalMWE>
- <https://tinyurl.com/enetCollectAdverbialMWE>