Measuring Distributional Similarity in Context

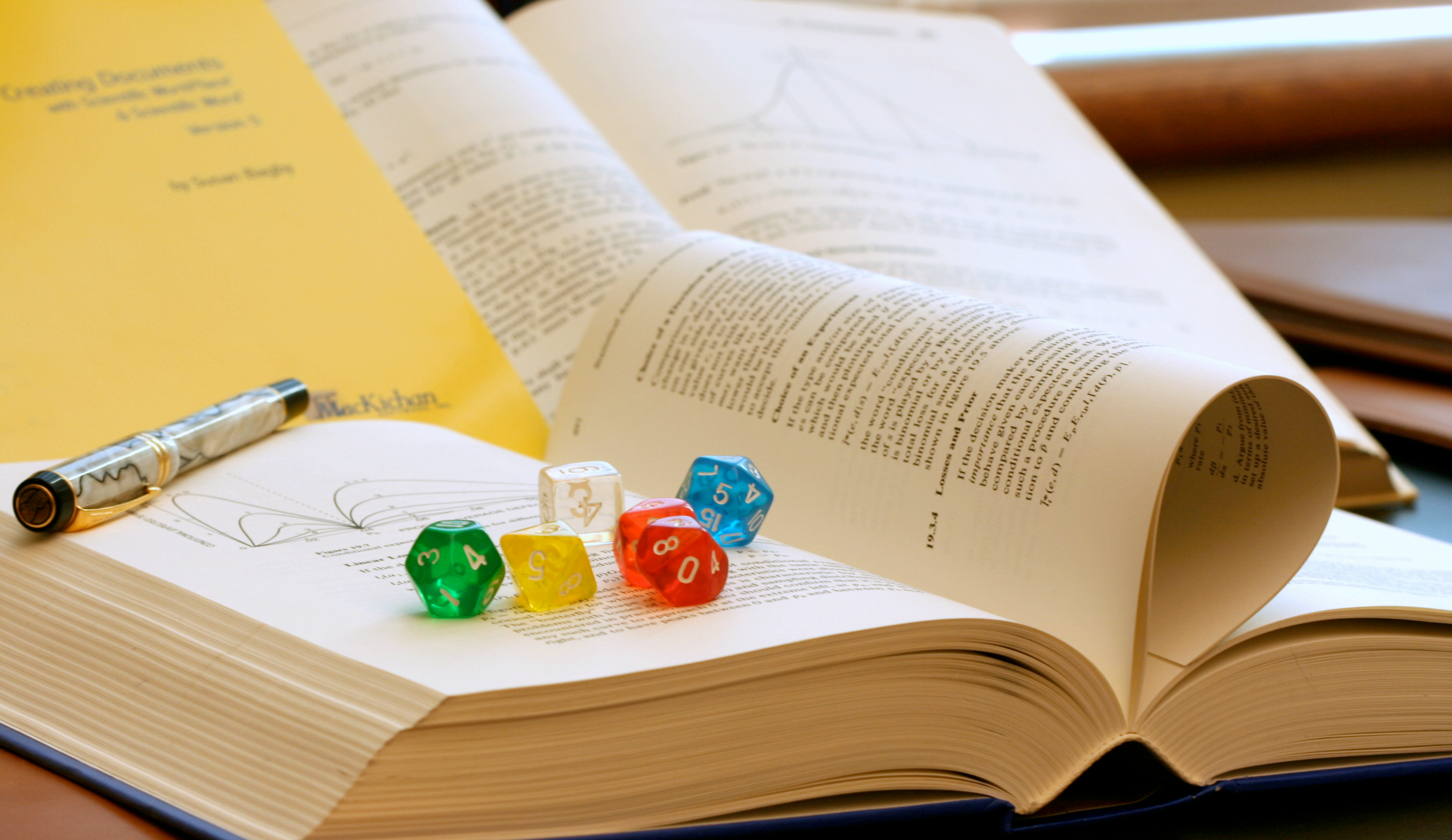
Georgiana Dinu, Mirella Lapata

Dmitrijs Milajevs

Vector-based models
identify
the different word senses
constantly and irrespectively
of co-occurring context

Indirect representation using vector operations

Previous work

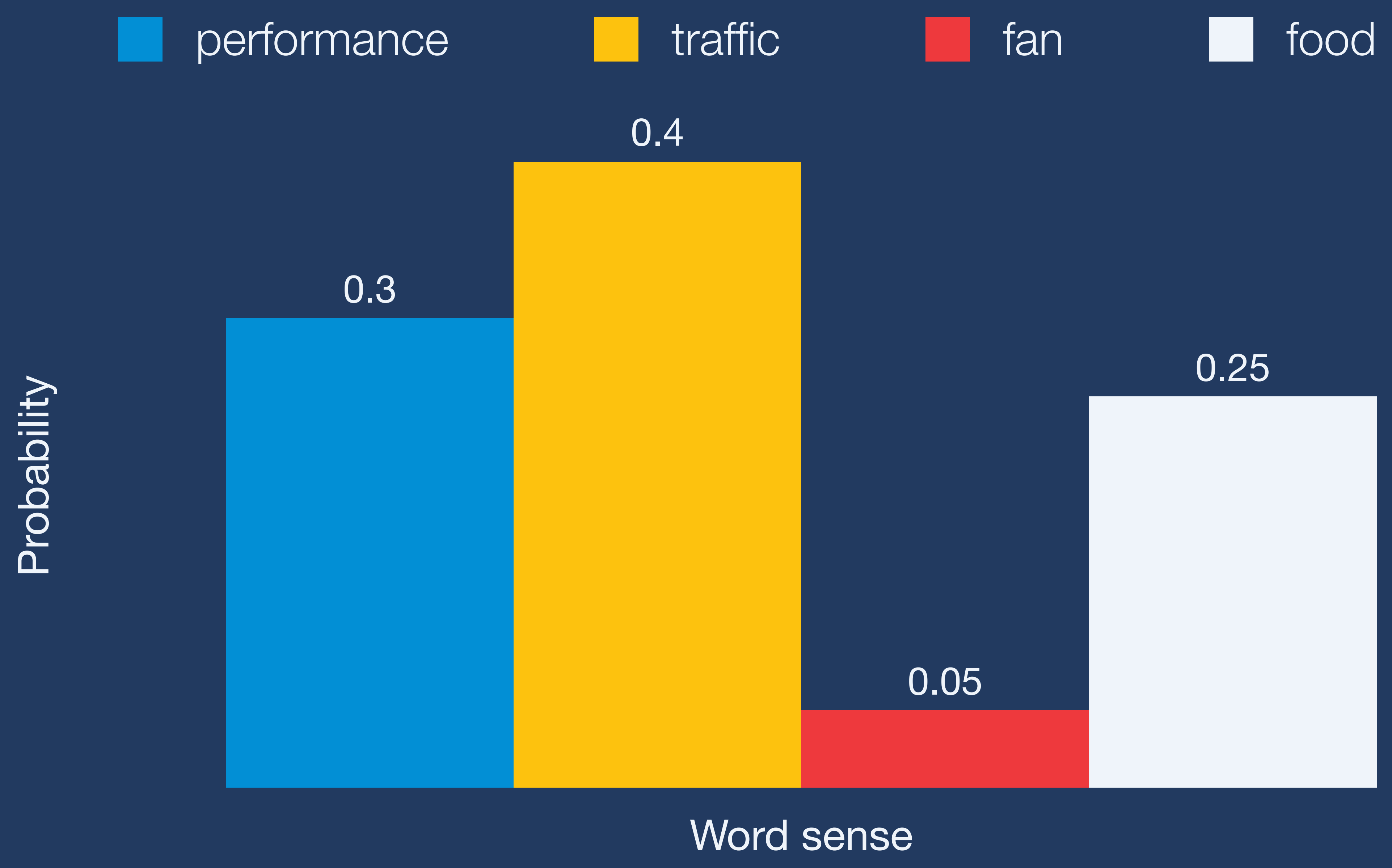


Direct representation using probability distributions

Solution

Meaning is a probability distribution over senses

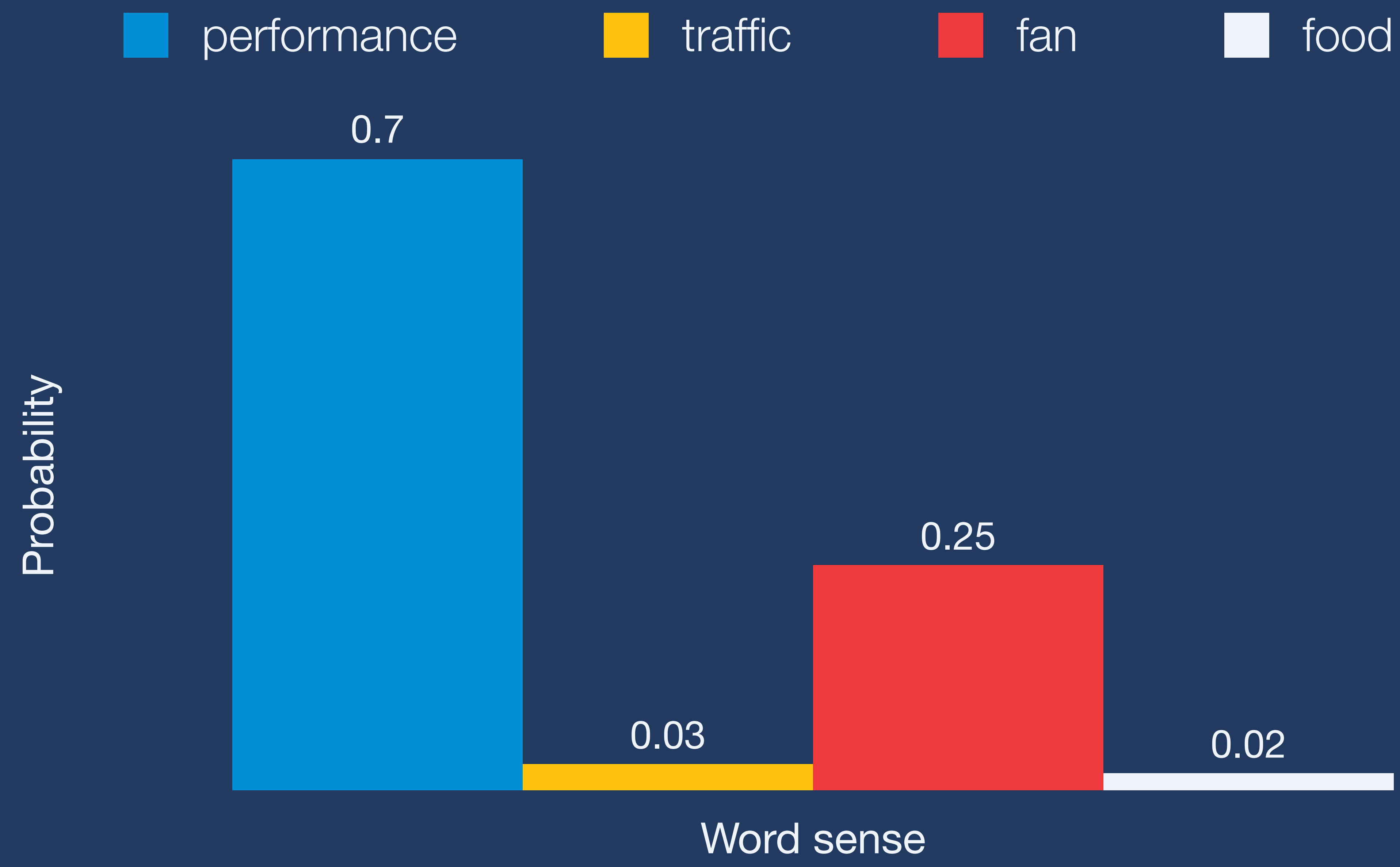
Sense distribution for jam



<P(performance|jam), P(traffic|jam), P(fan|jam), P(food|jam)>

A context feature directly modulates word’s sense distribution

Sense distribution for music jam



<P(performance|jam,music), P(traffic|jam,music), P(fan|jam,music), P(food|jam,music)>

Good ideas need good strategy to realize their potential.

Startup Quote!

REID HOFFMAN

FOUNDER, LINKEDIN

Model input: co-occurrence matrix

	jar	pet	my	rise
jam	100	5	32	10
ham	60	3	4	0
cat	8	94	120	11
dog	5	167	118	9
sun	0	3	30	145

- Rows are target words
- Columns are context features
- A cell (*i*, *j*) is the co-occurrence count of target t_i with the context feature c_j

Target words share a global set of latent senses

Meaning as a distribution over *K* senses

Target word

$$\mathbf{v}(t_i) = <P(z_1|t_i),\ldots, P(z_k|t_i)>$$

v(jam) = <P(performance|jam), P(traffic|jam), P(fan|jam), P(food|jam)>

Target word given a context feature

$$\mathbf{v}(t_i, c_j) = <P(z_1|t_i, c_j),\ldots, P(z_k|t_i, c_j)>$$

v(jam,music) = <P(p|jam,music), P(traffic|jam,music), P(fan|jam,music), P(food|jam,music)>

$P(z_k|t_i,\ c_j)$ estimation

$$P(z_k|t_i, c_j) = \frac{P(t_i, z_k)P(c_j|z_k, t_i)}{\sum_k P(t_i, z_k)P(c_j|z_k, t_i)}$$

$P(z_k|t_i, c_j)$ estimation

$$P(z_k|t_i,c_j) = \frac{P(t_i,z_k)P(c_j|z_k,t_i)}{\sum_k P(t_i,z_k)P(c_j|z_k,t_i)}$$

Assume, that target words and context features are conditionally independent given a sense

$$P(t_i,c_j|z_k) = P(t_i|z_k)P(c_j|z_k)$$

$P(z_k|t_i, c_j)$ estimation

$$P(z_k|t_i, c_j) = \frac{P(t_i, z_k)P(c_j|z_k, t_i)}{\sum_k P(t_i, z_k)P(c_j|z_k, t_i)}$$

Assume, that target words and context features are conditionally independent given a sense

$$P(t_i, c_j|z_k) = P(t_i|z_k)P(c_j|z_k)$$

$$P(z_k|t_i, c_j) \approx \frac{P(z_k|t_i)P(c_j|z_k)}{\sum_k P(z_k|t_i)P(c_j|z_k)}$$

$P(z_k|t_i)$ and $P(c_j|z_k)$ estimation

Latent sense induction

Approximate an input matrix

Non-negative matrix factorization

Non-negative matrix factorization

$$V \approx WH$$

100	5	32	10
60	3	4	0
8	94	120	11
5	167	118	9
0	3	30	145

Non-negative matrix factorization

$V \approx WH$

3.6	0.2	1	0.1
0	1.1	1	0.2

28	0
16	0
3	100
1	136
0	30

101	6	27	2
58	4	154	1
11	110	100	23
4	149	133	31
0	22	29	7

Non-negative matrix factorization

An iterative algorithm minimizes divergence between V and WH

$$D(V||WH) = \sum_{i,j} \left(V_{i,j} \log \left(\frac{V_{i,j}}{(WH)_{i,j}} \right) - V_{i,j} + (WH)_{i,j} \right)$$

Factor matrices interpretation

$$V \approx WH$$

$P(t_i, c_j)$

.26	2.43	2.12	.0
.16	.0	.32	1.37

.01	.02
.0	.0
.05	.01
.07	.0
.0	.11

.01	.02	.03	.07
.0	.01	.01	.01
.01	.17	.11	.02
.02	.16	.14	.0
.02	.0	.04	.15

Factor matrices interpretation

$V \approx WH$

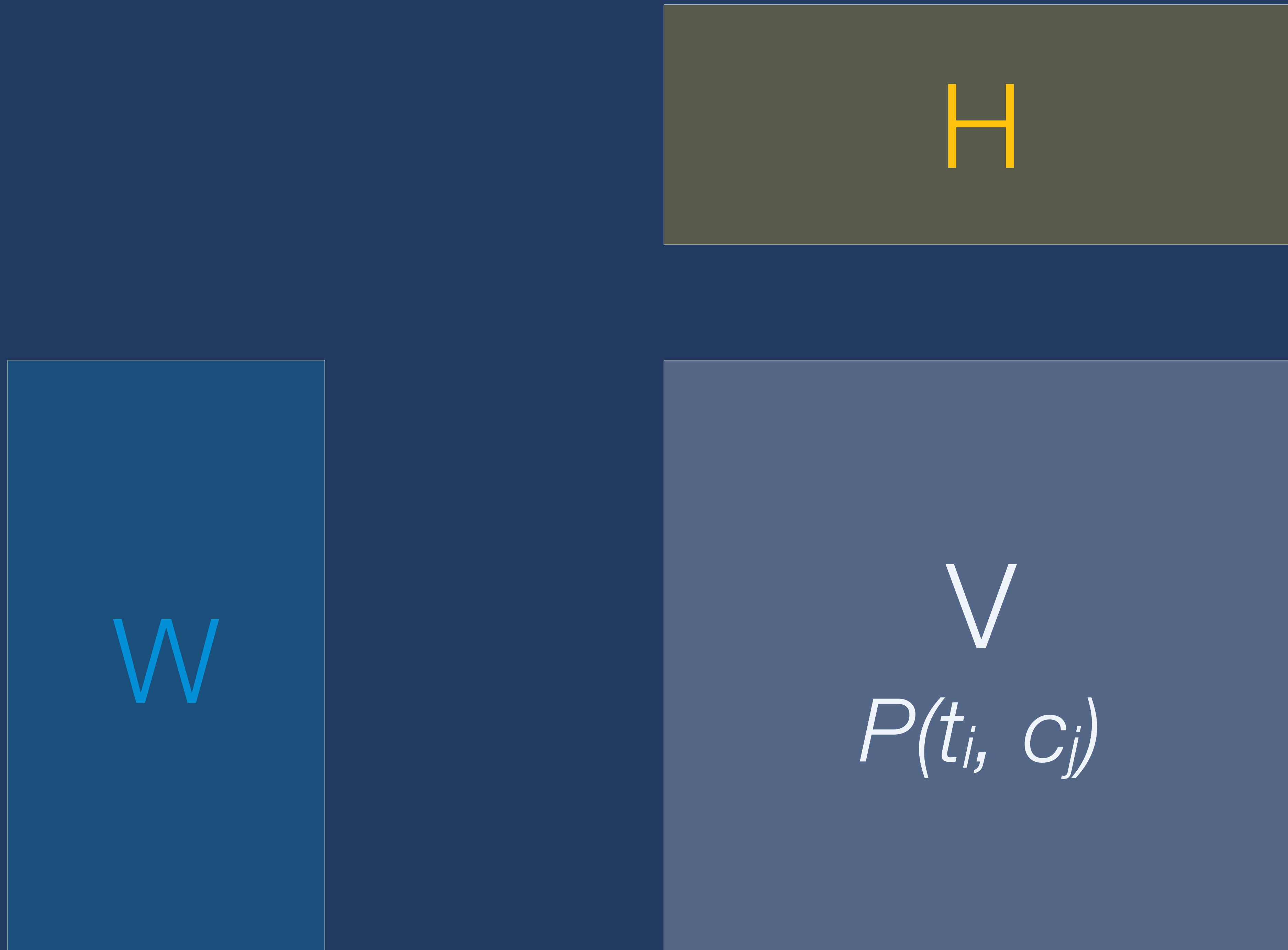
?

?

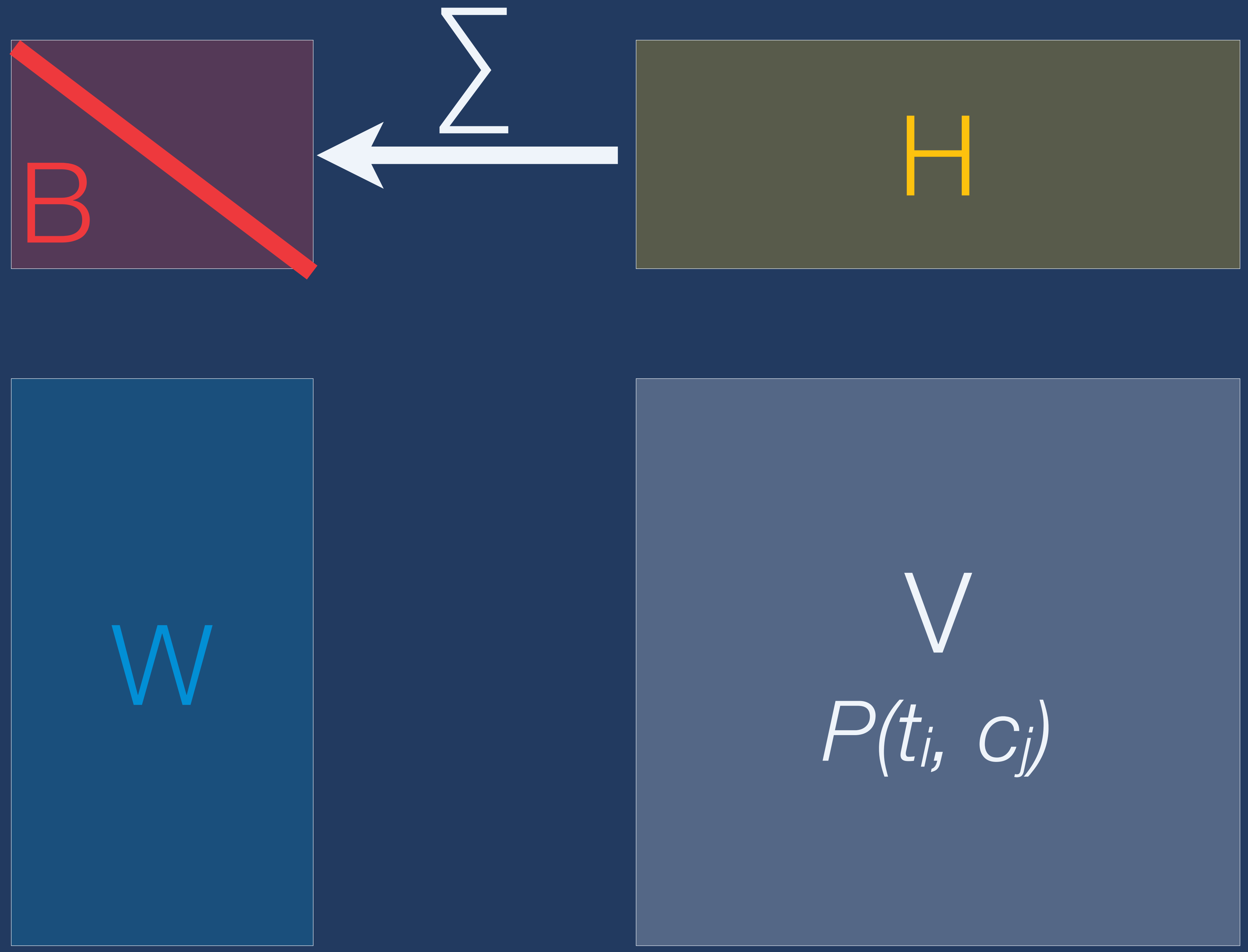
.26	2.43	2.12	.0
.16	.0	.32	1.37

.01	.02
.0	.0
.05	.01
.07	.0
.0	.11

.01	.02	.03	.07
.0	.01	.01	.01
.01	.17	.11	.02
.02	.16	.14	.0
.02	.0	.04	.15

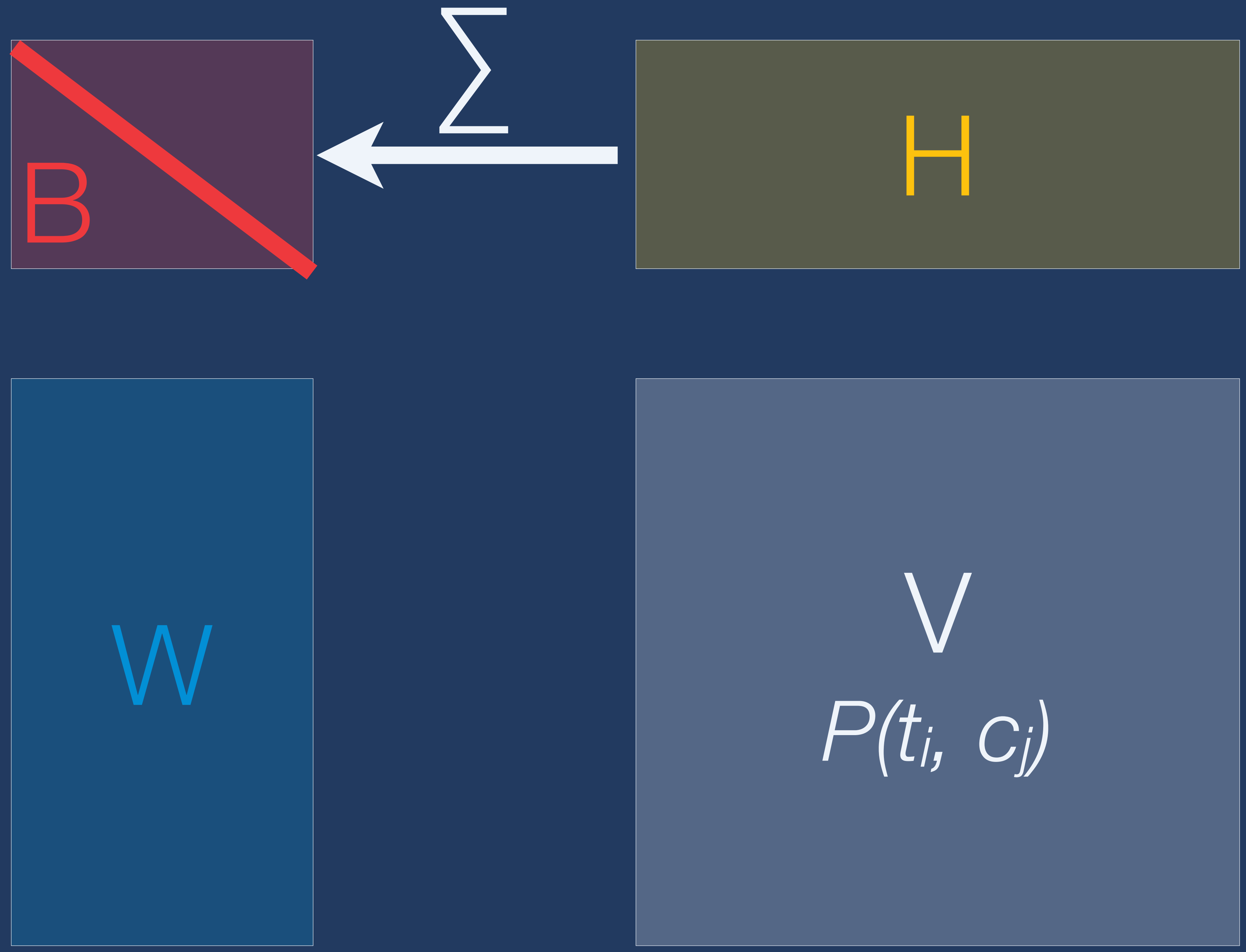


$V_{ij} = P(t_i, c_j)$ | V ≈ WH



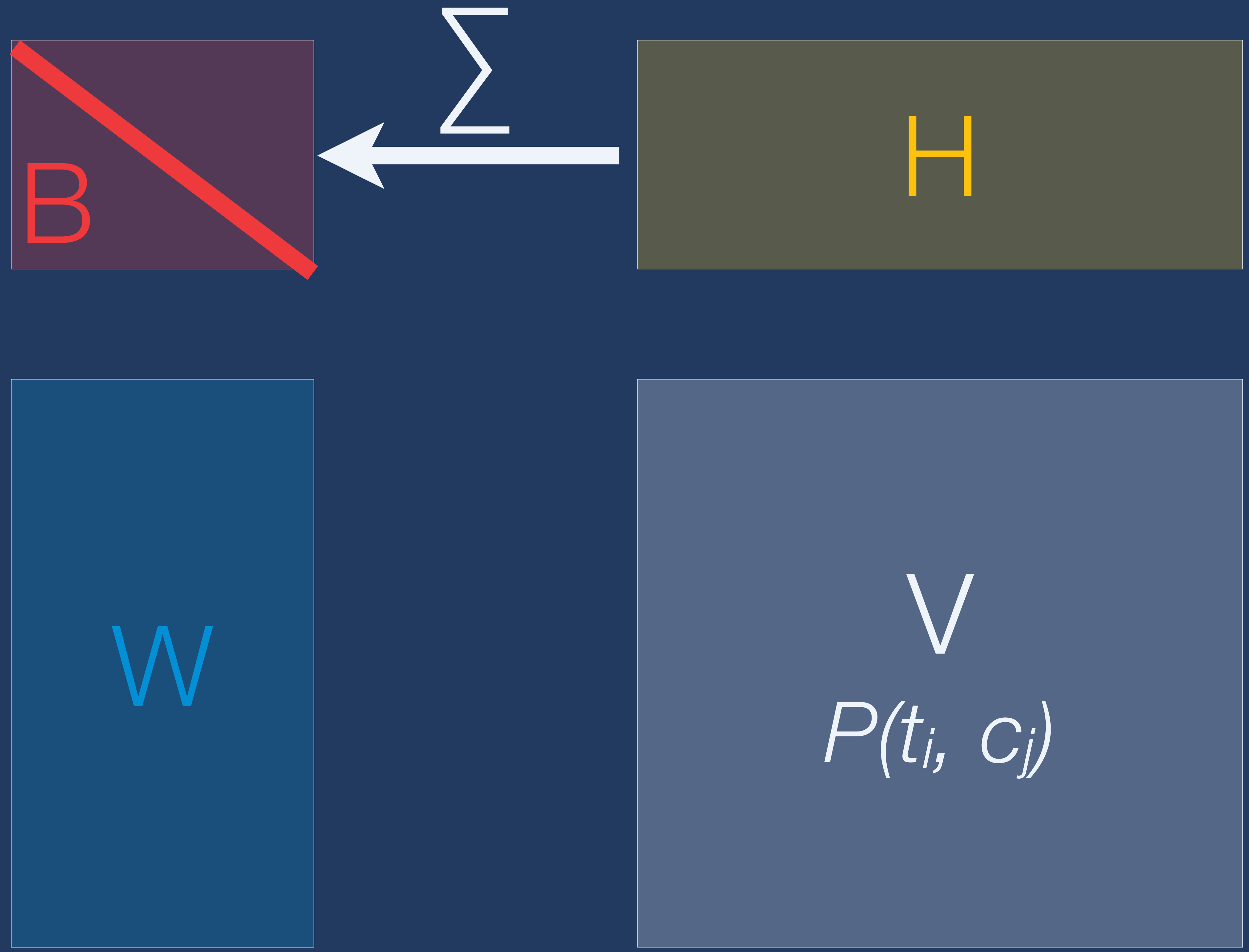
$$B_{kk} = \sum_j H_{kj}$$

B



$$(WB)_{ik} = \sum_j W_{ik} H_{kj} = P(t_i, z_k)$$

WB



$$(B^{-1}H)_{kj} = H_{kj}/\sum_j H_{kj} = P(c_j|z_k)$$

$B^{-1}H$

$P(t_i, z_k)$ and $P(c_j|z_k)$ estimation

Define: $B_{kk} = \sum_j H_{kj}$

$(WB)_{ik} = P(ti, z_k)$

$(B^{-1}H)_{kj} = P(c_j|z_k)$

$V = WH = WBB^{-1}H = (WB)(B^{-1}H)$

$V_{ij} = \sum_k (WB)_{ik}(B^{-1}H)_{kj} = \sum_k P(t_i, z_k)P(c_j|z_k)$

Initially, we wanted

$P(c_j|z_k)$ and $P(z_k|t_i)$

For now we have

$P(c_j|z_k)$ and $P(ti, z_k)$

$P(z_k|t_i)$ and $P(c_j|z_k)$ estimation

Define: $A_{ii} = \sum_k (WB)_{ik} = P(t_i)$

$(A^{-1}WB)_{ik} = (A^{-1})_{ii}(WB)_{ik} = P(z_k|t_i)$

$(B^{-1}H)_{kj} = P(c_j|z_k)$

$V = WH = A(A^{-1}WB)(B^{-1}H)$

$V_{ij} = \sum_k P(t_i)P(z_k|t_i)P(c_j|z_k)$

Matrix factorization: summary

Find factors W and H

Define diagonal matrices A and B

Rewrite WH as $A(A^{-1}WB)(B^{-1}H)$ to obtain required probabilities

Word similarity
Lexical substitution

Evaluation

Word similarity: sim(t, t') = sim(**v**(t), **v**(t'))

Model	Spearman ρ
SVS	38.35
LSA	49.43
NMF	**52.99**
LDA	**53.39**
LSA_{mix}	49.76
NMF_{mix}	51.62
LDA_{mix}	51.97

Judge similarity of words out of context

Compared with 353 word pairs judged by humans

Lexical substitution

Model	Kendall's τ_b
SVS	11.05
Add-SVS	12.74
Add-NMF	12.85
Add-LDA	12.33
Mult-SVS	14.41
Mult-NMF	13.20
Mult-LDA	12.90
Cont-NMF	14.95
Cont-LDA	13.71
Cont-NMF$_{mix}$	**16.01**
Cont-LDA$_{mix}$	**15.53**

Rank appropriate substitutions

200 target words, in 2000 sentences. Human provided substitution

Future work

Distinguish target words and contextual features

Future work

Compute collective influence of contexts

Future work

Avoid usage of a global set of senses

Future work

Conclusion

Word meaning is represented as a distribution over latent senses. Contexts modulate word meaning distribution.

NMF and LDA are used to induce the latent senses.

This method outperforms previously reported results in word similarity and lexical substitution.

References

G. Dinu, M. Lapata (2010): Measuring Distributional Similarity in Context
Proceedings of EMNLP-10

D.D. Lee, and S.H. Seung (2001): Algorithms for Non-negative Matrix Factorization
Advances in Neural Information Processing Systems

Artwork credits

- Unicorn and rainbow http://weheartit.com/entry/8790261
- Startup quote http://startupquote.com/post/3504489915
- Probability http://www.flickr.com/photos/aidanmorgan/3249101355
- Fields http://www.flickr.com/photos/drhundertwasser/1506349540
- Arrows http://www.etsy.com/listing/64265171/arrow-pin-set-colors-you-choose
- Clips http://weheartit.com/entry/9652436
- Measure http://camereon.deviantart.com/art/Control-Measurement-154121448
- Tears http://www.tylershields.com/2011/05/08/unicorn-tears/
- Difference http://stranger04.deviantart.com/art/difference-192772921
- Cookies http://foodaddict.me/post/5188000478
- Meanings http://www.learnsomethingeveryday.co.uk/#/2011/05/09