

Combining Labels with a Multi-View VAE

NAACL-HLT, 2019-06-03



Alexander Hoyle

University College London¹



Lawrence Wolf-Sonkin

Johns Hopkins University

Microsoft®

Research

Hanna Wallach

Microsoft Research



Ryan Cotterell

University of Cambridge²



Isabelle Augenstein

University of Copenhagen

¹Now at University of Maryland. ²Moving to ETH Zurich.

Problem: annotators' schemata are different



Roger Ebert



Jay Sherman



Idea: combine annotations across sources

**Rotten
Tomatoes**

Roger Ebert



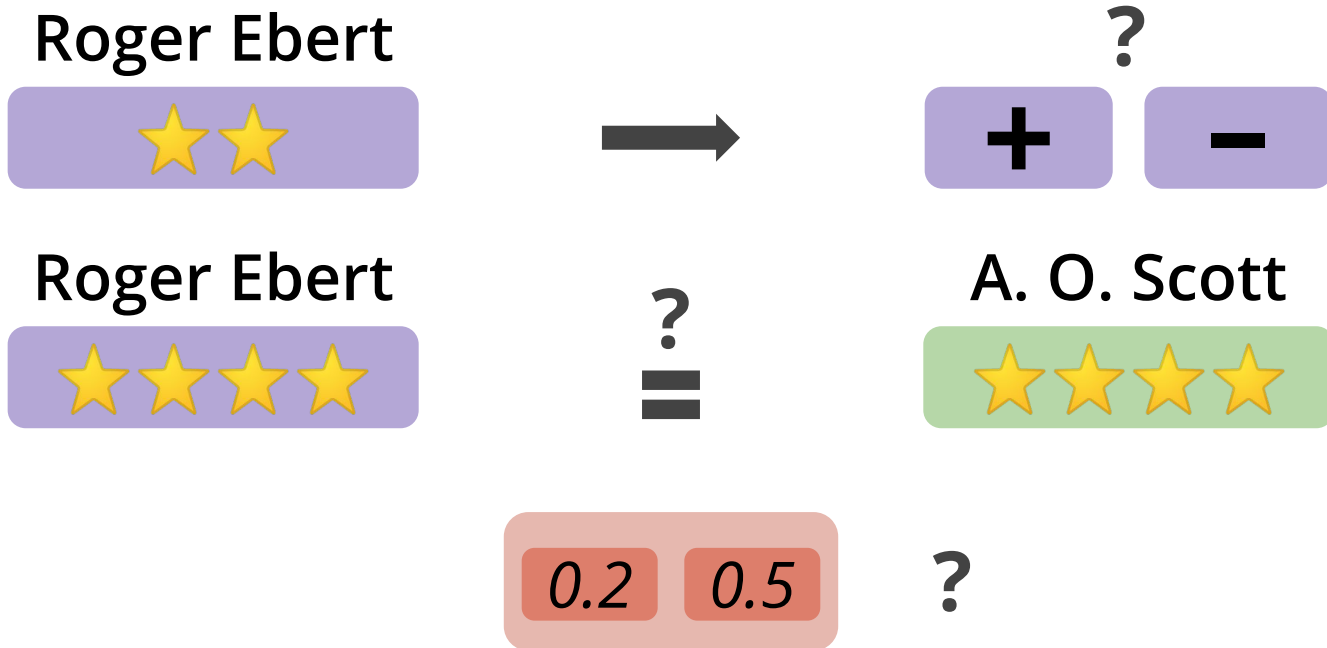
Jay Sherman



50%

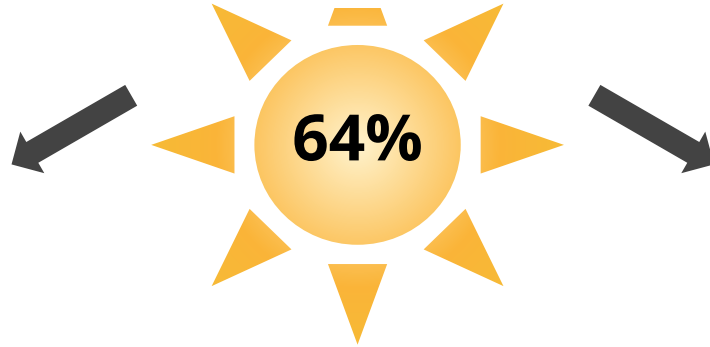
Aggregation increases coverage and reduces bias

...but can we aggregate consistently?



A generative story of reviewing

"True review"



Roger Ebert



Jay Sherman



This talk summarized in one slide

Combine disparate annotation schemes into one
...with generative modeling!

Merging Sentiment Lexica with a VAE

Sentiment lexica are also inconsistent

“good”

SentiWordNet

[0.672, 0.0]

MPQA

Positive

SenticNet 5

0.849

Hu-Liu

Positive

General Inquirer

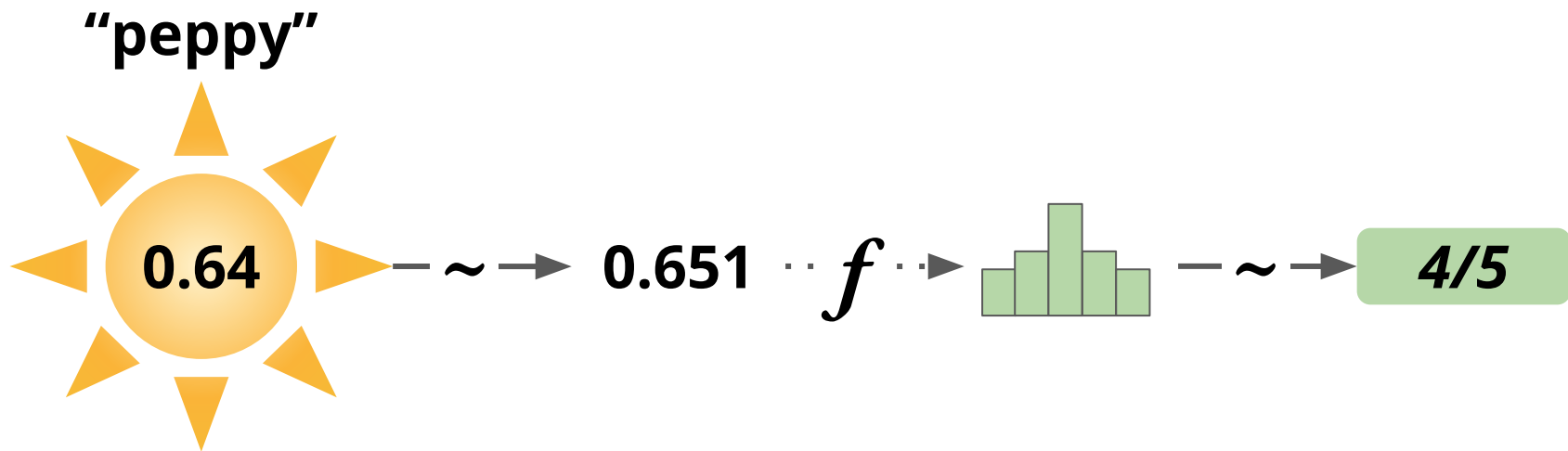
Positive

VADER

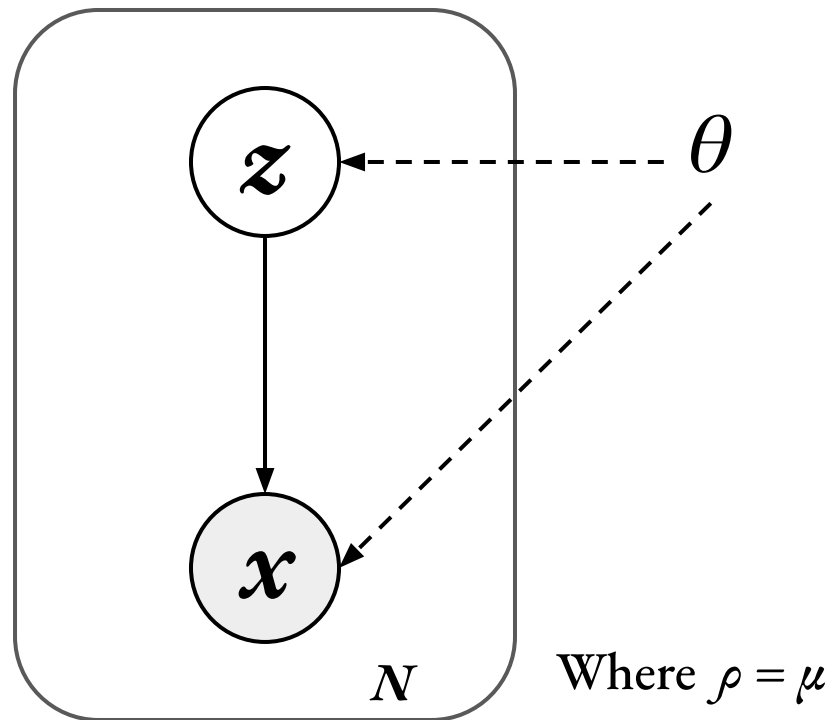
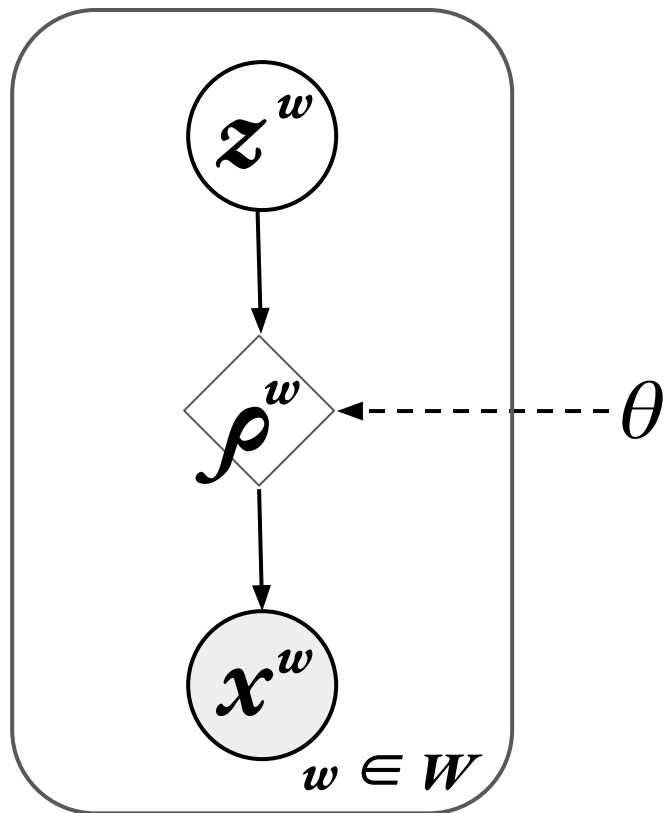
*[0, 0, 0, 0, 0,
4, 4, 1, 1]*

Generating labels is an imperfect process

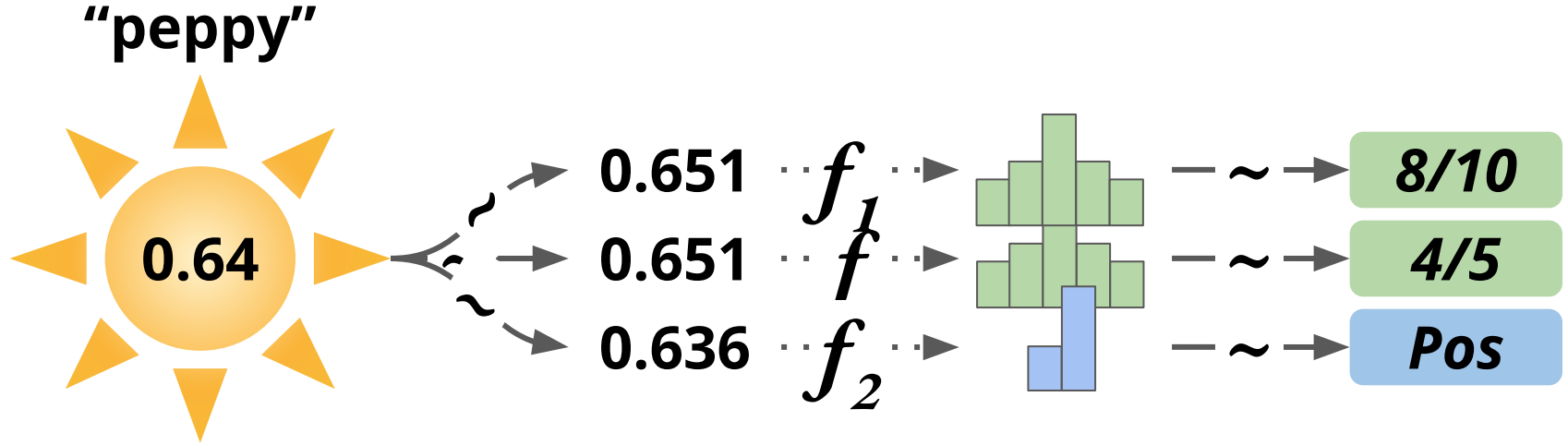
Suppose there exists a true, unobserved label



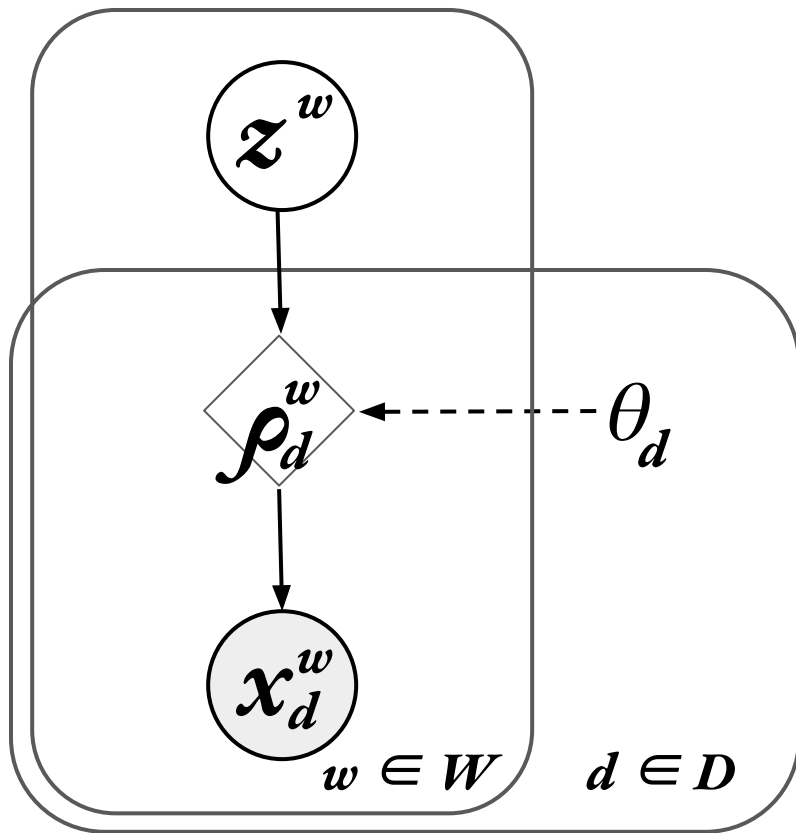
One latent variable per datapoint



Problem: observations don't have consistent scales



Solution: lexicon-specific emission distributions



$$z^w \sim \text{Dir}(\alpha^w)$$

$$\rho_d^w = f(z^w; \theta_d)$$

$$x_d^w \sim P_d(x_d^w | \rho_d^w)$$

Individual distributions are schema-dependent

 P_d

SentiWordNet

 $\mathcal{N}(\rho, 0.01 I)$

MPQA

 $\text{Bern}(\rho)$

SenticNet 5

 $\mathcal{N}(\rho_\mu, \rho_\sigma)$

Hu-Liu

 $\text{Bern}(\rho)$

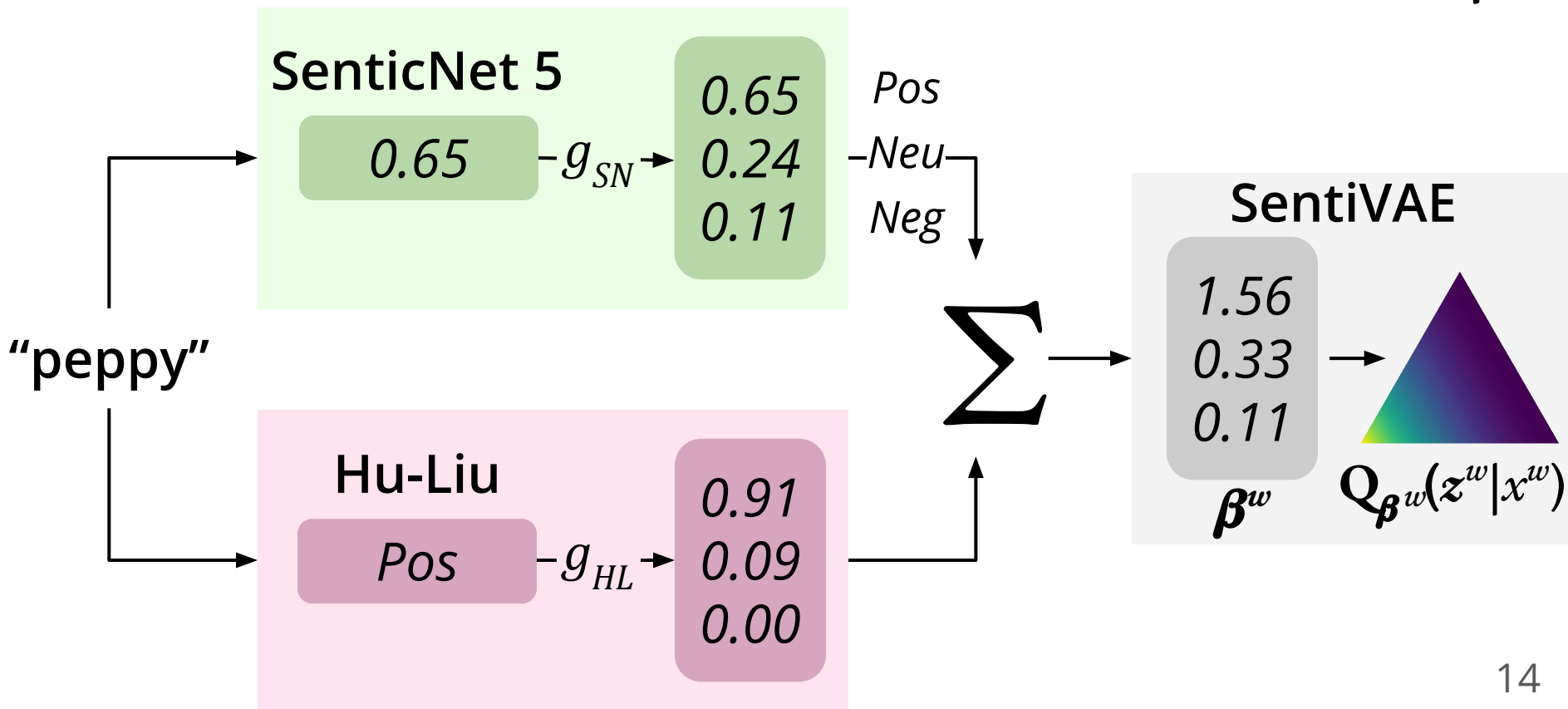
General Inquirer

 $\text{Bern}(\rho)$

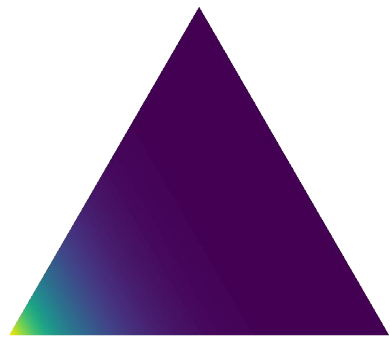
VADER

 $\text{Mult}(\rho)$

Encoder approximates intractable posterior $P(\mathcal{Z}|\mathcal{X})$

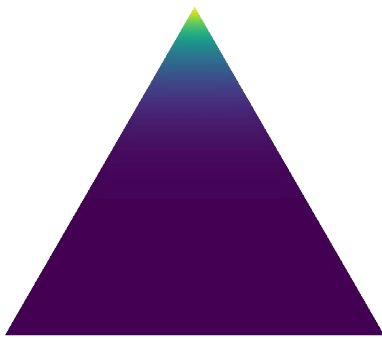


Distribution over latent representation is interpretable



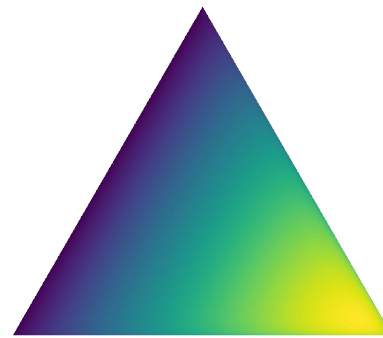
"Superb"

6.00, 1.00, 1.00



"Terrible"

1.00, 1.00, 6.99



"Portuguese"

1.07, 1.89, 1.05

Evaluation

Evaluation task: predict sentence sentiment

x_i

“The movie was good”

y_i

Pos

SenticNet

0.94

0.84

-

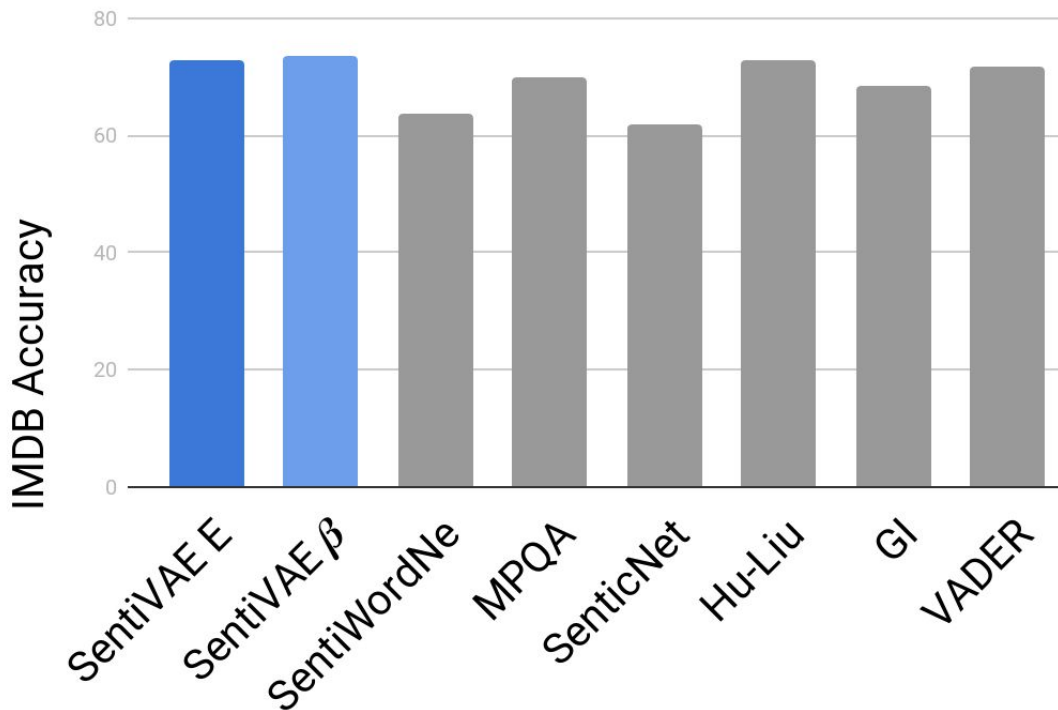
0.85

$\frac{1}{4}\Sigma$

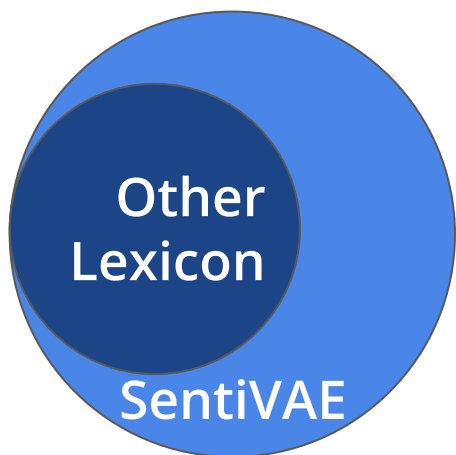
0.87

1

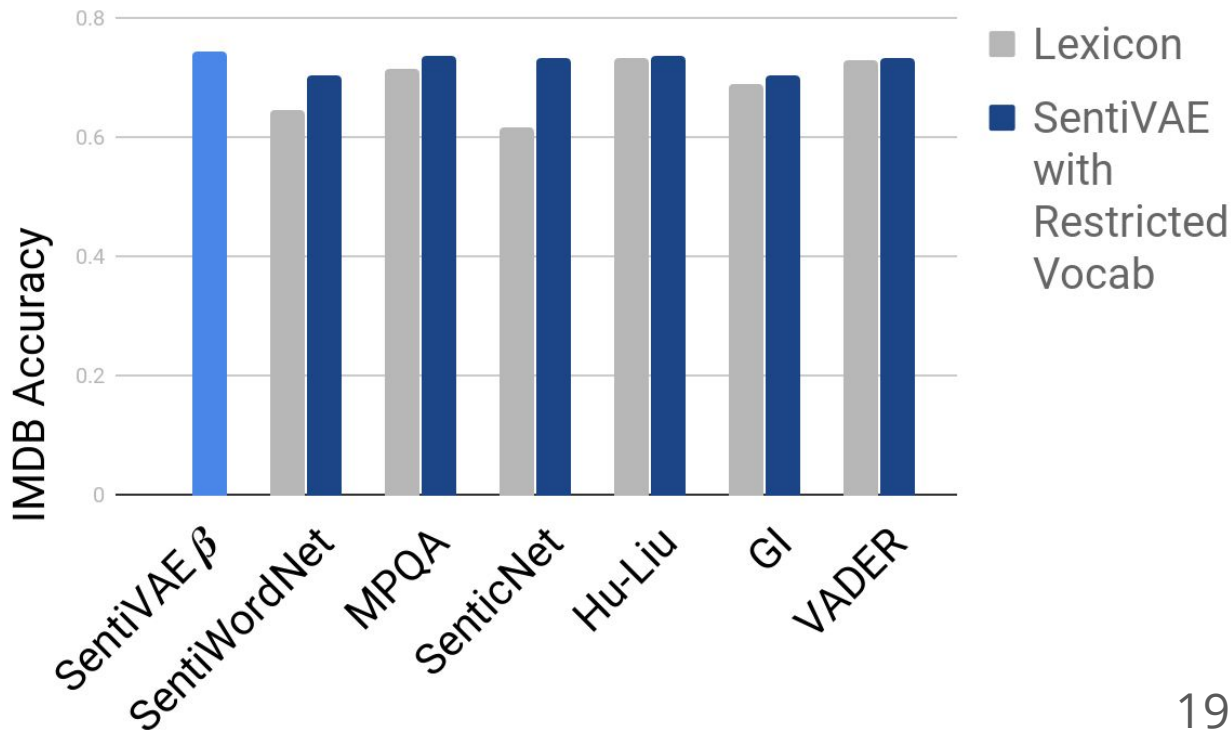
Our representation outperforms individual lexica



Performance owed to better representation in addition to greater coverage



**Vocabulary
Coverage**



For what other kinds of annotation schemes could this technique be useful?

Thanks

- Co-authors: Lawrence Wolf-Sonkin, Hanna Wallach, Ryan Cotterell, and Isabelle Augenstein
- University of Copenhagen for travel funding

Fin