Probabilistic Graphical Models for Credibility Analysis in Evolving Online Communities

Subhabrata Mukherjee
Max Planck Institute for Informatics, Germany

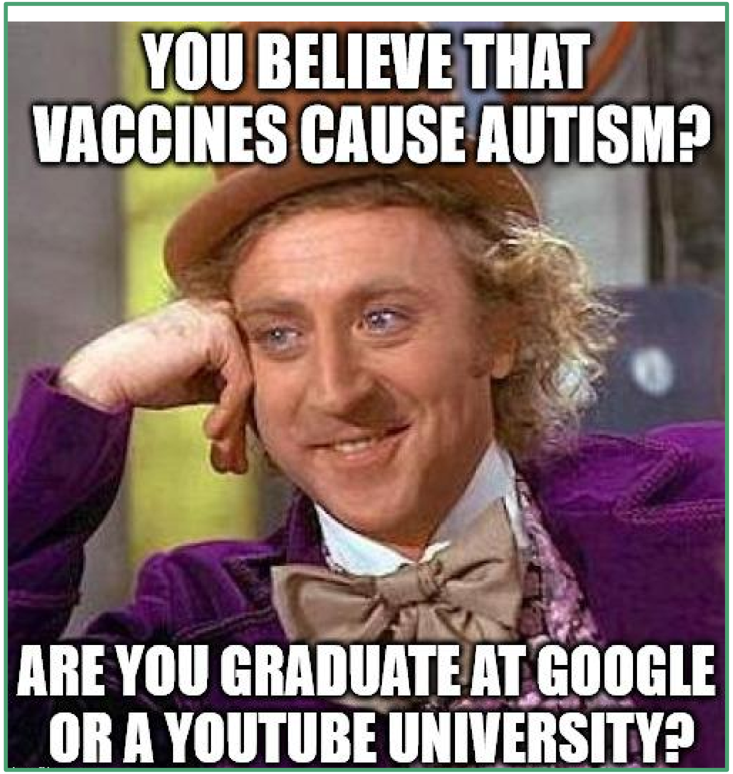
smukherjee@mpi-inf.mpg.de

Outline

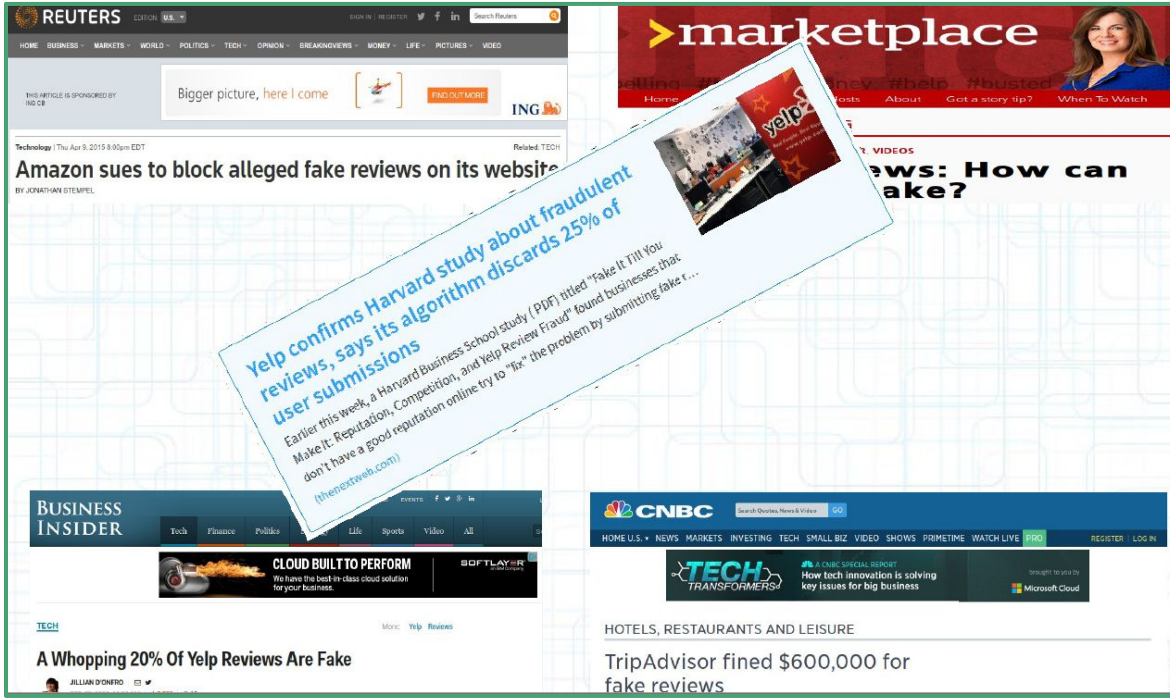
- Motivation
- Prior Work and its Limitations
- Credibility Analysis
  - ↘ Framework for Online Communities
  - ↘ Temporal Evolution of Online Communities
  - ↘ Credibility Analysis of Product Reviews
- Conclusions

Online Communities as a Knowledge Resource

- Online communities are massive repositories of knowledge accessed by regular users and professionals
  - ↘ 59% of adult U.S. population and half of U.S. physicians rely on online resources [IMS Health Report, 2014]
  - ↘ 40% of online consumers consult online reviews before buying products [Nielson Corporation, 2016]
- However their usability is restricted due to serious credibility concerns (e.g., spams, misinformation, bias etc.)



"Rapid spread of misinformation online" --- one of top 10 challenges as per The World Economic Forum



Misinformation for health can have hazardous consequences

Outline

- Motivation
- Prior Work and its Limitations
- Credibility Analysis
  - ↘ Framework for Online Communities
  - ↘ Temporal Evolution of Online Communities
  - ↘ Credibility Analysis of Product Reviews
- Conclusions

Truth Finding

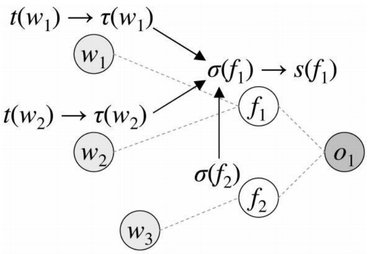
- Structured data (e.g., SPO triples, tables, networks)
- Objective facts (e.g., Obama_BornIn_Hawaii vs. Obama_BornIn_Kenya)
- No contextual data (text)
- No external KB, metadata

Table 1: Observation Tables

	$\mathcal{X}^{(1)}$		$\mathcal{X}^{(2)}$		$\mathcal{X}^{(3)}$	
Object	City	Height	City	Height	City	Height
Bob	NYC	1.72	NYC	1.70	NYC	1.90
Mary	LA	1.62	LA	1.61	LA	1.85
Kate	NYC	1.74	NYC	1.72	LA	1.65
Mike	NYC	1.72	LA	1.70	DC	1.85
Joe	DC	1.72	NYC	1.71	NYC	1.85

Table 2: Ground Truth and Conflict Resolution Results

	Ground Truth		Voting/Averaging		CRH	
Object	City	Height	City	Height	City	Height
Bob	NYC	1.72	NYC	1.77	NYC	1.72
Mary	LA	1.62	LA	1.69	LA	1.62
Kate	NYC	1.75	NYC	1.70	NYC	1.74
Mike	NYC	1.71	DC	1.76	NYC	1.72
Joe	DC	1.73	NYC	1.76	DC	1.72



Linguistic Analysis

- Unstructured text
- Subjective information (e.g., opinion spam, bias, viewpoint)
- External KB (e.g., WordNet, KG)
- No network / interactions, metadata

Research Questions

1. How can we *jointly* leverage users, network, and context for credibility analysis in online communities?
2. How can we model users' *evolution*?
3. How can we deal with *limited* data?
4. How can we generate *interpretable* explanations for credibility verdict?

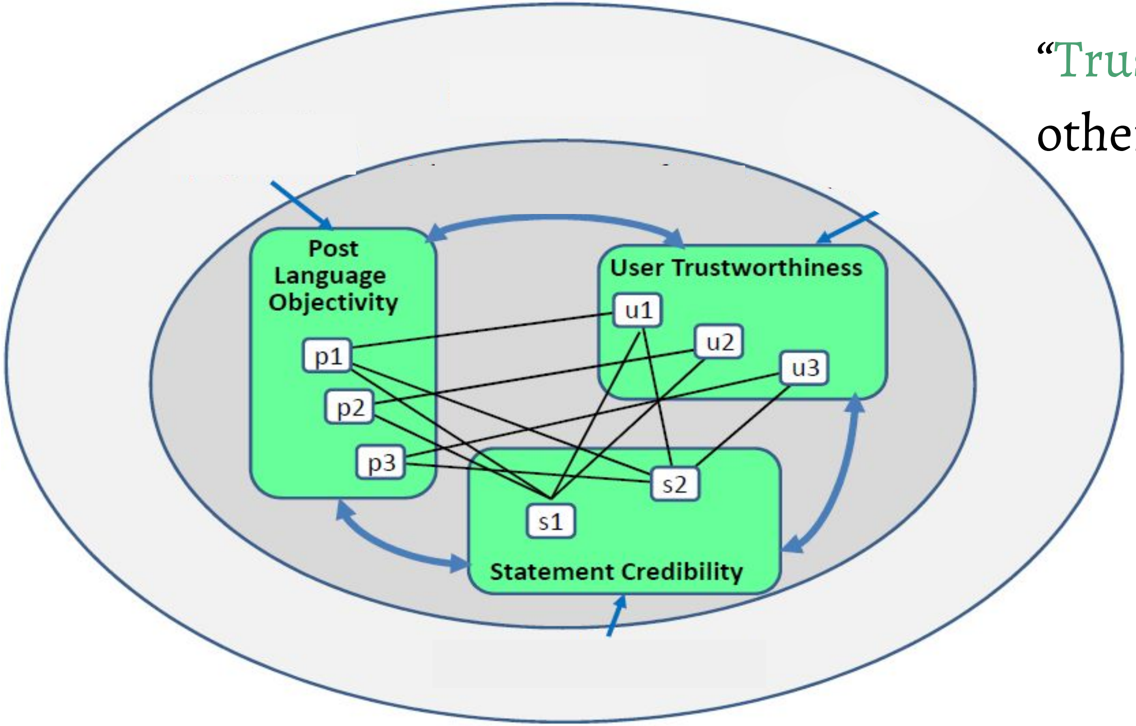
Contributions

- Credibility Analysis Framework for Online Communities
  - ↘ *Classification:* Health Communities [SIGKDD 2014]
  - ↘ *Regression:* News Communities [CIKM 2015]
- Temporal Evolution of Online Communities
  [ICDM 2015, SIGKDD 2016]
- Credibility Analysis of Product Reviews
  [ECML-PKDD 2016, SDM 2017]

Outline

- Motivation
- Prior Work and its Limitations
- Credibility Analysis
  - ↘ Framework for Online Communities
  - ↘ Temporal Evolution of Online Communities
  - ↘ Credibility Analysis of Product Reviews
- Conclusions

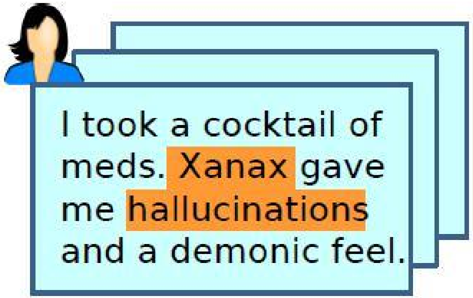
What is Credibility?



"A statement is credible if it is reported by a trustworthy user in an objective language"

"Trustworthy users corroborate each other on credible statements"

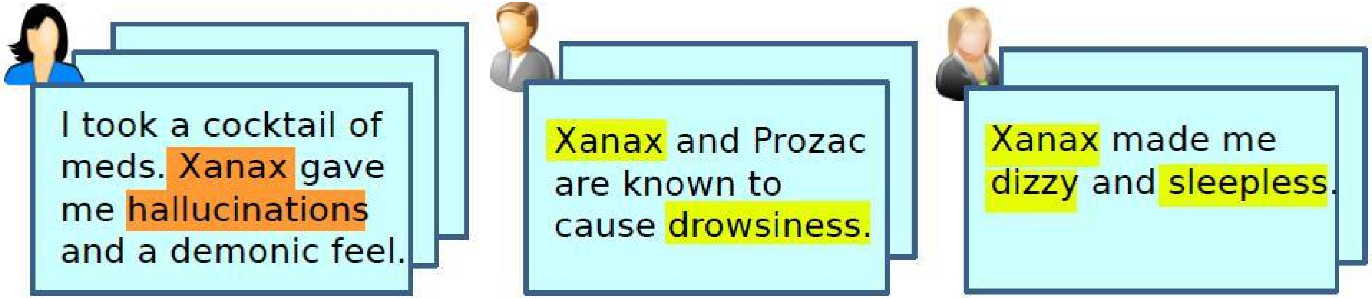
Credibility Analysis Framework for *Classification*



Problem: Given a set of posts from different users, extract credible statements (*subject-predicate-object* triples like DrugX_HasSideEffect_Y) from trustworthy users

Subhabrata Mukherjee, Gerhard Weikum and Cristian Danescu-Niculescu-Mizil: SIGKDD 2014

Credibility Analysis Framework for *Classification*



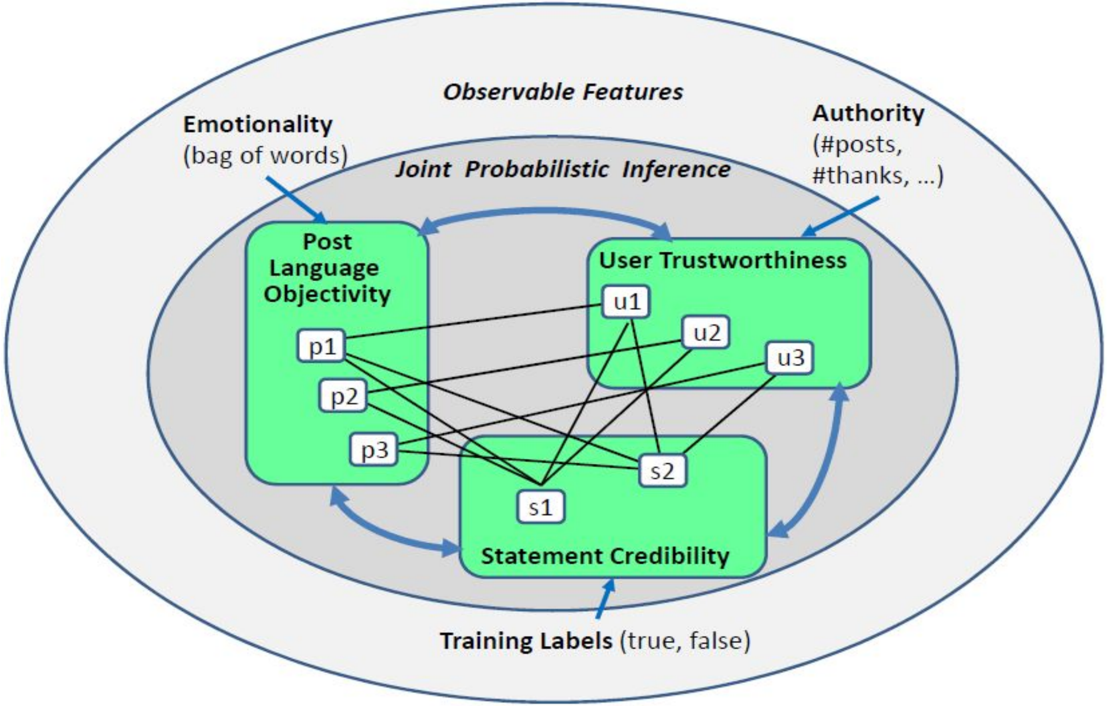
Problem: Given a set of posts from different users, extract credible statements (*subject-predicate-object* triples like DrugX_HasSideEffect_Y) from trustworthy users

Subhabrata Mukherjee, Gerhard Weikum and Cristian Danescu-Niculescu-Mizil: SIGKDD 2014

Network of Interactions: Cliques

➔ Each user, post, and statement is a random variable with edges depicting interactions. Variables have observable features (e.g, authority, emotionality).

➔ A *clique* is formed between each *user* writing a *post* containing a *statement*.



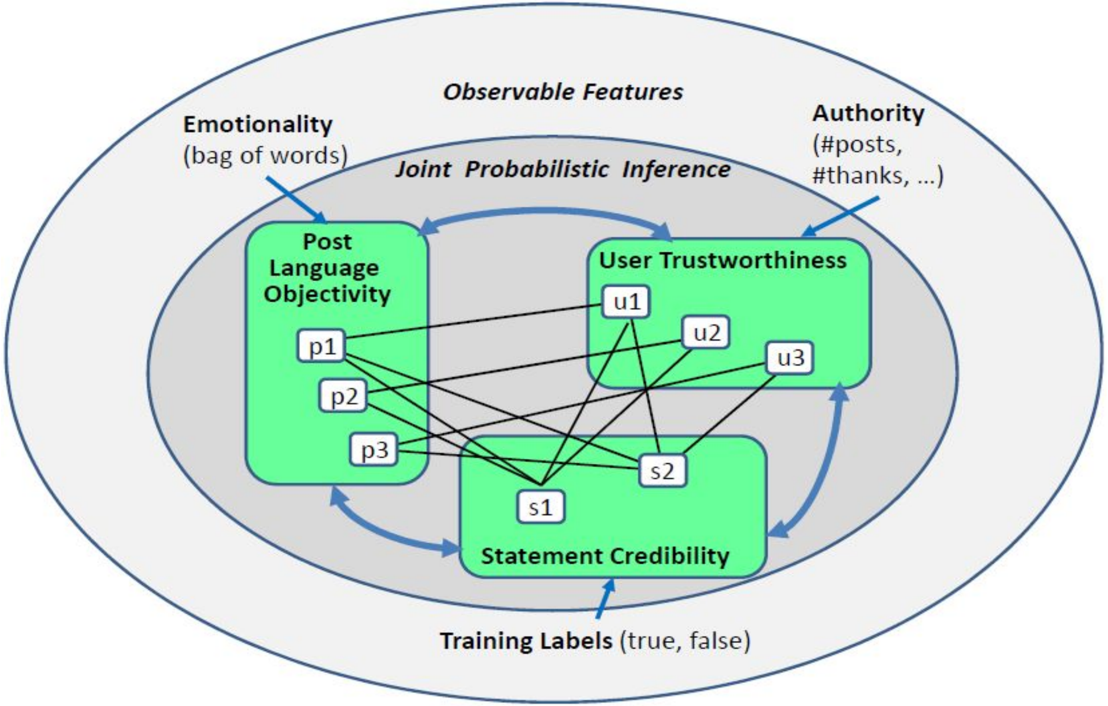
Statements: An IE tool generates candidate triple patterns like:

Xanax_causes_headache,
Xanax_gave_demonic-feel

Potentially thousands of such triples, with only a handful of credible ones

Network of Interactions: Cliques

Each user, post, and statement is a random variable with edges depicting interactions



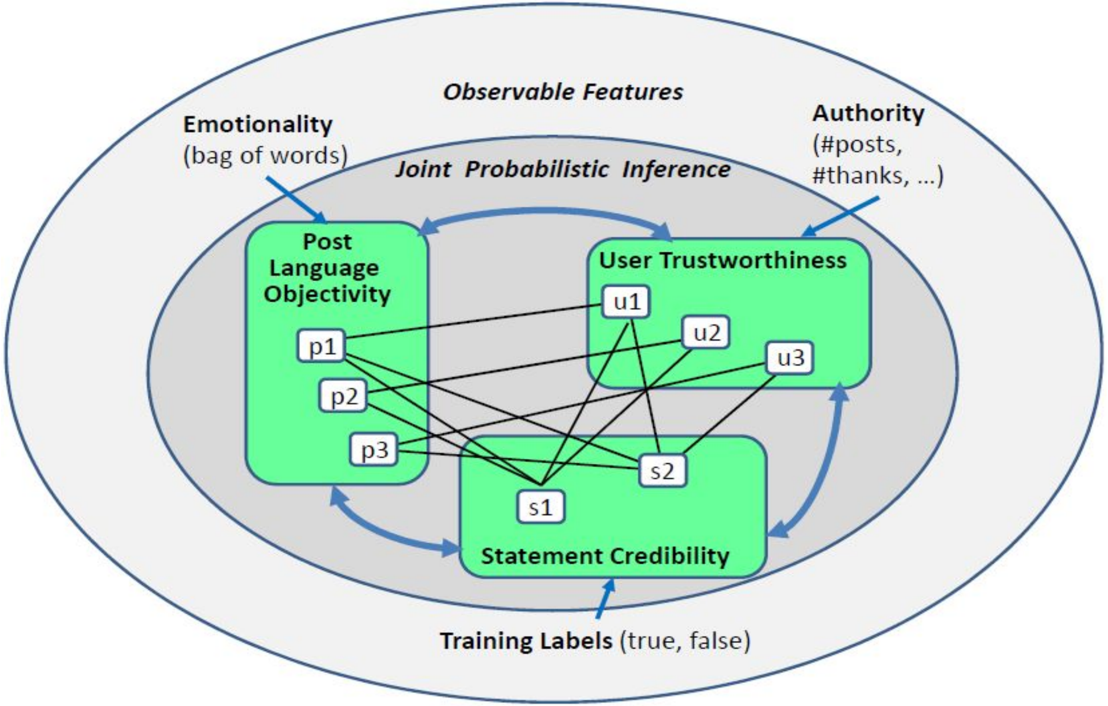
Statements: An IE tool generates candidate triple patterns like:

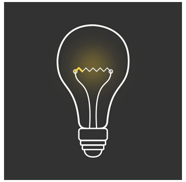
Xanax_causes_headache,
Xanax_gave_demonic-feel

Potentially thousands of such triples, with only a handful of credible ones

Idea: *Trustworthy* users corroborate on *credible* statements in *objective* language

Conditional Random Field to Exploit Joint Interactions (Users + Network + Context)





How to complement expert medical knowledge with large scale non-expert data?

Partial Supervision: Expert stated (top 20%) side-effects of drugs as partial training labels.

Model predicts labels of unobserved statements.

Semi-Supervised Conditional Random Field

1. Estimate user trustworthiness: $trust = \frac{\sum_{Statement} label(Statement = True)}{\langle Statement \rangle}$

2. Estimate label of unknown statements S_u by Gibbs Sampling:

$$Pr(S_u \mid Post, User, S_{labeled} ; W) \propto \prod_{\Phi \in Cliques} trust(user) \cdot \Phi(S_u, Post_{features}, User_{features}; W)$$

3. Maximize log-likelihood to estimate feature weights:

$$W' = argmax_W \sum_{S_u} Pr(S_u \mid Post, User, S_{labeled}; W) \cdot log\, Pr(S_u\ ,\ S_{labeled} \mid Post, User; W)$$

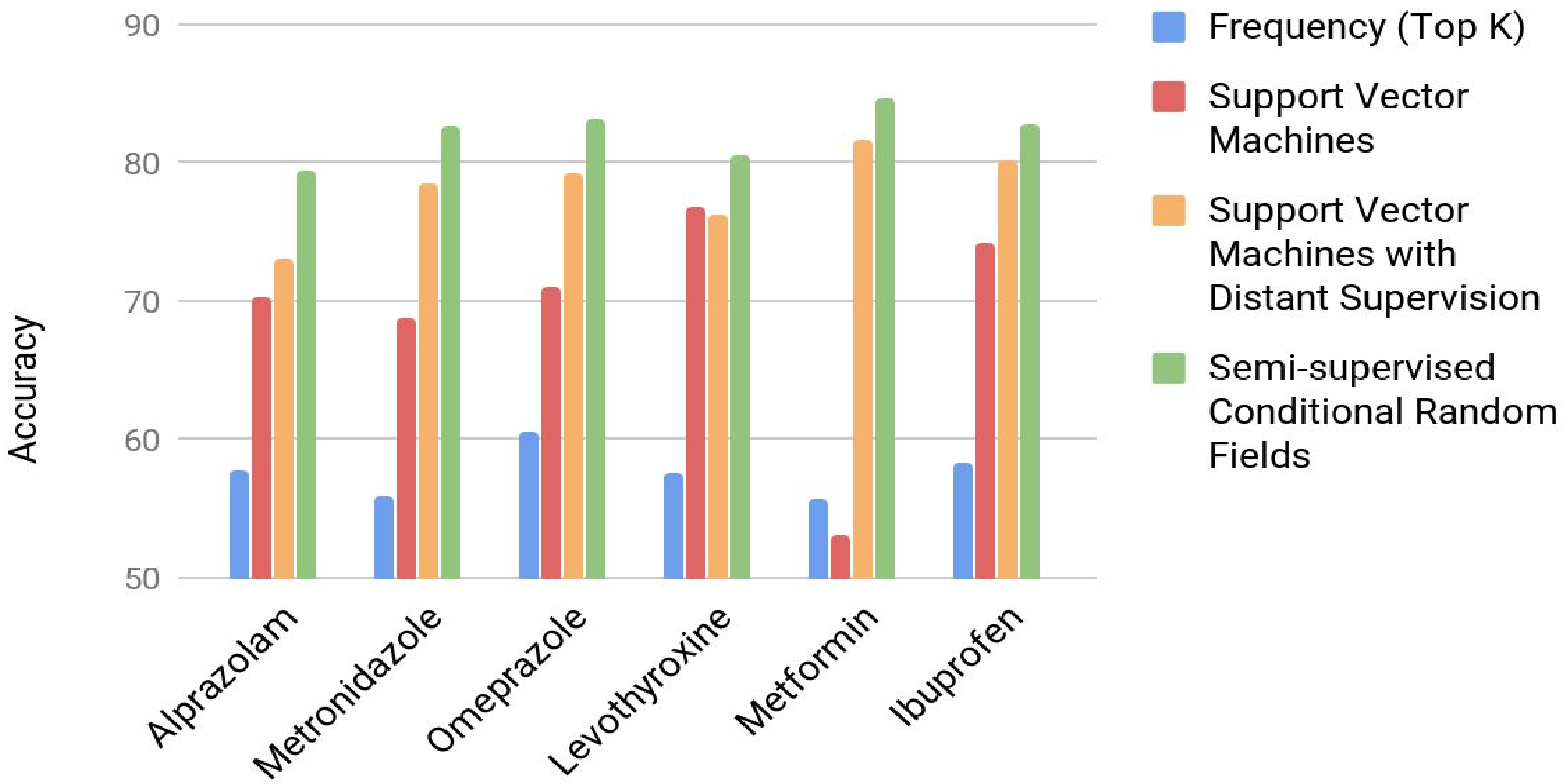
4. Apply E-Step and M-Step till convergence

Healthforum Dataset

- Healthboards.com community (www.healthboards.com) with 850,000 registered users and 4.5 million posts

- Expert labels about drugs from MayoClinic (www.mayoclinic.org)
  ↘ 6 widely used drugs for experimentation

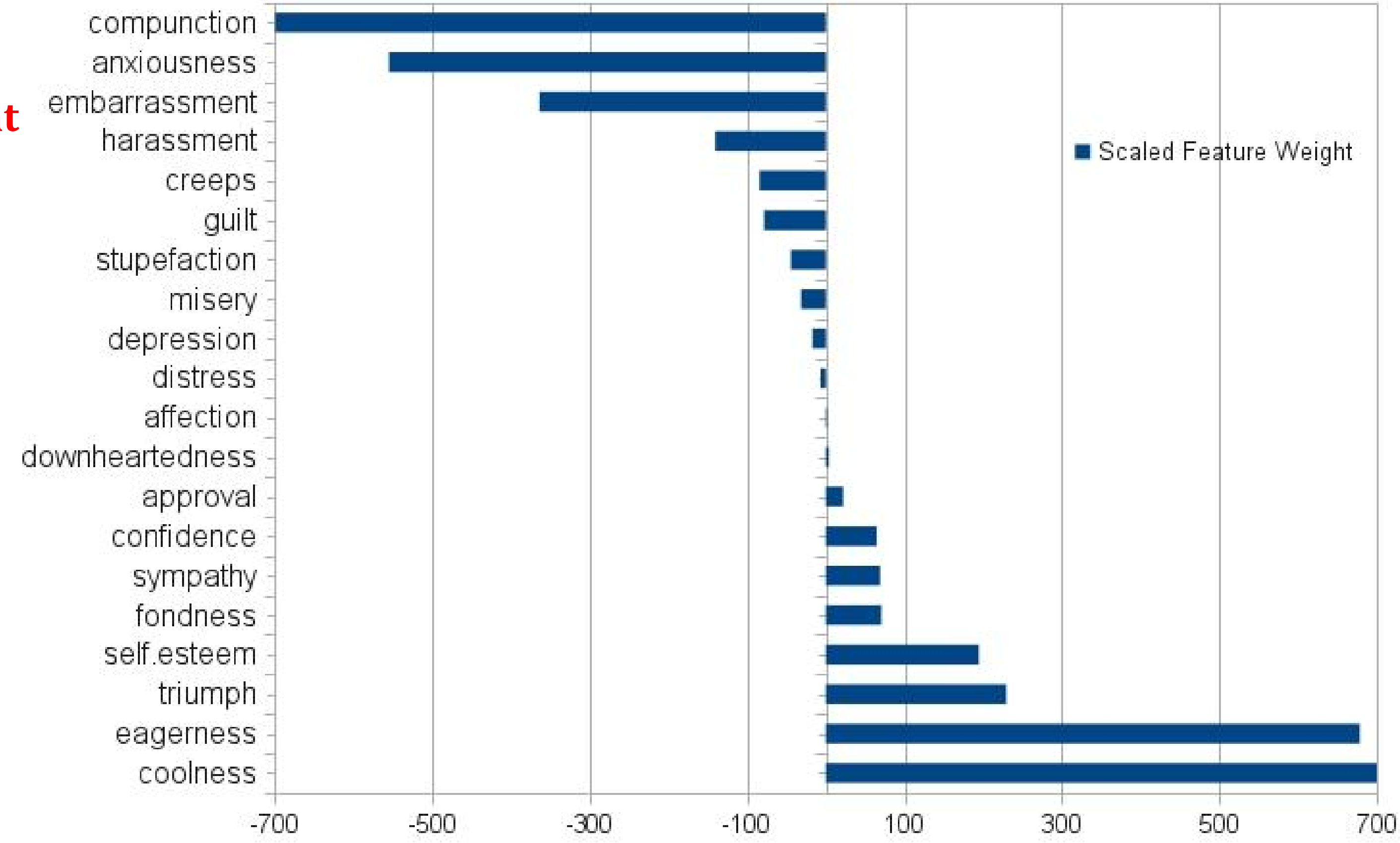
EXPERIMENTAL RESULTS



What constitutes credible language?

compunction
anxiety
embarrassment
misery
distress

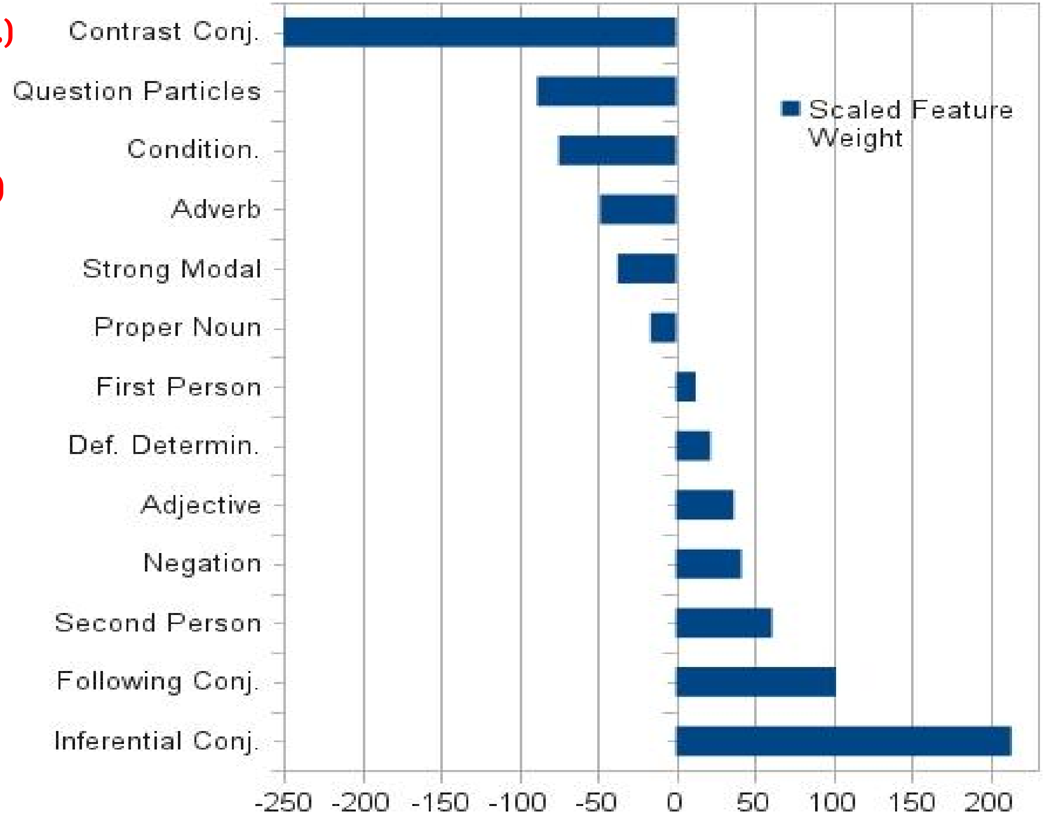
confidence
sympathy
self-esteem
eagerness
coolness



Affective Emotions

What constitutes credible language?

contrast (despite, though, ..)
question (what, why, ..)
conditional (if)
adverb (maybe, probably, ..)
modality (might, could, ..)



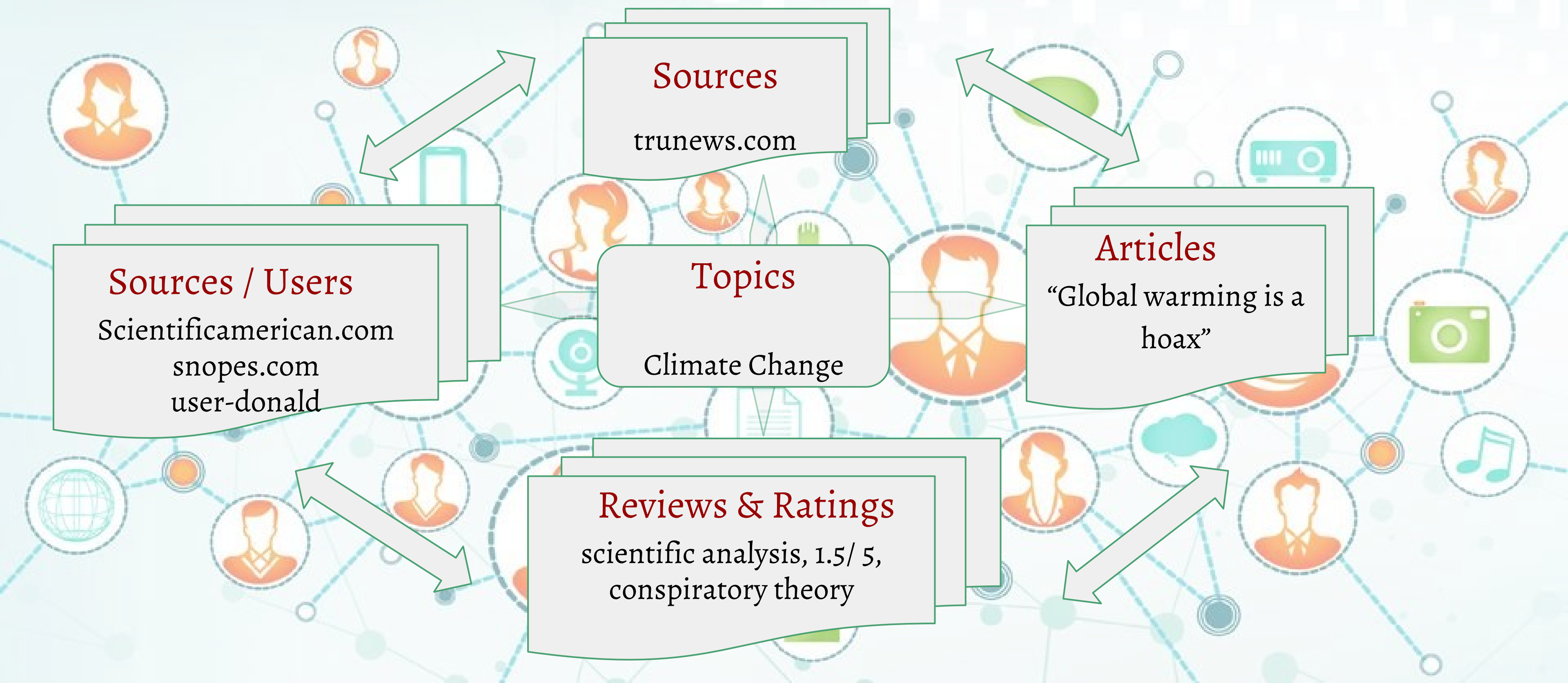
determiner (this, that,..)
negation (not, never, ..)
second person (you, ..)
conjunction (therefore, consequently, ..)

Discourse and Modalities

Credibility Analysis Framework for *Regression*

In many online communities users *rate* items on their quality

Credibility Analysis in News Communities



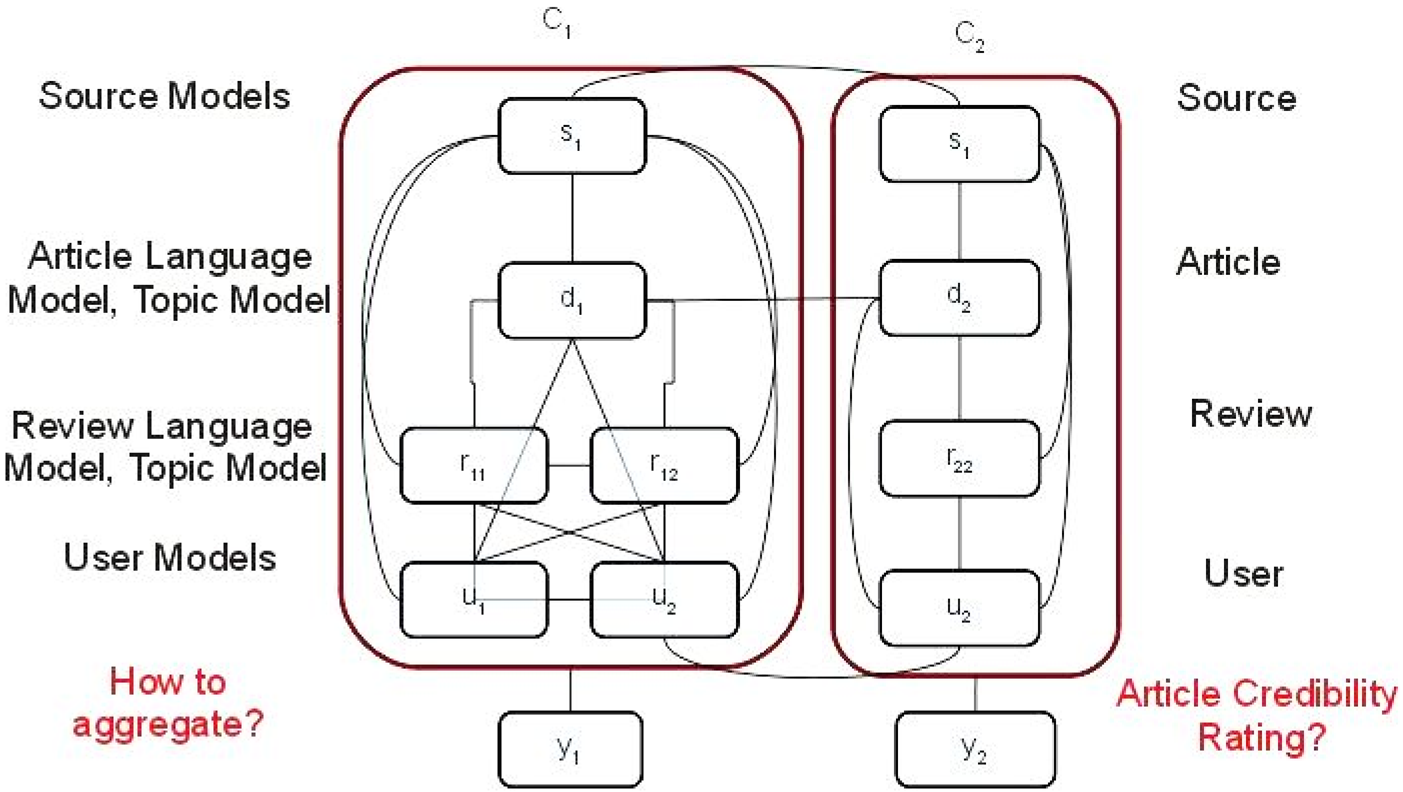
However, user feedback is often *subjective*; influenced by their bias and viewpoints

Credibility Analysis Framework for *Regression*

We use CRF to capture these *mutual* interactions in news communities (e.g., newstrust.net, digg, reddit) to *jointly rank* all of the underlying factors.

Idea: Trustworthy sources publish objective articles corroborated by expert users with credible reviews/ratings

Online Communities: Factors



Related to Ensemble Learning, Learning to Rank

How to incorporate *continuous ratings* instead of discrete labels in CRF ?

Probability Mass Function for discrete labels: $p(y|X) = \frac{exp(\Psi)}{\sum_y exp(\Psi)}$

Probability Density Function for continuous ratings: $p(y|X) = \frac{exp(\Psi)}{\int_{-\infty}^{\infty} exp(\Psi) dy}$

Energy Function to Combine All

partitions the user space

user expertise

error of predictor SVR

$$\psi(y, s, d, \langle u \rangle, \langle r \rangle) = -\sum_{u} \boxed{\alpha_u \mathbb{I}_u(d)(y - \mathrm{SVR}_u)^2}$$

$$- \sum_{s} \boxed{\beta_s} \mathbb{I}_s(d)(y - \mathrm{SVR}_s)^2 - \boxed{\gamma_1}(y - \mathrm{SVR}_L)^2 - \boxed{\gamma_2}(y - \mathrm{SVR}_T)^2$$

source trustworthiness **language objectivity** **topical perspective**

How to incorporate *continuous ratings* instead of discrete labels in CRF ?

- We show that a certain energy function for clique potential --- geared for reducing mean-squared-error --- results in multivariate gaussian p.d.f. !!!

$$P(y|X) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu))$$

- *Constrained* Gradient Ascent for inference

Predicting Article Credibility Ratings in Newstrust.net

Model	Only Title MSE	Title & Text MSE
Language Model: SVR		
Language (Bias and Subjectivity)	3.89	0.72
Explicit Topics	1.74	1.74
Explicit + Latent Topics	1.68	1.01
All Topics (Explicit + Latent) + Language	1.57	0.61
News Source Features and Language Model: SVR		
News Source	1.69	1.69
News Source + All Topics + Language	0.91	0.46
Aggregated Model: SVR		
Users + All Topics + Language + News Source	0.43	0.41
Our Model: CCRF+SVR		
User + All Topics + Language + News Source	0.36	0.33

Progressive decrease in mean squared error with more network interactions, and context

Take-away

- Semi-supervised and Continuous CRF to jointly identify trustworthy users, credible statements, and reliable postings in online communities

- A framework to incorporate richer aspects like user expertise, topics / facets, temporal evolution etc.

Outline

- Motivation
- Prior Work and its Limitations
- Credibility Analysis
  - ↘ Framework for Online Communities
  - ↘ Temporal Evolution of Online Communities
  - ↘ Credibility Analysis of Product Reviews
- Conclusions

Temporal Evolution

- Online communities are dynamic, as users join and leave; acquire new vocabulary; evolve and mature over time

- Trustworthiness and expertise of users *evolve* over time

How to capture evolving user expertise?

Illustrative Example for Review Communities

- Consider following camera reviews by the *same* user John:

"My first DSLR. Excellent camera, takes great pictures with high definition, without a doubt it makes honor to its name."

[Aug, 1997]

"The EF 75-300 mm lens is only good to be used outside. The 2.2X HD lens can only be used for specific items; filters are useless if ISO, AP,... . The short 18-55mm lens is cheap and should have a hood to keep light off lens."

[Oct, 2012]

Mukherjee et al.: ICDM 2015, SIGKDD 2016

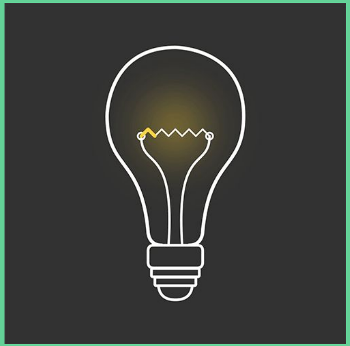
Illustrative Example for Review Communities

- Consider following camera reviews by John:

“My first DSLR. Excellent camera, takes great pictures with high

c ... .”

[Aug, 1997]



How can we quantify this change in users’ maturity / experience ?

How can we model this evolution / progression in users’ maturity?

“ ... ide. The 2.2X

I ... are useless if

ISO, AP,.... . The short 18-55mm lens is cheap and should have a hood to keep light off lens.”

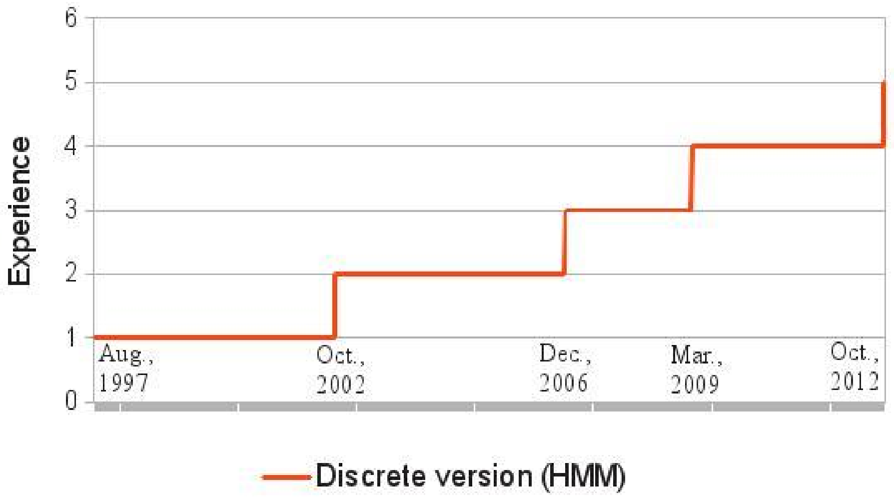
[Oct, 2012]

Mukherjee et al.: ICDM 2015, SIGKDD 2016

Prior Work: Discrete Experience Evolution

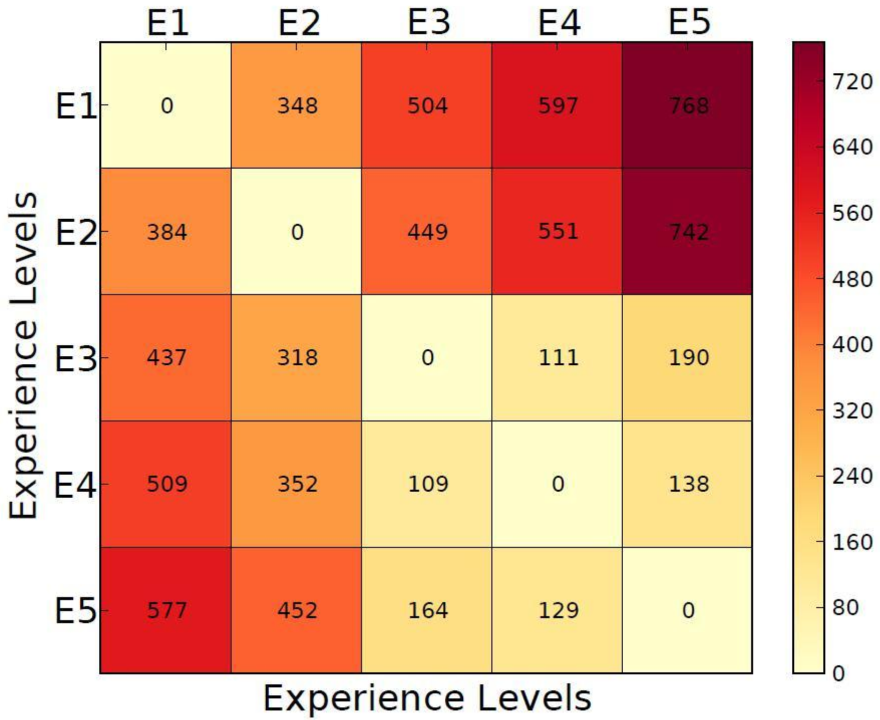
1. Users at similar levels of experience have similar facet preferences, and rating style (McAuley and Leskovec: WWW 2013)

2. Additionally, our work exploits *similar writing style* (Mukherjee, Lamba and Weikum: ICDM 2015)



Assumption: At each timepoint a user remains at the same level of experience, or moves to the next level

Language Model (KL) Divergence Increases with Experience

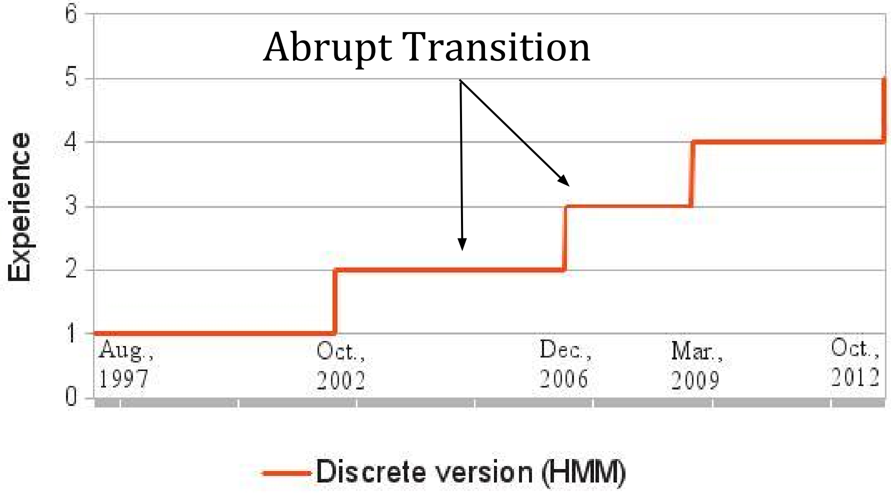


Experienced users have a *distinctive* writing style different than that of amateurs

Prior Work: Discrete Experience Evolution

1. Users at similar levels of experience have similar facet preferences, and rating style (McAuley and Leskovec, WWW 2013)

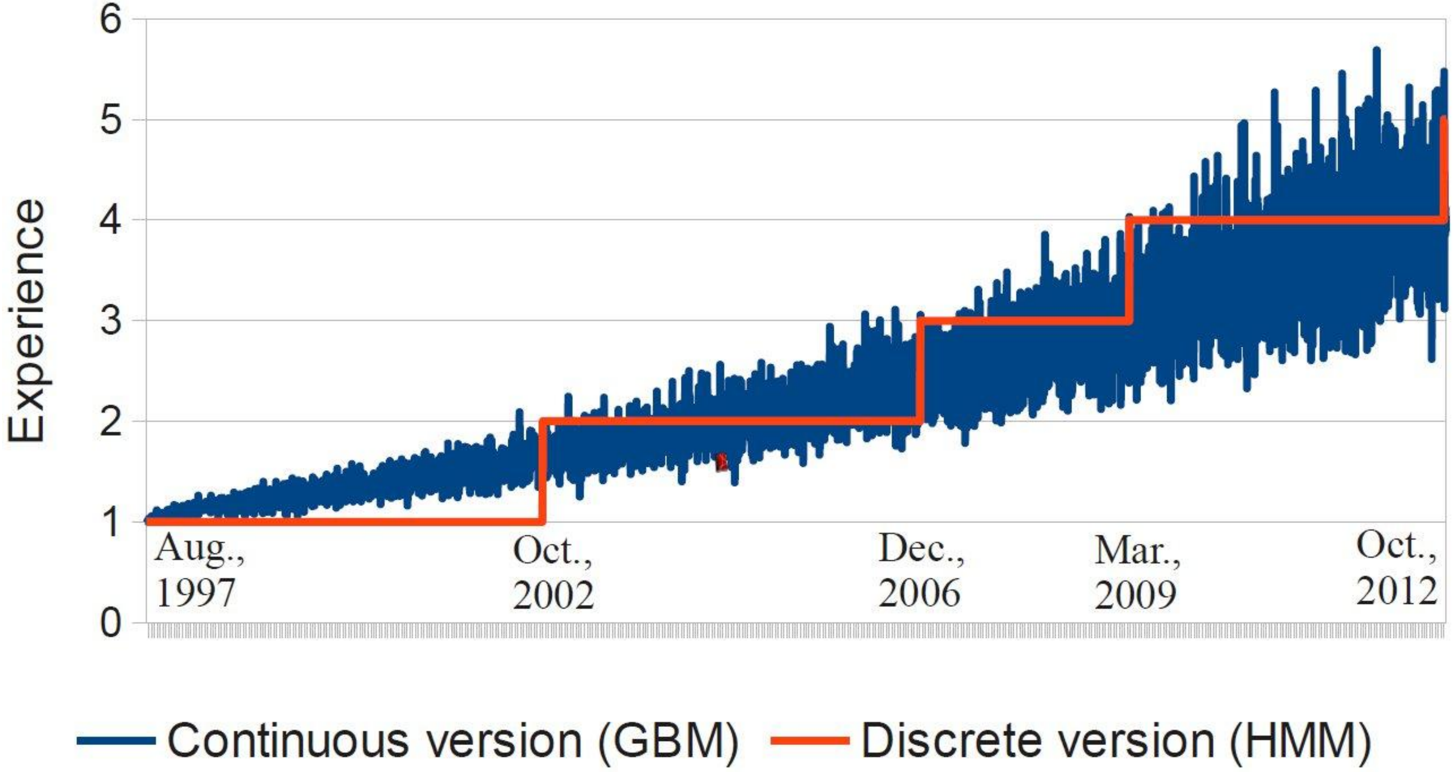
2. Additionally, our work exploits *similar writing style* (Mukherjee, Lamba and Weikum, ICDM 2015)



Assumption: At each timepoint a user remains at the same level of experience, or moves to the next level

Continuous Experience Evolution

(Mukherjee, Günnemann and Weikum, SIGKDD 2016)



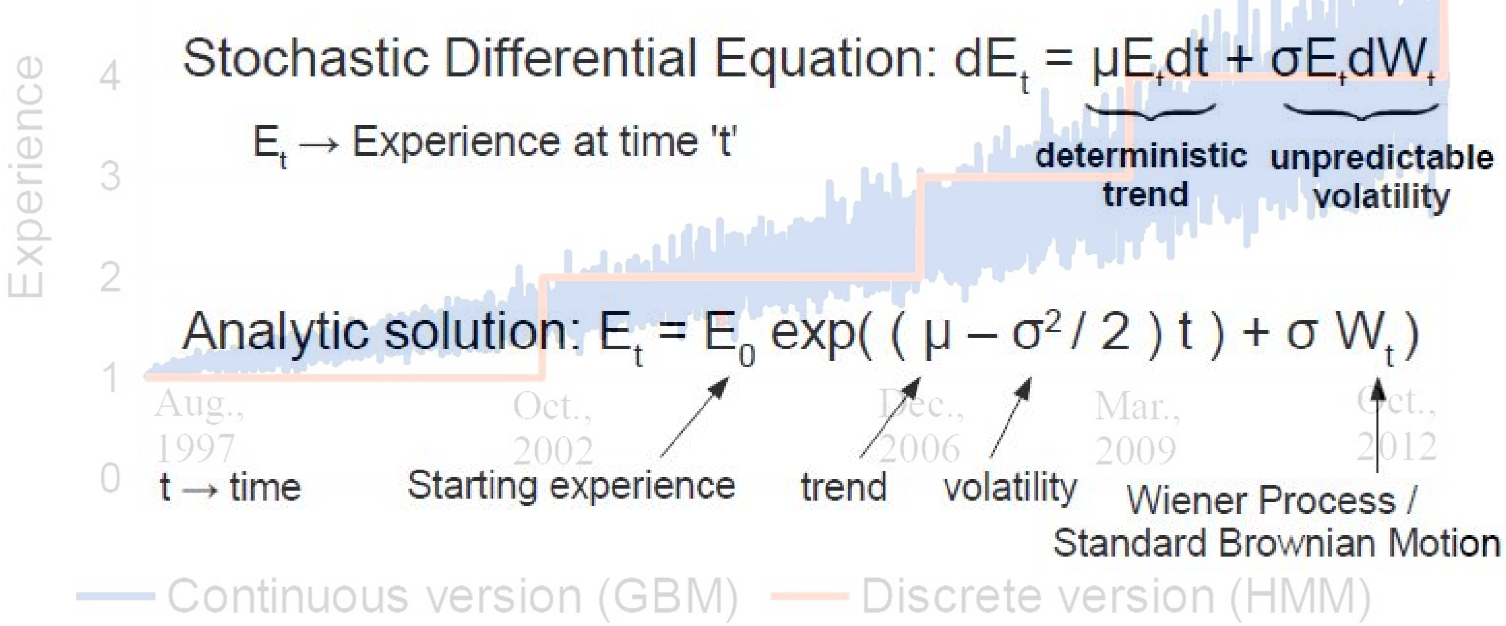
Continuous Experience Evolution: Assumptions

- ★ Continuous-time process, always positive
- ★ Markovian assumption: Experience at time *t* depends on that at *t-1*
- ★ Drift: Overall trend to increase over time
- ★ Volatility: Progression may not be smooth with occasional volatility

  E.g.: series of expert reviews followed by a sloppy one

Geometric Brownian Motion

We show these properties to be satisfied by the continuous-time stochastic process: *Geometric Brownian Motion*

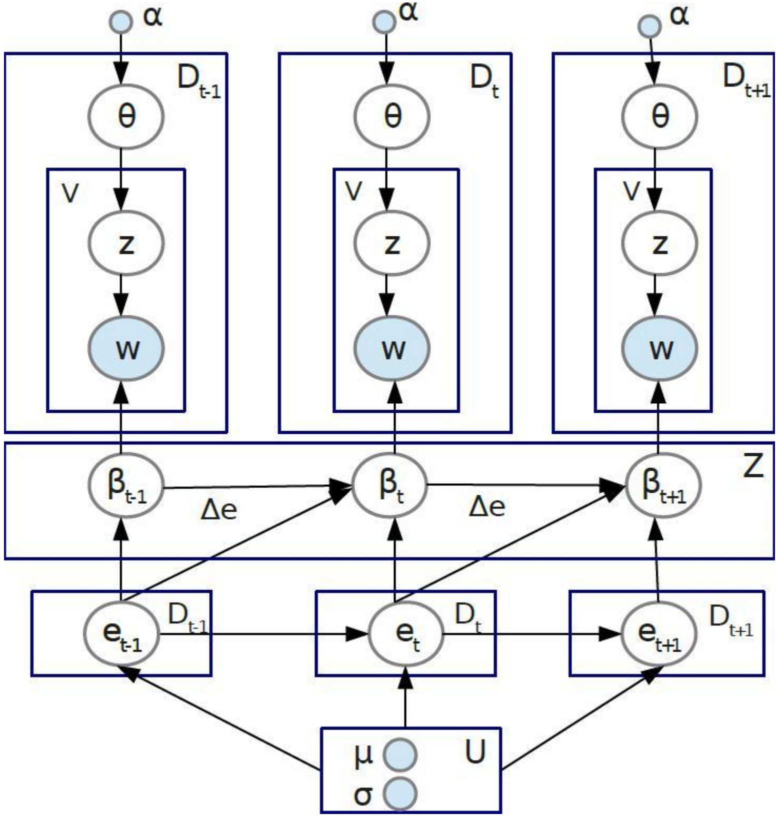


Language Model (LM) Evolution

- Users' LM also evolve with experience evolution
- Smoothly evolve over time preserving Markov property of experience evolution
- Variance of LM should change with experience change
- Brownian Motion to model this desiderata:

$$\underbrace{\beta_{t,z,w}}_{\text{LM at time 't'}} \sim \text{Normal}\left(\underbrace{\beta_{t-1,z,w}}_{\text{LM at time 't-1'}} , \sigma \cdot \underbrace{|e_t - e_{t-1}|}_{\text{Experience change}} \right)$$

Inference



Topic Model (Blei et al., JMLR '03)

+

Users (Author-topic model,Rosen-Zvi et al., UAI '04)

+

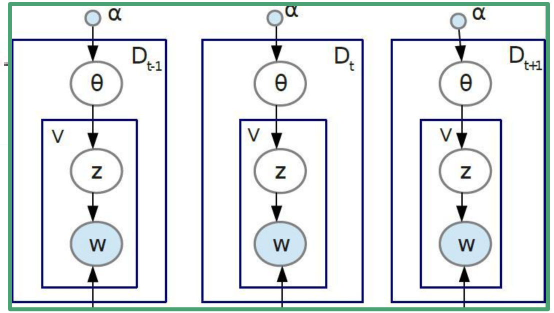
Continuous Time (Dynamic topic model, Wang et al., UAI '08)

+

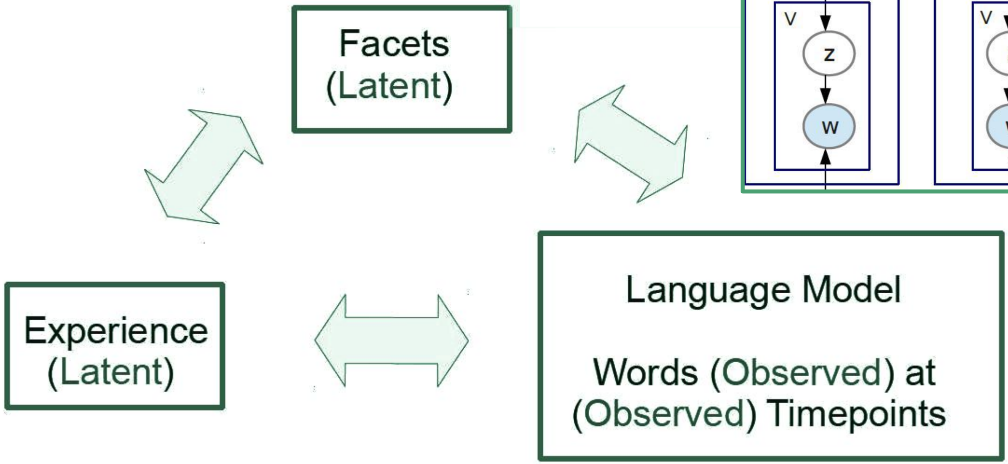
Continuous Experience (this work)

Sampling based Inference

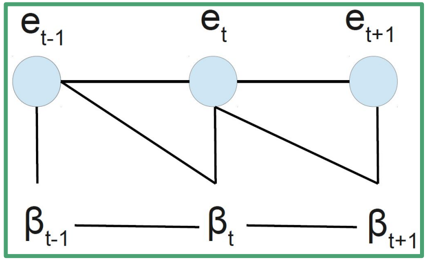
Gibbs Sampling for Facets



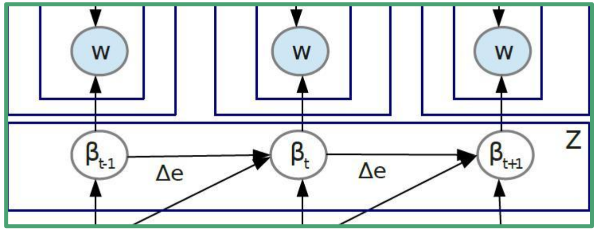
E.g.: The smell of grains a malts
on the nose with the slight hop aroma...
The taste of the beer is crisp



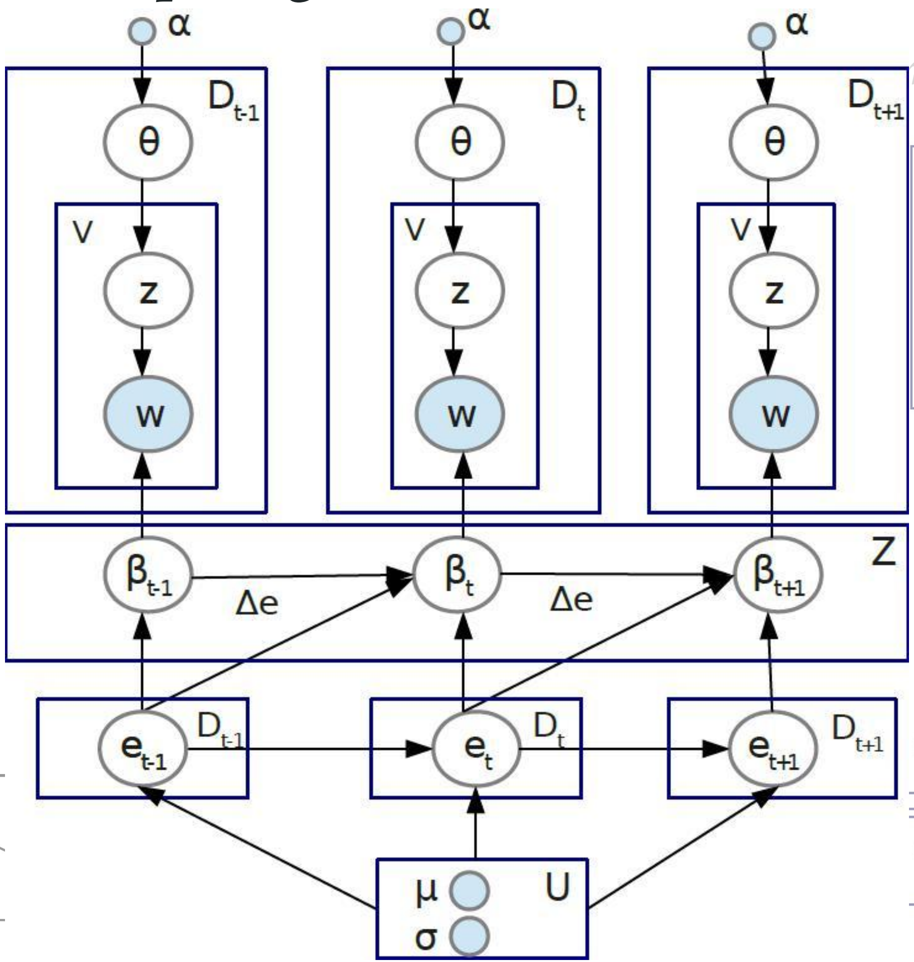
Metropolis Hastings for Exp. evolution



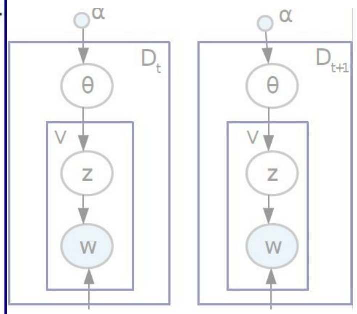
Kalman Filter for LM evolution



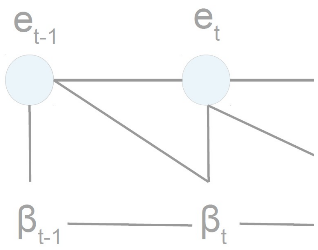
Sampling based Inference



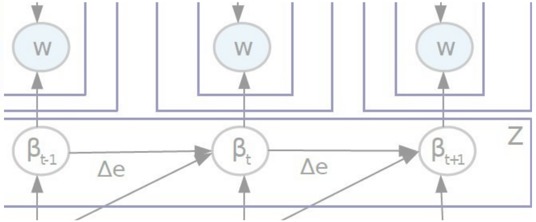
...mpling for Facets



Metropolis Hastings for Exp. evolution



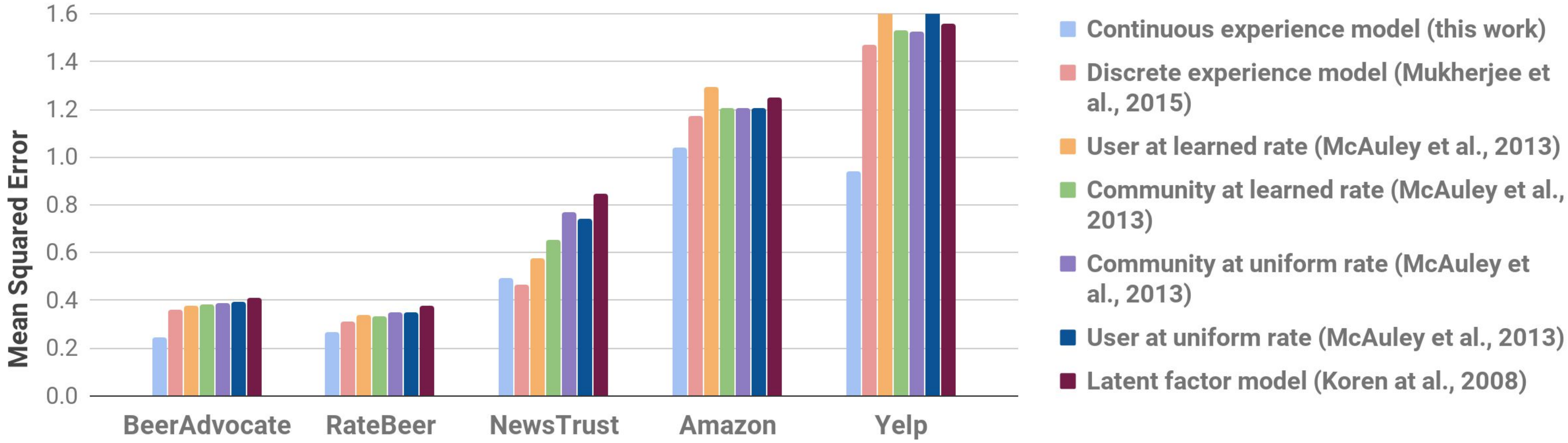
Kalman Filter for LM evolution



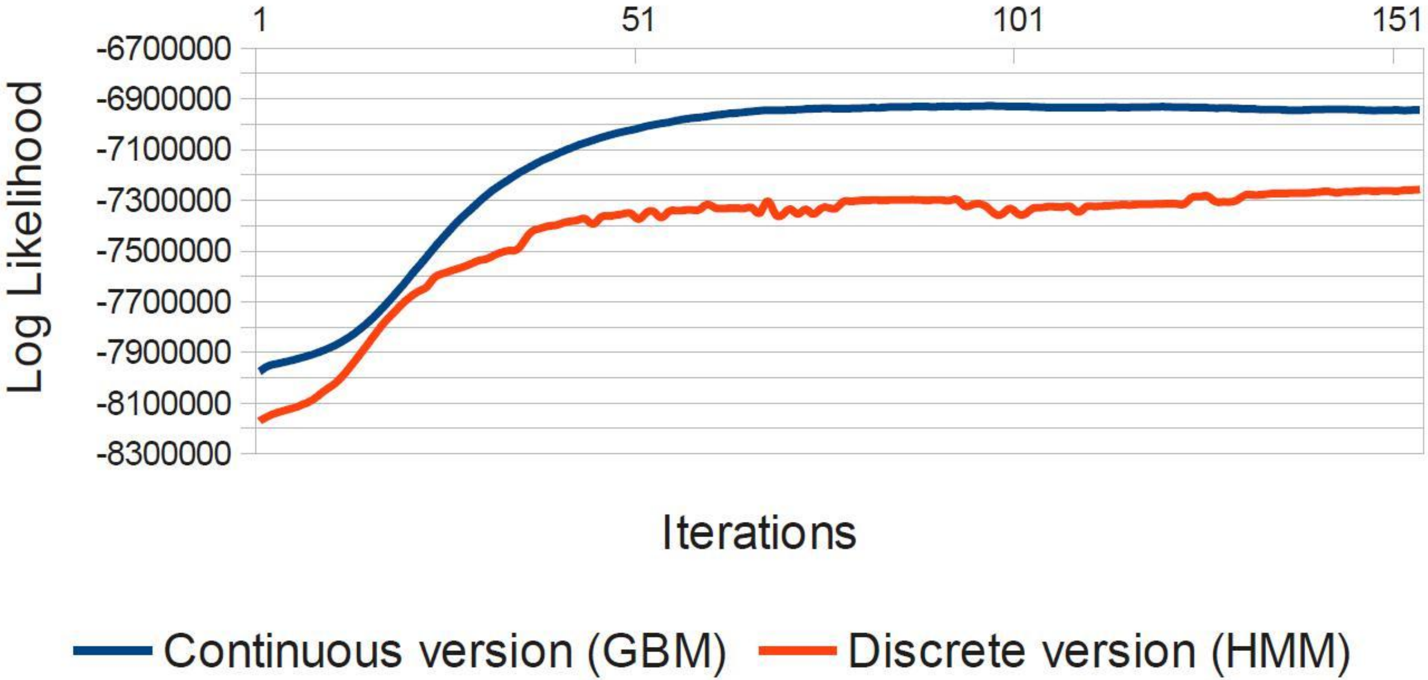
Dataset Statistics

Dataset	#Users	#Items	#Ratings	#Time (Years)
Beer (BeerAdvocate)	33,387	66,051	1,586,259	16
Beer (RateBeer)	40,213	110,419	2,924,127	13
Movies (Amazon)	759,899	267,320	7,911,684	16
Food (Yelp)	45,981	11,537	229,907	11
Media (NewsTrust)	6,180	62,108	89,167	9
TOTAL	885,660	517,435	12,741,144	-

Can we recommend items better, if we consider users' experience to consume them?



Log-likelihood, Smoothness, and Convergence



Interpretability: Top Words* by Experienced Users

	Most Experience	Least Experience
BeerAdvocate	chestnut_hued near_viscous cherry_wood sweet_burning faint_vanilla woody_herbal citrus_hops mouthfeel	originally flavor color poured pleasant bad bitter sweet
Amazon	aficionados minimalist underwritten theatrically unbridled seamless retrospect overdramatic	viewer entertainment battle actress tells emotional supporting
Yelp	smoked marinated savory signature contemporary selections delicate texture	mexican chicken salad love better eat atmosphere sandwich
NewsTrust	health actions cuts medicare oil climate spending unemployment	bad god religion iraq responsibility questions clear powerful

*Learned by our generative model *without* supervision

Interpretability: Top Words* by Experienced Users

	Most Experience	
BeerAdvocate	chestnut_hued near_viscous ... sweet_burning faint_vanilla woody_herbal citrus_hops mouthfeel	...
Amazon	aficionados minimalist underwritten theatrically unbridled seamless retrospect overdramatic	viewer entertainment battle actress tells emotional supporting
Yelp	smoked marinated savory signature contemporary selections delicate textu...	...
NewsTrust	health actions cuts medicare oil ... spending unemployment	...

Experienced users in the beer community use more "**fruity**" words to describe taste and smell of beers

Experienced users in the news community discuss about **policies and regulations** in contrast to amateurs interested on **polarizing** topics

*Learned by our generative model *without* supervision

Take-away

- Insights from Geometric Brownian Motion trajectory of users:
    - Experienced users mature faster than amateurs
    - Progression depends more on time spent in community than on activity
- Users' experience evolve continuously, along with language usage
- Recommendation models can be improved by considering users' maturity
- Learns from *only* the information of *users reviewing products at explicit timepoints* --- no meta-data, community-specific / platform dependent features --- easy to *generalize* across different communities

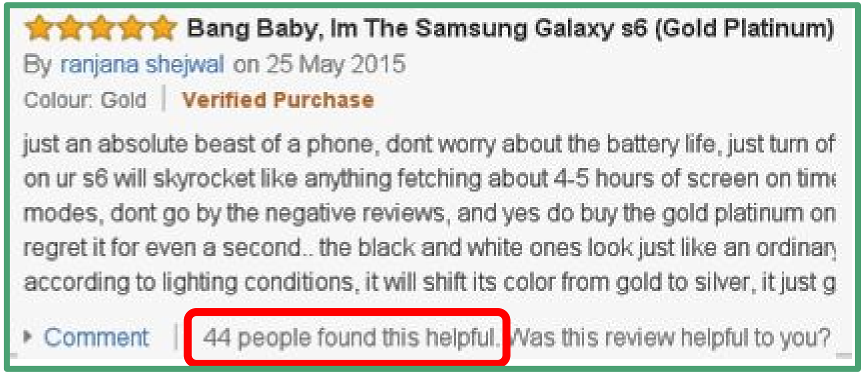
Outline

- Motivation
- Prior Work and its Limitations
- Credibility Analysis
  - ↘ Framework for Online Communities
  - ↘ Temporal Evolution of Online Communities
  - ↘ Credibility Analysis of Product Reviews
- Conclusions

Can we use this framework to find helpful product reviews?

Distributional Hypotheses

- Users with similar facet preferences and expertise are likely to be equally helpful.



- Reviews (e.g., camera) with similar facet-sentiment distribution (e.g., bashing “zoom” and “resolution”) are likely to be equally helpful.

Subhabrata Mukherjee, Kashyap Popat, Gerhard Weikum: SDM 2017

Consistency Analysis of Product Reviews

We analyze consistency of embeddings from previous models to detect fake / anomalous reviews with discrepancies like:

1. Rating and review description (*promotion/demotion*)

Excellent product… technical support is almost non-existent … this is unacceptable. [4]

2. Rating and Facet description (*irrelevant*)

DO NOT BUY THIS. I can't file because Turbo Tax doesn't have software updates from the IRS "because of Hurricane Katrina". [1]

3. Temporal bursts (*group spamming*)

Dan's apartment was beautiful, a great location. (3/14/2012)[5]

I highly recommend working with Dan and… (3/14/2012) [5]

Dan is super friendly, confident… (3/14/2012) [4]

Subhabrata Mukherjee, Sourav Dutta, Gerhard Weikum: ECML-PKDD 2016

Future Work

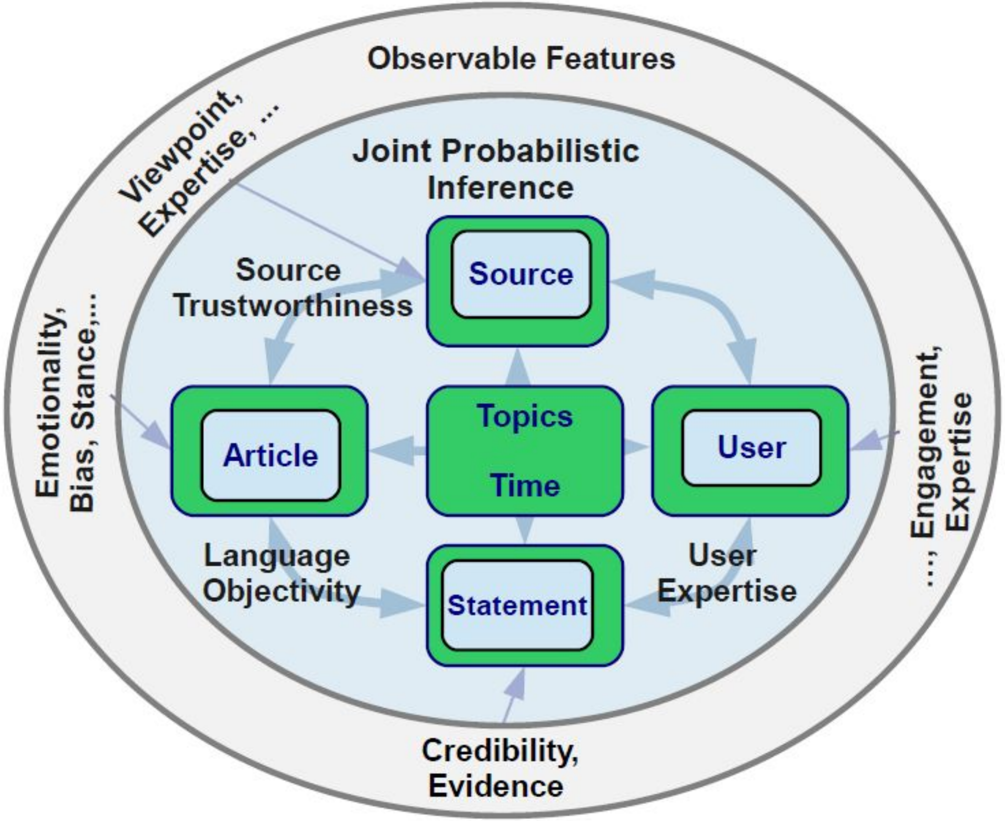
★ Going *beyond* topics and bag-of-words features / lexicons
Learning linguistic cues from embeddings

★ Incorporating *richer facets* like multi-modal interactions, stance, influence evolution etc.

★ *Applications* to tasks like Anomaly Detection, Community Question-Answering, Knowledge-base Curation etc.

Conclusions

1. How can we *jointly* leverage users, network, and context for credibility analysis in online communities?
2. How can we model users' *evolution*?
3. How can we deal with *limited* data?
4. How can we generate *interpretable* explanations for credibility verdict?



Interactional Framework for Credibility Analysis

Acknowledgments (1/3)

Gerhard Weikum

(Advisor)

Stephan Günnemann

Cristian Danescu-Niculescu-Mizil

Jannik Strötgen

Kashyap Popat

Sourav Dutta

Hemank Lamba

Collaborators and Co-authors

Acknowledgments (2/3)

Gerhard Weikum

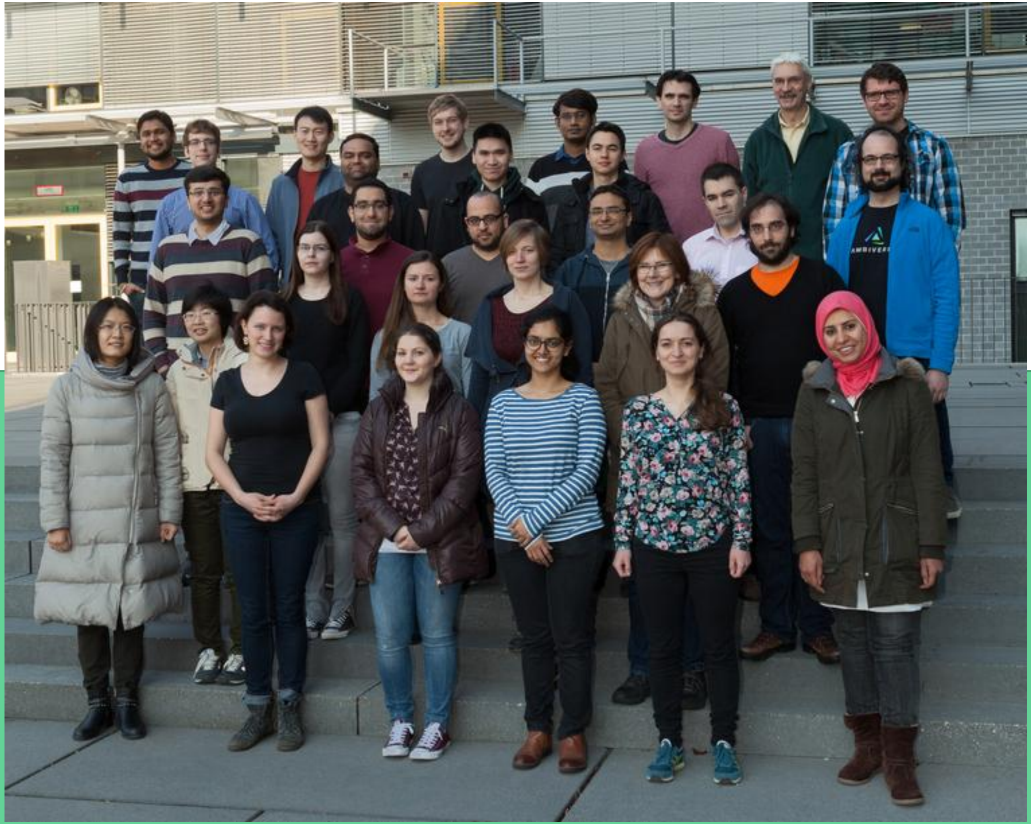
Jiawei Han

Stephan Günnemann

Dietrich Klakow

Dissertation Committee

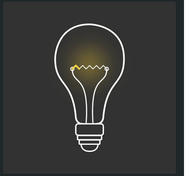
Acknowledgments (3/3)

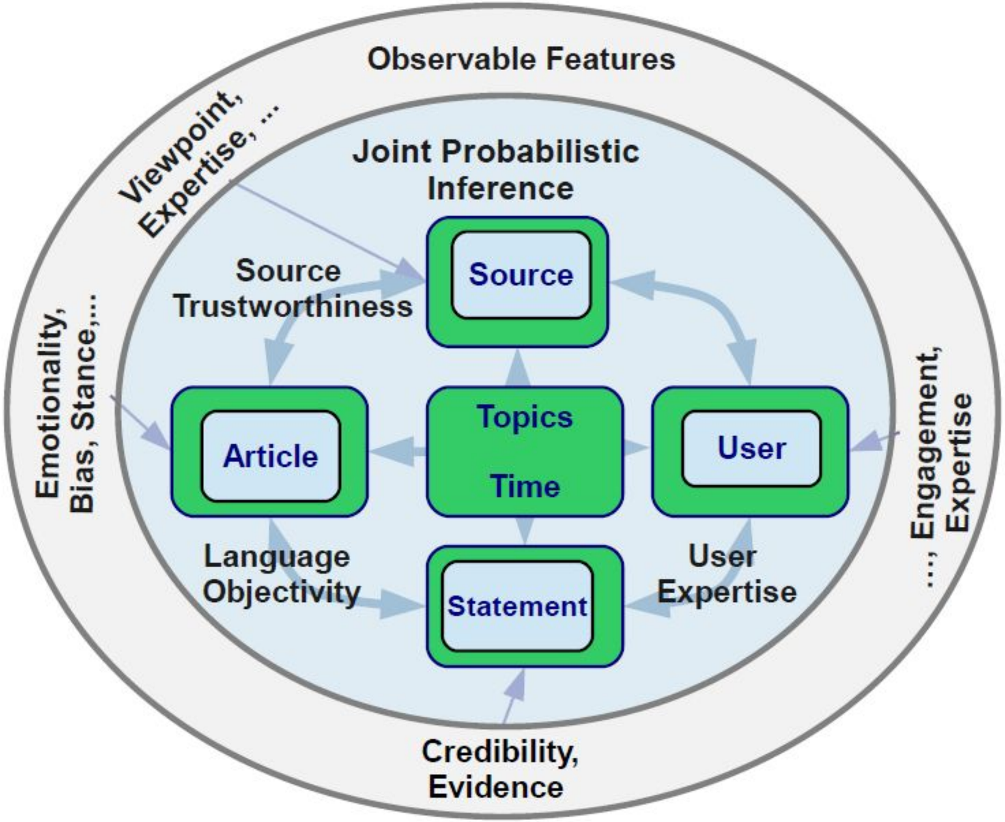


Databases and Information Systems Department at Max Planck Institute

THANKS!

1. How can we *jointly* leverage users, network, and context for credibility analysis in online communities?
2. How can we model users' *evolution*?
3. How can we deal with *limited* data?
4. How can we generate *interpretable* explanations for credibility verdict?





Interactional Framework for Credibility Analysis