Probabilistic Graphical Models for Credibility Analysis in Evolving Online Communities

By Subhabrata Mukherjee

subhabrata.mukherjee.ju@gmail.com

Outline

- Motivation
- Related Work
- Credibility Analysis
  - Health Communities
  - News Communities
  - Temporal Evolution & Review Communities
- Related Applications
- Conclusions

Online Communities as a Knowledge Resource

- Online communities are massive repositories of (untapped) knowledge, largely unstructured in nature
- Wealth of topic-specific communities and discussion forums about health, news, music, consumer products etc.
    - Half of US physicians rely on online resources (e.g., Youtube and Wikipedia) [IMS Health Report, 2014]
    - 40% online consumers would not buy electronics without consulting online reviews first [Nielson Corporation, 2016]

Online Communities as a Knowledge Resource: Credibility and Trust Concerns

- Noisy, unreliable, and subjective user-generated content
  - Rumors, spams, misinformation, bias
  - Yelp internally rejects 25% reviews as fake[1]
  - Only 34% of adult US population somewhat trust on social media information [PEW Research, 2016] and 80% do not trust major news networks [Gallup poll, 2013]

[1] https://www.yelpblog.com/2013/09/fake-reviews-on-yelp-dont-worry-weve-got-your-back

Traditional Knowledge Resources

Several large-scale Knowledge Bases (KBs) exist like YAGO, NELL, DBpedia, Freebase etc.

- These store millions of facts about people, places, and things (or, entities) (e.g., Obama_BornIn_Hawaii)
- High precision, low coverage --- store information mostly about prominent entities
- Require manual curation, or operate over structured data (e.g., Wiki infoboxes)
- Only recently efforts are put to combine open Information Extraction and structured KB (e.g., KnowledgeVault)

Prior Works on Credibility

Truth / Fact Finding

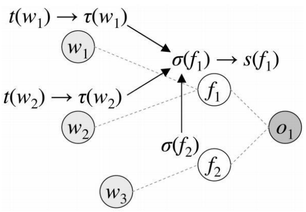
- Structured data (e.g., SPO triples, tables, networks)
- Objective facts (e.g., Obama_BornIn_Hawaii vs. Obama_BornIn_Kenya)
- No contextual data (text)
- No external KB (Knowledge Base) or metadata

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, or metadata

RQ: How can we complement expert KBs (traditional resources) with large scale non-expert data (online communities)?

RQ: How can we develop models that jointly leverage users, network, and context for knowledge fusion?

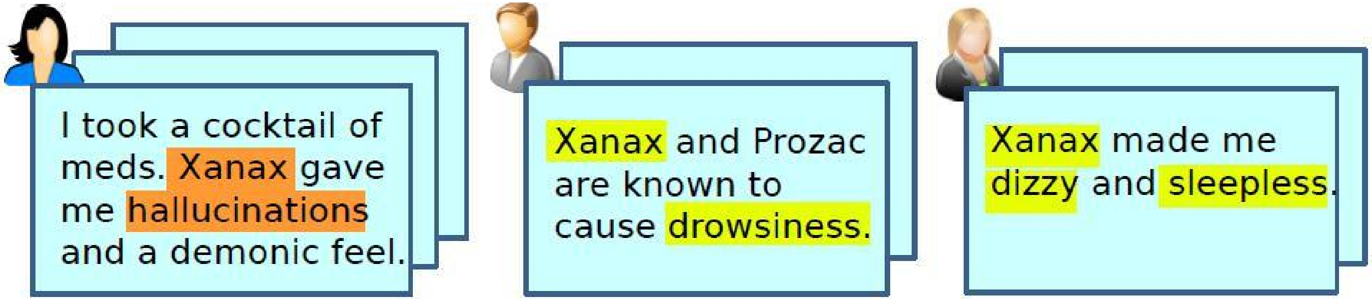
We study this with respect to some diverse online communities:

1. **Healthforums (e.g., Healthboards, Patients.co.uk)**
2. **News communities (e.g., Digg, Reddit, Newstrust)**
3. **Product review communities (e.g., Amazon, Yelp, Beeradvocate)**

Outline

- Motivation
- Related Work
- Credibility Analysis
    - Health Communities
    - News Communities
    - Temporal Evolution & Review Communities
- Related Applications
- Conclusions

Case-Study I: Identifying Credible Side-effects of Drugs from User-generated Posts in Healthforums

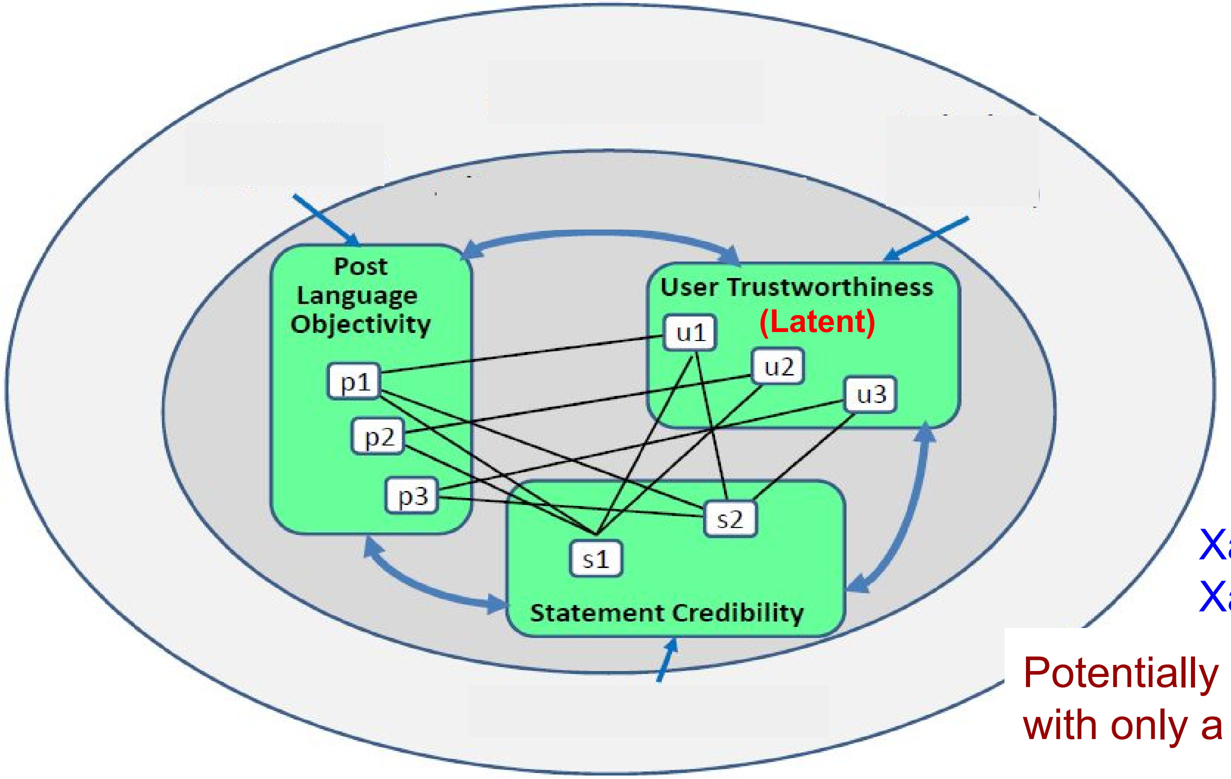


Problem: Given a set of posts from different users, extract credible SPO triples (DrugX_HasSideEffect_Y) from trustworthy users

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

Network of Interactions: Cliques

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



Statements: Assume access to an IE tool to generate candidate triple patterns like:

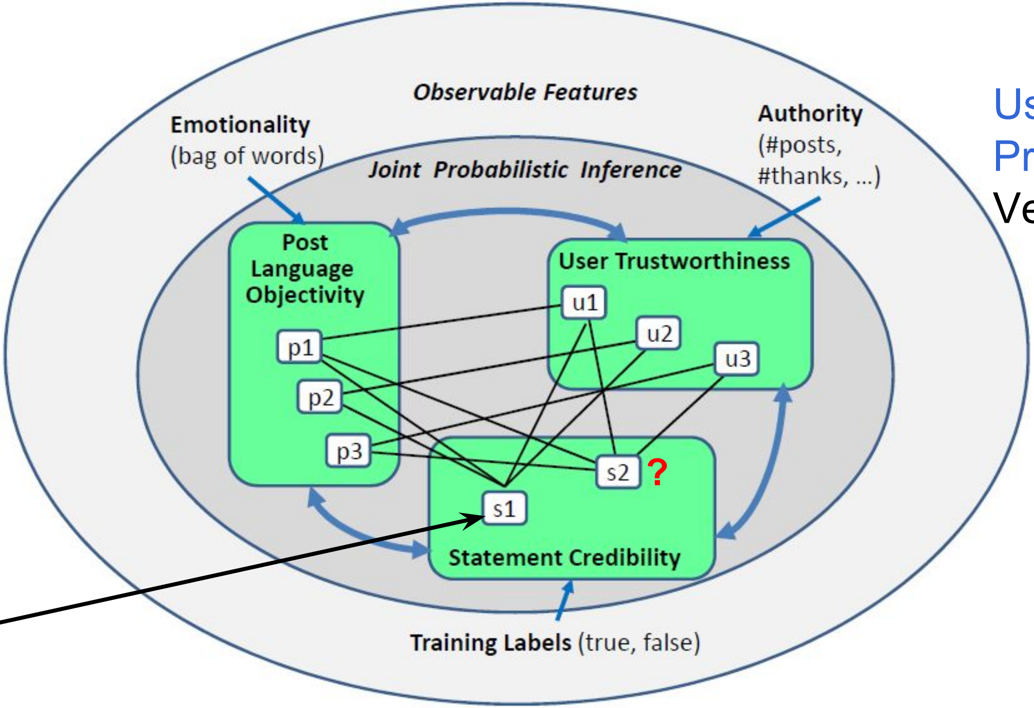
Xanax_causes_headache,
Xanax_gave_demonic-feel

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)

Linguistic Features from Context: Discourse, Modalities, Affective Emotion, Subjectivity, Negation etc.



User Features from Profile: Demographics, Verbosity, Activity etc.

Partial Supervision: Expert stated (top 20%) side-effects of drugs from MayoClinic used as partial training labels. Model predicts the most likely label assignment of remaining unobserved ones.

Semi-Supervised Conditional Random Field

1. Estimate user trustworthiness : $t_k = \frac{\sum_i \mathbb{1}_{S_{i,k}=\text{True}}}{|S_k|}$

2. E-Step : Estimate label of unknown statements by Gibbs' sampling :

$$Pr(S_i^U | P, U, S^L; W) \propto \prod_{\nu \in C} t_k \times \phi_\nu(S_\nu^*, p_j, u_k; W)$$

3. M-Step : Maximize log-likelihood to estimate feature weights using Trust Region Newton :

$$W^{(\nu+1)} = argmax_{W'} \sum_{S^U} q(S^U) \log Pr(S^L, S^U | P, U; W')$$

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 messages

- We sampled 15, 000 users with 2.8 million messages

Expert labels about drugs from MayoClinic portal

- 2172 drugs categorized in 837 drug families
- 6 widely used drugs used for experimentation

Healthforum Dataset

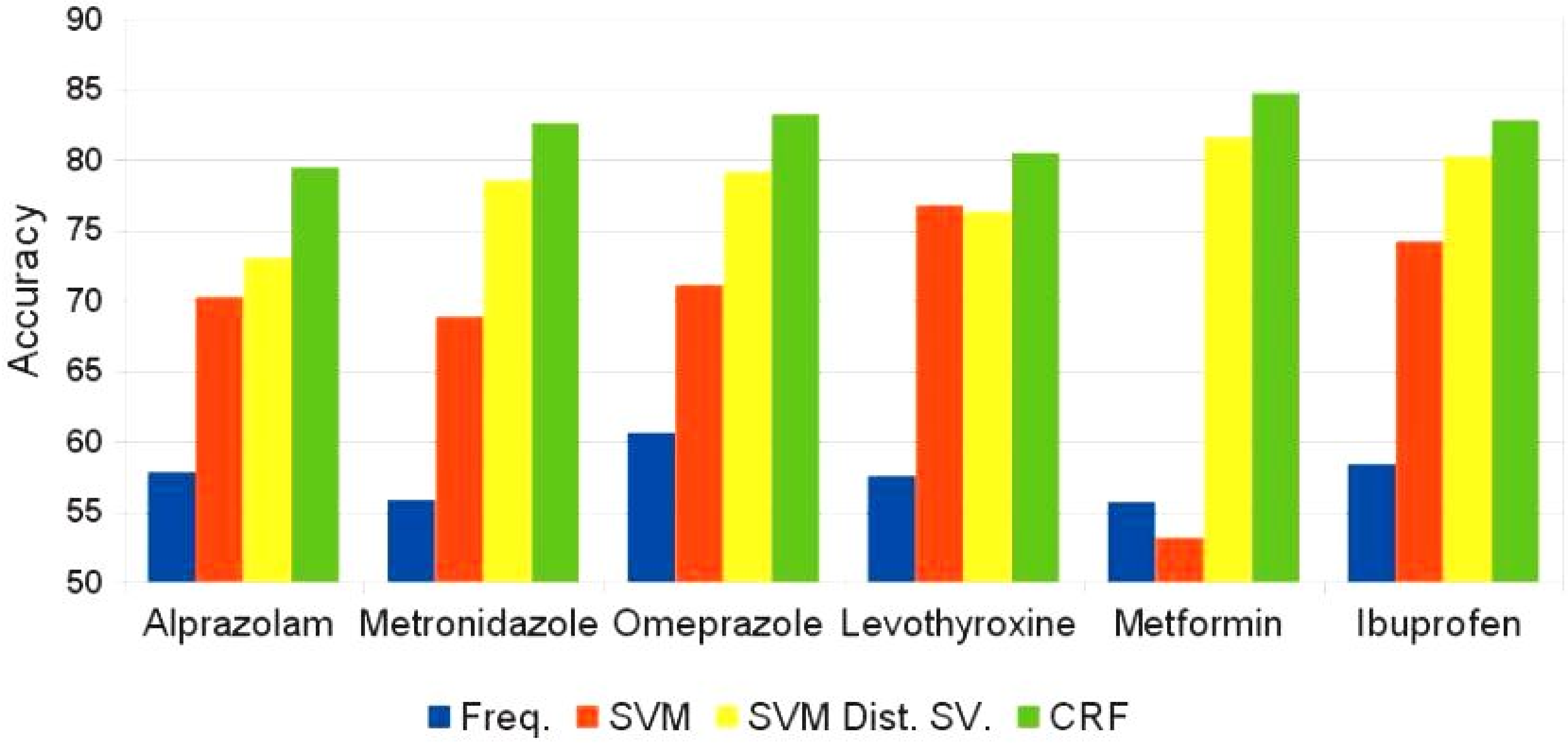
Healthboards.com community (`www.healthboards.com`) with 850, 000 registered users and 4.5 million messages

- We sampled 15, 000 users with 2.8 million messages

Expert labels about drugs from MayoClinic portal

- 2172 drugs categorized in 837 drug families
- 6 widely used drugs used for experimentation

Experimental Results



Take-away / Contributions

- Semi-supervised CRF to jointly identify trustworthy users, credible statements, and reliable postings from partial expert information
- A framework to incorporate richer aspects like user expertise, topics / facets, temporal evolution etc.

Outline

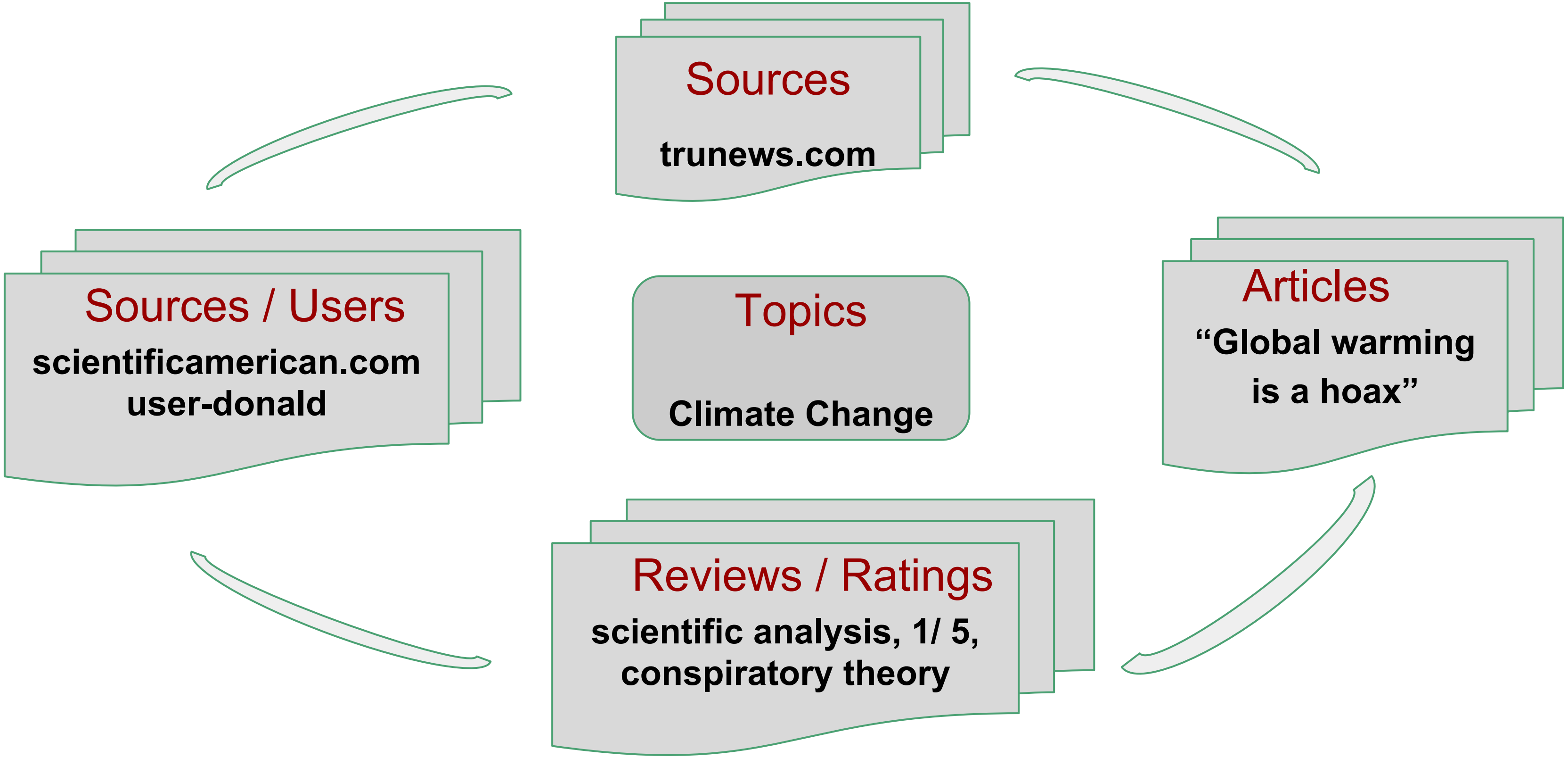
- Motivation
- Related Work
- Credibility Analysis
  - Health Communities
  - News Communities
  - Temporal Evolution & Review Communities
- Related Applications
- Conclusions

Case-Study II: Credibility Analysis in News Communities

- A news community is a (heterogeneous) news aggregator site (e.g., reddit.com, digg.com, newstrust.net)
  - Users can give explicit feedback (e.g., rate, review, share) on the quality of news
  - Interact (e.g., comment, vote) with each other
  - These interactions/feedback are biased by users' viewpoints on polarized topics

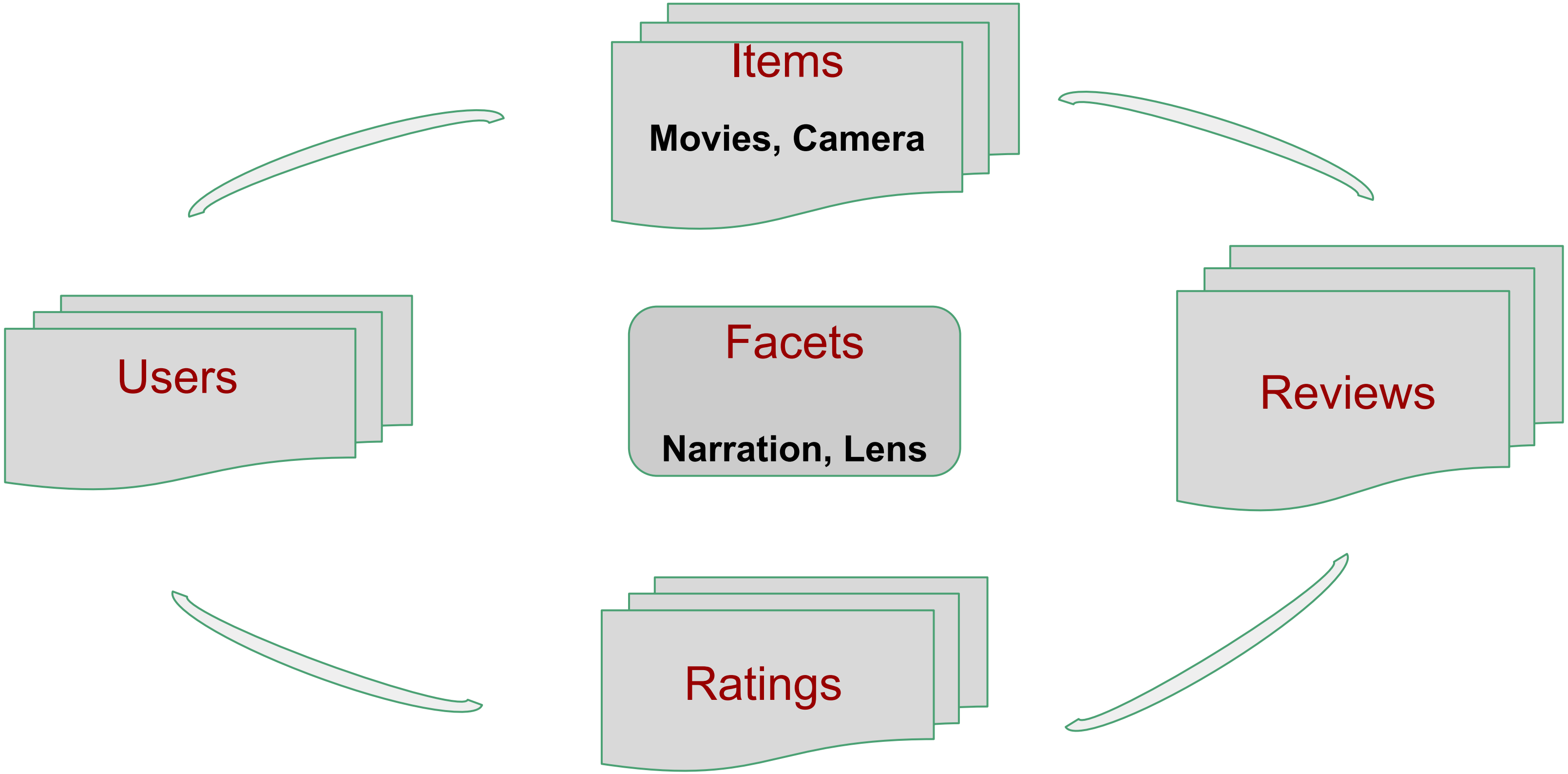
Subhabrata Mukherjee and Gerhard Weikum: CIKM 2015

News Communities: Interactions



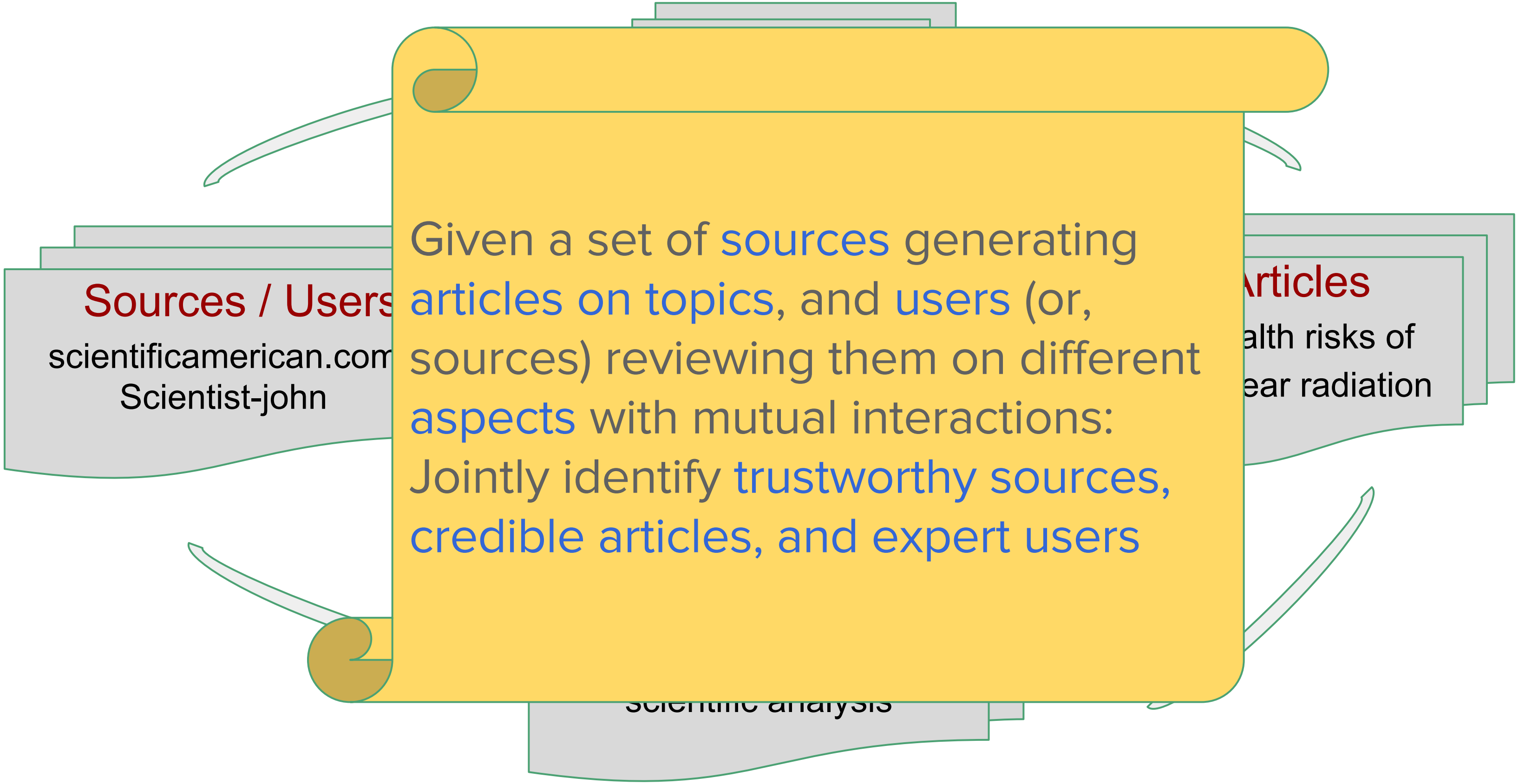
Idea: Trustworthy sources publish objective articles (on specific topics) corroborated by expert users with credible reviews/ratings, and the converse

Review Communities: Interactions



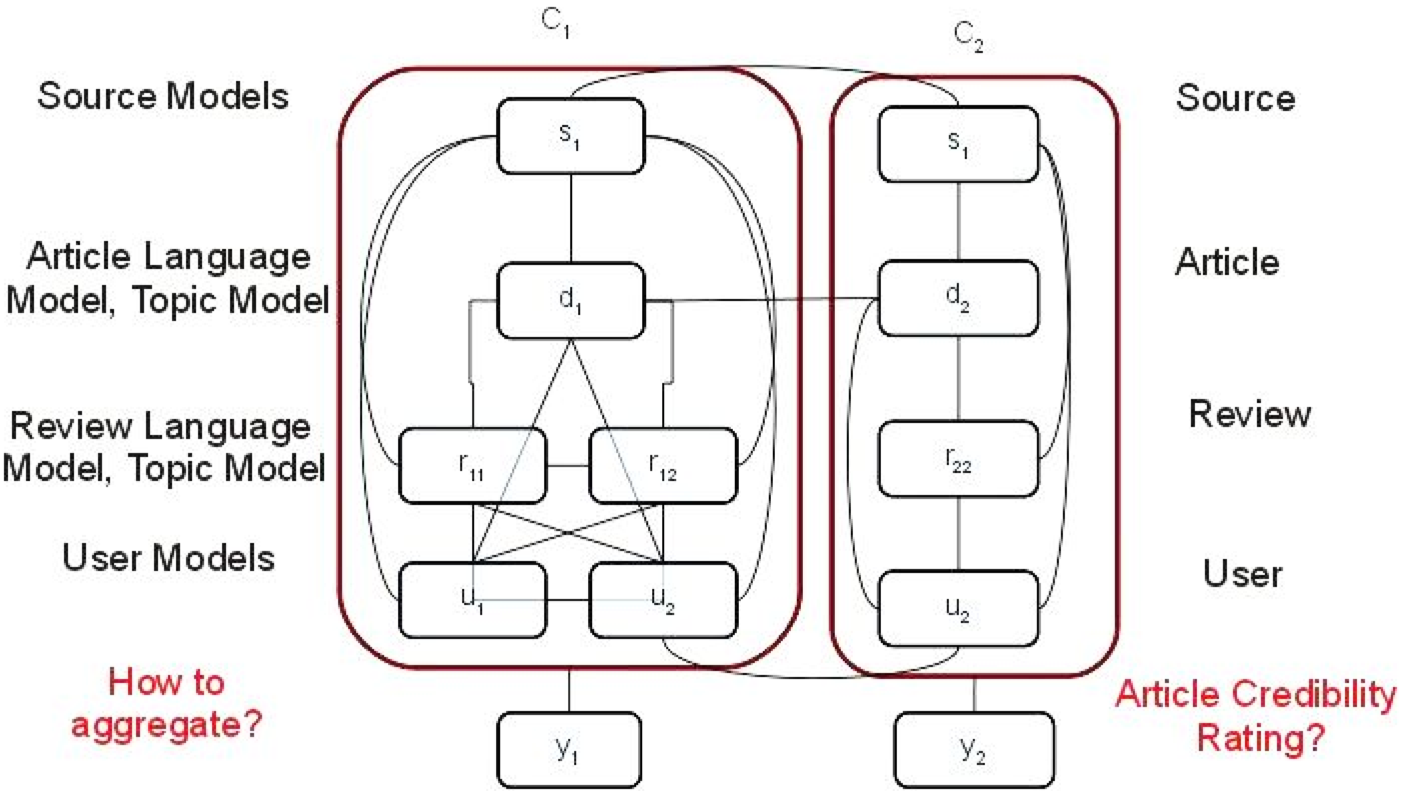
Idea: Expert users contribute credible reviews/ratings that highlight essential facets of items

Problem Statement



Given a set of sources generating articles on topics, and users (or, sources) reviewing them on different aspects with mutual interactions: Jointly identify trustworthy sources, credible articles, and expert users

Online Communities: Factors



Related to Ensemble Learning, Learning to Rank

We use CRF to capture these joint interactions

RQ: 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:

$$p(y|X) = \frac{exp(\Psi)}{\int_{-\infty}^{\infty} exp(\Psi)dy}$$

Continuous Conditional Random Field

- We show that a judiciously selected energy function for clique interactions 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 optimization problem with constraints on Σ

  Constrained Stochastic Gradient Descent for inference

Predicting Article Credibility Ratings

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 / Contributions

- Continuous CRF to jointly learn user & source expertise, article & review/rating credibility
- A generalized (extensible) framework for Credibility Analysis incorporating richer aspects like user & source expertise, and viewpoint on (latent) topics

Outline

- Motivation
- Related Work
- Credibility Analysis
    - Health Communities
    - News Communities
    - Temporal Evolution & Review Communities
- Related Applications
- Conclusions

Temporal Evolution

Online communities are dynamic, as users evolve and mature over time

Therefore, expertise and trustworthiness are not static concepts

RQ: How to capture evolving user expertise?

We study this w.r.t item recommendation task

Camera Reviews

by 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]

RQ1: How can we quantify this change in user maturity or experience ?

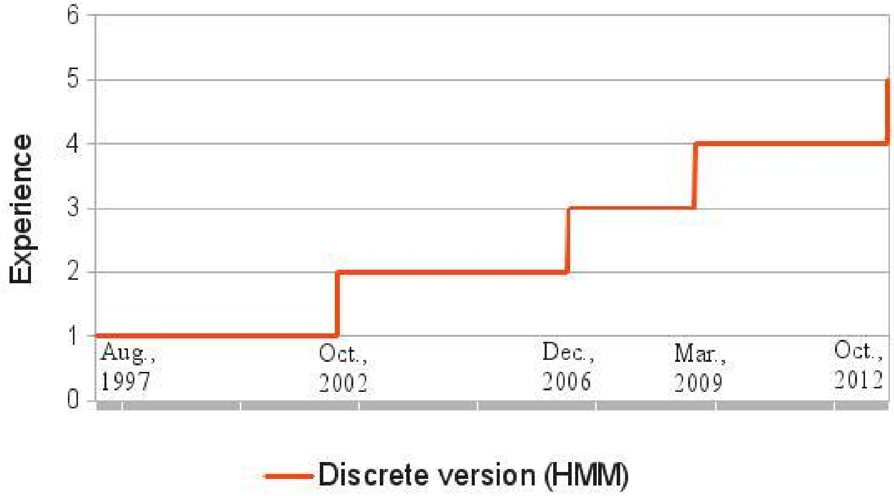
RQ2: How can we model this evolution or progression in maturity?

Subhabrata Mukherjee, Gerhard Weikum et al.: ICDM 2015, KDD 2016

Discrete Experience Evolution

Users at similar levels of experience have similar writing style, facet preferences, and rating behavior

[S. Mukherjee, H. Lamba, G. Weikum, ICDM '15]



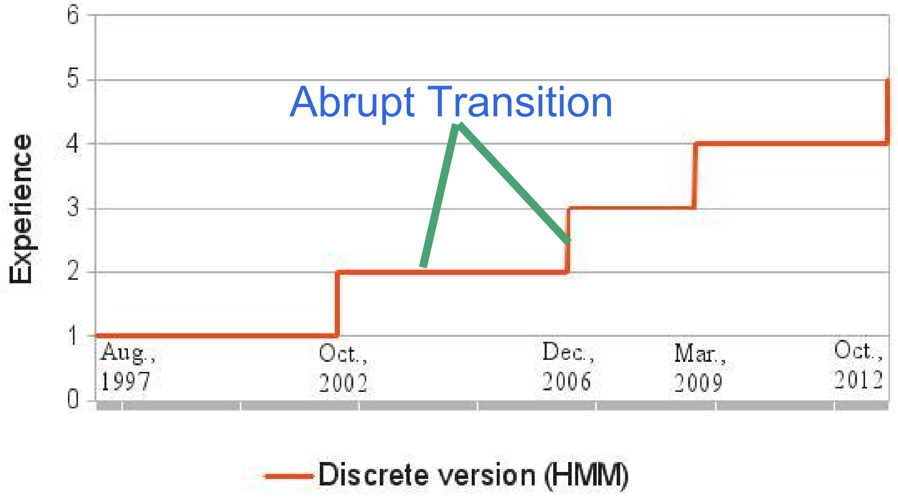
Assumption: At each timepoint (of writing a review) a user remains at the same level of experience, or moves to the next level

Discrete

Users at similar levels of experience have similar writing style, facet preferences, and rating behavior

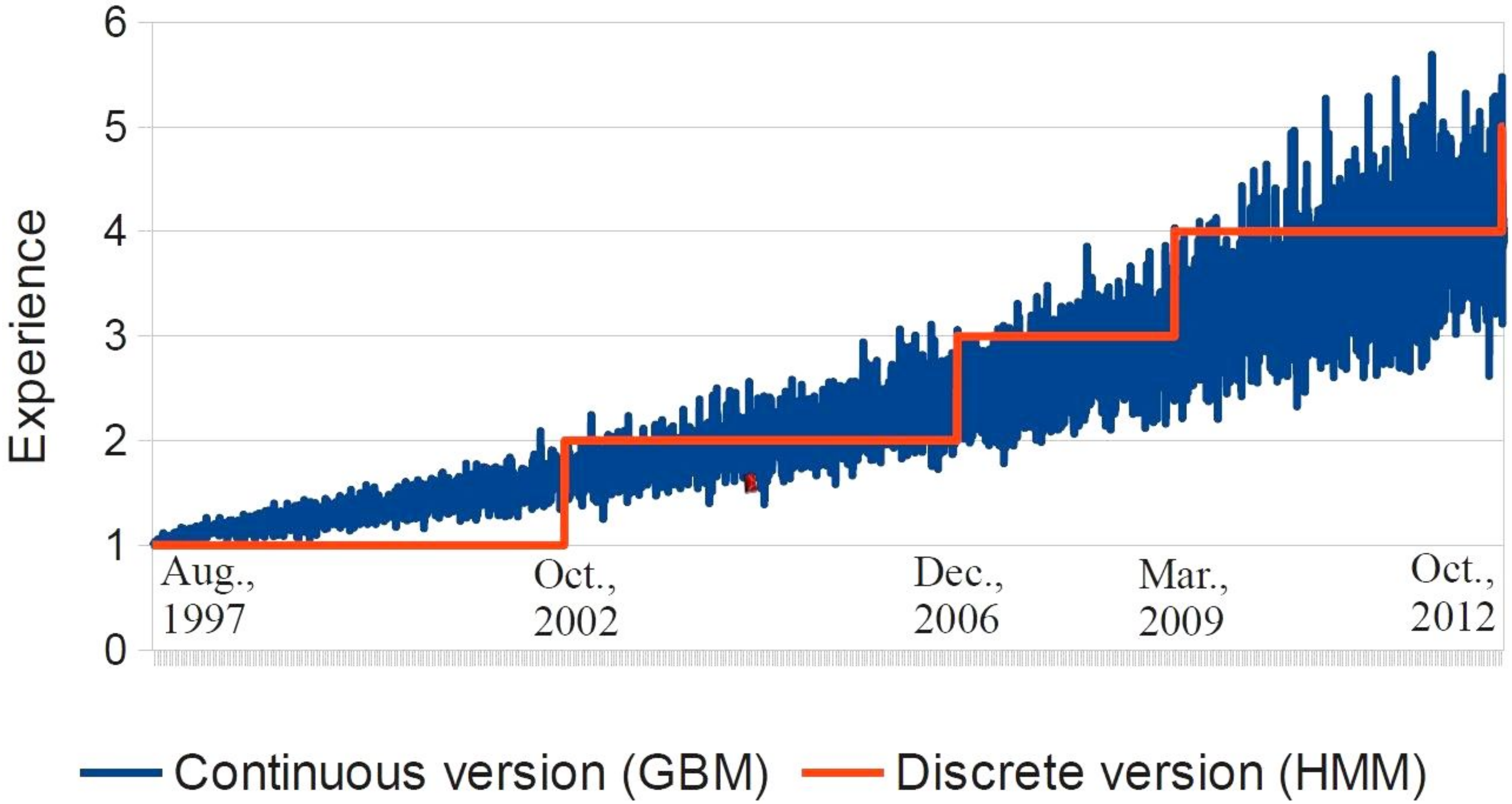
[S. Mukherjee, H. Lamba, G. Weikum, ICDM '15]

Experience Evolution



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

Continuous Experience Evolution (KDD 2016)



Continuous Experience Evolution: Assumptions

- Continuous-time process, always positive
- Markovian assumption: Experience at current time depends only on the latest observed experience
- 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

We show these properties to be satisfied by the continuous-time stochastic process:

Geometric Brownian Motion

Geometric Brownian Motion

Stochastic Differential Equation: $dE_t = \underbrace{\mu E_t dt}_{\text{deterministic trend}} + \underbrace{\sigma E_t dW_t}_{\text{unpredictable volatility}}$

$E_t \rightarrow$ Experience at time 't'

Analytic solution: $E_t = E_0 \exp((\mu - \sigma^2 / 2) t) + \sigma W_t)$

$t \rightarrow$ time

Starting experience

trend

volatility

Wiener Process / Standard Brownian Motion

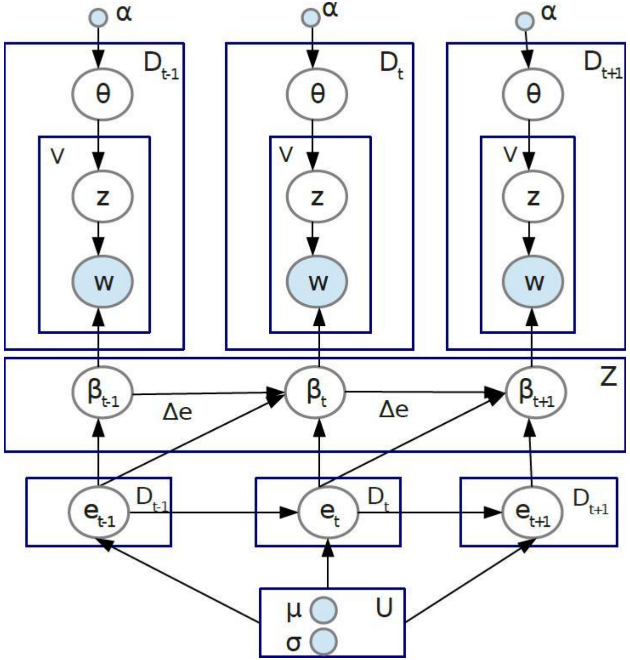
We show these properties to be satisfied by the continuous-time stochastic process:

Geometric Brownian Motion

Language Model (LM) Evolution

- Users' LM also evolves with experience evolution
- Smoothly evolve over time preserving Markov property of experience evolution
- Variance 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)

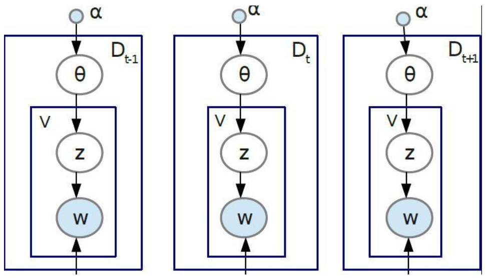
+ Users (Author-topic model,
Rosen-Zvi et al., UAI '04)

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

+ Continuous Experience (this work)

Sampling based Inference for High Dimensional Data

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

Facets
(Latent)

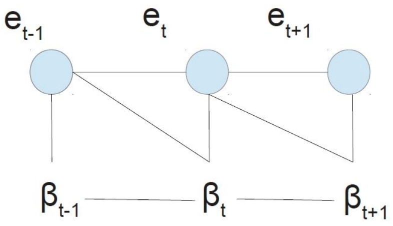
Metropolis Hastings for Exp. evolution

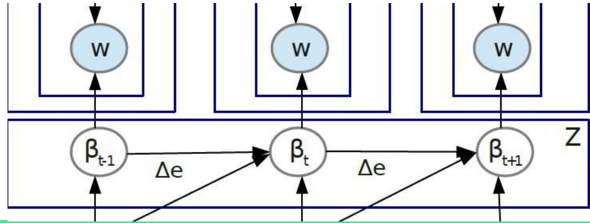
Experience
(Latent)

Language Model

Words (Observed) at (Observed) Timepoints

Kalman Filter for LM evolution

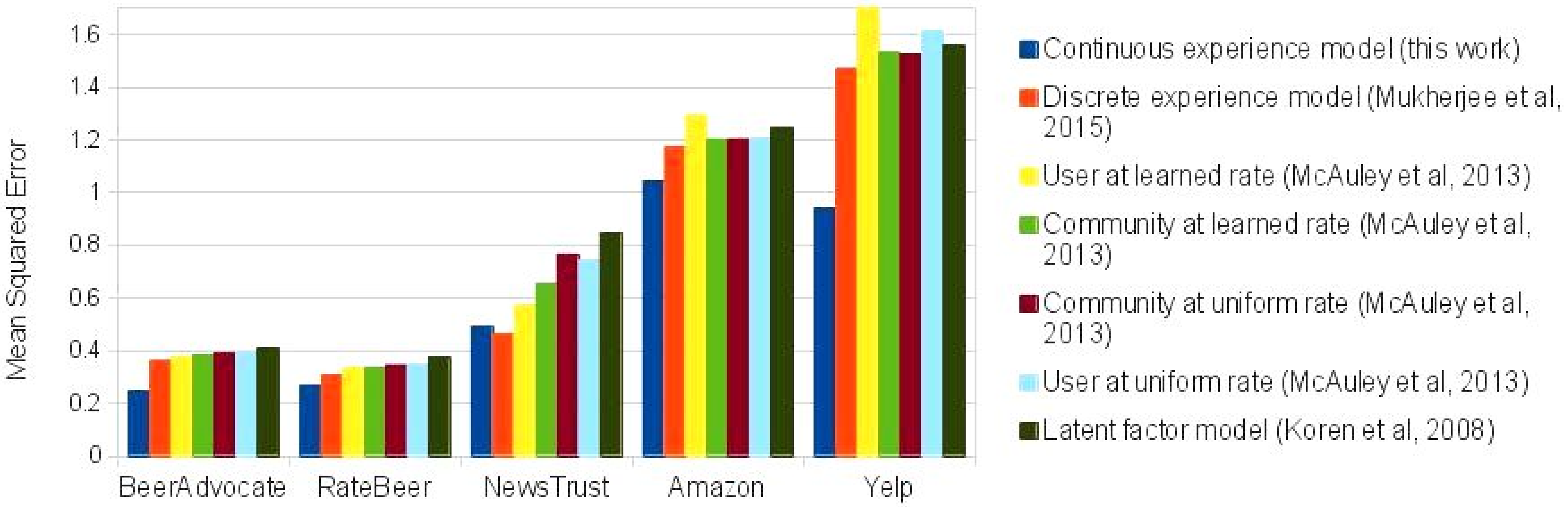




Datasets

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	-

RQ: Can we recommend items better, if we consider user experience ?



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

Insights from GBM Trajectory of Users

- Experienced users mature faster than amateurs, exhibit a higher variance
- Progression depends more on time spent in community than on activity

Take-away / Contributions

- Users' experience evolve continuously in nature, along with their language usage
- Recommendation models can be improved by explicitly considering user experience
- Finally, we propose a Brownian Motion based stochastic model to capture the above phenomena

Outline

- Motivation
- Related Work
- Credibility Analysis
  - Health Communities
  - News Communities
  - Temporal Evolution & Review Communities
- Related Applications
- Conclusions

RQ: Can we use this framework for finding helpful product reviews?

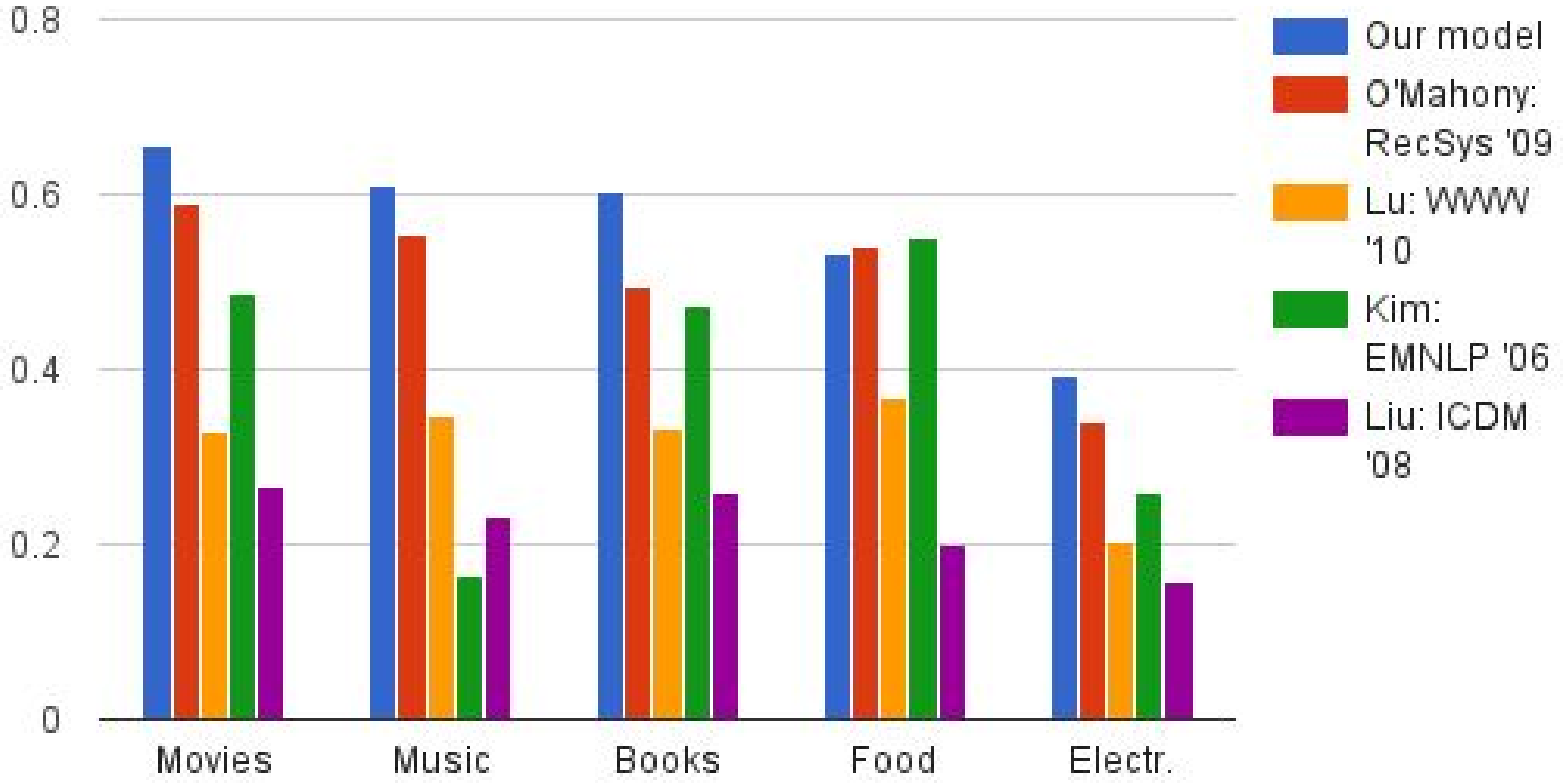
Distributional Hypotheses

- Reviews (e.g., camera reviews) with similar facet distribution (e.g., focusing on “zoom” and “resolution”) for items are likely to be equally helpful.
- Users with similar facet preferences and expertise are likely to be equally helpful.

Experiments: Datasets from Amazon

Factors	Books	Music	Movie	Electronics	Food
#Users	2,588,991	1,134,684	889,176	811,034	256,059
#Items	929,264	556,814	253,059	82,067	74,258
#Reviews	12,886,488	6,396,350	7,911,684	1,241,778	568,454

Ranking Task: Spearman Rho of our model vs. baselines.



Snapshot of Inconsistencies

We can use a similar idea to detect fake / anomalous reviews using consistency analysis of latent semantic factors

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

Excellent product-alarm zone, technical support is almost non-existent because of this i will look to another product. 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*)

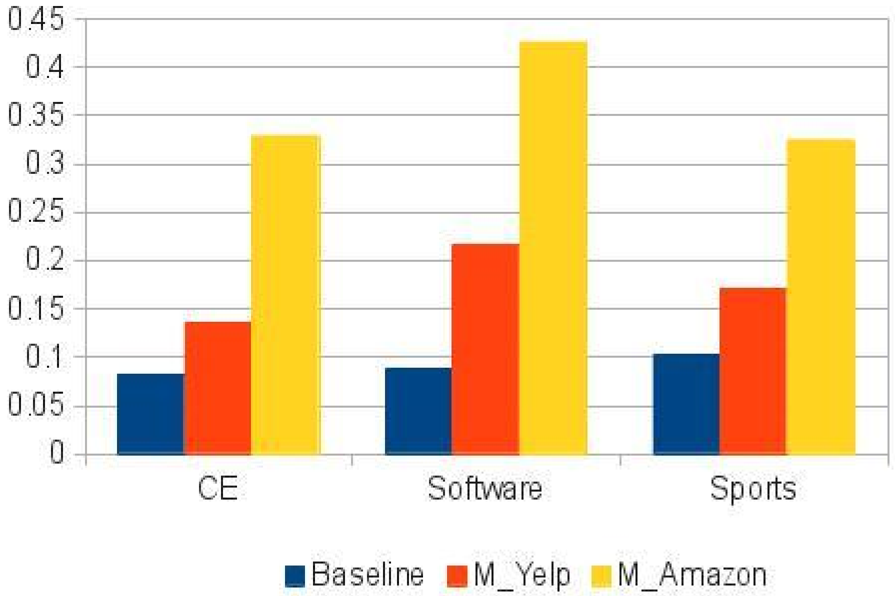
Dan's apartment was beautiful, a great location. (3/14/2012)[5]
I highly recommend working with Dan and... (3/14/2012) [5]
Dan is super friendly, confident... (3/14/2012) [4]
my condo listing with no activity, Dan stepped in (4/18/2012) [5]

Subhabrata Mukherjee, Sourav Dutta, Gerhard Weikum: ECML 2016

Transfer Learning: Yelp to Amazon

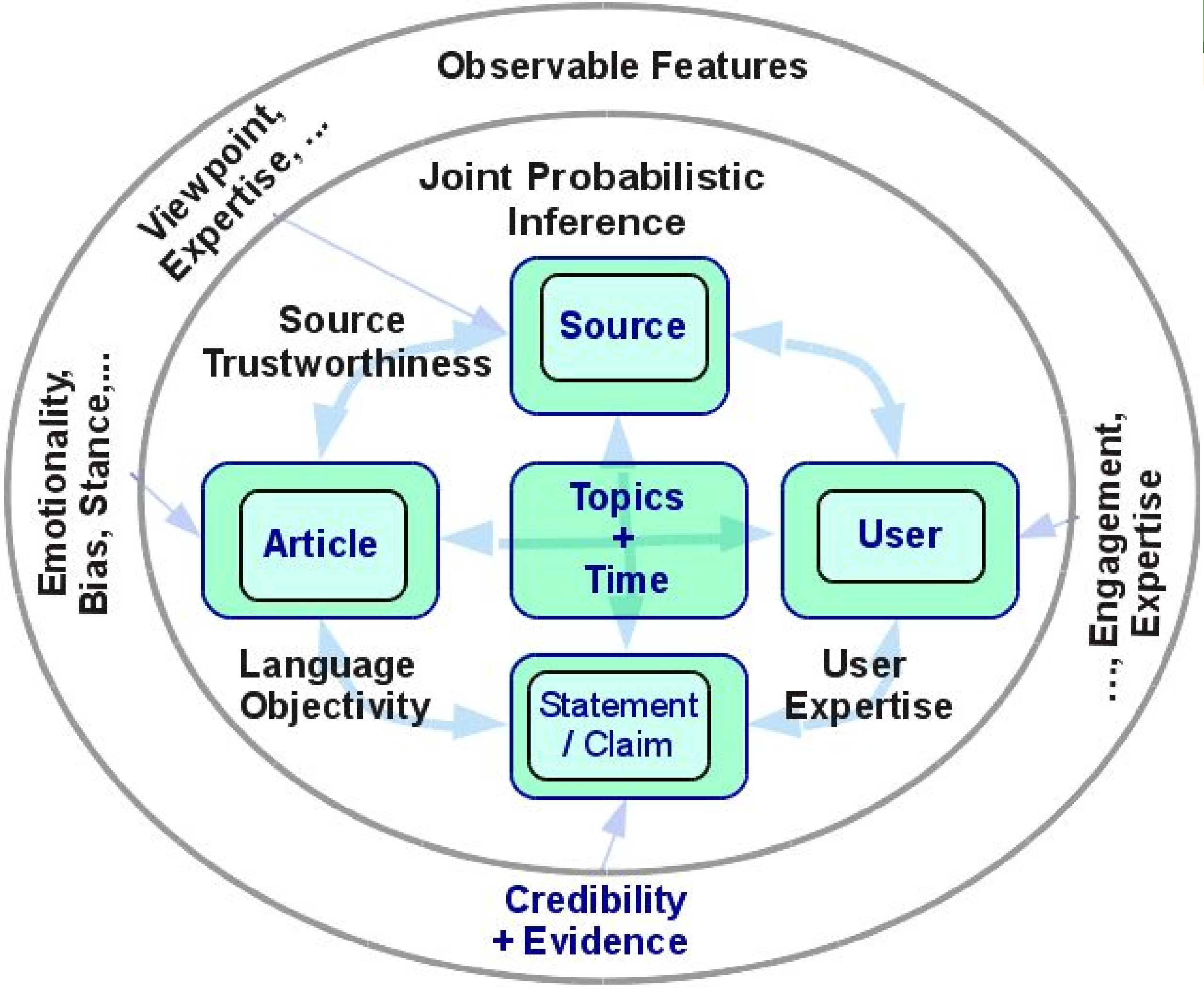
M_Yelp: Trained on Yelp and tested on Amazon with parameter tuning
M_Amazon: Trained and tested on Amazon using Ranking SVM
Training: Reference ranking based on #sales volume of items in Amazon

Kendall-Tau Rank Correlation



Credibility Analysis Applications

- Knowledge-base curation
- Crowd-sourcing applications / aggregation / community question & answering
- Truth-finding
- Expert-finding
- Opinion & Sentiment Mining, Recommendation
- Anomaly, Fraud, Rumor Detection



Take-away:

Credibility Analysis as a Complex Interactional Process