Exploring Latent Semantic Factors to Find Useful Product Reviews

———

Subhabrata Mukherjee, Kashyap Popat, Gerhard Weikum
smukherjee@mpi-inf.mpg.de
Max Planck Institute for Informatics

SIAM International Conference on Data Mining (SDM), 2017

Objective: Find useful / helpful product reviews in online communities

The headset are great. This is one of the best bundled headset with mobile p

There is no expandable memory.

Awesome for s6 to have theme customization. This is good as now u dont ha

Finger print scanner is good and almost spot on.

Battery life is the only section where phone falls behind. Samsung has quick mode and get through the day.

The phone build quality is good. Phone is eye catcher especially the gold unit phone.

S6 seems to have done everything right which was messed up in S4 and S5.

▸ Comment | 8 people found this helpful. Was this review helpful to you?

★★★★★ Bang Baby, Im The Samsung Galaxy s6 (Gold Platinum)
By ranjana shejwal on 25 May 2015
Colour: Gold | **Verified Purchase**

just an absolute beast of a phone, dont worry about the battery life, just turn of on ur s6 will skyrocket like anything fetching about 4-5 hours of screen on time modes, dont go by the negative reviews, and yes do buy the gold platinum on regret it for even a second.. the black and white ones look just like an ordinary according to lighting conditions, it will shift its color from gold to silver, it just g

▸ Comment | 44 people found this helpful. Was this review helpful to you?

2

Tasks

1. Prediction

   Predict the helpfulness score of a review as (x/y): 'x' number of users found the review helpful out of 'y' number of users

2. Ranking

   Rank the reviews for any item based on the helpfulness scores

3

Review 1

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

Review 2

"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."

Review 1

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

Review 2

"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

Review 2 is more *helpful & informative* than Review 1

Review 1

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

Review 2

"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

Review 2 talks about important facets of the camera

Review 1

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

Review 2

"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

Review 2 seems to have been written by an expert user

Both camera reviews by *same user*

"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]

Temporal Evolution: User 2 evolved into an expert and more helpful user now

Both camera reviews by *same user*

"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]

How do we find whether a review is written by an expert user?
How do we model the progression in expertise of a user?

Discrete

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

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

Expertise Evolution

Discrete version (HMM)
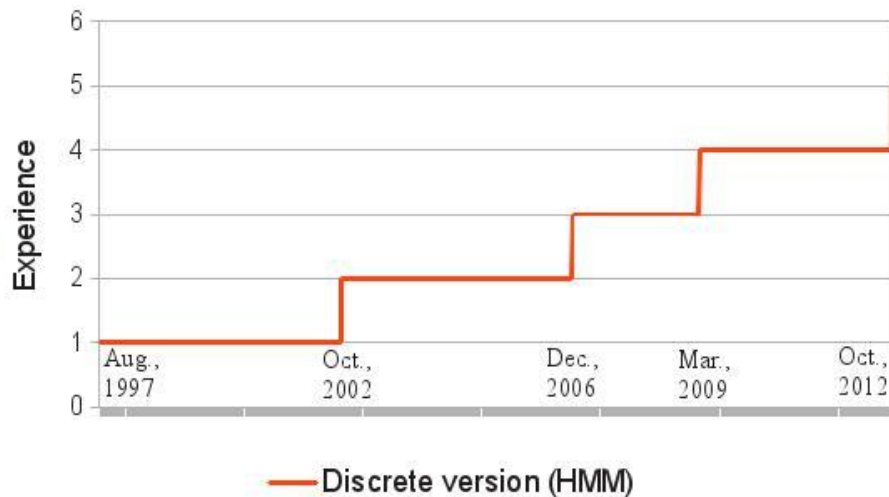
Assumption: At each timepoint (of writing a review) a user remains at the same level of (latent) expertise, or moves to the next level

Continuous Expertise Evolution with Geometric Brownian Motion
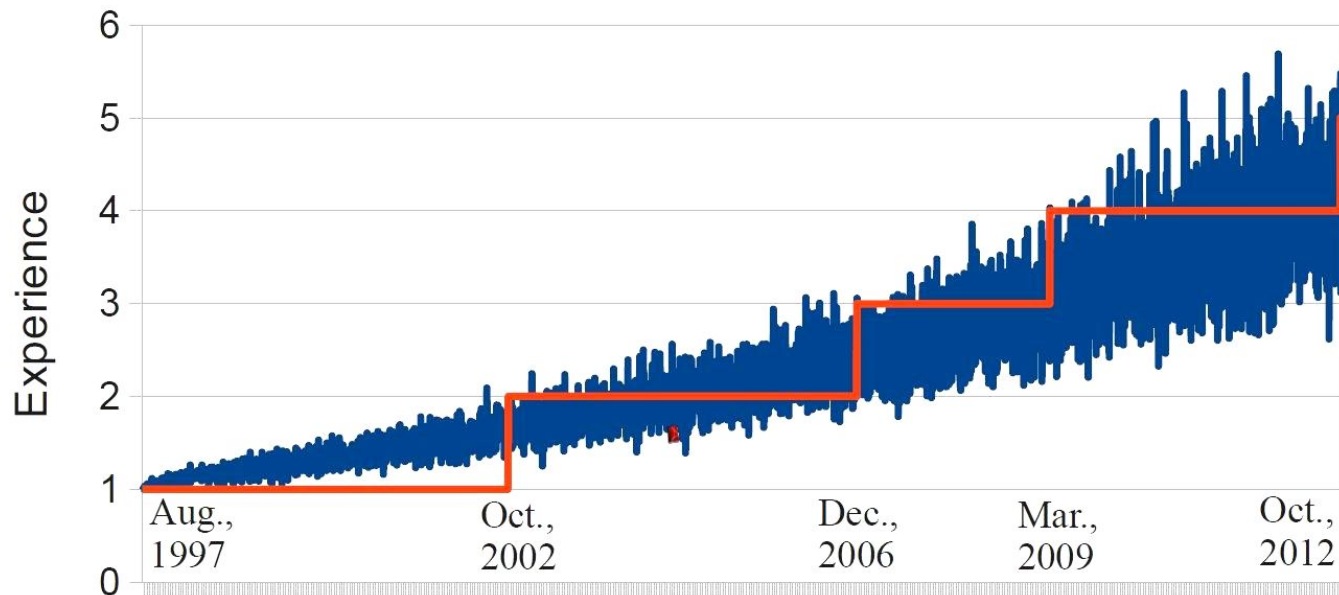
S. Mukherjee, S. Guennemann, G. Weikum, KDD 2016

Can we use similar principles to find useful / helpful product reviews?

Distributional Hypotheses and Semantic Factors

- Reviews (e.g., camera reviews) with similar facet distribution (e.g., both focusing on "zoom" and "resolution") are likely to be equally helpful.
- Helpful reviews focus on the important facets of an item.
- Users with similar facet preferences and expertise are likely to be equally helpful.

Consistency Factors

- **Prior user reputation** (mean helpfulness votes received by her reviews)
- **Prior item prominence** (mean helpfulness votes received by the item's reviews)
- **User rating deviation from community** rating on an item
- **Global rating deviation** (rating bias)
- **Timeliness or "Early-bird" bias** (temporal offset from the first review on the item)

Joint Model

EF 75-300 mm lens is only good to be used outside. The 2.2X HD lens... sexy screen on the edges. Display is the best... Phone is thin built with a good grip despite its size.

Input: { UserId, ItemId, Review, Rating, Timestamp }

- No community-specific characteristics, user profile, item metadata etc.

Facets (Latent)

Expertise (Latent)

(Observed) Words at explicit timepoints in reviews with helpfulness scores
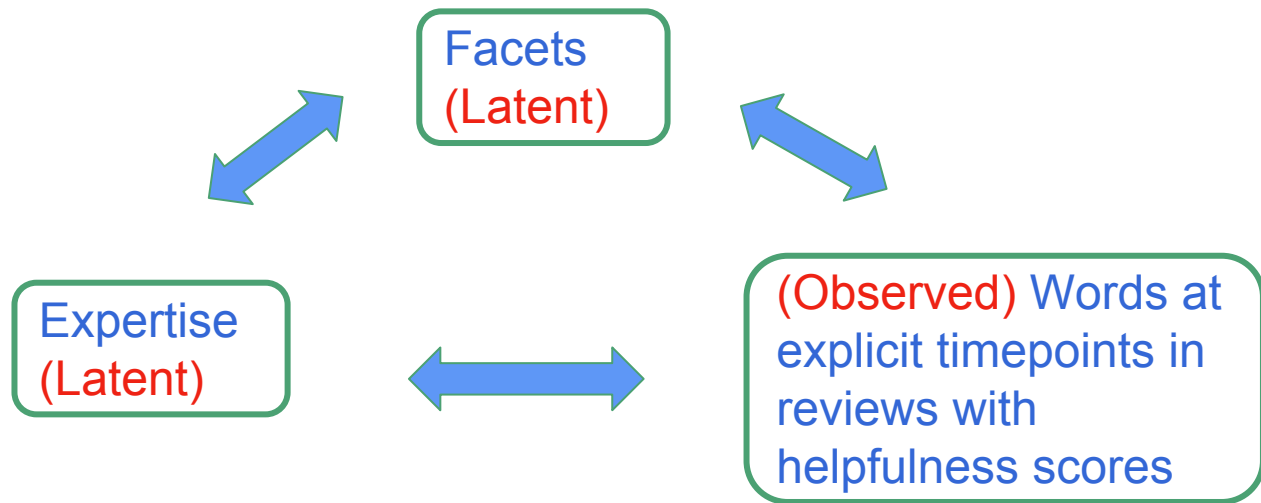
Inference

- Let ξ be a tensor of dimension E X Z (E is the number of latent expertise levels and Z is the number of latent facets). $\xi_{e,z}$ depicts the opinion of users at (latent) expertise level 'e' about (latent) facet 'z'

- Distributional hypotheses intrinsically integrated in ξ

Inference

h(u, i) ➜ helpfulness score of a review by user 'u' on item 'i' at time 't'

$$\widehat{h}(u, i) = f(\beta_u, \beta_i, |r - \overline{r_u}|, |r - \overline{r_i}|, |r - \overline{r_g}|, b_t, \xi; \Psi)$$

Observed consistency factors   Unobserved

$$\Psi^* = argmin_\Psi \frac{1}{|U|} \sum_{u,i \in U,I} (h(u, i) - \widehat{h}(u, i))^2 + \mu ||\Psi||_2^2$$

Parameter $\psi_{e,z}$ depicts the importance of facet 'z' for users at expertise level 'e' in helpful reviews

Generative Model
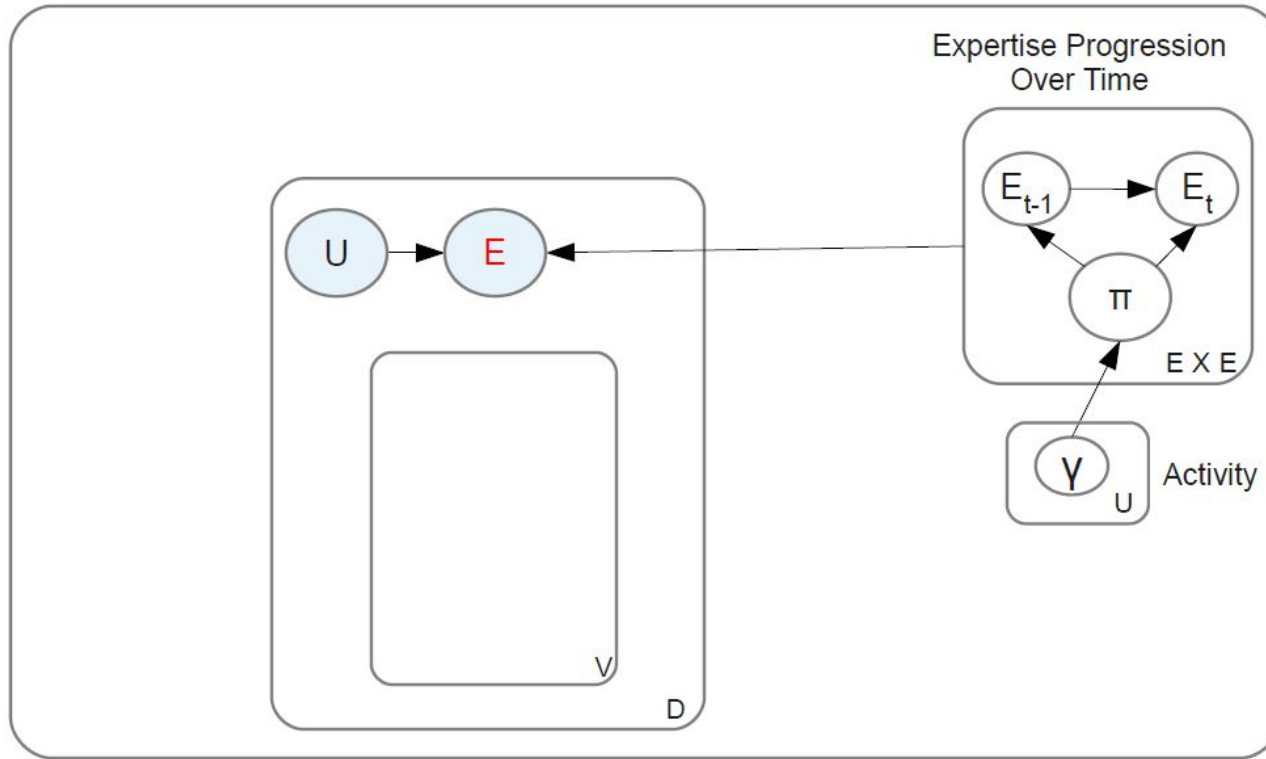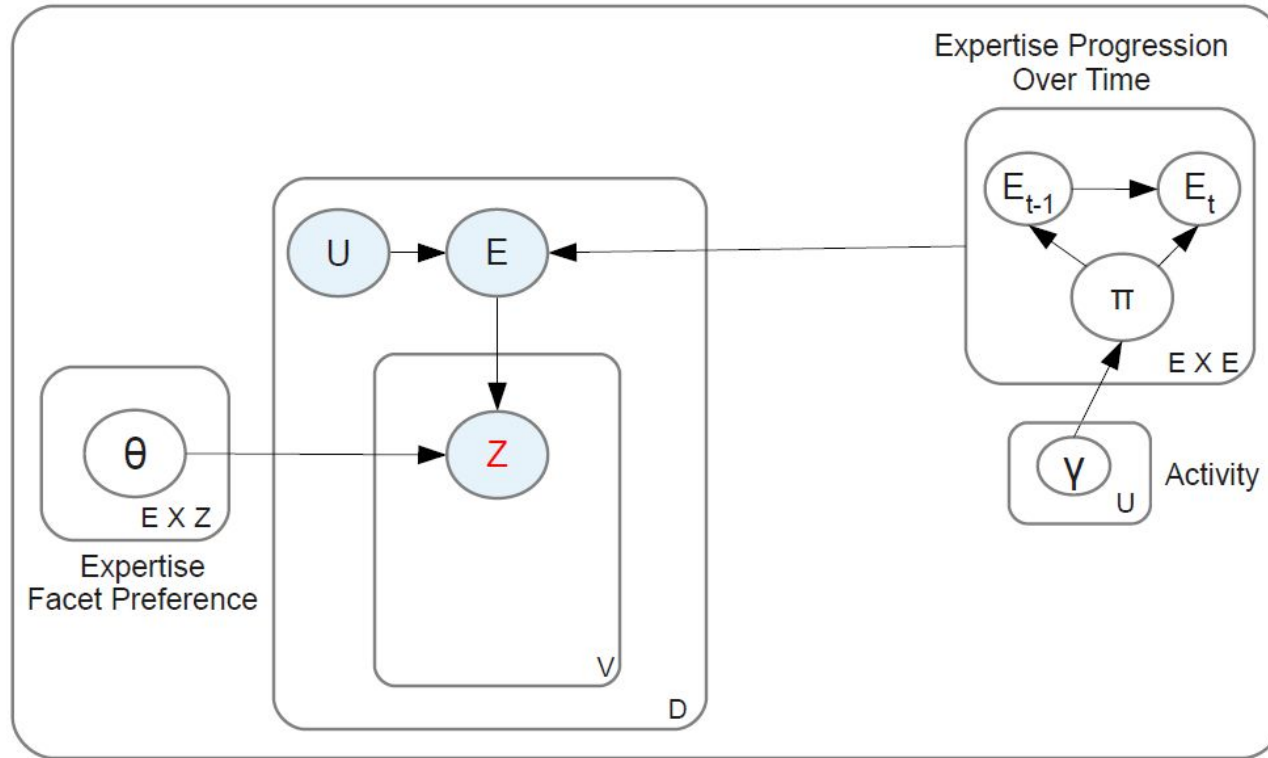
Generative Model
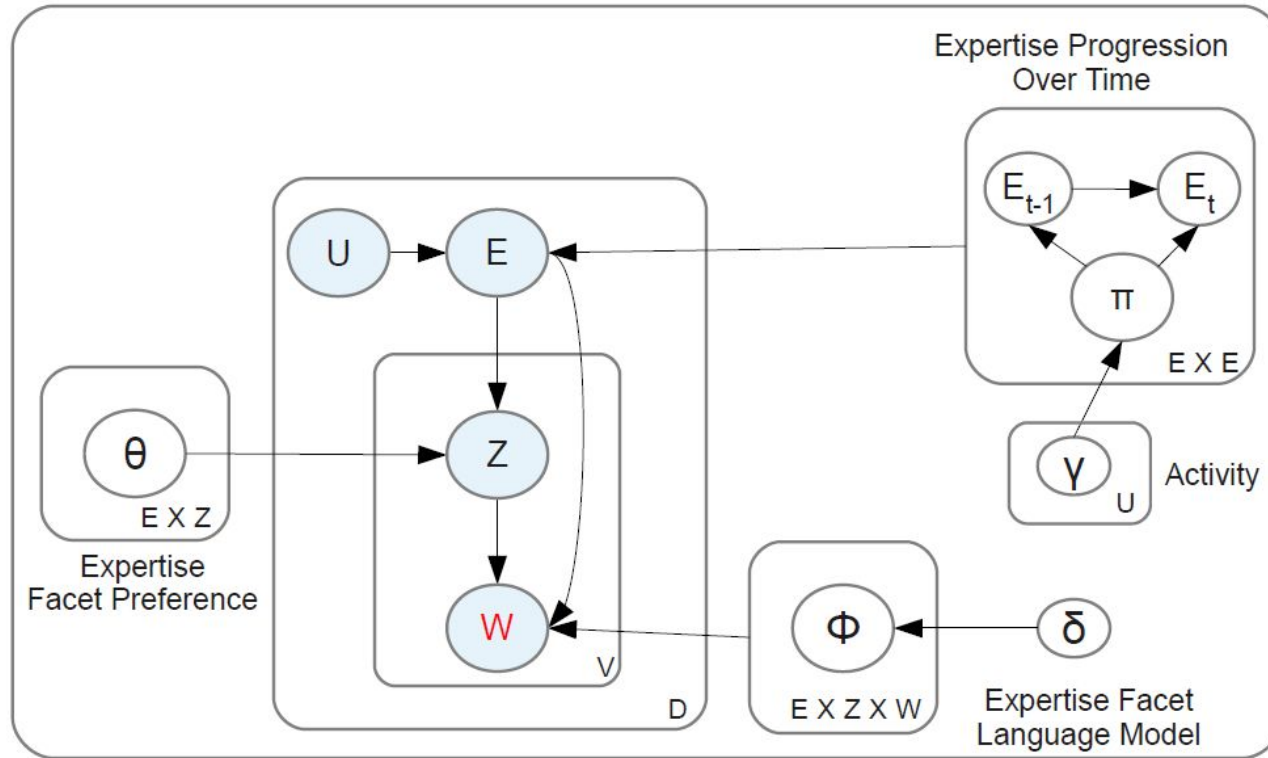
Generative Model
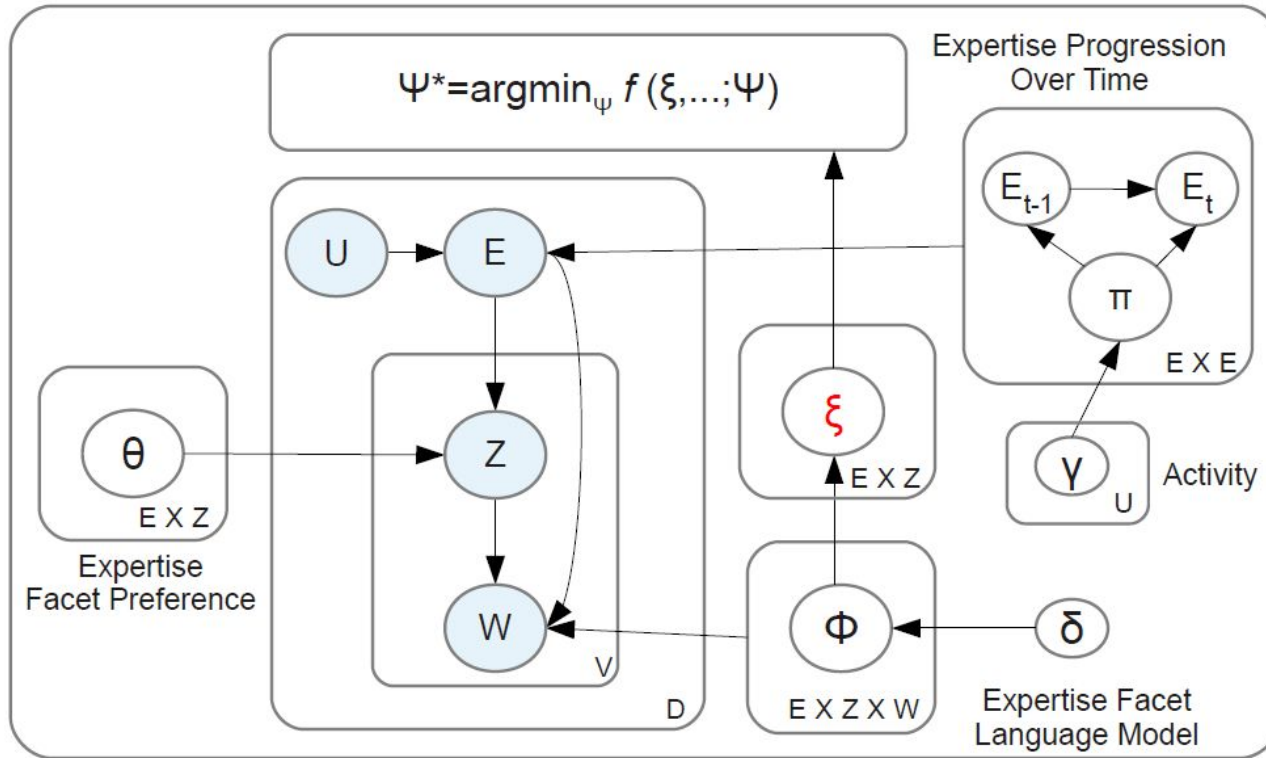
Generative Model

Generative Model
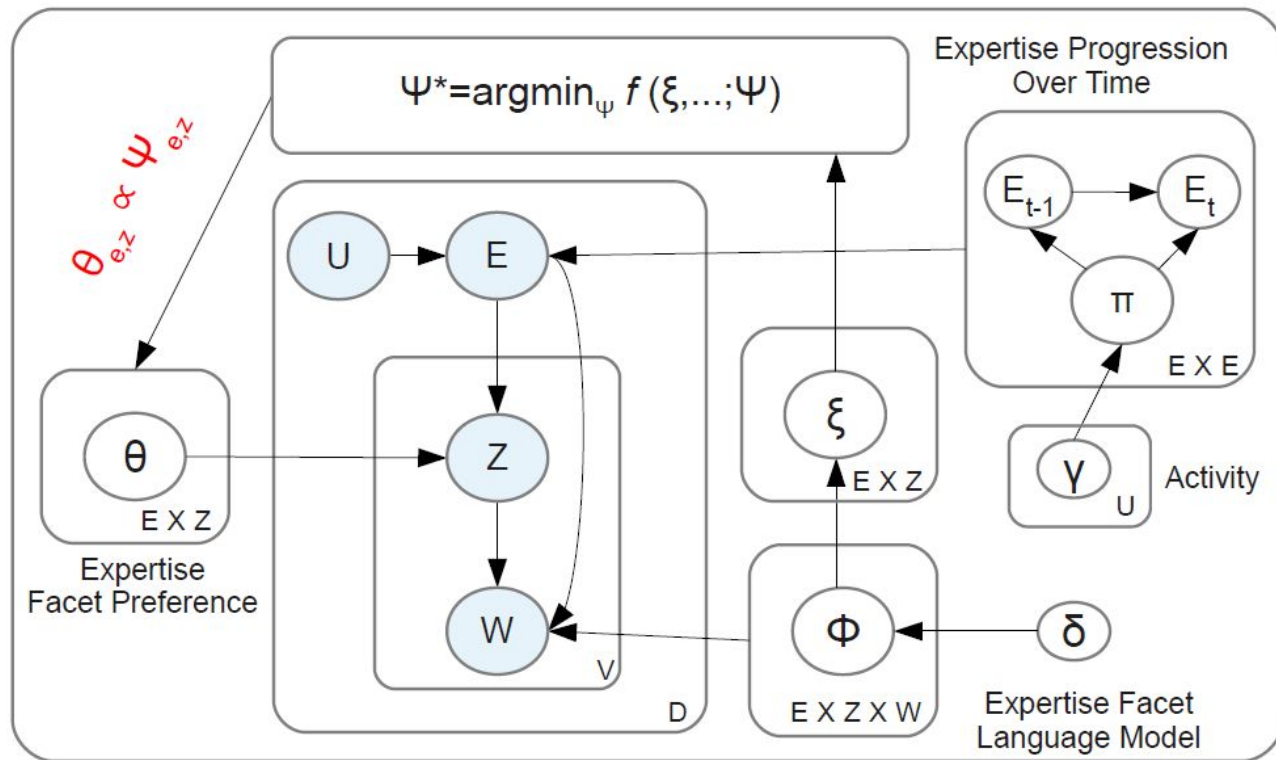
Experiments: Datasets from Amazon

| 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.
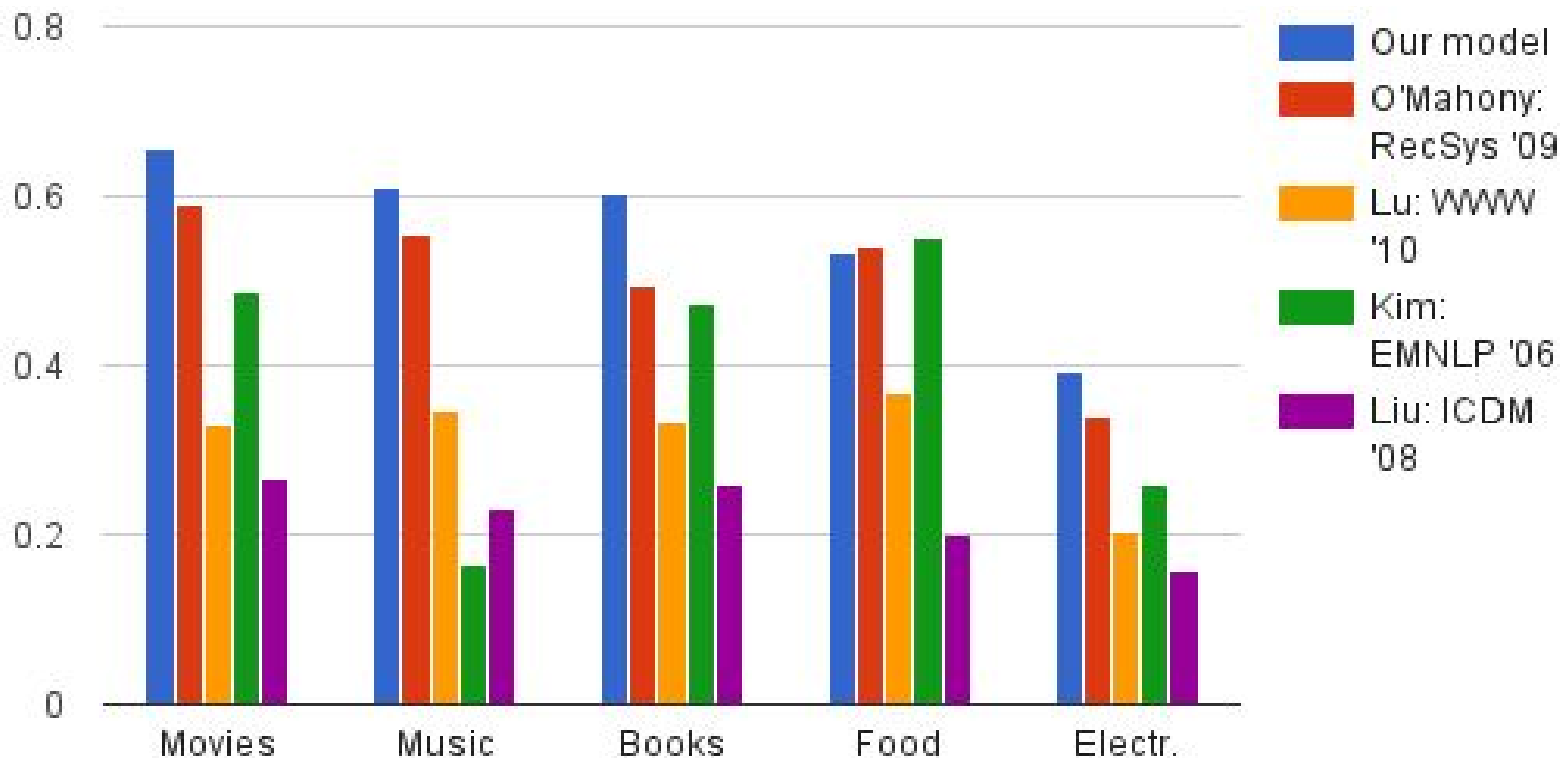
Increase in Log-likelihood per Iteration

Movies   Music   Books   Food   Electronics
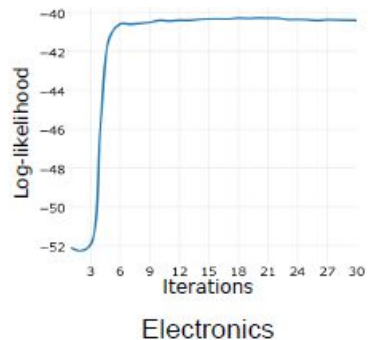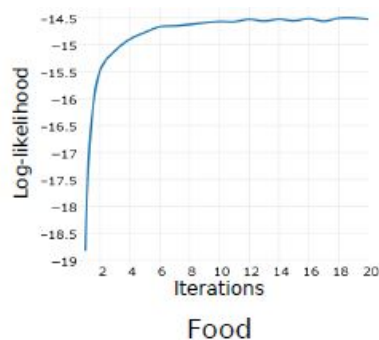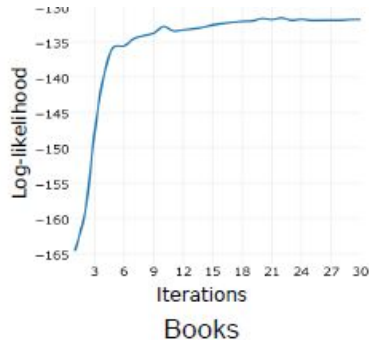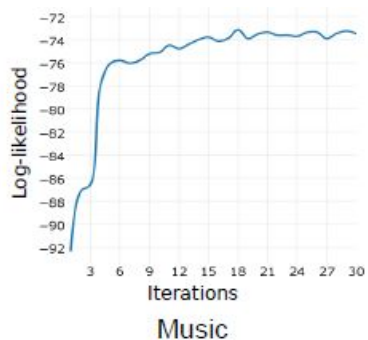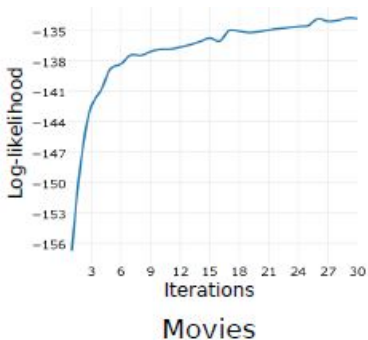
Facet Preference Divergence with Expertise ($\xi_{e,z}$)

Movies   Music   Books
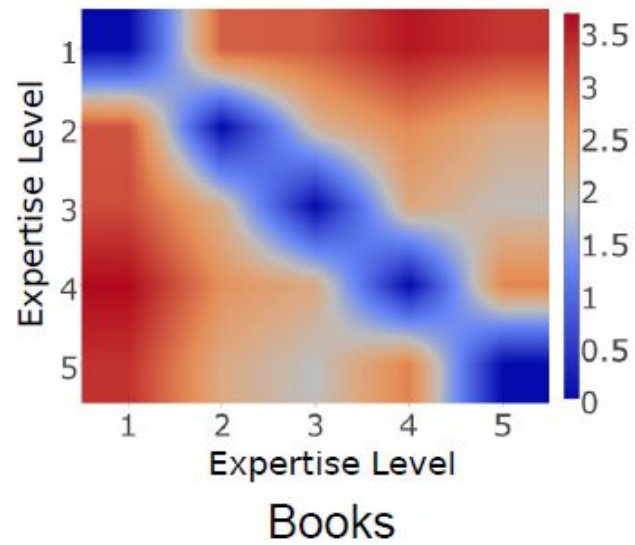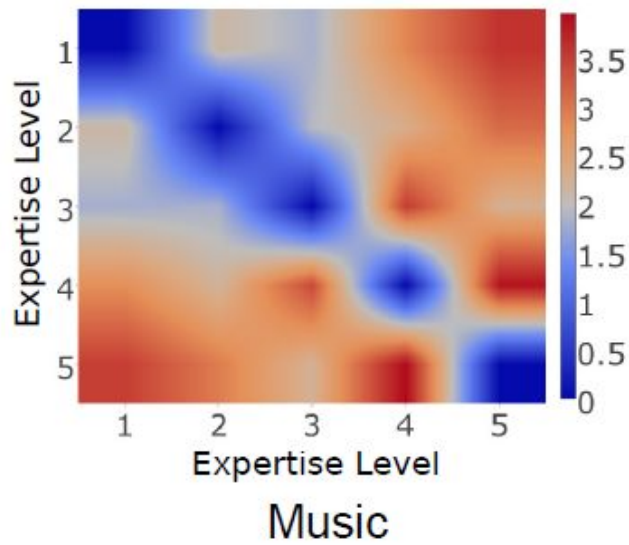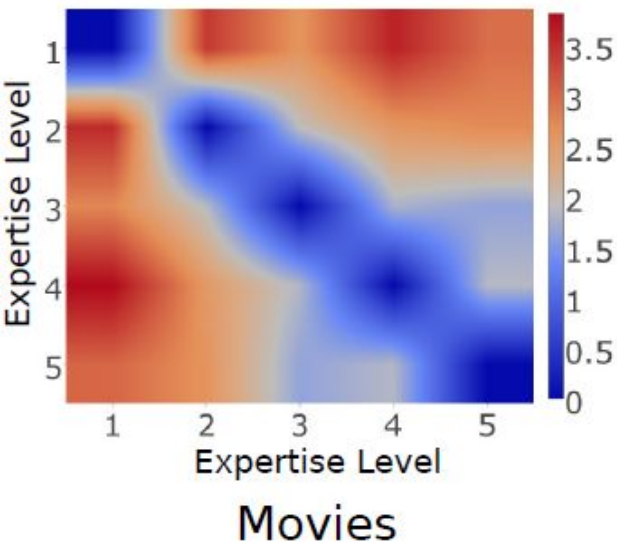
Language Model Divergence with Expertise ($\phi_{e,z,w}$)
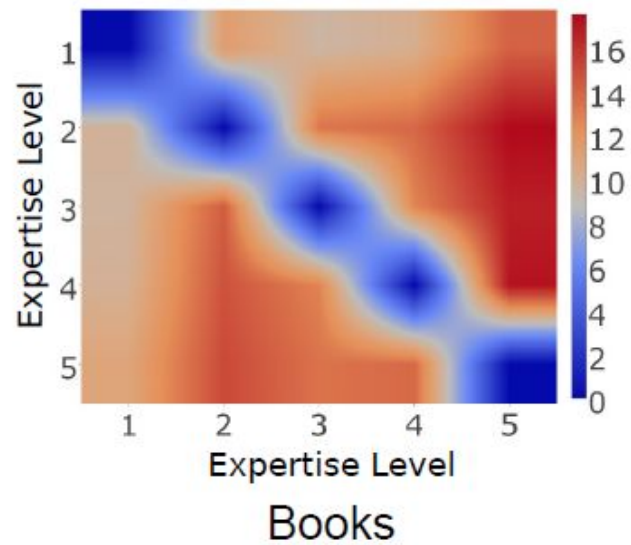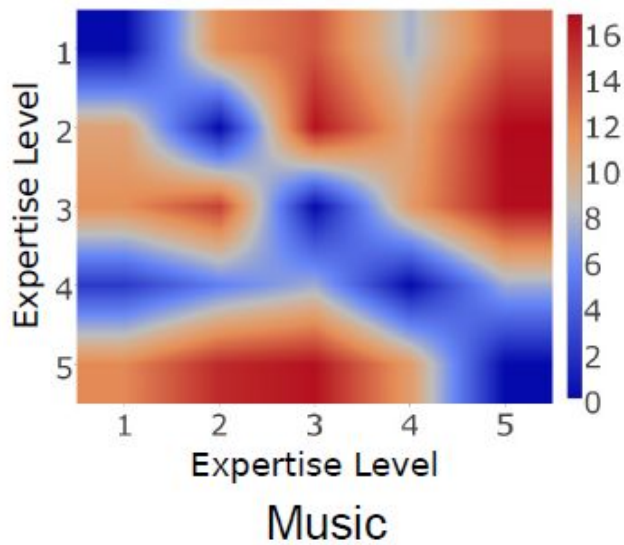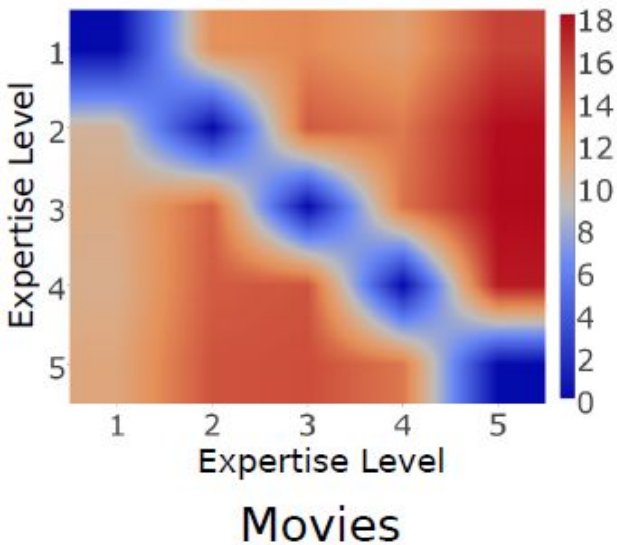

Movies

Music

Books

Interpretability: Top Words for *Most* Helpful Reviews

- **Music:** album, lyrics, soundtrack, touch, songwriting, features, rare, musical, lyrical
- **Books:** serious, complex, content, illustrations, picture, genre, literary, witty
- **Movies:** scene, screenplay, depth, justice, humanity, packaging, perfection, flicks
- **Electronics:** adapter, wireless, computer, sounds, camera, range, drives, mounted
- **Food:** expensive, months, clean, texture, spicy, odor, processed, packs, weather, sticking, caused, scratching, sensation, sipping, smelled

Interpretability: Top Words for *Least* Helpful Reviews

- **Music:** will, good, favorite, cool, great, genius, earlier, notes, attention, place
- **Books:** will, book, time, religious, liberal, material, interest, utterly, moves, movie
- **Movies:** movie, hour, gay, dont, close, previous, features, type, months, meaning
- **Electronics:** order, attach, replaced, write, impressed, install, learn, tool, offered
- **Food:** night, going, haven, fat, avoid, sugar, coffee, store, bodied, graham, variety
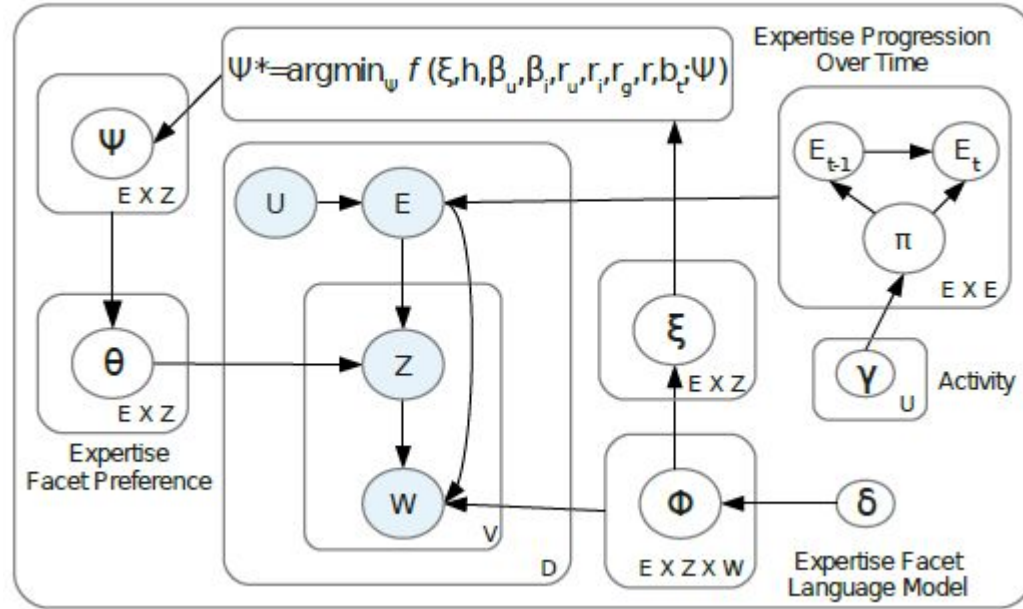
Take-away

A *joint* analysis of semantics, consistency, and user expertise to find useful product reviews