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 phones.

There is no expandable memory.

Awesome for s6 to have theme customization. This is good as now u dont have to 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
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
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
“My first DSLR. Excellent camera, takes great pictures with high definition, without a doubt it makes honor to its name.”

“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 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]

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)
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.

(Observed) Words at explicit timepoints in reviews with helpfulness scores
Let $\xi$ be a tensor of dimension $E \times Z$ ($E$ 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 $\xi$.
Inference

$h(u, i) \rightarrow$ helpfulness score of a review by user ‘u’ on item ‘i’ at time ‘t’

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

**Observed** consistency factors **Unobserved**

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

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

<table>
<thead>
<tr>
<th>Factors</th>
<th>Books</th>
<th>Music</th>
<th>Movie</th>
<th>Electronics</th>
<th>Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Users</td>
<td>2,588,991</td>
<td>1,134,684</td>
<td>889,176</td>
<td>811,034</td>
<td>256,059</td>
</tr>
<tr>
<td>#Items</td>
<td>929,264</td>
<td>556,814</td>
<td>253,059</td>
<td>82,067</td>
<td>74,258</td>
</tr>
<tr>
<td>#Reviews</td>
<td>12,886,488</td>
<td>6,396,350</td>
<td>7,911,684</td>
<td>1,241,778</td>
<td>568,454</td>
</tr>
</tbody>
</table>
Ranking Task: Spearman Rho of our model vs. baselines.
Increase in Log-likelihood per Iteration
Facet Preference Divergence with Expertise ($\xi_{e,z}$)
Language Model Divergence with Expertise ($\phi_{e,z,w}$)
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
A joint analysis of semantics, consistency, and user expertise to find useful product reviews.