FLAME: Combining Aspect Based Opinion Mining and Collaborative Filtering

Yao Wu, Martin Ester
Simon Fraser University, Canada
February 4, 2015
Introduction

Background

- Abundant user-generated reviews available.
- Reviews can help users make better decisions.

Challenges

- **Information Overload** – It’s impossible for users to read all the reviews.
- **Preference Diversity** – People have different opinions towards the same products.
Aspect Based Opinion Mining (ABOM)

- Joint learning of the aspects and the sentiments in reviews.
- E.g., in the above example, **Aspects**: performance, display, value and size. **Sentiments**: 5 stars on display.
1 Introduction

- Aspect-based opinion mining
  - Aspect identification
  - Opinion/sentiment/rating prediction

---

**Apple iPhone 5**

5 star: Everything great!! Attractive exterior, large display, long-lasting battery life.
4 star: I like it very much. It has a beautiful appearance and big screen, and its battery lasts longer.
4 star: The glass on the front is strong enough, and I never had to worry about getting scratched.
3 star: It has a bigger screen, very good-looking, but just too pricey…
3 star: Only newbies and wannabe Apple fans pay expensive price soon to be outdated!
1 star: It is really overpriced. DO NOT BUY IT!

---

**Sentiment Analysis System**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Keywords (aspect terms/opinion words)</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exterior</td>
<td>exterior, appearance, attractive, beautiful, good-looking</td>
<td>5</td>
</tr>
<tr>
<td>Screen</td>
<td>display, glass, screen, large, strong, big, bigger</td>
<td>4</td>
</tr>
<tr>
<td>Battery life</td>
<td>battery life, battery, long-lasting, longer</td>
<td>4</td>
</tr>
<tr>
<td>Price</td>
<td>price, pricey, expensive, overpriced</td>
<td>1</td>
</tr>
</tbody>
</table>
Personalized Latent Aspect Rating Analysis

Example

- The *food* of the same restaurant might be delicious for some users but terrible for others.
- When facing a large number of reviews expressing various opinions, users have no idea with whether the *food* of a restaurant meets his own expectation.

Motivation 1

- learn a user’s personalized preferences on different aspects (e.g., *food*) from his past reviews.
- predict his preference on the aspects of a given item by mining the opinions by other users with similar preferences.

Collaborative Filtering ?!
1 Introduction

- Collaborative filtering
  - Predict a user’s interests by collecting preferences from other users
Collaborative Filtering (CF)

Collaborative Filtering

- predict a user’s interests by collaboratively collecting preferences from many other similar users
- But they only take the overall ratings as input. Two users who have assigned the same 4-stars to a restaurant might have significantly different reasoning.

Motivation 2

- Text reviews provide rich information to understand preferences of users at a finer granularity.
- Aspect-based sentiment scores are not explicitly specified by users, but implicitly expressed in the reviews.
Motivation

To sum up ...

Perform Aspect Based Opinion Mining and Collaborative Filtering together.

- Infer latent aspects and aspect ratings.
- Learn users’ preferences on different aspects.
- Predict latent aspect ratings for users on new items.
3 Model

- Factorized Latent Aspect ModEl (FLAME)
  - Combine aspect-based opinion mining and collaborative filtering
3 Model

User aspect distribution
\[ \eta_u \sim \mathcal{N}(0, \sigma_\eta I) \]

User latent factor
\[ \phi_u \sim \mathcal{N}(0, \sigma_u^2 I) \]

Global aspect distribution
\[ \eta_0 \sim \mathcal{N}(0, \sigma_\eta I) \]
3 Model

\[ \phi_{i,a} \]

\[ \phi_u \]

Item aspect distribution

\[ \eta_i \sim \mathcal{N}(0, \sigma_{\eta} I) \]

Item aspect latent factor

\[ \phi_{i,a} \sim \mathcal{N}(0, \sigma_{i,a}^2 I) \]

\[ \eta_0 \]

\[ \eta_u \]

\[ \eta_i \]

FLAME
3 Model

Aspect word distribution

$$\beta_a \sim \mathcal{N}(0, \sigma_\beta I)$$

Aspect-rating word distribution

$$\gamma_{a,r} \sim \mathcal{N}(0, \sigma_\gamma I)$$

Combine them, get a new language model

$$\alpha_{a,s}[j] = \frac{\exp(\beta_a[j] + \gamma_{a,s}[j])}{\sum_{l=1}^{V} \exp(\beta_a[l] + \gamma_{a,s}[l])}$$
3 Model

\[
\theta_d[a] = \frac{\exp(\eta_0[a] + \eta_u[a] + \eta_i[a])}{\sum_{a' = 1}^{A} \exp(\eta_0[a'] + \eta_u[a'] + \eta_i[a'])}
\]
3 Model

Document-aspect rating distribution

$$\varphi_{d,a}(\mathbf{r}) = \frac{\mathcal{N}(\mathbf{r} | \phi_u^T \phi_{i,a}, \sigma^2_{r,a})}{\sum_{r'=1}^{R} \mathcal{N}(\mathbf{r'} | \phi_u^T \phi_{i,a}, \sigma^2_{r,a})}$$

$$r_{d,a} \sim \mathcal{N}(\phi_u^T \phi_{i,a}, \sigma^2_{a})$$

$$p(r_{d,a} = r)$$

FLAME
3 Model

\[ r_d \sim \mathcal{N}\left( \sum_a \theta_d[a] \mathbb{E}[r_{d,a}], \sigma_r^2 \right) \]
3 Model

\[ a_t \sim \text{Multi}(\theta_d) \]

\[ s_t \sim \text{Multi}(\varphi_d, a_t) \]

FLAME
3 Model

\[ w_n \sim \text{Multi}(\alpha_{a_t, s_t}) \]
Learning Parameters

We adopt a mixture of maximum a posteriori (MAP) point estimates and Bayesian inference to learn the latent variables.

- Lower bound of the joint probability

\[
\mathcal{L} = \sum_d \left( \langle \log p(r_d | \phi_u, \phi_{i,a}, \theta_d) \rangle + \sum_{t \in d} \left( \langle \log p(a_t | \theta_d) \rangle \\
+ \langle \log p(s_t | \phi_d, a) \rangle + \sum_{n \in t} \langle \log p(w_n | a_t, s_t, \beta, \gamma) \rangle \right) \\
+ \sum_{u} \left( \langle \log p(\phi_u | \sigma_u) \rangle + \langle \log p(\eta_u | \sigma_\eta) \rangle \right) \\
+ \sum_i \left( \langle \log p(\phi_i | \sigma_i) \rangle + \sum_a \langle \log p(\phi_{i,a} | \sigma_{i,a}) \rangle \\
+ \langle \log p(\eta_i | \sigma_\eta) \rangle + \langle \log p(\eta | \sigma_\eta) \rangle \right) \\
+ \sum_{\alpha} \langle \log p(\beta_\alpha | \sigma_\beta) \rangle + \sum_{\alpha} \sum_{\beta} \langle \log p(\gamma_{\alpha,\beta} | \sigma_\gamma) \rangle \\
- \sum_d \sum_{t \in d} \left( \langle \log q(a_t | \pi_t) \rangle + \langle \log q(s_t | \lambda_t) \rangle \right) \right)
\]

- Coordinate Ascent-like method to optimize different groups of variables alternatively

- Newton’s method and L_BFGS are used to for variables that do not have closed form solution.
4 Experiment

• Data sets

<table>
<thead>
<tr>
<th></th>
<th>TripAdvisor</th>
<th>Yelp</th>
</tr>
</thead>
<tbody>
<tr>
<td># Users</td>
<td>9,419</td>
<td>6,944</td>
</tr>
<tr>
<td># Items</td>
<td>1,904</td>
<td>3,315</td>
</tr>
<tr>
<td># Reviews</td>
<td>66,637</td>
<td>115,290</td>
</tr>
<tr>
<td>Density</td>
<td>0.37%</td>
<td>0.50%</td>
</tr>
<tr>
<td># Sentences Per Review</td>
<td>12.60 ± 8.64</td>
<td>11.67 ± 7.80</td>
</tr>
<tr>
<td># Words Per Sentence</td>
<td>7.50 ± 3.76</td>
<td>6.47 ± 4.64</td>
</tr>
</tbody>
</table>
4 Experiment

- Perplexity

<table>
<thead>
<tr>
<th></th>
<th>TripAdvisor</th>
<th>Yelp</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA-A</td>
<td>1012.80</td>
<td>767.24</td>
</tr>
<tr>
<td>LDA-AR</td>
<td>918.07</td>
<td>728.00</td>
</tr>
<tr>
<td>D-LDA</td>
<td>771.05</td>
<td>621.24</td>
</tr>
<tr>
<td>FLAME</td>
<td><strong>733.12</strong></td>
<td><strong>590.46</strong></td>
</tr>
</tbody>
</table>

LDA

D-LDA
4 Experiment

- Aspect rating prediction on TripAdvisor

<table>
<thead>
<tr>
<th></th>
<th>PMF</th>
<th>LRR+PMF</th>
<th>FLAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.970</td>
<td>1.000</td>
<td>0.980</td>
</tr>
<tr>
<td>$\rho_A$</td>
<td>N/A</td>
<td>0.110</td>
<td>0.195</td>
</tr>
<tr>
<td>$\rho_I$</td>
<td>0.304</td>
<td>0.177</td>
<td>0.333</td>
</tr>
<tr>
<td>$L_{0/1}$</td>
<td>0.210</td>
<td>0.238</td>
<td>0.196</td>
</tr>
</tbody>
</table>

$$\rho_A = \frac{1}{D} \sum_{d=1}^{D} \rho(s_d, s'_d)$$

$$\rho_I = \frac{1}{U \cdot A} \sum_{u=1}^{U} \sum_{a=1}^{A} \rho(s_{Iu,a}, s'_{Iu,a})$$

Zero-One Ranking loss ($L_{0/1}$)

PMF

LRR
Aspect Identification

(a) Location

(b) Location 2-star

(c) Location 5-star

(d) Service

(e) Service 2-star

(f) Service 5-star
Figure: Aspect Weights. *Global* represents the overall aspect distribution on the corpus. *user-1* and *user-2* are the aspect weights of two randomly sampled users, and *item-1* and *item-2* are the aspect weights of two sample items.
Other Applications

**Personalized Review Recommendation**
Pick the reviews by users with similar tastes.

**Recommendation Explanation**
More persuasive recommendation explanations by the predicted aspect ratings and some selected reviews written by similar users.