Discovering Facets in Reviews

1 The Original Model in ICDM Paper

Suppose the review corpus $(\mathcal{R}, \mathcal{V})$ consists of reviews $\mathcal{R} = \{r_1...r_R\}$ and ratings $\mathcal{V} = \{v_1...v_R\}$. Next assume that each review r_i is divided into sentences $s \in r_i$, and that each rating v_i is divided into $\mathcal{K}$ aspects $\{v_{i1}...v_{iK}\}$, e.g., ratings on feel, look, smell, taste, overall.

Then the probability for an entire corpus is:

$$P^{(\theta,\phi)}(aspects|\mathcal{R},\mathcal{V}) = \prod_{i=1}^{\mathcal{R}} \prod_{s \in r_i} P^{(\theta,\phi)}(aspects(s)|s, v_i) \tag{1}$$

where:

$$P^{(\theta,\phi)}(aspects(s) = k|s, v) = \frac{1}{Z_s^{(\theta,\phi)}} \exp \sum_{w \in s} \{\theta_{kw} + \phi_{kv_{ik}w}\}. \tag{2}$$

where θ_{kw} is the aspect weights and $\phi_{kv_{ik}w}$ is the sentiment weights.

The normalization constant Z_s is

$$Z_s^{(\theta,\phi)} = \sum_{k=1}^{K} \exp \sum_{w \in s} \{\theta_{kw} + \phi_{kv_{ik}w}\}. \tag{3}$$

In this model the θ parameters correspond to the topic component, and the ϕ parameters correspond to the topic-specific rating component. So examining the significant weights of θ_k should reveal words that correspond to the k-th aspect, e.g., "taste". Likewise, examining the significant weights of $\phi_{k,v}$ should reveal words that correspond to used to describing a rating of v for the k-th aspect, e.g., "delicious".
This model makes a few key assumptions:

- For every review, the ratings for all aspects are observed.
- Conditioned on the ratings and the an individual sentence in a review, modeling the aspect of that sentence is independent of other sentences in the review.

2 Modification-hidden ratings

In general, we cannot assume access to ratings for every aspect. In fact, most platforms only allow for giving an "overall" rating, and the aspects must be discovered purely from the reviews themselves. This model assumes that we only observe the overall rating and overall rating is model as the average of hidden rating vectors$\{v_{i1} \dots v_{ik}\}$. So, we have:

$$v_i = v_{overall} = \frac{1}{K} \times \sum_{k=1}^{K} v_{ik} \tag{4}$$

The objective function is:

$$\begin{aligned} P^{(\theta,\phi)}(aspects = k, ratings = \vec{v_{ik}}|\mathcal{R}) \quad = \quad & \prod_{i=1}^{\mathcal{R}} \Bigg\{ \Big\{ \prod_{s \in r_i} \frac{1}{Z_s^{(\theta,\phi)}} \exp \sum_{w \in s} \{\theta_{kw} + \phi_{kv_{ik}w}\} \Big\} \\ & \times \frac{1}{Z_{\vec{v}_{ik}}} \exp \Big(-\lambda \| \frac{1}{k} \sum_{k} v_{ik} - v_i \|^2 \Big) \Bigg\}. \end{aligned} \tag{5}$$

Then we can perform optimization by alternately optimizing:

$$t_s = \arg\max_k \sum_{w \in s} \{\theta_{kw} + \phi_{kv_k w}\} \tag{6}$$

$$v_k = \arg\max_v \Big\{ \sum_{s \in r_i} \delta(t_s^i = k) \sum_{w \in s} \phi_{kv_k w} - \lambda \| \frac{1}{k} \sum_k v_{ik} - v_i \|^2 \Big\} \tag{7}$$

$$(\theta, \phi)^{(t+1)} = \arg\max_{\theta,\phi} \log p^{(\theta,\phi)}(aspects, ratings|\mathcal{R}, \mathcal{V}) - \Omega(\theta, \phi) \tag{8}$$

where t_s is the topic assignment for one sentence and v_k is the rating for kth topic in review i, λ is a weighting factor to balance the squared loss and sum of ϕ. $\Omega(\theta, \phi) = \|\theta\|_2^2 + \|\phi\|_2^2$

When calculating the derivative, the objective function is then to minimize

$$F = -\log p^{(\theta,\phi)}(aspects, ratings|\mathcal{R}) + \Omega(\theta, \phi) \tag{9}$$

The derivative is calculated as follows:

$$\frac{\partial F}{\partial \theta_{kw}} = 2\theta_{kw} - \sum_{i=1}^{\mathcal{R}} \sum_{s \in r_i} \Big\{ 1 - \frac{1}{Z} \exp \sum_{w \in s} \{\theta_{kw} + \phi_{kv_{ik}w}\} \Big) \Big\} \tag{10}$$

$$\frac{\partial F}{\partial \phi_{kv_k w}} = 2\theta_{kw} - \sum_{i=1}^{\mathcal{R}} \sum_{s \in r_i} \Big\{ 1 - \frac{1}{Z} \exp \sum_{w \in s} \{\theta_{kw} + \phi_{kv_{ik}w}\} \Big) \Big\} \tag{11}$$