Jointly Modeling Aspects, Ratings and Sentiments With Temporal Dynamics

Rishabh Misra
CSE Department
University of California, San Diego
r1misra@ucsd.edu

Tushar Bansal
CSE Department
University of California, San Diego
tbansal@ucsd.edu

***Abstract*—In this work, we propose an extension to the paper Jointly Modeling Aspects, Ratings and Sentiments for movie recommendation (JMARS) by carefully incorporating temporal dynamics into the model and empirically showing that it provides an improvement of around 0.9% on two datasets. We replicate the quantitative results and qualitative analysis provided in the original paper and furthermore, provide some analysis from the extended model.**

I. Introduction

Model proposed in JMARS belongs to the family of integrated models that use ratings and reviews to extract a wealth of information. Authors provide a statistical model and demonstrate in experiments that their approach excels at recommending movies while simultaneously providing meaningful analysis of the interests and aspects relevant for users and movies.

JMARS provides a principled extension of the factorization models commonly used for recommendation. It assumes that each user (and each movie) has an aspect distribution of interest and reviews are generated by drawing from the product of movie and user aspects. Hence, reviews inform both about the content of a movie and also about the interests of a user. In addition, JMARS models words into five groups: Background, Item specific, Aspect, Aspect sentiment, and General sentiment. Thus, it provides a joint model of user activity, movie content, ratings, reviews, and a detailed language model of the reviews.

II. JMARS

Users are assumed to have a given interest distribution θ_u in terms of aspects they write and care about. Moreover, they are also assumed to have biases b_u regarding what can be considered to be a reasonable baseline with regard to their choice. Likewise, items contain a number of aspects, as indicated by θ_m and a bias b_m. Furthermore, JMARS considers that user's expectations v_u match the item's properties v_m, when viewed under the angle of a specific aspect, as captured by M_a. These aspect-specific ratings of a item by a user r_{uma} are then aggregated, based on the user's priorities to obtain an aggregate rating r_{um}.

As for the actual review text, JMARS assumes the following: reviews contain words drawn from a baseline language model of words typically occurring in reviews ϕ_0, there are positive and negative sentiment words, as indexed by ϕ_s, where $s \in \{positive, negative\}$ and finally, there are aspect specific word distributions ϕ_a, again colored by sentiment s, i.e. ϕ_{as}. Depending on whether a user appreciates a particular aspect of a item, as indicated by r_{uma}, he will generate positive or negative sentiment words (or simply neutral ones). Moreover, there are item-specific words, such as the name of the main protagonists, the title, and other named entities that are bound to occur in a review, regardless of the user. This approach of mixing between five different components summarizes the strategy for JMARS.

A. Model Overview

1) Matrix Factorization with Aspects: As in PMF [5], JMARS assume that users u and items m are characterized by latent factor vectors v_u and v_m respectively, that are drawn from zero-mean spherical Gaussian priors.

$$v_u \sim N(0, \sigma_u^2 I)$$

$$v_m \sim N(0, \sigma_m^2 I)$$

Ratings per aspect is then calculated as:

$$r_{uma} = v_u^T M_a v_m + b_u + b_m + b_0$$

Here b_u and b_m are biases for users and items respectively and b_0 is a common bias. The idea is that while v_u and v_m encode the general profile, the matrix M_a emphasizes the aspect specific properties. That is, while items may be overall good, they may or may not excel quite as much in specific aspects. JMARS assumes Gaussian priors with fixed mean and precision on real-valued parameters. Specifically, this means that each element of M_a, v_u, v_m, b_u, b_m follows a Gaussian distribution with zero mean and a fixed precision.

2) Two factor model: JMARS also contains an exponential additive model in terms of θ_u and θ_m, which are defined as follows:

$$\theta_u \sim N(0, \sigma^2_{useraspect}\mathbf{I})$$

$$\theta_m \sim N(0, \sigma^2_{itemaspect}\mathbf{I})$$

Moreover, the joint aspect distribution is given by:

$$\theta_{um} \propto exp(\theta_u + \theta_m)$$

or

$$p(a|\theta_u, \theta_m) = \frac{exp(\theta_{ua} + \theta_{ma})}{\sum_{a'} exp(\theta_{ua'} + \theta_{ma'})}$$

Finally, JMARS assumes that the extent of discussion in a review and the relative importance of a aspect coincide. Under this assumption, the final rating can be calculated as:

$$\hat{r}_{um} = v_u^T \left[\sum_a p(a|\theta_u, \theta_m) M_a \right] v_m + b_u + b_m + b_0$$

Here $\hat{r}_{um}$ is the predicted review rating, and the observed rating r_{um} is generated using $N(\hat{r}_{um}, \epsilon^{-2})$.

3) Language Models: JMARS assume that the review language model is given by a convex combination of five components. The 5 components are defined as:

Background Words: The words that are uniformly distributed in every review are considered background words. For example, in the case of product reviews, these words include “size”, “wear”, etc.

Aspects: These are words associated with specific aspects. For example, “color”, "elastic" and "care" are all aspect words related to the "material" aspect.

Aspect Sentiments: These words usually come with a specific aspect to express positive or negative sentiments. For example, words such as “loose”, “stretchable” usually appear with a discussion of the “fit”.

Sentiment: These are the words that express general sentiment. For example, sentiment words such as “great”, “bad”, or “worse” do not really convey any aspect specific content.

Item Specific: Words such as the name of the clothing articles in an item, or any term that appears only for the item are considered item-specific. Background and Item specific words provide less information about item quality.

Crucial to the mixture between these models is the use of a switch variable which chooses between the above types. JMARS accomplish this via π, the switching distribution.

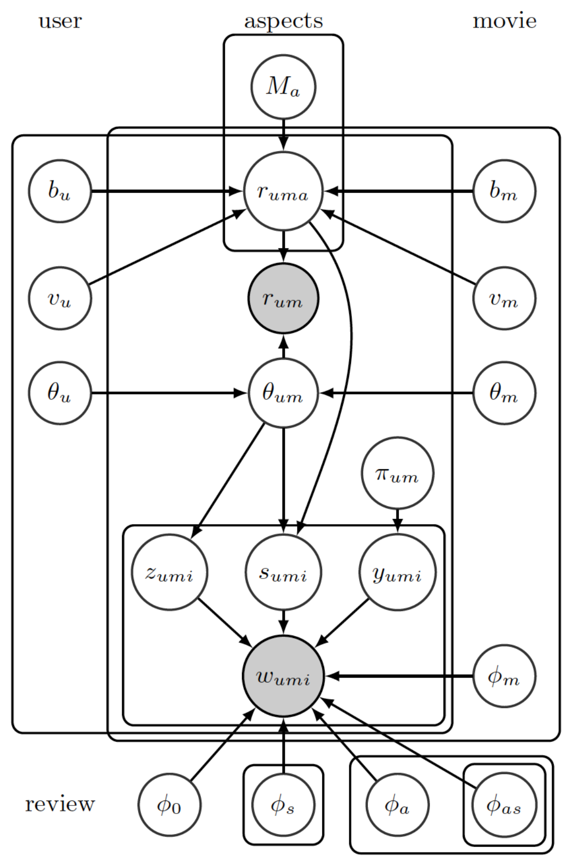

Fig. 1: JMARS Model

- **Switching distribution** π_{um}: For each user u and item m, π_{um} is drawn from a Dirichlet prior. Subsequently, for each word i in the review, y_{umi} is drawn from a multinomial distribution defined by π_{um}, that is
$$\pi_{um} \sim Dir(\gamma) \text{ for } \pi_{um} \in P_5,$$
$$y_{umi} \sim Mult(\pi_{um}).$$
- **Aspect** z_{umi}: Aspect for each word in the review is sampled from θ_{um}, i.e. $z_{umi} \sim Mult(\theta_{um})$.

- **Aspect sentiment** s_{umi}: When s_{umi} is an aspect-specific sentiment, its sentiment is determined by the aspect-specific rating via a logistic link function.

$$p(s_{umi}|r_{uma}, z_{umi} = a) = \frac{1}{1 + e^{-s_{umi}(c.r_{uma}-b)}}$$

  In other words, the propensity of picking a $+ve$ or a $-ve$ sentiment word are related to the aspect specific rating r_{uma}.
- **Aggregate sentiment** s_{umi}: When s_{umi} is an general sentiment, this is entirely analogous to above. The only difference is that we draw s_{umi} from the aggregate rating $\hat{r}_{um} = \sum_a \theta_{uma} r_{uma}$.
- **Language models** ϕ_0, ϕ_s, ϕ_a, ϕ_{as}, ϕ_m: Each of the language models is a multinomial distribution with a Dirichlet as a conjugate. That is, we assume that

$$\phi_0 \sim Dir(\eta_0),$$
$$\phi_s, \phi_{as} \sim Dir(\eta_{sentiment}),$$
$$\phi_a \sim Dir(\eta_{aspect}),$$
$$\phi_m \sim Dir(\eta_{item})$$

  where the value of each element in η depends the part-of-speech tag of the corresponding word.
- **Emission model**: The final piece in our approach is to model how the actual words are being generated.
  - Based on y_{umi}, we first decide which of the five model types to pick.
  - If y_{umi} is aspect specific, select ϕ from the aspect models using aspect z_{umi}.
  - If y_{umi} is aspect-sentiment specific, inspect s_{umi} for a matching sentiment for aspect z_{umi}.
  - If y_{umi} is sentiment specific, inspect s_{umi} for the corresponding sentiment.

  Likewise, authors choose the baseline model ϕ_0 or the item specific model ϕ_m as needed.

B. Learning

The goal here is to learn the hidden factor vectors, aspects, and sentiments of the textual content to accurately model user ratings and maximize the probability of generating the textual content. Hence the objective is the negative log posterior, defined as

$$\mathcal{L} := -\log p(\mathcal{R}, \mathcal{W}|\Upsilon, \Omega).$$

where $\mathcal{R}, \mathcal{W}$ denote the ratings and words respectively and γ and Ω are the Gaussian and Dirichlet hyperparameters. Unfortunately, inference in this problem is intractable in its direct formulation. Instead, authors resort to a hybrid inference procedure combining sampling and variational optimization. That is, they use Gibbs-EM, an inference method that alternates between collapsed Gibbs sampling and gradient descent, to estimate parameters in the model. After collapsing out the parameters pertaining to the language model, terms cease to be conditionally exchangeable, hence L cannot be decomposed further. That said, all relevant terms decompose for the purpose of the inference algorithm and we have:

$$\mathcal{L} \text{ ``=''} \sum_{r_{um} \in \mathcal{R}} \left[\epsilon^{-2}(r_{um} - \hat{r}_{um})^2 - \log p(w_{um}|\Upsilon, \Omega)\right].$$

The first term denotes the prediction error on user ratings. The second term denotes the probability of observing the text conditioned on priors.

E-Step: In the E-step, Gibbs sampling is performed to learn the hidden variables $\{y, z, s\}_{umi}$ by fixing the values of θ_{um} and all $\{r_{uma}\}_{a=1}^{A}$ updated in the gradient descent step. Dirichlet-Multinomial conjugacy allows Gibbs sampling to work by sampling on the individual tuple of $\{y, z, s\}_{umi}$, collapsing out all the language models ϕ.

For the word in the i-th position of the review written by user u for item m, we jointly sample its switching variable y_{umi}, topic z_{umi} and sentiment s_{umi}, conditioned on its Markov blanket. Let w = w_{umi} and d denote the set of variables $\{umi\}$.

Here $C^{w}_{y=4,d,m}$ denotes the number of times that w is sampled as a item-specific word in item m excluding the current word assignment; all the other Cs are defined in the same way. **I**$(\cdot)$ is a indicator function that returns 1 if the statement is true and 0 otherwise. In other words, we effectively have a big switch statement distinguishing 5 cases.

M-Step: In this step, we use gradient descent to learn the set of parameters $\Theta = [\{v_u, b_u, \theta_u\}_{Uu=1}, \{v_m, b_m, \theta_m\}_{Mm=1}, \{M_a\}_{a=1}^{A}]$ by fixing values of $\{y, z, s\}_{umi}$. The objective function is modified as follows:

$$
\begin{aligned}
&p(y_d = y, z_d = z, s_d = s | y_{\neg d}, w, \theta_{um}, \Omega) \\
&\propto \frac{C^y_{\neg d} + \gamma}{\sum_{y'=1}^5 C^{y'}_{\neg d} + 5\gamma} \cdot \left[\frac{C^w_{y,\neg d} + \eta^w_0}{\sum_{w'=1}^V C^{w'}_{y,\neg d} + \eta^{(\cdot)}_0} \right]^{\mathbb{I}(y=0)} \\
&\cdot \left[\frac{C^w_{y,\neg d,s} + \eta^w_{\text{sentiment}}}{\sum_{w'=1}^V C^{w'}_{y,\neg d,s} + \eta^{(\cdot)}_{\text{sentiment}}} p(s|\hat{r}_{um}) \right]^{\mathbb{I}(y=1)} \\
&\cdot \left[\frac{C^w_{y,\neg d,z} + \eta^w_{\text{sentiment}}}{\sum_{w'=1}^V C^{w'}_{y,\neg d,z} + \eta^{(\cdot)}_{\text{sentiment}}} \cdot \theta_{umz} \cdot p(s|r_{umz}) \right]^{\mathbb{I}(y=2)} \\
&\cdot \left[\frac{C^w_{y,\neg d,z} + \eta^w_{\text{aspect}}}{\sum_{w'=1}^V C^{w'}_{y,\neg d,z} + \eta^{(\cdot)}_{\text{aspect}}} \cdot \theta_{umz} \right]^{\mathbb{I}(y=3)} \\
&\cdot \left[\frac{C^w_{y,\neg d,m} + \eta^w_{\text{movie}}}{\sum_{w'=1}^V C^{w'}_{y,\neg d,m} + \eta^{(\cdot)}_{\text{movie}}} \right]^{\mathbb{I}(y=4)}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}' = &\sum_{r_{um} \in \mathcal{R}} \left[\epsilon^{-2} (r_{um} - \hat{r}_{um})^2 - \log p(\{w, y, z, s\}_{um} | \Theta) \right] \\
&- \log p(\Theta | \Upsilon).
\end{aligned}
$$

The first term is the squared error which allows us to reduce the squared error from predicted rating. The second term's goal is to maximize the likelihood of generating all the observed $\{y, z, s, w\}_{u,m}$ variables obtained from Gibbs sampling. The final term is the Gaussian prior of all the parameters. Our is then to minimize the following objective function, decomposed from the above equation.

Let $\mathcal{L}'_{um}$ be the objective for a single rating and review texts, that is:

$$
\mathcal{L}' = \sum_{r_{um} \in R} \mathcal{L}'_{um} logp(\Theta | \gamma)
$$

We expand the likelihood contribution of a given (user, item) pair $\mathcal{L}'_{um}$ as follows:

$$
\begin{aligned}
\mathcal{L}'_{um} = &\epsilon^{-2} (r_{u,m} - \hat{r}_{u,m})^2 \\
&- \log p(\{w, z, s\}_{um} \mid \theta_u, v_u, b_u, \theta_m, v_m, b_m, M_a) \\
= &\epsilon^{-2} (r_{u,m} - \hat{r}_{u,m})^2 - \sum_s N^{y=1}_{u,m,s} \log p(s \mid \hat{r}_{um}) \\
&- \sum_a \sum_s N^{y=2}_{u,m,a,s} \log p(s | r_{uma}) - \sum_a N^{y=3}_{u,m,a} \log \theta_{uma}.
\end{aligned}
$$

where $N^{y=1}_{u,m,s}$ is the number of times general sentiment s appears in user u's review in item m, and $N^{y=2}_{u,m,a,s}$ is the number of times the aspect sentiment s appears under aspect a, and $N^{y=3}_{u,m,a}$ is the number of times aspect a appears in the review. We then compute the first derivatives of $\mathcal{L}'$ with respect to the variables. We optimize $\mathcal{L}'$ using gradient decent (the original paper uses L-BFGS but the convergence using this seemed slow for us).

III. Extension

The extension we propose to this method is to add temporal dynamics to the current model. The JMARS considers static user-item biases and latent factors across time. However, our hypothesis is that user latent factors, biases and interest distribution for different aspects might change with time.

Introduction of temporal dynamics in recommender systems has been a popular concept. In [3], incorporating the monotonically increasing time function lead to significant improvement to the model. We take inspiration form [3] and incorporate temporal dynamics in a similar fashion in our model. Specifically, we change the parameters in the original model as follows:

$$
\begin{aligned}
b_u &= b_u + \alpha_{bu} * dev(t) \\
v_u &= v_u + \alpha_{vu} * dev(t) \\
\theta_u &= b_u + \alpha_{\theta u} * dev(t) \\
\text{where } dev(t) &= sign(t - t_u) |t - t_u|^\beta
\end{aligned}
$$

where t_u is the mean of the times user u posted the reviews and β is a constant. We substitute the above variables in JMARS equations to capture a changing user preference over time. The α values are learned using the EM algorithm like other parameters and β is tuned on a separate validation set.

IV. Data

The original paper runs experiments on the IMDB movie dataset but since the IMDB data does not contain the time of the review, we decided to opt for the Amazon datasets [2] from two categories: Clothing and Instant Video. Clothing data is for 1,981 users, 1,962 items and contains 11,935 reviews. Instant video data is for 2000 users, 1643 items and contains 14355 Reviews. Each dataset has a density of around 0.3%. Since the inference on JMARS is computationally expensive and time intensive, and we spent most of our time implementing the method, we use a relatively small dataset to get some quick results.

We use a separate and even smaller dataset containing 476 users, 483 items and 1923 reviews to tune the hyperparameters like β, ϵ, learning rate and others. The validation was performed on the 20% split of this data.

V. Experiments and Quantitative Results

Our implementation of JMARS and the extension is available on Github[6]. For the evaluation, we perform a random 80-20 split based on reviews for each dataset. The selected value of β was 0.000085, mean of all gaussian parameters was set to 0 and standard deviation to 0.01. The β value is so small because we used the unix timestamp provided in the Amazon data as is, which is a big number. After tuning the hyperparameters, we perform 50 iterations of the proposed EM algorithm. In each step, we perform one iteration of Gibbs Sampling (E-step) and 10 iterations of gradient decent (M-step). The analysis was performed with aspects and latent dimension set as $A = \{6, 12\}$ and $K = 5$ respectively. Although, the original paper performed 500 iterations on a large dataset, computational cost and time constraint restricted us to perform only 50 iterations.

The tables below describe the results we obtained on the Amazon clothing and instant video datasets for various parameter settings after 50 iterations of EM algorithm.

Clothing Data	without time	with time	improvement
Baseline	1.1505	1.1420	0.74%
JMARS (A=6; K=5)	1.1251	1.1152	0.88%
JMARS (A=12; K=5)	1.1244	**1.1150**	0.84%

TABLE I: Performance of various models on Amazon Clothing data

Instant Video Data	without time	with time	improvement
Baseline	1.1269	1.1170	0.88%
JMARS (A=6; K=5)	1.0945	**1.0843**	0.93%

TABLE II: Performance of various models on Amazon Instant Video data

The baseline model in the above table does not consider the language models into account and just optimizes the latent factors. That is, we do not perform any Gibbs sampling step and it could be considered as a simple latent factor model. We also tried to run HFT [2] by obtaining code from author's website but we could not get the method to converge. The minimum MSE for clothing dataset we got from HFT is 1.23, but since we couldn't get it to converge, we don't claim the correctness of this result.

From the results, we notice that when we change number of aspects from 6 to 12, there's not much improvement. However, when we add parameters corresponding to temporal dynamics into the model, we get an improvement of around 0.9% even though the increase in the number of parameters is far less as compared to increasing the number of aspects. This proves that the improvement is just not because of being able to learn more things by increasing the parameters. Having tested this hypothesis on the clothing data, we do not perform any experiment for A=12 on video data given the computational cost and time constraint.

The reason for the improvement of our extension could be attributed to the fact the model is somewhat able to learn temporal pattern in user reviews. However, we think that the improvement is not significant because of two reasons:

- The data we use is relatively small with few reviews for each user so the temporal dynamics might not learned properly.
- The time function we use is a linear function, which might not work the best in capturing the temporal dynamics across different aspects. Something like binning might work better for our use case however binning would substantially increase the number of parameters in the model.

VI. Qualitative Results

We also run several experiments to analyze what the model is learning. The model seems to be effective in learning and segregating different words in different language models. Here we just provide qualitative results from the clothing data. Results from Instant Video dataset are of similar nature. All the results provided are for when the number of aspects were 6 and when we performed only 50 iterations of Gibbs EM (as opposed to 500 iterations performed in the original paper) on our proposed extension.

A. Language Models

1) Background Words: Below are some of the background words we obtained from the reviews:

Price	Product	Picture
Size	Fit	Wear
Quality	Purchase	Material

As we can see, these words are some of the most common words in clothing item review.

2) Aspects: The model learns 6 different aspects from the item reviews. Below are two of the most relevant aspects that we could discern clearly:

Aspect 1: Material/Color

Color	Material	Elastic
Light	Care	Weather

Aspect 2: Size/Fit

Tight	Wear	8/10 (shoe sizes)
Inch	Comfort	Chest

Other aspects were not very clear which we believe could be because of the limited number reviews we considered and being able to perform only 50 iterations.

3) Aspect Sentiments: Following are the respective aspect sentiment words corresponding to the above mentioned aspects:

Aspect 1: Material/Color

Great	Design	Soft
Quality	Durable	Cheap

Aspect 2: Size/Fit

Doesn't/Don't	Shrink	True
Thick	Small	Loose

As we can notice, most of these words communicate sentiment only in the context of the identified aspects.

4) General Sentiments: Below, we mention some of the most relevant top positive sentiment words that the model learned:

Like	Comfort	Nice
Well	Love	Buy
Good	Great	Pretty

Additionally, following are some of the most relevant top negative sentiment words that the model learned:

Problem	Waste	Flaw
Review	Nothing	Worst

As we can observe that the model was quite successful in identifying the general sentiment words but don't really relate to any specific aspect.

5) Item Specific: Following are few words that we discovered from the item specific topic in the language model.

Item 1:

Shoe	Clarks	Merrell	Timberland

Item 2:

Necklace	Jewellery	Vintage	Pendant

Item 3:

Bag	Compartment	Pocket	Purse

B. *Effect of Temporal Dynamics*

Introduction of temporal dynamics in the model allowed us to learn temporal changes in user preferences. Following is one of the examples where we can observe two reviews, that are 10 months apart, from a user focusing on different aspects of an item:

Date: 06/11/2013

My hubby is hard on his shoes, so I like to find him good ones at a **reduced price**, such as these. he likes the **fit and feel** of New Balance, so these will be his next pair when his current ones are too tattered to wear anymore. Good **grippy sole** for our rocky western trails, and decent laces that shouldn't break with his hard use.

Date: 04/02/2014

Thanks to another reviewer I got the **green** ones instead of the **raspberry**. The **green insoles** have just the right arch support for my plantar fasciitis-ridden feet. I am glad to have them in my everyday Merrell slip on shoes. These insoles are **not too soft, but soft enough**, and after just one day of wear I don't notice them at all, which is perfect. Based on the last pair (I had the **raspberry**) I expect about a year from these, but will happily accept a longer wear time from them.

Both the reviews are for a (different) pair of shoes. We can clearly observe that in the first review the user is more concerned about the fit, comfort, utility of the shoes. However, the second review is more focused on the color and feel of the material and doesn't really talk about the fit or the utility of the shoes. We could observe similar patterns for users with many reviews spread across time and had high values of $\alpha_{\theta u}$'s. However, the learning was noisy for users with less reviews or reviews with small variance across time. We believe this problem could be resolved if we consider a large dataset and prune it such that each user has more than specified number of reviews.

VII. Challenges

A. *Non-triviality of extensions*

The JMARS consists of many small component models. Adding or changing one of the models can lead to changing the whole premise of the problem and since the model is generative, we would have to consider the impact of the change on all the other components. Adding temporal dynamics was relatively not that hard but we also tried to look into various hierarchical modelling techniques to incorporate hierarchical nature of discovered aspects in our model but adding hierarchy in an unsupervised fashion in this model was a challenge.

B. Implementation

We use the code from [4], but the provided code was not by the official authors and was incomplete, inefficient and incorrect at many places. We ended up using only around 20% of the provided code and spent most of our time implementing the original method correctly. Given so many small components in the model, debugging the original code and re-implementing the model correctly was very time consuming for us.

C. Computational complexity

After we implemented the original method, we found that performing inference was very slow in JMARS mostly because lots of computation needed to be done for each word we had in the reviews. This restricted us from experimenting extensively on large datasets and showing more refined results.

VIII. Future Work

We obtain very interesting insights when we add the temporal dynamics to the model. We believe that the results could be further improved for larger datasets with more reviews per user as it would be easier to learn the temporal variables. We would also want to try other functions for capturing time like binning. Another area of improvement, as mentioned in the midterm presentation, is to add hierarchical structuring to the language models. This could really help in improving how the aspects are being learned. For example, an aspect ‘material’ could have sub-aspects as ‘colors’, ‘feel’, ‘washable’ etc.

References

[1] Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). Diao, Qiming and Qiu, Minghui and Wu, Chao-Yuan and Smola, Alexander J and Jiang, Jing and Wang, Chong. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014. 193-202.

[2] Julian, McAuley, and Jure Leskovec. "Hidden factors and hidden topics: understanding rating dimensions with review text." Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013.

[3] Koren, Yehuda. "Collaborative filtering with temporal dynamics." Communications of the ACM 53.4 (2010): 89-97.

[4] Balani, Nihal and Kumar, Revant. "JMARS", GitHub repository, https://github.com/nihalb/JMARS

[5] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In W. Cohen, A. McCallum, and S. Roweis, editors, International Conference on Machine Learning, volume 307, pages 880–887. ACM, 2008.

[6] Misra, Rishabh and Bansal, Tushar. "t-jmars" Github repository, https://github.com/bansaltushar92/t-jmars

[7] R. M. Bell and Y. Koren. Lessons from the Netflix prize challenge. SIGKDD Explorations, 9(2):75–79, 2007.

[8] A. Lazaridou, I. Titov, and C. Sporleder. A bayesian model for joint unsupervised induction of sentiment, aspect and discourse representations. In Annual Meeting of the Association for Computational Linguistics, pages 1630–1639, 2013

[9] J. J. McAuley, J. Leskovec, and D. Jurafsky. Learning attitudes and attributes from multi-aspect reviews. In International Conference on Data Mining, pages 1020–1025, 2012.