The questions are ordered as per the paper order by class and the paper presentation schedule in class. Scroll down for Week 3 Wednesday papers.

Week 3 Monday Papers:

Paper 1: “Time Weight Collaborative Filtering (Ding and Li, 2005)”

1) Why does it use MAE not MSE? MSE is better to get the solution faster.
2) What advantage is the MAE better than MSE?
3) It use K-means to cluster with the rating information , is any other better approach?
4) It use decay factor to model it, does it mean the model cannot catch the seasonal effect? E.g. in winter, people would buy A, in summer people would by B.
By Kaung Hsuan Lee

1. The clusters are defined by the rating information which could be highly inaccurate, Would it be advantageous of having clusters which are personalized to a user / or based on item similarity / user assigned or learnt tags ?
2. In the particular dataset for movies, would it be more advantageous to use some level of timing information / preferences from similar users - which could lead to better recommendations (e.g recently released movies of same genre / actor) ?
3. In general items also evolve over time (which is not the case for this particular dataset) - which could lead to change in preferences. For eg. manufactures claim them keep improving their products (for e.g. toothpaste) Could this be factored in somehow ?
4. Although this could be out of the scope of this paper - the number of clusters needs to be known a-priori for K-Means. In a hypothetical situation where we could track item evolution - this could be a drawback

By Balasubramaniam Srinivasan

1. **Time function:** Exponential function decays rapidly near x=0 than logistic function. There would be a high interest of a user in an item when the item is new and we can expect the interest to remain high for some time instead of decaying rapidly. Logistic function captures this behavior. They did not report any experiments to validate that exponential is better than logistic.
2. **clustering:** K-means doesn't account for temporal effects in items' ratings. An item rating may change over time. Also, clustering algorithm uses only item rating for clustering and item-to-item similarity like cosine similarity are applied later to find `n` nearest neighbors. How accurate would be the clustering based on item's rating? or How about applying K-means on Metric Embedded space to find clusters?

3. **Boredom:** They try to capture user getting bored of a group of items as time goes by. This means similar products (clustered as one centroid) get the same value of T0. This model wouldn't capture per item boredom. Suppose of I got bored of a thriller movie, this doesn't necessarily mean I got bored of all similar movies or all thriller movies.
- By Kulshreshth Dhiman

1. **MAE Calculation and T0 estimation**: When calculating the MAE, the equation uses sum(i) pi-qi/N, here pi is mentioned as the predicted rating, however qi is mentioned as the user true rating. The predicted rating is the rating which is predicted using the decay function and is the rating of the item with respect to the current date. However, I do not understand what is the time frame considered for the user true rating. For e.g, if the user true rating is 10 years old, how is it relevant in calculation of MAE, and estimating T0.

2. **Simiarity Function**: While computing similarity between items using cosine similarity or pearson correlation, we don't take into account the time when the item was rated. How can we encompass time parameter in the item feature vector and then compute similarity?

3. **Conditional Probability-Based Similarity**: In computing the conditional probability based similarity between different items, the formulae estimates it by MLE using freq(i,j) / freq(i). If freq(i,j) is the frequency of items i and j bought together. There is no notion of bought together mentioned, bought together on the same cart? or bought together in history of purchase item?. If the latter, I fail to understand how it captures similarity.

- By Digvijay Karamchandani

1. **Rate decays for different items in same cluster:** In the paper, authors perform K-means clustering on the items using user ratings. The complete cluster of items then uses the same value of rate of decay or T0. But there can be multiple cases when the rate of decay is different for the items with similar ratings?

2. **Using Beta function to model Temporal effects of time:** In the paper, the authors experimented with exponential and logistic functions with the assumption that the importance of the reviews would decrease over time. But there is a possibility that the importance of the ratings can increase and then decrease.

Hence, beta functions might be able to better model the temporal effects of time on the preference prediction.

3. **T0 calculation**: In the paper, the authors have calculated the range of T0 by brute force and then performed a grid search over the actual value of T0. I think we could also try calculating T0 by minimizing the MAE using gradient descent or other methods.

- By Kriti Aggarwal

1. Instead of fixing the time function as exponential or logistic, wouldnt it be better to make the system learn the function itself instead of learning the T0 value. If a model does that, then it would automatically capture the personalized opinions of the user over time and would eliminate the use of K-means clustering inorder to capture the lack of interest of user on a movie.

2. The model focuses on the similarity between the items and doesnt actually capture the user characteristics. To predict an the opinion of user(i) on item(j) it just uses the similarity of all the items that have been purchased by user(i) previously and does not take into consideration all the other users who have bought item(j). I really dont see how this model can be called as a collaborative filtering model. Maybe I have missed some points in the paper.

3. The paper deals this problem by trying out values of T0 which is not a very good solution for a learning problem. I am not able to find how the system is actually learning any parameter. If they are trying to learn T0, wouldn't minimizing the error between predicted and target ratings be enough? Since we have a time function that's supposed to take care of the model's temporal fluctuations. What is the need of K-Means in this problem?

By Vignesh Gokul

1. **The model clusters the items and then learns temporal trend for that cluster. But say for example, a user liked thriller movies, and then rated a couple of thriller movies poorly (Does that necessarily mean that the user does not like thriller anymore? Could it be that these 2 movies were just poorly rated by majority of the users? If that's the case then wouldn't it be better to take into account global rating of movies in deciding whether user has stopped liking thriller or is the dislike just for a particular movie)**

- **Rahul Dubey**

Paper 2: Collaborative Filtering with Temporal Dynamics (Koren,2010)

1. **Offline vs Online Model:** How can models like *timeSVD* be updated with the streaming of new ratings? Do we periodically bring the database offline and update the bias distributions? Considering the accuracy of these distributions, could we safely assume that these distributions won't change for earlier time windows and perhaps only the most recent time windows need be updated? Could we model this online, iteratively and efficiently? (Kiran Kannar)

2. **User bias Modelling:** The user bias is modelled as b_u(t) = b_u + a_u * dev_u(t) where dev_u(t) is signed time distance between dates *t* and *t_u,* the mean date of ratings by u.
    a. Is this interpretation valid? The dev_u(t) seems to be some decay/relevance mechanism over the average rating date, which doesn't seem intuitive. The mean date of ratings can be updated in a streaming model efficiently (constant time). However, a single new rating forces the updation of every bias term, which is counter-intuitive. Periodic offline updations could reduce the cost, but still would need to update each bias term across all time periods. (Kiran Kannar)
    b. Could we do better? One method to do so is to consider a monotonic increase in user knowledge about the item, and therefore model the evolution of expertise (and avoid bias modelling across time) See *From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise through Online Reviews* (Kiran Kannar)

3. **Sudden drifts/ "Spike" Modelling:** In the offline model of Netflix movie rating data, the average number of times a user rated was 40 different days. This is a property of the dataset/domain. In a dataset of restaurant ratings, could we still assume such a small number of different days of rating on average? If not, the extra parameter *b_u,t* rises linearly with time and is no longer a constant number. However, the number of spikes would still be constant. *b_u,t* would no longer be an accurate distribution. Could we model these spikes differently? Could you point to any existing research that has worked on this? (Kiran Kannar)

4. Fixing the temporal item parameters as bins does help in simplifying the problem. It does make sense for movies but can this be generalized to other datasets like amazon where the changes might not be that gradual? We might want to capture a finer resolution for items as well. (Tushar Bansal)
5. Future values of parameters: Since we cannot predict the suture values of parameters c_u(t_ui), b_{u,t_ui} to c_u and 0. However, setting c_u(t_ui) to c_u will only capture the average scaling factor for a user. Cases with abrupt changes (like shown in the beginning of paper) won't be fully incorporated in the future ratings. Can we take a better estimate here like setting them to the latest 't' value? (Tushar Bansal)
6. For time changing baseline parameters- bu(t) is measured by the time distance between the current date and when the user rated the item which can capture the short lived temporal effects

but what if the user's activity changes very gradually or we have data very few points? (Akanksha Grover)

7. I was just curious about the effect of b_u,t term on the RMSE(reduced to 0.9605) - between linear and linear+ models. Are the daily user effects (user mood, multi-user accounts etc.) more significant on this dataset or is it in general? (Akanksha Grover)
8. In Temporal Dynamics for Neighborhood models, for item-item modeling, what is meant by using two sets of item-item weights -wij and cij, one related to rating values and other disregards rating values, considering only which items were rated. Also what is the intuition behind the exponential decay function (for the user) for finding accurate item-item weights ? (Akanksha Grover)

1) It uses gradual concept effect, whether it will not work for the shopping event for Christmas shopping season. E.g. people will only buy a lot of things related to Christmas before or on Christmas Day, but they will not buy these things after Christmas, It seems it is not graudal concept, there is some jump point, how to modify the model to fit it or does this model already handle this situation?
2) If we would like to add trust friend or untrust foe in the model, how to add them in this model?
   By Kuang Hsuan Lee

1. The paper says that modeling user bias on day-of-the-week basis is not useful. Although, I feel in the context of Netflix, one could see difference in watching trends if data was aggregated for weekday vs weekend. That could have been considered as a factor while creating temporal user bias vectors.
- Rahul Dubey

Paper 3 : Latent Factor Transition for Dynamic Collaborative Filtering (Zhang et al., 2014)

1 - In the paper, transition matrix is modeled to capture time-invariant patterns. To quote an example given in the paper "if user 'i' increasingly prefers the movies directed by James Cameron over time, the entry (j, j) in B_i will have a value larger than 1, assuming that the jth latent factor corresponds to James Cameron". That is, they learn the uni-directional preferences (positive or negative) from the data which do not change

with time. Isn't that a wrong way to model temporal dynamics given that our aim for modeling the time is to capture how the preferences fluctuate from time to time?

For example, let's assume initially I liked James Cameron (let's call this period T1) because he directed good movies but then I started hating him for some reason (period T2). Since, my interest towards James Cameron is time dependent, transition matrix would not learn this. Maybe, this is the reason this model is not able to perform significantly better than timeSVD.

2 - Looking at the results, it seems that method works well in relatively Sparse dataset than dense dataset. What could be the reason for that?

3 - Authors say that they are proposing a temporal model, however, it seems more like a sequential model as the preferences are found by sequentially transforming matrix B. Is it really temporal? Also, could linear interpolation for updating B work better (i.e. a*(current_update) + (1-a)*(previous_value))?

By Rishabh Misra

1) The proposed time-independent transition matrix between items is specific to each user, but globally calculated across time. This seems to only be able to capture monotonic change in preferences toward a latent factor. Is there any reason other than to avoid overfitting that the model only considers a monotonic preference change?
2) This model seems to suffer from an extreme case of cold start problem because of the need to build a personalized transition matrix for each user, which is similar to the transition cube in the FPMC model. The FPMC model overcomes this by factorizing the transition cube and using BPR. The proposed model does the first step, but not the second. Does BPR not perform well when applied to a model that uses explicit ratings?

By Siddharth Dinesh

1) How do modify the model to catch the seasonal effect?
2) If there are the noise in the data very earlier, whether the noise will keep affecting the later status, which means it will propagation the noise?

3) If some user will have the rating jump situation(e.g. Yesterday the user is happy, he gives each one high rating, today he is sad and he gives each one low rating), does the model regards it as the noise or catch this effect?

By Kuang Hsuan Lee

1. The transitions are based on discrete time windows and not on gradual change. What are the pros and cons of that? Won't we lose transition information if the windows are created at wrong times, as oppose to having decaying function?

-Rahul Dubey

Paper 4: "Who, What, When, and Where: Multi-Dimensional Collaborative Recommendations Using Tensor Factorization on Sparse User-Generated Data (Bhargava et al., 2015)"

1. Authors use an algorithm to infer the kind of activities reflect from the photos, how would the accuracy of these inferred activities affect the result?
2. Since the tensor created is 99.999% sparse, does that mean the predicting results rely more on the 4 2-dimensional matrices?
3. The value in the tensor's cell represents the frequency of the current user being at the current location performing the current activity at the current time, it is normalized based on the total number of photographs for each user, would it be better to be normalized based on total number of activities(combine photos from same location, same time, same activity)
4. Time related data is only extracted from Flickr and make the tensor, would it be possible to add time dimension in Location x Activity matrix and Activity x Activity matrix?

By Yiwen Gong

1. It may be hard for the factorization model to generate latent representations of time bins that have never or appeared in the training data. Can we do anything to dynamically choose bin size? it is not clear how to choose the epoch length parameter due to the asynchronous nature of the user-item interactions/time normalization between users.
2. There does not exist a coupled matrix including the time dimension - could it help to tie in a coupled activity x time matrix?
3. As a tensor factorization model, it may have difficulty modeling surrounding sequential contexts in a particular feature dimension e.g., the previous and successive locations of

a check-in location, which might be important. Is there a way to implicitly impose this kind of sequence learning in the factorization objective?

By Chester Holtz

Paper 5: "TribeFlow: Mining & Predicting User Trajectories (Figueiredo et al., 2016)"

1) It used semi-markov chain model, what is the advantage to use model it? And could we model it as markov chain? If not, what is the condition make us cannot use markov chain? If yes, what is the disadvantage to use markov chain instead semi-markov chain?
2) It transfers non-stationary to stationary. Is it possible for this method to transfer the stock price market because the stock price market is also non-stationary and in this paper, we have user, item, time; the stock price market we also have company, price, time. Is it possible?
3) If we would like to add the effect by trust friend, how to model that with the current framework?

By Kaung Hsuan Lee

Week 3 Wednesday papers

Paper 2: "On the Temporal Dynamics of Opinion Spamming – Case Studies on Yelp (Kc and Mukherjee, 2016)"

1) It use VAR model, if the time series is nonstationary, could we use VAR model also? Or change to use other model?
2) Does it mean that if the new owner of restaurant knows this paper, he will generate the fake review from the first day, because if there are no many reviews, the first reviews will affects users most, if the owner do this way, whether it means this model will not work?
3) How do yelp know the review is fake or truth? Only the writer of the fake review knows that, right?

By Kaung Hsuan Lee

- It was unclear to me how the fake/spam reviews of the Chicago restaurants in the Yelp dataset were revealed to be fake/spam. Is there any other information available to reveal how the fake reviews were identified, and confirm the accuracy of the dataset?
- Regarding section 4.1 on early spamming: how did the dataset owners or authors confirm it was the restaurants themselves that did the spamming?
- Regarding Table 4 Non-text features: what are the funny and cool counts in the reviews?
- From Stephanie Chen

Paper 3: Personalized Itinerary Recommendation with Queuing Time Awareness (Lim et al., 2017)

1 - The proposed algorithm might not work well if the data is sparse or if we have to recommend new items, because the estimates of popularity, interest and queuing time are the average over the available data. Could this be improved somehow, e.g. by incorporating the context information about item and learning parameters over them to decide how much importance they should be given?

2 - Also, could the method be improved by learning latent parameters for the user preference, place popularity and queuing time from the data? This way we could decide which aspect drives the recommendation of the user. Eg. Some people might not prefer to wait too much irrespective of place popularity of interest or vice versa.

3 - Also, if this method was used by many people, it'll recommend same itinerary which might defeat the purpose of taking queueing time into account as everybody would try to go the same recommended place and create crowd. Could recent recommendation be taken into account somehow so that following recommendations are different and diverse?

By Rishabh Misra

1. The authors use geotagged photos to calculate interest and relevancy. This method is prone to inaccuracies due to problems in measuring location accurately. Wouldn't it be easier to calculate relevancy in terms of the number of times a user has visited attractions of a certain category, relative to his/her total attraction visits (as in one of the related works)? Are there any advantages in using the first method over the alternative?
2. The problem is formulated to propose an itinerary based on a time budget and starting and ending points. The authors do not discuss or prove the existence of a solution. Is a solution guaranteed to exist? If not, can the algorithm be altered to provide a 'closest' solution of some sort?
3. The queueing times are calculated using historic data. If multiple people use the algorithm to form an itinerary with the same starting and ending points, it is likely that the itineraries for many may be similar. This would lead to queueing times that are significantly higher than the estimates. Can global or dynamic factors be accounted for in some manner? Would the problem have to be reformulated or can the algorithm be modified to accommodate for this?

By Shreyas Udupa

How can a similar approach described in the paper (like the PersQ algorithm) be adopted for attractions that are not in theme parks, like museums, beaches or places in a city that are further apart.(say when a user plans a trip to city). Can we really expand the problem to other use cases? What will be some problems that we would encounter? (Eg. travelling times and distances may vary a lot between location)

Also in case of user-attraction interest, it may also be possible that the user takes more photos of an attraction if he/she seldom goes there, which may not be an indicator of the interest level of the user, or it is an attraction where you cannot click pictures.

By Akanksha Grover

Paper : "Optimizing the Recency-Relevancy Trade-off in Online News Recommendations (Chakraborty et al., 2017)"

1. Why does he use OLS(ordinary least squares)? Because the data is time series, which means the current item is highly correlated to the previous item. Also OLS must

have Homoscedasticity assumption, but it seems the data is not this case. As time goes, the variance will decrease. Is any other better model to fit it?
By Kaung Hsuan Lee

2. A crucial time period is the time gap between *time of birth* and *time of discovery/first click*? The longer the time period, one may never see the article. However, depending on the virality aspect of the article, once there is a discovery, the article may have a large future impact. Does the paper consider this? If not, how can we incorporate this time gap and its uncertainty more effectively? (Kiran Kannar)

3. How do you choose the value Δt, i.e. the time interval for recent-impact methods? The paper mentions that Twitter choose 15 minutes, NYT a day etc. How arbitrary is this? Is the arbitrariness based simply on "it works, but no idea how, and don't care"? (Kiran Kannar)

4. In the normative argument, isn't the recommendation question of WHOM to recommend more important than WHAT to recommend? Identifying *opinion leaders* who can be knowledge *hubs* clearly seems to be an important task (similar to influential nodes in the graph) (Kiran Kannar)

5. Could you talk more about the underlying layers of virality in predicting the lifetime impact? The authors do mention some existing work where the virality is considered coarsely. It looks like they completely ignore this. (Kiran Kannar)

6. We are comparing performance in two different settings - one where we use implicit feedback (in terms of number of clicks) and the other with explicit feedback (in terms of number of retweets). Wouldn't it better to use separate strategies for the two different tasks? Additionally the datasets also appear to be from 3 different time frames (Although the author's model demonstrates good performance on all the three datasets) - could we have another parameter which captures new reader evolution trends? (Balasubramaniam Srinivasan)
7. Based on Q4 above, would it be in fact better to consider this problem as 'expanding a bipartite graph' and performing link prediction over a given time frame / 'future dense subgraph identification' (in an attempt to understand the global user base better, as the set of influential nodes might not remain the same in a time evolving graph)? Using graphical models though suffers from a limitation - where we must know the maximum possible users (user nodes) - but again we can account for growth in user base in other ways. (Balasubramaniam Srinivasan)
8. This may be out of the scope of the paper - but articles get updated quite frequently these days (the link to the article remains the same though)- although the author explicitly states that he doesn't have all the meta-information for all the three datasets - would a model incorporating this effect perform better ? Does this mean we update the 'time of birth' of the article / other parameters? (Balasubramaniam Srinivasan)

9.

Paper 5: Application: User Session Identification Based on Strong Regularities in Inter-activity Time (Halfaker et al., 2015)

1. Do you think that it is a valid assumption to have a single global session threshold for all users? The paper makes a case for a global threshold but doesn't compare the approach with any non-global threshold approach. Can we have user/segment specific thresholds obtained by same method for users/segments with enough data points? (Tushar Bansal)
2. How good do you think are the results of this paper? The paper doesn't perform any quantitative evaluation of the proposed method. Do we have any information about the state-of-art and how this method performs in comparison? (Tushar Bansal)
3. The approach only uses the distribution of data-points to learn the model. Can we use some other features (like user actions in the session etc.) to improve this model and get better clusters? (Tushar Bansal)

4. The authors do not seem to distinguish between different categories of users on StackOverflow - user, admin, moderator and contributor. A lot of users simply have short sessions of bursts of *viewing*. Contributors (question/answer) could be longer. This granularity is missing and assumptions that a user rarely posts more than one answer. (Kiran Kannar)

5. While I like the idea that writing questions and answers on StackOverflow generally take longer time,
    a. This should not be a validation to find a cutoff valley point.
    b. A lot of user experience on SO is marred by rigid rules and harsh moderation due to which users spend an inordinate amount of time crafting their content. New users feel especially unwelcome when joining Stack Overflow. See https://hackernoon.com/the-decline-of-stack-overflow-7cb69faa575d and https://meta.stackoverflow.com/questions/251758/why-is-stack-overflow-so-negative-of-late
    (Kiran Kannar)

6. A lot of the paper covers observations and conjectures on observations that don't have statistical guarantees confirmed by quantitative measures (as asked by Tushar above). It looks like the authors try to fit their results to their predefined notions on Activity Theory. See last page "While it's hard to say conclusively, we suspect that the "short_within" ……" I believe this reduces the quality of the contribution of the paper. (Kiran Kannar)

7. Does a global activity threshold work for other games which have longer *within* sessions and perhaps longer *between* sessions times? The generality seems limiting. (Kiran Kannar)
8. The authors seem to distinguish users without ID by using IP and user-agent, it might be inaccurate especially for public network or mobile users. (Yiwen Gong)
9. I also have a doubt on how the webviews are determined for Wikimedia. Was the event only triggered by loading a new part of the page? If so, it will be a problem if user scrolls up. Plus, if a user is not scrolling the webpage, it does not mean he/she is not looking at it. (Yiwen Gong)
10. I wonder will the user session of some social networking apps differentiate from the 1 hour rule? I think it would be better for the authors to identify a particular system feature which complies their observation.(Yiwen Gong)