Questions for <paper>:

"Factorizing Personalized Markov Chains for Next-Basket Recommendation (Rendle et al., 2010)"

About the lemma1 : it said x is invariant.

So does it mean the preference of the given user is constant for time?

Also, if the preference is changed by season, maybe for winter or summer the preference is different, how to modify the formula?

It perform stochastic gradient descent on these bootstrap samples. What kind of situation it will not converge or it must converge? And this approach is better in any situation than stochastic gradient descent on all data? Does it mean we can do it in neural network in any application to speed up?

By Kuang Hsuan Lee

"Personalized Ranking Metric Embedding for Next New POI Recommendation (Feng et al., 2015)"

It used norm-2 distance, how about other norm? And what kind of the situations are better for norm-2 or norm-1?

I think the (D P u,l) and (D S l c,l) already contain some information for the geographical influence, for example, if the locations are far away, it already affect the relationship between two locations and the relationship between the location and the user. Besides, why the geographical influence makes same effect on both?

By Kuang Hsuan Lee

"Translation-based Recommendation (He et al., 2017)"

Why does it use AUC measure the performance, what is the advantage?

It used the 'average' behavior—to make predictions for these users. Is it possible to use different 'average' behavior for the different gender or different locations? If we use too many 'average' behavior, what is the disadvantage?

It maps into latent space, is it possible to combine some information into the latent space? If we want to add location or gender or other information. How to do that?

By Kuang Hsuan Lee

"Playlist Prediction via Metric Embedding (Chen et al., 2012)"

Is any example to explain the The "entry vector"U(s) and the "exit vector" V (s)?

Why regularize (s,s) ?

It used Euclidean metric, and it means it assume it is symmetric? But in the real world, it would not. How to change it?
It makes assumption: user preferences independent of the sequence context. Is that assumption reasonable? If not, could we add covariance matrix or is any other approach to solve this situation?
By Kuang Hsuan Lee

Questions for "[Modeling User Consumption Sequences (Benson et al., 2016) | pdf](#)"

In this paper, as far as I understand, authors have trained their model by finding parameters which increase the likelihood of the data and evaluated their model by comparing their model simulation with the data or what % of likelihood is captured by the model. Wouldn't this be seen as overfitting the data given that they have not tested the performance of their model on unseen data?

They have also made several assumptions which increase the likelihood of the data and are prone to overfitting.

1. (Page 5, right column, first para) "From analyzing the data, we found that B = 20 is appropriate for the music and video data sets, and B = 60 is appropriate the clicks and check-in data. After making this assumption, the number of items per session follows a power law with exponential cutoff with maximum value B, the between session times follow a power law with minimum value B" -> They seem to be choosing B which fit their assumptions regarding the distribution followed by session length, intra-session gap, and inter-session gap.
2. (Page 6, left column second para) - The probability of selecting a particular novel item is a simple maximum likelihood estimate. Wouldn't that be problematic when data is sparse?
3. (Section 6, first para) - "We restrict ourselves to just personalizing the recency weights w because they are the largest component of the model (both in terms of number of parameters and effect on likelihood)." -> Again this is done by observing likelihood of the data.
4. Also, Authors say their model is generative but didn't mention what priors they have used. Also, objective function posed just maximizes the likelihood. Isn't it a discriminative model?

By Rishabh Misra