Personalized Key Frame Recommendation (Chen et al., 2017) | pdf

1. In this paper we model the interaction between the image content and latent factors linearly. However, various literature in this course have suggested trying modeling multi modal interactions non linearly (similar to the paper on early alzheimer's detection in the 7th week).
2. Attention networks would be a good way to model which words in the review to pay attention the most in the LSTM architecture. This would help in learning the embeddings of the user and item better.
3. Ground truth is based on the sentiment of the opinion using the Stanford NLTK toolkit. So the benchmark seems somewhat ambiguous for this reason. For e.g, we tend to use slangs/sarcasm on these sites, the opinions of such cases may not be captured by the toolkit.

Digvijay Karamchandani

1. Word2Vec could be used as input to the LSTMs. This would help in reducing the complexity of the model. Does learning the semantic embeddings help in improve the model in any way?
2. The image features having been directly used from the Imagenet pretrained model. Wouldn't it be a better idea to add fully connected layers for the features to be fine-tuned for our use case.
3. The user mentions about the long tail issue, i.e, many users have <10 comments overall. The paper does not mention how they try to resolve the issue. With such sparse data, does modeling comments to determine user representation actually help?

Sudhanshu Bahety

1. Does this model make a compromise on the context of the video and the user preference? Basing the keyframes on the sentiments of the TSC is vague because a particular user might have liked that keyframe because of the previous context for which he might have had a negative rating.
2. The CNN model is a pretrained Imagenet model which is very biased towards dogs and fishes. Wouldn't it be better if the model is trained from scratch?
3. It would have been better if there had been an additional subjective evaluation as just seeing if the predicted keyframe (without any context) is liked by an user or not is not a measure of how good this model is. (By Vignesh Gokul)

1. The assumption for this paper is that a user's interested frames are those he/she leaves comment on. But in realistic scenarios, the number of a user's

commented frames are always less than the interested frames. Do you think this could be a major problem for the base of this work.

2. A large number of user's comments are typed when watching the video. So there will be a gap between the time the interested frame is shown and the time a user publishes the comment. There also will be delays due to internet problems. If during this time, the scene of the video changes dramatically, then the extracted visual features of the frame would be not accurate. How will this problem be dealt?
3. The author of this paper used element wise multiplication in combining frame latent factor and frame image feature. I think this is not convincing because element wise multiplication could erase important feature information of both the latent factor and the image feature. And also the latent factor represents different information from the image feature. Is there any method to do this more elegantly? For example, using fc layers before combining them?

-Siyu Jiang

1 - Authors divide each movie into L frames irrespective of the movie length and did not describe how they chose these frames. First, this would result in missing some potentially good frames and second, long movies be at more disadvantage. To mitigate this, I think selecting the number of key frames dynamically based on length of the movie would be more beneficial. What do you think?
2 - Authors model the text likelihood using RNN for each user and key frame in their training data. This data also contains negative samples (those where user sentiment is negative and those where user has not commented). Based on this, isn't the equation 8 wrong, given that it also models the text likelihood of the comments that are not available? If it is, how could we handle the cases where comments are not available?
3 - Do you think the model is scalable given that there are lots of components are parameters that are to be learn? Also, how could cold start problem be addressed?

- By Rishabh Misra

1. The approach described by the authors relies on some previous methods to generate L key frames from preselected L movie shots. Their model then, ranks these frames and presents them to a user. But given that we now have TSC data for movies, can we devise a new method to first select candidate key frames for a given movie by leveraging this TSC Data for that movie (for example, selecting most commented frames, etc)? How would such an approach affect our results?

2. The authors consider polarity of comments in their model. Would it be a good idea to, in addition, use topic detection based approach and use detected topic information in the comments as signals to improve the recommendation of key frames for a given user? What changes would be required in the model to achieve that?

3. The authors use a model trained on ImageNet dataset to extract image features. Given that we have the TSC data, Is it possible to use this TSC data (comments on frames/images) in addition to, or instead of, ImageNet for this purpose (to get the image features relevant in context of movies)? What effect will such an approach have on the quality of the image features and further, the overall results?

By Nitin Kalra

1. This model only select some shots of the video to represents the content in this section and use a CNN model to process it. Could this part be improved by using some model facing the video processing (like T-CNN)?
2. This model is highly depend on time. The TSC and video shot are both related to some specific time interval. I have general questions that how can we take the advantage of time for other recommendation tasks.
3. The user vector interferes every steps in the preference-aware LSTM and the structure gives a result for every word. Could this part can be better using some user personalized text summarization mode?

- Wen Liang

1. They use element-wise dot product of frame latent and image latent. Wouldn't it be better if we take dot product of image latent and user latent vector, since that can capture user-image interaction.
2. Why a sentence/comment generation model is used here? This model essentially trying to capture sentiments of a comment. Would a sentence classification model be a better fit here?
3. How are the comments available during the testing? They wanted to predict key frames for an unseen movie for a user. This means user never commented on the test video.
4. For CNN model to extract visual features, would it make sense to use CNN output layer instead of deep layer of size 4096 as this would be biased to object classification task for ImageNet data.

- Kulshreshth Dhiman

1. How representative are the frames commented on by users (or time-synchronized comments as expressed in the paper) of the actual importance of the frames? Is there some way to tell if different frames would be useful in different settings, e.g. which frame is more likely to get a user to click on a video vs. comment on it?
2. Are some of the baselines discussed in the paper at an inherent disadvantage, because these models are not intended to model video features? Are there baseline algorithms implemented on similar datasets that would be a better fit for this task?
3. Currently a positive example occurs when a user comments on a particular frame of the video. Are there other ways to collect data about important frames in the video that users may not have commented on? For example, would the thumbnails of videos the user is interested in also be valid training data?

- Rajiv Pasricha

An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace (Chen et al., 2016)

1. Users may prefer to go for less reputed retailers if items are cheap and prefer to go for reputed sellers / amazon fulfilled sellers when items are expensive. Is that taken into account here?
   (Rahul Dubey)

1. The machine learning method used in this paper is very naive, only RF is used. What other methods could be applicable to this problem?
2. It seems that while the authors were constructing the feature vector, they missed an important one which is delivery time, these info are available from the 'New Offers' page, but authors claimed that they could not measure the shipping time.
3. Do you think the second from lowest price is mostly useless? Would it be better to have a target price which is from the buy box winner?

(Yiwen Gong)

1) The authors don't have ground truth data if the seller is really employing algorithmic pricing they just make an intelligent guess which could be wrong in the first place. How much do you think this affects their observations ? **(Sai Chaitanya Kolasani)**

2) For sellers using lowest price match strategy is it possible by other arbitrary users to misuse this system ? A user can create a bogus seller account and set a very low price which the algorithm automatically matches. **(Sai Chaitanya Kolasani)**
3) What is the effect of item category on the buy box algorithm ? For low cost items users may want the lowest price even if the seller is not reputable. So do you think adding item category as one more feature improves the accuracy ? **(Sai Chaitanya Kolasani)**

Sai Chaitanya Kolasani

1. I think the difference between algo and non-algo sellers can be affected by some other properties. How could we exclude those effects?
2. The paper only discussed about the changes frequencies, rank, number of feedback, Buy Box winning and other outcomes. The analysis is highly limited. Do you think they are good evidence to conclude that algo pricing has great power?
3. How could we use similar way to guide some new sellers to win the market?

(Wen Liang)

1. What is the intent behind selecting 2nd lowest price as a target price in detecting dynamic pricing?
2. For products where there is no Amazon as a seller, the authors are looking only at detecting lowest and 2nd lowest. Is that sufficient criteria?
3. In the buy box algorithm, the feature vectors price difference to the lowest and price ratio to the lowest may indicate the same thing and both are highly weighted. Their correlation is not mentioned, do you think it is a good idea to include both separately?
4. Whether a product is new or used is not considered in the buy box algorithm as well as in the dynamic price detection. Even with the lowest price, a used product almost never makes it to the buy box and sellers of new products seem to be competing at a price point different/higher from the used product prices.

- *Sejal Shah*

1. How do you think new sellers could be compared with the existing sellers? In this analysis, they have excluded sellers with insufficient feedback.
2. Since the number of items sold by Amazon as a “seller” would be far larger than that of other third party sellers, do you not think this would bias the analysis?
3. This paper says that “algorithmic sellers are more likely to win the Buy Box at all ranks except for the top one”. This is confusing and not very well explained. They say that algorithmic sellers win the Buy Box even though they do not offer the

least prices, due to their feedback and sales volumes. Does this mean that bigger sellers (in terms of number of items sold), would entirely dominate the Buy Box and eliminate smaller sellers?
Aditi Ashutosh Mavalankar

1. In the paper on real-time navigation we studied that too many changes to recommendation of products lead to confusion and considerable decrease in the performance of the recommender. Here, on the contrary, we are seeing a recommender (seller recommender) system which is refreshing very frequently. Do you think frequent changes to Buy-Box seller could lead to user dissatisfaction? Or is the fluctuation a factor only for products recommended and not sellers?
2. The feature “Positive feedback” talks about if the seller has positive feedbacks to previous sells (maybe binary of numerical). Do you agree that most sellers that have a chance to compete for the winning spot, tend to have higher positive feedbacks? So, shall it be more effective to model if the seller has any negative feedbacks or missed shipments etc?
3. The conclusions about business strategies of algorithmic users (and the fact that they have high selection rates in Buy-Box) point that most big sellers (logically makes sense as well because of the resources they have) follow algorithmic selling. In a recommendation system this might create a bias towards the sellers visible to the user. This is a major problem when recommending which products are visible, but not exactly studied if it is an issue with which sellers are shown. Except the notion of free and fair trade, do you think it has any other repercussions with respect to the choices available to a buyer?

-Dhruv Sharma

1. **The data used in this paper are the best sellers. Why are the authors using only popular items as data? Would this create bias in the results?**
2. **Is there any specific reason for only using random forest as the classifying algorithm? Do you think other models are need for this study?**
3. **The ultimate goal of dynamic pricing is to sell products to consumers. Based on the results in this paper, how do you think the consumer’s behaviors/data could be incorporated into the study?**

-Siyu Jiang

Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market (Salganik and Watts, 2008)

1. If the study was intended to measure the effects of 'social influence' and 'quality' only, would it have been better to present the songs and artists in terms of numbers instead of names to prevent the effect of 'familiarity'?
2. The authors measure the number of times a song has been listened to, but do not have a threshold on the duration to increase count. Do you think that the results would have been different if they had employed such a measure?
3. This study uses songs to study market behavior. Considering that songs expect lesser time investment from users, can we expect the behavior and results to be the same in a marketplace of movies or books?
4. Can we expect behavior to be similar if each download had to be paid for, or if users were exposed to ratings and reviews as well?
   (Shreyas Udupa)

1. The listeners can only see how many times a song has been downloaded. It would be interesting to observe how the listeners would behave if they could also see the number of times a song has been played. If a song has been downloaded almost every time it has been heard, it would account for quality for a new listener.
2. As there is no distinction in the music by genre and other music attributes, the popularity (number of downloads) is largely driven by the initial listeners' subjective preference. Do you think including these attributes would affect the results?
3. The experiment is conducted over a couple of months, it is hard to generalize to long term popularity. Plenty of music, books, movies get popular after decades of their release and fared either average or poorly on their release.

- *Sejal Shah*

1. **How do you think the results of this experiment could provide insights for building recommender systems?**
2. **This paper was written about ten years ago, when there is not as much accesses and information of new/less popular songs as today. With today's personalised recommender systems creating playlists for individuals, do**

you think the result of this paper is still helpful for building recommender systems?**

3. **Whether a song will be download and how will it be rated largely depends on the user's' mood when listening to the song. In this paper, all experiments are done in a lab setup. Do you think this would create bias for the result?**
-Siyu Jiang

Modeling the Assimilation-Contrast Effects in Online Product Rating Systems Debiasing and Recommendations (Zhang et al. 2017)

1 - In the paper, authors have only considered historical ratings to debias the recommender system. What could be the challenges if we want to incorporate reviews or temporal dynamics into this model? One challenge I could think of for incorporating the reviews is that if we we model a particular review considering all the previous reviews, the approach would be very expensive given that we would have to model each word of all the previous reviews.
2 - Although, data's representation in various plots seem to be mimicking assimilation-contrast effect well, in each of the plots, we have small bump in upward direction before the point where prior expectation is equal to the average next rating and small bump in downward direction after the point where prior expectation is equal to the average next rating, which is not explained by the assimilation-contrast effect. Could this pattern be useful, if somehow modeled?
3 - The idea of debiasing seems good, but would it work well in real recommendation scenario where people refer the ratings/reviews before purchasing? What would make them trust the recommender system over the ratings which other users provide?
By Rishabh Misra

1. Would it be better to use the helpfulness count of a review / rating (rather than aggregating all previous ratings) and understanding its influence on the bias- rather than just the number of historical ratings?
2. Is the exponential decay a good idea? Wouldn't users be influenced by the earliest ratings / reviews a bit more?
3. By using this model, would external influence (outside the interface) also be classified as the influence of historical ratings?

- **Balasubramaniam Srinivasan**

1. Some aspects of the model seem to be pretty arbitrary, such as the magnifying function and kernels. Is there a way to devise a more general form for these functions or learn them from the observed data?
2. Is it possible to compare the predictions generated by the proposed model with baseline predictions? For example, is it possible to use something like an A/B test to see whether debiasing recommendations leads to more successful results in a live system?
3. What kind of rating progressions are observed for cold start users and items? Does the Assimilate-Contrast theory still hold when an item has not received very many ratings?

Rajiv Pasricha

1. This process of debiasing the historical ratings is happening in the back end of the recommendation system, but to the user, there is really no difference except seeing a slightly different list of recommendations. Therefore can we maybe incorporate into the actual avg ratings for a given item (for example displaying a 'debiased' rating on the side of the actual rating)?

2. Even though Amazon or Trip Advisor most likely did not have prevalent bot users, however, in some shopping websites using recommendation systems, there might be bot users and their fake reviews and ratings. How would the debiasing process react to these type of historical ratings?

3. Unless for a particular product with very few historical ratings, the intrinsic quality/rating would not make a big difference on the choice of the user on whether to purchase or not. Therefore is there any way the model can be changed to debiase products with few or fluctuating historical ratings because those are the ones that the customers are not sure whether to buy or not.

- Yan Cheng

When do Recommender Systems Work the Best? The Moderating Effects of Product Attributes and Consumer Reviews on Recommender Performance

1 - In the paper, authors say that the impact of recommenders are moderated by other factors such as types of items sold, item attributes (price etc.), and consumer-generated reviews. This could be true in general but amount of impact depends a lot the interest of users. For example, authors mention that price of the item moderates the effect of recommender systems, but for a user, who like to buy expensive products, recommender systems could still be very useful. Do you feel that personalized recommendation would be a good thing to study in a manner similar to what author study in this paper?

2 - The study in the paper is based on conversion rate i.e. percent of product views that result in purchases. Do you think this metric is a little strict to measure the effect of recommender systems? The main goal of recommender systems is to let users discover items which they might be interested in. For any website other than e-commerce (like youtube etc.), a click would be strong signal of interest. Based on that, do you think monitoring the click behavior would be a better choice for measuring the impact of recommender systems?

3 - In the paper, authors try to study whether the recommender system is complementary or supplementary to review ratings and study their effect on conversion rate. However, this does not take into account that recommender systems could be used as an exploration tool. Is there any way this could be studied using the data mentioned in the paper?

By Rishabh Misra

1. Though the authors find two clear clusters for utilitarian and hedonic, we could see many cases where a utilitarian product for one may be hedonic for other. Could we look into personalizing the selection of utilitarian/hedonic products for different users.
2. Can we try to generalize this analysis (utilitarian vs hedonic and search vs experience) for websites with not such clear distinction between categories like netflix, where all movies might fall into a single category.
3. How concrete is the coefficient analysis given the fact that we are not very aware of the correlations between different variables present in the model?

Tushar Bansal

1. Nowadays there isn’t a single recommender system on the webpage of a product. Alongwith ‘who bought this also bought’, there is also ‘frequently bought together’, ‘also viewed’, sponsored recommendations, recommended lists, etc. Do you think the analysis will be more informative and relatable if multiple of these systems are considered?

2. Alongwith knowing whether recommender system boosts conversion, it would be interesting to see if there is additional purchase due to recommendations. The authors bring up the number of products purchased variable, however don't use it further in their analyses.
3. What do you think about the use of a linear probability model for interpreting interactions of product attributes with recommendation treatment? Would something like a stepwise logit or a tree model do a better job?

- *Sejal Shah*

1. The authors included three Control Attributes in the discussion. Why did the authors include them? What do you think the functionality of these attributes are? Should the authors explore more into these attributes or not?

2. The authors mentioned that they only used purchase-based collaborative filtering system in the paper. How feasible would it be to perform similar analysis on other recommender systems such as a personalized recommender system taking into account of user attributes such as previous purchase history, ratings given and etc, or even a recommender system that takes into account the relationship between the product attributes with the user attributes?

3. Besides Ecommerce/online shopping, can the same type of study be expanded into other areas where recommender systems are used so that the relevant user would have an idea of when and how to implement recommender systems?

- Yan Cheng

Using navigation to improve recommendations in real-time (Wu et al., 2013)

1. How is negative implicit feedback taken into account? Say when a horizontal scroll happens in a row, or a row is skipped?
2. How about interactions based on thumbs up, thumbs down, adding to playlist, and checking out more details?
3. How is the scenario handled when video is started but user stops it and never views it again. Hence, this indicates negative feedback.

(Rahul Dubey)

1. It is stated in the paper that, rows will be fixed if they have been seen, seems that during runtime, only rows that are not on the top will be more optimized, what happen to the rows that user are more interested in? They have to dive into the entire category each time? On the other side, if user is not interested in the first five of a certain row but he/she scroll horizontally on that row, the next videos in that row should also be resorted accordingly.
2. Do you think the inference model would be simpler if Netflix is able to incorporate 'dislike' functionality? As browsing might not be very accurate, especially when user are actually more interested in the videos hidden behind the first 5 of each row.
3. Why the browsing history is not being carried over to the next session? What features do you think could be used next time?

(Yiwen Gong)

1. The authors mention that this model is tightly coupled with the UI design. The layout shown is only used in a limited number of devices like appletv, roku etc., For the general netflix website what do you think is the right way to capture user navigation without using sophisticated technology like eye tracking ? **(Sai Chaitanya Kolasani)**
2. The model doesn't consider any negative feedback from navigation. For example the user might not like one particular row. Do you think these effects can amplify undermining the recommendations ? **(Sai Chaitanya Kolasani)**
3. Can the model be improved by higher order interactions ? For example if the user opened a item from a particular row to view its details or added to his watch list the weights can be readjusted. **(Sai Chaitanya Kolasani)**

- **Sai Chaitanya Kolasani**

1. The online update is a very effective way to solve cold start problem for this Netflix movie dataset. Do you think it is also a good way for other datasets with lower sparcity?
2. Could it also be effective to add additional basic info or social features?
3. This model takes the advantage of impression fatigue and repeating strategy. Does this advantage work for other recommedation application?

(Wen Liang)

1. The authors mention that scrolling updates the model. However, it is quite common for users to scroll down and then go back up, if for example, they missed out on a few rows. How do you think this model can handle that?
2. How do you think global impression fatigue can be incorporated into this model?

Aditi Ashutosh Mavalankar

1. Do you think the solution is highly tied to the current UI prevalent in Netflix like systems (video/TV recommendations)? If so, we have studied before (such as the gaze prediction paper) how other signals can be used for the task. How do you compare the effectiveness of this approach against any of them?
2. The paper talks about the fact that a recommendation system initially selects the candidate of videos to be shown and the navigation inference is used to rank them accordingly. It then talks about how it provides signals for alleviating the cold start problem. The question here is, doesn't the major onus of providing the correct set of candidates still rely overf the backend recommender?
3. The evaluation takes into account Precision metric but again the precision of the model is highly dependent on the candidate videos.
    a. SHould the paper evaluate the difference in precision before and after using navigation?
    b. What are some other evaluation scheme you think (except MRR) which could be beneficial here?-what about CTR?

Dhruv Sharma

Exploring the filter bubble: the effect of using recommender systems on content diversity

1. Are there domains or situations where a collaborative filtering approach to recommendations would actually increase the speed of the natural narrowing down of diversity?
2. The authors assume that people take up recommendations for their appealing content. In a physical library containing hard copies of books, the only form of persuasion that a user could be influenced by was the ordering of genres of books in the library. Most libraries use the Dewey decimal system which is an unbiased, alphabetical ordering of books on a shelf. Considering that most online movie and book recommendation systems do not have a comparable unbiased ordering for their catalog, is it a fair assumption that it is the appealing content that makes people take up recommendations, rather than just the front page effect?
3. In the news domain, most recommendations are provided using a more-like-this or related articles approach. What approaches can be incorporated into a similar article recommendation system to increase the diversity of the recommended articles?

Siddharth Dinesh

1. The authors use the tag-genome system for measuring diversity. But with that level of specificity would it be harder to get a good measure of diversity ? For example tags like zombies, serial killer, nudity, revenge, franchise could all belong to both a serious thriller movie and a parody/comedy movie which are not really the same. For these purposes would it be better to go for more generalized tags ?
2. Don't you think that the threshold for someone in the "Following" group is too low at just taking 1 recommendation per block ? This way the majority of movies they are watching are non-recommended and hence maybe affecting the stats ? Wouldn't it be more relevant to analyse effect of taking different amount of recommendations ?
3. Isn't the measure of diversity used in the paper temporally local ? How about measuring diversity across time ? Like they can be getting the same 4-5 types of movies being recommended in each block and still be seen as diverse by the evaluation method described in the paper ?

- Ajitesh Gupta

1. The authors use people who do not generally use recommender system as a control group to compared with rate user. But would it be possible that people don't use recommender system(the ignoring group) may have higher standard in making decisions while people use recommender system(the following group) may be more lenient while making choices. Also, with higher standard in choice making, people don't use recommender system may be more strict on rating and thus not likely to give high ratings. Will those affect the authors' conclusions?
2. The authors mention that the purpose of this research is to examine the effect of recommender system on users "throughout their lifecycles", then is the research data be illustrative enough? Since the data were collected within the time period from 2008 to 2010, how do they prove that these data were evident enough to reflect users' long term decision making behaviors.
3. People may be more strict on choosing movies while they may be more lenient on making shopping decisions. Will it be more persuasive to conduct this study on commodity shopping rather than on movie choosing, given that many people may avoid certain kind of movies?

- Yan Cheng

- This is more of a philosophical question: what if people don't actually care whether their recommended movies, items, etc. put them in a "filter bubble"? What if some users want to stay in said bubble and only view movies, buy items that are similar to what they already enjoy, and they're fine with that? Would forcing recommendations to be diverse and to get users out of their bubbles actually be a way of manipulating them?
- I appreciate that the authors included a time limit of 3 months to account for "the reality that as time passes, the likelihood of a causal link between the recommendation and

consumption diminishes”; is there any similar common sense type accommodation you think the authors should have included?

- Stephanie Chen

Online popularity and topical interests through the lens of instagram (Ferrara et al., 2014) | pdf

1. The paper selects top 2100 hundred users who participates in weekly contest. Thus, the paper misses the analysis on vast majority of users who do not take part in the contest which might reflect the general trend in the instagram. Shouldn’t the paper consider these users as well?

Kriti Aggarwal

1. To model the user, we use tf-encoding of the hashtags. Can we use word2vec representation of the hashtags and then represent the users using some combination of word2vec?

Sudhanshu Bahety

1. The paper inferences that the the breadth of topical interests has little correlation with user popularity, i.e, whether a user is specialized in a particular topic or not has little corresponding variance in popularity. Isn’t this counter intuitive ?. The common observation is that users with most popularity tend to be frequent in posting about specialized topics, e.g. Food and Lifestyle models, sports anchors, celebrities of different domains etc. It would be better if the hypotheses and inferences could be judged on a metric.

2. The topical clustering uses only tags. Can we use other metrics like followers/followee to cluster the users? Moreover, what insights would we get based on these cluster?

Digvijay Karamchandani.

1. There are a lot of “fake” likes and comments exist on social media, which is used for advertising, propaganda or other purposes. Does this paper provide an insight on how to find these kind of “fake” comments?
2. The users who join the photographic contest are already users who are very active on social media. Doing analysis only on these users and their posts creates a bias for the result. Do you think incorporating less active users would be a better option?

3. The author claims a large majority of users adopting only few tags and the user vocabulary sizes are limited is a reflection of the information economy principle. But could this be due to the automatic tag generation method of the social media?

- Siyu Jiang

1. The observations were made on data collected over a very short period of time(1 month) this doesn't capture any dynamic behavior. For example some tags like #christmas are popular only during particular times of the year. Is it better to collect information spanning longer duration and will it change these observations significantly ? **(Sai Chaitanya Kolasani)**
2. The collected media and tags are from photographic contest. Don't this fact biases the results ? For eg: Users already winning these contests generally posts more as there is an incentive but it may not be general human behaviour on this network.**(Sai Chaitanya Kolasani)**
3. The results show users with very broad interests and specialized users are equally likely to be popular. What might be the explanation behind this ? Is it just because users with broad interests tend to post more and thereby may become popular ? **(Sai Chaitanya Kolasani)**

(Sai Chaitanya Kolasani)

1. What could be other crawling techniques which can capture general characteristics of Instagram?
2. This paper samples users from a contest where each contest has specific topic. Does it limit the number of tag topics? What could be the reason for large #topics not affecting their results?
3. A user can belong to multiple clusters. Would a Fuzzy C-means clustering where user can belong to multiple multiple clusters be used here instead of k-means?
4. Authors earlier tried crawling from a tourist place. They got poor clustering coefficient. Isn't this a bad strategy as people coming to a tourist place may come from distant places.

- Kulshreshth Dhiman

1. The paper describes an approach that leverages weekend hashtag project, using 72 most popular such hashtags and randomly picking 2100 users that participated in at least one of them as seed nodes, to form RIN. What could be impact of using this RIN on the results, given that such an approach might narrow the scope of tags considered and also, reduce diversity of instagram users considered?

2. How would the approach, analysis and results discussed in the paper change in case Instagram also offered a share button to its users?

3. Authors find out that actual user vocabulary size is limited, with majority of media being labelled with just a few tags. Further, this might suggest that users cannot keep track of all tags emerging on the platform. From perspective of Instagram, how can this insight be leveraged to enable the users to be able to discover, learn and keep track of new kind of tags, hence increasing user engagement on the platform?

By Nitin Kalra

Post Processing Recommender Systems for Diversity

1. The author only used two datasets related to movies (Netflix Prize and MovieLens) where the inherent diversity is very small (limited categories of movies). What do you think this post-processing technique would perform on a much diverse dataset such as from Amazon and etc where recommendation diversity comes from the fact that items might be used together with other items of different categories?

2. The authors did not explain the exact process of how to apply the graph optimization problem to post-processing the results of a CF recommendation system. Was it explained in the full version of the paper instead?

3. During the quality evaluation step, the authors purposely filtered out the items which received a rating of 1 or 2. How do you think this step would make a difference in the test results?

- Yan Cheng

1. Can you discuss some intuitions why the different learning procedures provide such different results? Min-cost-flow problem should be convex problem to solve via linear programming, so why are the results so different?
2. I am also confused why they need to do such aggressive preprocessing like filtering out low rating movies. Doing this may spoil some assumptions the original mf model makes.
3. I am curious how their recommendation algorithm responds to changes in the input. How can the authors measure the robustness of their model under some input noise?

- Chester Holtz

1. How would this task perform - when the system inherently has a graph which does not obey power law?

2. While this may improve sales diversity it does not directly translate to increased revenue? Would it possible to incorporate this into the model as well?
3. This kind of strategy works well for movies domain as people would want to explore a bit? Is this generalizable to other domains (especially when the user item interactions can't be treated as iid)?
- **Balasubramaniam Srinivasan**

Do you think the results would have changed much if the authors had used the optimization approach rather than the reranking approach? (Section 3)
What are your thoughts on making the trade-off between discrepancy and precision? Do you agree with what the authors decided to do in Section 5.3?
How well do you think the filtering of the original data worked, and do you think it skewed the results?
- Stephanie Chen

Exploiting Socio-Economic Models for Lodging Recommendation in the Sharing Economy

1. For any travel related recommendation system, it makes a lot of sense to somehow model things like season (or even weekends) separately, as they consist of different patterns as oppose to regular bookings. Do you think incorporating features to take seasonality into account would have helped?
2. The evaluation is done on two popular cities: NYC and London. Do you think the model will perform as well in locations which are not so popular? t291

(Rahul Dubey)
1. The model proposed in the paper uses 176 hand crafted features and use gradient boosting regression to find the ranking of lodgings for the users. Would a deep learning model be a better option to capture the non-linear interactions in the data?

2. The paper tries to model various socio economic factors like perceived value, perceived authenticity. In the evaluation the paper compares two methods, a simple baseline model 'Popularity' that just takes one feature maximum booking into account, when compared to the BPRMF which is a complex collaborative filtering model. But both the methods are shown to perform equivalently. Does that suggest that MRR might not be the right evaluation metric for this paper? Does that pose a requirement to use some other evaluation metrics to validate the results?
- Kriti Aggarwal

1. The paper is about the importance of socio economic factors on the rankings for the users. The paper mentions that these features were not taken into account for other

models and they used hand crafted features for the same. But in the evaluation the paper mentions that even if many of the features are removed from the model , the performance of the model is not impacted. Does this suggest that these features might actually not impact the user rankings?

2. The paper mentions several subjective features including perceived value and perceived authenticity. For modeling each of these features the paper utilizes hand crafted features. But the modeling of perceived value is quite subjective and depends on the psychology of the user. For eg. the user might have inherent bias for some lodgings location, brand, safety etc. Would these factors help in improving the perceived value of the model?

- Digvijay Karamchandani

1. The paper only captures sentiment from the EWOM (Electronic word of mouth) while there are multiple features that can be captured from the reviews. Would LDA help in modeling the topics in the user reviews?
2. The model has not captured the seasonal effect. For eg. the preferences of the users for a particular location would be dependent on the current weather and temperature conditions. Would the adding of seasonal effect enhance the model's accuracy?

- Sudhanshu Bahety

3.

1. Can you discuss some effects of their proposed normalization and data preprocessing on the interpretability of the performance results?
2. They apply some sentiment score as a feature, but why can't the authors just directly use some of the review text - it should benefit from the empirical robustness of their model.

- Chester Holtz

1. Is there any way in which the model could be fed with known events that could probably cause a fluctuation such as that described in section 5.1? In other words, could we not have two separate models for regular and seasonal/one-time events?
2. Empirical Feature Efficiency (EFE), which they propose in this paper as a new metric, is in fact, not new, as far as I know. Ablation experiments are used in more or less the same manner, and this metric does not add any value to existing knowledge about ablation experiments. Please correct me if I am missing out on anything, but do you think this metric is novel?
3. How do you think the results would differ if cities other than London and New York were taken? For example, consider exotic locations, which are often visited by many tourists. The criteria there would be different, and users might possibly give more importance to the surrounding areas, instead of, say, price. How do you think the SEM and features would change?

Aditi Ashutosh Mavalankar

1. How do you think the results would vary if the location that the study was conducted for was a more rural location where places of interest are farther away and many of the hand-crafted features for the perceived value element would end up being 0?
2. The study is conducted on data from Airbnb. However, there are many competing lodging marketplaces which all affect the perceived risk for a customer. How do you think additional features such as the difference in price between a hotel room and an airbnb room would affect the results of this study?
3. What additional hand-crafted features do you think would add value to the interpretation of this study? I imagine this is why the authors refrain from a deep learning model to automatically learn features.

The Role of Social Networks in Information Diffusion (Bakshy et al., 2012)

1. Would it be better to analyze the clusters of friends where the URL is shared rather than just looking at the ties. A person might be more/less inclined to share if the friends who share it belong to separate groups (school, family etc.).
2. What should be a good criteria if we were to select the URLs to show among weak ties? Should the ones with best strength be selected or we should aim for more diversity in results?
3. Can we use this analysis to predict the virality of URLs. For example a URL being shared more between weak ties might have better chances of being viral than the one being shared between strong ties?

-Tushar Bansal

- Are there any factors you've encountered personally on social media that the authors didn't take into consideration in the paper? (e.g. if you're already close w/ someone, you may share information via text message, rather than on Facebook)
- Do you agree with the authors' definition of a strong tie (e.g. being tagged in a photo together just one time)?

- How can shares of articles via connections in other venues other than the News Feed be measured and factored in here?
- Stephanie Chen

1. The authors put a check on the fact that the subject did not visit the url in the past 2 months through facebook ? Would it be more beneficial (if possible) to consider only urls that are recent ? because otherwise the urls being shared could be quite old like songs, blogs etc and hence the user might have already seen it ?
2. The authors consider the interactions between users to determine weak/strong ties. Do you think taking into account the nature of these interactions can have a significant effect ? Like a couple of users could have had a long chat, comments on common posts but those interactions can be hostile due to different ideologies. So obviously content shared by one won't be shared by the other.
3. The threshold for a strong tie is kept at 1 interaction. Wouldn't that be too low of a threshold especially considering the previous question in mind ?
4. The authors do not take into consideration the fact that the article might have been shared by private message or a friend's wall. Do you think this might have affected the results significantly? because in personal experience I have seen quite a lot of article sharing behaviour by those 2 methods.

- Ajitesh Gupta