The Startup Ecosystem and a Preference Based Recommendation System

Zhangyi (David) Fan

Advised by: Professor Prasanna Tambe
MDATS 599: Masters Practicum
University of Pennsylvania

May 8, 2021

Referencing historical marketing literature on hierarchical choice models, the paper derived a generalizable ML model with the necessary features that can approximate a multi-level structure. The final model perform better than conventional model and random benchmark across both seasoned individual and new individual who has had no application history. In addition, the paper, using model diagnostic techniques such as variable importance plot, found that, while the order of individual preferences did not change drastically, some aspects of job listings did appear more important to users now more than before. For instance, the user in the pre COVID world appears to be more concerned about the compensations offered by the jobs along with job location. They also generally appear to be more aligned with their historical patterns compared to those in the post COVID world.

1 Introduction

Recommendation systems have been one of the most studied and impactful technological innovations in the past 20 years, and its usage has sieved through all product-based retailing domains. From a firm perspective, the recommendation system provides the firm a way to tailor its services effectively and efficiently for appropriate customers, while from the consumer side, a relevant recommendation could drastically decrease search time, increase conversion rates and allow them to see more products that align with their interests.

This paper specifically investigates building a recommendation system for a two-way job platform (analogous to LinkedIn). Whereas for other products, recommendations can often be done by teasing out consumer loyalty for a specific product (like a certain type of coffee or fruit), jobs differ from other product offerings in that an individual, either by platform constraint or their own volition, would only apply to one job at most once. Given such constraints, the emphasis on a job recommendation model requires extrapolating from previously absent choices, which is arguably more challenging.

Methodologically, the paper lends itself from conventional marketing literature that uses structured models, most often the multinomial logit model, which allows consumers to have varying preferences along specific product traits. The parameter from the model naturally lends itself to interpretation-based tasks by allowing modelers to make statements about how different product traits affect conversion rate. Furthermore, Variants of these models allow one to incorporate consumer characteristics to these preference parameters. By being able to estimate consumer preferences based on their characteristics, one will be able to effectively counteract the known cold start problems by providing reasonable estimates for these new consumers.

Beyond the general problem illustrated above, another objective of interest emerged as a result of COVID. During the pandemic, individual's preference for jobs have likely shifted. For instance, due to the rise in unemployment as a result of COVID-layoffs, the salary elasticity of individuals might have shifted downwards with more supply in labor market and smaller demand. Moreover, the transition of most jobs to a remote setting has likely lessened the effect of location proximity on driving consumer conversions and possibly may have amplified the effect of popular sought-after locations such as San Francisco. By fitting a model that predicts consumer behavior in an interpretable way, one could leverage such a model to provide a diagnostic for how consumer preferences may have shifted before and after the pandemic.

The following sections of the paper will be structured as follows. Section 2 will provide a review of traditional techniques and a deep dive of the model development. Section 3 will discuss the dataset used. Section 4 will explain the model evaluation. Section 5 will discuss the fitting methodology employed in the paper. Section 6 will display the model results and fit. Section 7 will discuss the inferences from the model and Section 8 will conclude the paper.

2 Model Deep Dive

This section will expound upon the conventional form of the recommendation model used in computer science, the choice models used in marketing, and the reparametrized choice model that will ultimately be used in this paper.

2.1 Classic Recommendation Models

Conventionally, recommendation systems have been done in two main directions. First are the directions inspired by collaborative filtering. While many more variants of such models have emerged (such as latent factor collaborative filtering which leverages dimension reduction), as a method, collaborative filtering poses an inductive assumption that an individuals that applied to similar jobs as other individuals tend to be more alike. This assumption, however, can fall when there exists huge heterogeneity across individuals where not a lot of commonality directly occurs, or when an individual simply has a rather unique preference that is not easily generalizable.

The other direction, content-based recommendation, often uses embeddings to jointly represent individuals and the products in question. This model then later recommends jobs, whose embedding are closest to the individual's embedding. The inductive assumption posed by this method suggests that individuals are going to apply to the similar jobs that they have applied to in the past. This method, in and of itself, does not permit any form of information sharing across individuals, and assumes that an individual's propensity can only be characterized by his/her own application.

In both scenarios, the model proposition is valid to some extent but does not capture the full extent of the story nor the full extent of the data. There must be some form of balance between allowing for information sharing across individuals and the explicitly expressed preference of an individual.

In addition, conventional recommendation models such as the ones listed above are typically black-box in nature. These methods investigate past consumer choices or choices of similar consumers then use a distance-based metric to filter for jobs. Since these models leverage past consumer choices, for a completely new consumer with no recorded choices, the model will fail to make tailored recommendations since it does not account for relevant traits of the consumers a priori. This is also commonly known as the cold start problem. Moreover, these models are incapable of extrapolating consumer preferences apart from providing individual specific predictions. Ideally, however, by building a model that can predict consumer choices, one would also like to infer consumer specific traits in terms of the kind of jobs that a consumer may prefer. These inferences, when coupled with other metrics such as customer lifetime value or mean conversation rate, could function not only as a prediction-based tool but also one that can help firms target the best consumers or help its subsidiary job providers target applicants.

2.2 Marketing Choice Models

On the other end of the spectrum, models that map individuals to product choices have been widely studied in the context of marketing. For instance, one could imagine building a recommendation model using a simple logistic model specified as follows:

$$
\begin{gathered}
Utility = BX + error \\
P(Y=1) = e^{BX}/(1+e^{BX}) \\
X = Job\ Characteristic\ Vector \\
B = Preference\ Weight\ Vector
\end{gathered}
$$

Assuming such a model is built on different individuals, a logistic regression of such set up implies a homogenous covariate effect across all individuals. In other words, the preference weight vector is fixed for everyone. To put it into similar terms as before, a pure logistic regression implies perfect information sharing where all individuals share similar behavior. To relax such an assumption, one can embed in additional heterogeneity to the model as follow.

$$
\begin{gathered}
Utility = B_i X + error \\
P(Y_i=1) = e^{B_i X}/(1+e^{B_i X}) \\
X = Job\ Characteristic\ Vector \\
B_i = Preference\ Weight\ Vector\ for\ Individual\ I \\
B_0 = Population\ Preference\ Weight\ Vector \\
Sigma_0 = Population\ Preference\ Variance\ Vector \\
B_i \sim Normal(B_0, Sigma_0)
\end{gathered}
$$

This model creates a flexible scenario where an individual weight preference is heterogenous and can then be updated based on their individual choices.

Compared to the vanilla logistic regression model, this model, being Bayesian in nature, allows individual's explicit preferences and their choices to affect their individual parameters. However, such an update is subject to some shrinkage towards the population preference. The exact form of such Bayesian update is often fixed depending on the model choices; however, there has been ongoing criticism about the validity of such a constraint structure (e.g., are people all Bayesian updaters in a similar fashion). Naturally and implicitly, however, the population preference in this case is the same for everyone. As a result, this model would suffer the same cold start problem that the other more conventional recommendation models have faced. To expound, in this model, albeit allowing for heterogeneity, all new customers will simply be treated the same as they are all assumed to have the same prior. This exact phenomenon, however, can be tackled through a multi-level Bayesian model seen as follows.

$$
\begin{gathered}
Utility = B_i X + error \\
P(Y_i = 1) = e^{B_i X}/(1 + e^{B_i X}) \\
X = Job\ Characteristic\ Vector \\
Z_i = Demographics\ or\ Pre- \\
interaction\ Characteristic\ Vector \\
T = Demographics\ Weight\ Vector \\
B_0 = Population\ Preference\ Weight\ Vector \\
Sigma_0 = Population\ Preference\ Variance\ Vector \\
B_i \sim Normal(TZ_i, Sigma_0)
\end{gathered}
$$

The model above, effectively assumed that the prior of the preference can be set by some demographic or pre-interaction characteristics of an individual, as opposed to being set as the same across the population. A multilevel model structured like above can link pre-interaction traits that we observe of customers to their expressed preference in addition to accounting for customer specific heterogeneity. This makes it into a viable solution for the cold start problem mentioned earlier, while also preserving interpretability.

2.3 Reparametrize Hierarchical Models

.

Despite having been experimented in the space of recommendation model, this approach too has its limitations that make it less viable in the real-world setting. Firstly, as a Bayesian model, it requires immense amount of computational power and typically takes much longer to train than other forms of reduced form model in a single layer. Secondly, noticing the model structure, one could note that the integration of non-linear features or feature interaction in a model like this requires it to be specified a priori. Thirdly, the criticism regarding the rigidity of the Bayesian updates further casts doubts on the model generalizability.

Ideally, one would like to experiment with a model archetype that can approximate a multilevel hierarchical model but accommodates non-linear transformation and flexible update while also allowing for faster training speed. To do so, one would need to re-structure the hierarchical model above. First, let us consider a homogeneous multi-level model, where the individual preferences are solely decided by pre-interaction characteristics.

$$
\begin{gathered}
Utility = B_i X + error \\
P(Y_i = 1) = e^{B_i X}/(1 + e^{B_i X}) \\
X = Job\ Characteristic\ Vector \\
Z_i = Demographics\ or\ Pre- \\
interaction\ Characteristic\ Vector \\
T = Demographics\ Weight\ Vector \\
B_i = TZ_i
\end{gathered}
$$

On could rewrite this equation into a single layer model:

$$
\begin{gathered}
Utility = TZ_i X + error \\
P(Y_i = 1) = e^{TZ_i X}/(1 + e^{TZ_i X}) \\
X = Job\ Characteristic\ Vector \\
Z_i = Demographics\ or\ Pre- \\
interaction\ Characteristic\ Vector \\
T = Demographics\ Weight\ Vector
\end{gathered}
$$

Note, that a two-layered hierarchical model that extrapolates pre-interaction traits to set as the parameter for user preference can simply be extrapolated into a linear model built on top of interaction terms as opposed to simple product covariates. Now consider the heterogenous model case:

$$
\begin{gathered}
Utility = B_i X + error \\
P(Y_i = 1) = e^{B_i X}/(1 + e^{B_i X}) \\
X = Job\ Characteristic\ Vector \\
T = Demographics\ Weight\ Vector \\
Z_i = Demographics\ or\ Pre- \\
interaction\ Characteristic\ Vector \\
Sigma_0 = Population\ Preference\ Variance\ Vector \\
B_i \sim Normal(TZ_i, Sigma_0)
\end{gathered}
$$

The above model can also be structured into a different configuration:

$$
\begin{gathered}
Utility = B_i X + error \\
P(Y_i = 1) = e^{B_i X}/(1 + e^{B_i X}) \\
X = Job\ Characteristic\ Vector \\
T = Demographics\ Weight\ Vector \\
Z_i = Demographics\ or\ Pre- \\
interaction\ Characteristic\ Vector \\
Sigma_0 = Population\ Preference\ Variance\ Vector \\
B_i = TZ_i + delta \\
delta \sim Normal(0, Sigma_0)
\end{gathered}
$$

The delta term here can largely be referred to as the weight by which the actual data changes the prior preference for an individual. It is assumed to be centered at zero, but it effectively functions as a term for posterior update. Looking more closely with some abstraction one will realize that the preference parameters, being updated in a Bayesian fashion, are dependent on the prior mean TZ_i, the prior variance Sigma0 and the best estimate of the parameters from the individual's data. Note that there is no closed form solution for such preference calculation. However, we know for certain that it is a function of an individual's past positive and past negative behaviors. In other words, the formulation, above, can be thought of as:

$$B_i \, TZ_i + delta$$
$$delta \; f(PB_i, NB_i)$$
$$Utility = (B_i + f(PB_i, NB_i))X + error$$
$$P(Y_i = 1) = e^{(B_i + f(PB_i, NB_i))X} / (1 + e^{(B_i + f(PB_i, NB_i))X})$$
$$X = Job\ Characteristic\ Vector$$
$$T = Demographics\ Weight\ Vector$$
$$Z_i = Demographics\ or\ Pre-interaction\ Characteristic\ Vector$$
$$PB_i = Past\ positive\ behavior\ of\ an\ Individual\ I$$
$$NB_i = Past\ negative\ behavior\ of\ an\ Individual\ I$$

subjected to some functional constraint being imposed upon delta. The paper further notes that Sigma0 factor into the formulation above as a constant for shrinkage that controls how much influence the data has on the individual's parameters. This often dictates the form of Bayesian update, and is the term of rigidity. However, if one assumes that Sigma0 is not a population parameter but could in fact vary per individual (to achieve varying update speed), then one can remove the above constraint imposed on delta.

What we can then realize here is that such a hierarchical model approach illustrated above can be nicely simplified into the following:

$$Utility = TZ_iX + f(PB_i, NB_i)X + error$$
$$P(Y_i = 1) = e^{TZ_iX + f(PB_i, NB_i)X} / (1 + e^{TZ_iX + f(PB_i, NB_i)X})$$
$$X = Job\ Characteristic\ Vector$$
$$T = Demographics\ Weight\ Vector$$
$$Z_i = Demographics\ or\ Pre-interaction\ Characteristic\ Vector$$
$$PB_i = Past\ positive\ behavior\ of\ an\ Individual\ I$$
$$NB_i = Past\ negative\ behavior\ of\ an\ Individual\ I$$

This just means that if a model could approximate such interaction effect, while also capable of including a user's historical preference into account, it will be able to approximate such a hierarchical Bayesian model. That is, one simply need a flexible model g, such that

$$P(Y_i = 1) = g(Z_i, X, PB_i, NB_i)$$

where the model g can extrapolate the relationship listed above. One can imagine common machine learning model such as Random Forest or Gradient Boosting Tree. That can perform this job adequately (i.e., account for the appropriate interaction and extrapolate a flexible functional form behind positive and negative individual behavior). In addition these models will be able to further extrapolate non-linear and even more complicated interaction relationship that may exist between the dependent variables and outcome due to its non-parametric nature. Thus, a conventional machine learning model, with the appropriate characteristics and input data, should be able to co-opt the benefit of these multi-level models and even further expand upon it.

The paper will take advantage of such derivation and apply a conventional machine learning model, with the covariates, to an empirical dataset and compare it against conventional model in the recommendation space.

3 Dataset

The framework illustrated above is applied to a dataset of a job platform primarily focused on start-up listings. The dataset includes the jobs that an individual applied to off of a search. From a conceptual framework perspective, a model for job choice should naturally consider who the users are and what the job is. However, beyond that, as we have seen through the other conventional models, one also needs to consider a user's historical preference along with possible information sharing across individuals.

3.1 Features

In line with the general framework, the following dimensions are considered for modelling.

- **User characteristics**: The dataset includes user level characteristics including their years of experience, their primary roles (preferred roles), their primary location (preferred location),
- **Job characteristics**: The dataset includes the minimum and maximum salary and equity that the job offers, the experience required for the job, the associated role id of the job and the location of the job.
- **Startup characteristics**: The dataset includes information about the market that the star-up is in (e.g., food delivery, AI and etc.), the size of the start-up, and the degree of funding that the startup has received.

In addition to the natural characteristics presented in the dataset, some additional features were structured to be added into the dataset in Lieu of the new model set up.

- **Time since the job has been posted**: There is likely a lagging effect in terms of the appeal of job for individuals as time goes by. This includes possible concern around the job validity.
- **Social contagion effect**: While the word-of-mouth effect cannot be directly extracted from the dataset, the measures used by collaborative

filtering can serve as a good approximation to represent the most popular jobs for users with similar preferences (which in the case of individual with no history, this would be the jobs that are most popular across everyone).
- **User heterogeneity**: Lending from best practices, while the baseline covariates allow the model to deal with the cold start problem, user heterogeneity still certainly need to be accounted for in the model. As a feature, this will include the average of the attributes for the jobs that an individual has applied to in the past and the number of jobs he or she has applied to.

Mapping this set of features back to the model formulation earlier

3.2 Dataset Structure

The datasets are then partitioned into a pre-covid period and a post covid period to be independently calibrated and later compared. The time frame considered pre-covid period is drawn between 02-05 to 03-01 with the post-covid period decided between 05-01 and 06-30. Where 1000 individuals were randomly selected in both the pre and post covid periods to be analyzed.

In addition, note that the model in nature is a choice model where we assume an individual actively opts into a specific job because they prefer it. To train such a model one would naturally need to also have negative samples. Since the quantity to be estimated is P(Conversion | Exposure), ideally such negative samples are to be obtained from job listings that we know an individual has been exposed to but actively chose not to apply to. Given that such data is unavailable from the dataset, the paper approximates such behavior by randomly drawing jobs the individuals did not apply to but are active in the day with at least 10 applications.

To further explain this decision, one can view jobs in four different ways. Firstly, the jobs that an individual applied to, the jobs that an individual saw but did not applied to, the jobs that did not match any of an individual's search therefore was not applied to, and the jobs that an individual would have applied to if he had seen the job. In this case, we would ideally want to label the jobs in the first group with a label of 1 and the job in the second a label of 0. However, realistically the job that belongs to the third group will likely yield a negative label since they do not even fit the user's search criteria.

Furthermore, these jobs can also further help calibrate the model by allowing the model to discern the obvious cases that may not be present in the search-based dataset. For instance, for a software engineer, the only job points presented in the search-based dataset will probably be software engineering job since those are the only jobs that this individual will actively searched for. However, if a model was only trained on a dataset that contained software engineering data, it may suffer from the Lucas critique, when it attempts to give a prediction for an instance that it has never seen before. In fact, in that case, the model may simply disregard the differing roles that is present across different jobs (say business development and software engineering), as the training dataset never offered any signal that discriminates across roles, instead it will primarily reference other desirable characteristic i.e. high salary, that a user may be interested in, this in turn may cause the model to recommend irrelevant jobs to the user. By accounting for jobs in the third group, this will allow the model to learn the jobs that does not match the user's search preference at all.

The fourth group may be tricky as we would ideally want to exclude them out of the dataset. The approach mentioned above offers a way to crudely approximate for this. Note, what we are assuming in the fourth group is that the jobs that exist there matches the individual's preference but did not appear of a search and therefore did not have itself exposed to the individual. Given that typically the search engine for similar individuals is optimized in a similar manner, the jobs that did not appear for one individual likely wouldn't have appeared for another with similar preference. Under such assumption, using a filter of 10 applications guarantees that these jobs are not dormant jobs that will not appear of off a user's search, but only includes those jobs that are highly ranked (which either falls in the 2nd or 3rd bucket) and will appear in the results section of a search.

Mapping this set of features back to the model formulation earlier

$$P(Y_i = 1) = g(Z_i, X, PB_i, NB_i)$$

The Zi in this case would include the user demographic variables, while X will include the job and start associated characteristics (along with added metrics such as days elapsed and popularity). Note that in this dataset that we structured only positive behavior is considered (using the historical application data). If we did in fact observe negative application it may be worth including them too, but since the negative application in this case is randomly sampled, they'll be rather similar for everyone and hence needs not be directly considered.

4 Evaluation Criteria

The evaluation for a recommendation engine is not straightforward. As opposed to a conventional machine learning model where one can simply look at F1-Score and accuracy metrics, recommendation engine requires a more intricate statistic. This primarily stems from the fact that for recommendation models the criterion shouldn't be whether an individual applied to a job, but instead the relative ranking of the job that gets recommended to the individuals.

Additionally, we need to note that the purpose of the recommendation engine is that in the real world, people don't apply to all the jobs they would've wanted to apply to. So, if a recommendation model predicts that a user should apply to this job and the user did not apply to it, it may not necessarily be a flaw of the model but instead be an opportunity for the firm.

However, that said, a good recommendation model should nonetheless be able to approximate user preference well. That is to say, while a user may have not applied to all the jobs they would have wanted to apply to, the recommendation model should be able to pick up the jobs that a user actually applied to and place these jobs in a higher relative ranking than other jobs. Given this, the model metric that this paper proposes to gauge the model is a metric where the precision is held constant, and the model recall is judged. Specifically, the metric will be the percent of jobs recommended to a user that a user applied to. This will be gauge in a varying degree where a model will be allowed to recommend 1,3, and 5 jobs. As opposed to soliciting all active jobs in a day, the jobs panel will be the random jobs solicited earlier for computational efficiency.

5 Methods

To benchmark how a non-linear approximation of a hierarchical model will do, the paper will be comparing the proposed machine learning model with the conventional model used in the recommendation space.

Specifically, the following models were ran:

- **Random baseline**: where a random job is recommended to the individual
- **Simple Collaborative Filtering Model**: where the user preference is based on whether they've applied to a specific job. For every day the jobs with the highest preference score will be recommended.
- **Simple Content Embedding Model**: where the user's embedding representation includes the averages of their past application in terms of the job specific features. For every day the jobs with the lowest distance from the historical preference will be recommended.
- **Random Forest Model with partial information**: where only demographic, pre-interaction information are applied to the model. The job recommendation will be based on the application probability indicated by the model
- **Random Forest Model with full information**: where demographic, pre-interaction information, collaborative filtering score to account for social contagion and past preferences for user level heterogeneity are applied to the model. The job recommendation will be based on the application probability indicated by the model

Table 1: Pre COVID table

	% applied with 1 recommendation		% applied with 3 recommendations		% applied with 5 recommendations	
	Seasoned Individual	New Individual	Seasoned Individual	New Individual	Seasoned Individual	New Individual
Random	4.2%	4.0%	4.1%	3.8%	4.3%	4.1%
Collab Filter	3.4%	3.0%	3.3%	3.0%	3.2%	2.9%
Content Based	1.7%	1.5%	2.2%	1.9%	2.4%	2.1%
RF partial	6.4%	6.25%	7.6%	7.4%	7.7%	7.4%
RF Full	12%	11.9%	9.4%	9.3%	8.3%	8.1%

The dataset was portioned by the first weeks and the last for both the pre- and post-COVID periods. The first week's worth of data will be used as the training set, and the last week the test datasets. The Random Forest model will simply be trained on the training data and be used to predict for the testing data. The Collaborative filtering model will be trained on a sequential manner where the prediction for time period t, will reference all the data points before period t. This rolling approach will also be employed for the content embedding model.

6 Results

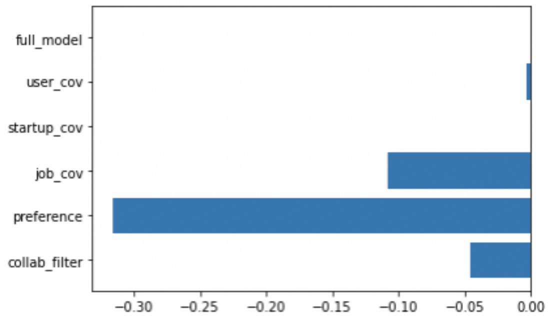
The model performance for the pre-COVID period can be found on table 1.

Surprisingly, the collaborative filtering model and content-based model were unable to beat the random baseline. This can largely be attributed to the limited sample and rather short time frame chosen. Since only 1000 individuals were chosen, the information sharing which collaborative filtering relies on ceases to work. Similarly, since an individual only averages 2 application in this dataset, the content-based approach will not end up with enough information to provide credible information. Unsurprisingly, those models also performed poorly for individual whose first purchase happened during the test-period of the model.

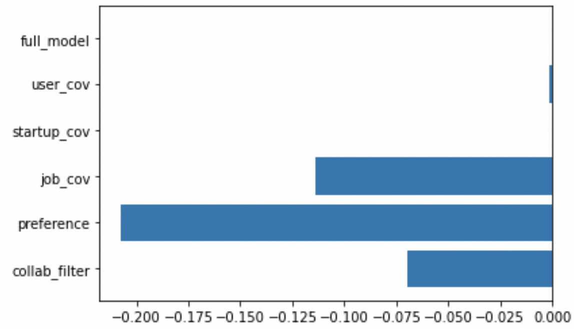
The Random Forest model with only demographic variables were able to perform quite desirably, its performance doubled that of the random baseline, indicating that there exist information sharing across individual that can be captured using simply pre-interaction traits. It is, however, worth noting that the Random Forest with preference-based information and collaborative filtering feature performed significantly better than the former model, quadrupling the random baseline, providing evidence that the integration of heterogeneity is equally valuable for a model set up such as this. What was particularly interesting is that this indicates that while the collaborative filtering and content based approach stand-alone has limited value, when it is combined into a model that allows them to interact with other features and accommodate for more complex extraction, the feature's power seems to have been further exploited. Both models performed favorably for both seasoned and new individuals, adding further validity to the models.

	Table 2: Post COVID table					
	% applied with 1 recommendation		% applied with 3 recommendations		% applied with 5 recommendations	
	Seasoned Individual	New Individual	Seasoned Individual	New Individual	Seasoned Individual	New Individual
RF partial	5.8%	5.6%	15.5%	4.9%	4.6%	4.4%
RF Full	12.8%	12.9%	32.0%	31.7%	5.8%	5.0%

Plot 1: Pre COVID variable importance plot



Plot 2: Post COVID variable importance plot



The Random Forest models were also then ran on the post COVID dataset for further analysis. The results generally align with the results before, except for the sudden surge of percentage applied with 3 recommendations, which is likely a result of random variation.

7 Analysis and Inferences

Given that the model in question, is now a simple Random Forest model, standard model-free inference techniques such as partial dependence, accumulated local effect, or individual conditional effect can all be applied to analyze the model. This paper specifically examines the use of variable importance plot to gauge how different features included in the model have shifted in their perceived importance before and after COVID.

The first analysis investigates how the different form of features affects the model. The feature groups included refers to the general feature grouping in the dataset section. Specifically, the plot uses a permutation based variable importance, where the value measured is the absolute decrease in AUC-ROC of the model when the variables selected are randomized. The variables investigated include user demographic information, startup information, job specific information, preference-based information and the social contagion (collaborative filtering variable).

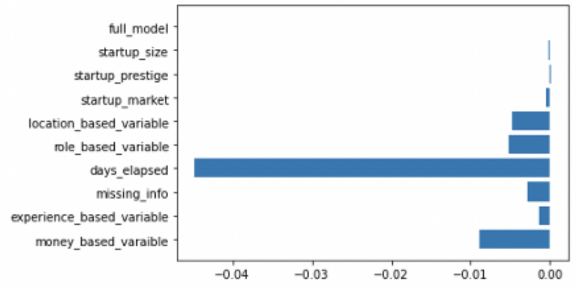
First, generally, across both periods, the preference-based variable dominates in terms of importance. This is largely in line with what we expected based on the model performance section. At large, the effect of heterogeneity drastically trumps the homogeneous assumption imposed by the user demographic variables. In addition to the heterogeneity and homogeneity interplay, collaborative filtering-based variable and job-based variable also serves as a relatively strong predictor for application. Interestingly but not surprisingly, the ordering of covariate significance stayed the same across both periods. However, the numerical effects did appear to change. Specifically, one can note that the numerical quantity of preference-based variable drastically decreased in the post-COVID period whereas the collaborative filtering variable's value increased.

Intuitively, this serves to indicate that an individual's historical preference post-COVID is less predictive to their action compared to pre-COVID. This is believable as in the post-COVID space, individuals who are more in need of a job may be more experimental in the kind of jobs they apply to as opposed to being more selective and confined to a specific set of standards. The collaborative filtering variable further adheres to this analysis, since in departing from an individual's fixed preference, they are more likely to flow towards the more popular jobs that either appears in their feed or have been brought to their attention by their social circles.

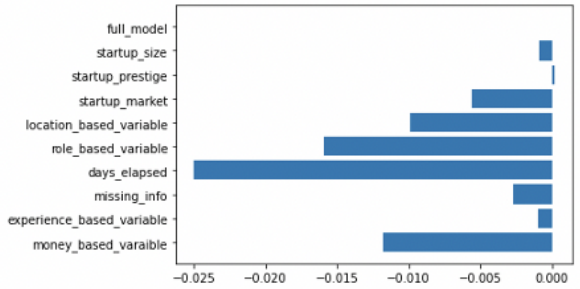
In addition to more general variable groupings, the paper further explored the importance of more specific variables and feature set in the dataset. The following feature set includes:

- **Startup size**: number of individuals employed by the start up.
- **Startup prestige**: the amount of funding that the startup received
- **Startup market**: the market that the start up is in
- **Location based variable**: The location of the jobs, the location of the user and the user's historical preference in terms of job location
- **Role based variable**: The role of the jobs, the primary role of the user and the user's historical preference in terms of job role
- **Days elpased**: number of days since the job was posted
- **Missing info**: amount of missing information for a specific job listing
- **Experience based variable**: The years of experience required by the job and the experience level indicated by the user

Plot 3: Pre COVID variable importance plot



Plot 4: Post COVID variable importance plot



- **Monetary based variable**: The compensation offered by the job (both salary and equity)

The overall trend is slightly more ambiguous when variables are broken down into the more detailed format. There are nonetheless a couple of key takeaways from this analysis.

First, the relative ordering in terms of variable importance has shifted. Compensation-based variables appear to now be less important compared to role-based variable. As opposed to this indicating that money is not important for individuals, this is likely an indication that individual is less selective about the exact compensation that the job offers but instead focusing more on whether they are compatible with the job. The ultimate cause driver for this phenomenon likely ties back to earlier analysis where individuals are more focused on getting the job as opposed to exploring for opportunities.

Second, there appears to be an increase in importance of the startup based variable relative to pre-COVID period. Specifically, the startup size and startup market variables both have seen a boost in their relative importance. This is likely an artifact of COVID making bigger start-ups and start-ups with pandemic-agnostic operations (e.g. food delivery) more desirable to work at than others because they contribute to increased job security.

Thirdly, the pandemic's overall effect in reducing location barriers through offering remote jobs could have possibly contributed to the location variable now being less important compared the role-based variable. However, the exact nature of the relationship can be rather ambiguous as certain areas may have had a boost in popularity due to the pandemic (such as San Francisco).

8 Conclusion & Next Steps

This paper had two objectives. Firstly, it aims to devise a recommendation model that can deal with the cold start problem. Motivated by models from quantitative marketing, the model derived that a generalizable ML model with the necessary features can be an approximation for such a hierarchical model and can then be used to tackle the aforementioned problem. This was generally a successful attempt as evidence by how the final model was able to perform much better than conventional model and the random benchmark across both seasoned individual and new individual who has had no application history.

Secondly, the paper seeks to build a model with interpretable insights which can be used specifically to analyze job adoption before and after COVID hit. Using model diagnostic techniques such as variable importance plot. The model found that, while the order of individual preferences did not change drastically, some aspects of job listings did appear more important to users now more than before. For instance, the user in the pre-COVID world appears to be more concerned about the compensation offered by the jobs along with job location. They also generally appear to be more aligned with their historical patterns compared to those in the post-COVID world.

That said, the paper and the methodology proposed can still certainly be improved in several different ways.

The model estimation could be improved with an added level of data on the exact exposure of an individual to better gauge their preference. The evaluation criteria, while valid, could also be admittedly made better using mechanisms such as online testing as opposed to offline methods. More model-free analysis could also be added to reveal specific trends that have changed over the course of COVID.

While user-level heterogeneity is present in the model proposed, the model does not account for any form of loyalty-based variable that may capture phenomenon such as start-up loyalty, where an individual is a big fan of a particular start up because of uncaptured factors about the startup such as their culture and mission. This could be further accounted for in the nature through an added fixed effect as start-up indicator.

The general proposed solution of incorporating pre-interaction features doesn't necessarily have to take form in a hierarchical way that the model tries to presume. Instead, advances in the space of variational encoder can be used as a multi-model representation for both individual level characteristic and actual application preference. One can imagine an underlying latent vector that is both a function of demographics and past preference. This approach is worth considering for future research in this space as well.

9 Bibliography

Bakhshandegan Moghaddam, Farshad Elahi, Mehdi. (2019). "Cold Start Solutions For Recommendation Systems." 10.13140/RG.2.2.27407.02725.

Jessica Gallant, Kory Kroft, Fabian Lange, and Matthew J. Notowidigdo (2020), "Temporary Unemployment and Labor Market Dynamics During the COVID-19 Recession", NBER Working Paper No. 27924

Rossi, P.E., McCulloch, R.E., and Allenby, G.M. (1996), “The Value of Purchase History Data in Target Marketing”, Marketing Science, Vol. 15, No. 4, pp. 321-340.

Xiaoyuan Su, Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", Advances in Artificial Intelligence, vol. 2009, Article ID 421425, 19 pages, 2009. https://doi.org/10.1155/2009/421425

Zisopoulos, Charilaos Karagiannidis, Savvas Demirtsoglou, Georgios Antaris, Stefanos. (2008). "Content-Based Recommendation Systems."