A. Hyperparameter Tuning

Since MVECF is based on WMF, we need to consider the same set of hyperparameters of ordinary WMF. The hyperparameters of WMF are the latent dimension l, the regularization parameter λ, and the confidence level c_{ui} of each observation $y_{ui} = 1$.

We performed a grid search within the following ranges $l \in \{10, 30, 50\}$, $c_{ui} \in \{5, 10, 20, 40\}$, and $\lambda \in \{0.0001, 0.001, 0.01\}$. The evaluation criteria for choosing the hyperparameter was MAP@20. The final hyperparameter values chosen for our experiments are $l = 30$, $c_{ui} = 10$ and $\lambda = 0.001$. For SGD update of MVECF, we chose the learning rate α to 0.001, because the convergence was too slow for smaller learning rates and MVECF with large λ_{MV} did not converge in some of our datasets for larger learning rates.

For $\mathrm{BPR_{nov}}$ of [47], it is a modification of BPR model [28]. BPR uses matrix factorization to predict ratings by minimizing $\sum_{u,i,j \in \mathcal{D}_s} -\log \sigma(\hat{y}_{ui} - \hat{y}_{uj})$, where $\sigma(\cdot)$ is the logistic sigmoid function, and $\mathcal{D}_s$ is a set of triples (u, i, j) with $y_{ui} = 1$ and $y_{uj} = 0$. [47] defined a new set of triples $\mathcal{D}_s^{dist} = \{(u, i, j) : (u, i, j) \in \mathcal{D}_s \text{ and } \mathrm{dist}(i, j) < \tau\}$, where $\mathrm{dist}(i, j)$ is the distance between items i and j, and τ is a distance threshold. They trained $\mathrm{BPR_{nov}}$ model with (u, i, j) sampled from $\mathcal{D}_s^{dist}$ with probability β and sampled from $\mathcal{D}_s$ with probability $1 - \beta$ to recommend items that are distinct from the items in the user's preference history. Therefore, the hyperparameters of $\mathrm{BPR_{nov}}$ are τ, β, learning rate α, regularization parameter λ, and latent dimension l. We set latent dimension $l = 30$ and $\beta = 0.8$. We performed a grid search within range $\lambda \in \{0, 0.00001, 0.0001, 0.001\}$ and $\alpha \in \{0.0001, 0.001, 0.01\}$ for $\mathrm{BPR_{nov}}$. The chosen values are $\alpha = 0.001$ and $\lambda = 0.00001$. Finally, τ was set to 0.9 because it showed the best balance between novelty and recommendation performance among 0.8, 0.9, and 1.0.

For graph based ranking models (LightGCN, UltraGCN, and HCCF) and those models with new sampling methods, we used the same hyperparameters reported in each paper and set the learning rate to 0.001, and the embedding size to 32.

B. Train and Validation Loss

Figure 6 shows the train loss and the validation loss of $\mathrm{MVECF_{WMF}}$ and $\mathrm{MVECF_{reg}}$ in one of the yearly sub-datasets (year 2015) of CRSP data. We can easily see that both models converge quite smoothly. In particular, we can see that $\mathrm{MVECF_{WMF}}$ shows faster convergence compared to $\mathrm{MVECF_{reg}}$. It implies that restructuring MVECF into an ordinary WMF form makes enhancement in the computational efficiency. Similar results are found in all other datasets.

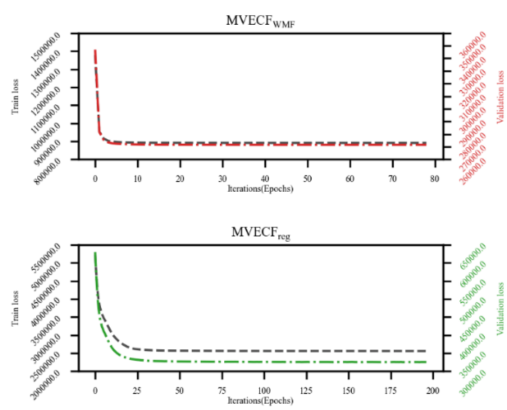

Figure 1. Train and validation losses of $\mathbf{MVECF_{WMF}}$ (top) and $\mathbf{MVECF_{reg}}$ (bottom).