Actively learning a Bayesian matrix fusion model with deep side information

Yangyang Yu & Jordan W. Suchow {yyu44,jws}@stevens.edu
School of Business
Stevens Institute of Technology
Hoboken, NJ 07030, USA

Abstract

High-dimensional deep neural network representations of images and concepts can be aligned to predict human annotations of diverse stimuli. However, such alignment requires the costly collection of behavioral responses, such that, in practice, the deep-feature spaces are only ever sparsely sampled. Here, we propose an active learning approach to adaptively sample experimental stimuli to efficiently learn a Bayesian matrix factorization model with deep side information. We observe a significant efficiency gain over a passive baseline. Furthermore, with a sequential batched sampling strategy, the algorithm is applicable not only to small datasets collected from traditional laboratory experiments but also to settings where large-scale crowdsourced data collection is needed to accurately align the high-dimensional deep feature representations derived from pre-trained networks. This provides cost-effective solutions for collecting and generating quality-assured predictions in large-scale behavioral and cognitive studies.

Keywords: Bayesian Matrix Factorization, Deep Learning, Active Learning

1. Introduction

In cognitive research, Bayesian probabilistic models typically serve two principal roles: one as a hypothesis positing how individuals draw inferences from their observations of the environment, and the other as a tool enabling scientists to learn from observations of human behavior (Vul et al. (2014); Griffiths et al. (2008); Mamassian et al. (2002)). Our work acts as an intermediate approach that bridges these two uses of Bayesian models. We use Bayesian Probabilistic Matrix Factorization (BPMF) with deep-side information to align a machine representation of entities to human behavioral responses to those entities, such that the model serves as both a model of people's mental representations and as a predictive model of their behavior. This method enables accurate inference of behavioral responses by generating low-rank predictions of perceptual response matrices. It offers a feasible model structure to align machine vision systems with human visual perception. For instance, it can integrate the bimodal information from facial imagery and psychological attributes and yield predictions of people's impressions of human faces.

BPMF has proven effective in consolidating multi-source information and predicting missing responses while constructing confidence intervals (Salakhutdinov and Mnih (2008); Adams et al. (2010)). In traditional laboratory experiments, it has been successfully applied to predict individuals' perceptual outcomes based on human interpretable features (Zhang et al. (2020)). Nevertheless, implementing BPMF in large-scale behavioral prediction tasks with multi-modal object features presents two primary challenges. First, high-quality predictions

require advanced machine-generated features outperforming traditional human-defined ones. Machine learning algorithms can generate a vast array of new objects based on prior ones, boosting diversity and realism while minimizing bias among stimuli. However, employing deep learning algorithms to extract informative, high-dimensional features from various objects linearly escalates computational costs with the number of data features. This complicates the integration of BPMF with deep learning methods, restricting it to sparse sampling from the deep-feature space. The second challenge arises from the scarcity of essential information for reliable predictions, owing to the highly sparse response matrix, a scenario exacerbated in human-subjects research. Here, data collection, constrained by budget and resources, needs to deal with large participant populations and potentially extensive question lists. Budgetary limits might restrict the number of questions posed, and elongated experimental instruments could yield inaccurate responses as participants may resort to mental shortcuts (Krosnick (1991)). Consequently, large-scale experiments often only capture responses to a minor portion of the total instrument from each participant. Thus, to effectively employ BPMF with deep side information on large-scale behavioral prediction tasks, developing a data sampling strategy to effectively target the most informative data points is critical, ensuring satisfactory predictive outcomes despite these constraints.

Active learning is a data acquisition technique that can interactively identify the most informative samples to efficiently create a training data set. Although this training set can be compact, it possesses a powerful predictive capacity. Active learning has been widely employed to tackle problems associated with accuracy in sparse matrix completion (Elahi et al. (2016); Chakraborty et al. (2013a)). One of its key strengths is the capacity to accurately infer the complete response distribution from a limited selection of samples, obviating the need to query the majority of the response matrix.

Here, we propose an active learning method for a BPMF model using uncertainty (Sugiyama and Ridgeway (2006)) and k-Center Greedy (Sener and Savarese (2017)) sampling strategies, showing enhanced learning efficiency compared to passive learning. We further examine the effect of varying Markov chain Monte Carlo(MCMC) simulation chain lengths on the active sampling performance to optimally integrate active learning into the BPMF model framework. The estimation of posterior uncertainty brings a cost associated with the number of posterior MCMC samples collected, presenting a trade-off between slow precise estimates and quick, less accurate ones. We investigate this trade-off by adjusting the number of MCMC samples for posterior uncertainty estimation in model parameters and measure its impact on algorithm performance within a fixed computational budget.

In this paper, we outline an active learning framework for the deep Bayesian matrix factorization model, train it on a large behavioral dataset—One Million Impressions dataset (Peterson et al. (2022)), and validate the model's learning efficiency and performance improvement via an effective active strategy. Lastly, we demonstrate our method's promising predictive results by adaptively querying a subset of the data for training.

2. Related work

We categorize the related work into three sections.

Bayesian Probabilistic Matrix Factorization: BPMF acts as a bridge between human cognition and mathematical inductive inference. Based on assumptions regarding cogni-

tive problems, it integrates prior knowledge from multiple channels, guiding participants to update their beliefs using mathematical forms. In the context of visual perception modeling, BPMF uses a matrix structure to combine 2-way side information - visual objects and observers' features, over priors. Through MCMC simulations, Bayesian posteriors are computed and linked to perceptual inference outcomes (Mamassian et al. (2002); Kersten et al. (2004)). A balance between the number of warm-ups and posterior samples is needed to avoid significant fluctuations and overkill in MCMC simulation performance (McElreath (2020)). It helps the BPMF achieve satisfactory predictive results through MCMC.

Modeling Perception Using Deep Features: Human perceptual tasks often benefit from multi-modal data fusion like visual and linguistic inputs. Prior research (Peterson et al. (2022); Zhang et al. (2018)) highlights machine-generated deep features' superiority over traditional human-interpretable attributes (e.g., image color or size) due to their comprehensive high-dimensional representations and self-generating capability for novel predictions. Utilizing pretrained networks like StyleGAN2 (Karras et al. (2020)) for images and Sentence-BERT (Reimers and Gurevych (2019)) for text, significantly lowers data-collection costs while enhancing prediction quality in the BPMF framework.

Active Learning: Inspired by sparse matrix completion research (Settles (2009); Chakraborty et al. (2013b)), we employ active learning to estimate a complete response distribution for each participant across diverse objects from a sparsely filled response matrix. This method expedites the alignment of machine inferences with human cognition by selectively querying the most informative data points, proving crucial in large-scale behavioral and cognitive experiments with limited resources.

3. Methods

The details of this approach and notation are articulated in two sections below.

3.1 Deep Bayesian Probabilistic Matrix Factorization

First, in predicting first impressions, pre-trained deep networks create face and trait features. Using indices j for face images, h for traits, and i for participants, Peterson et al. (Peterson et al. (2022)) extracted deep face features f_j via StyleGAN2, while we used Sentence-BERT for deep trait features t_h. This results in two latent spaces for response matrix R: a 512-dimensional image feature space $\mathcal{F}$ for faces and a 300-dimensional linguistic space $\mathcal{T}$ for traits, represented by conditioned, unit-variance, multivariate normal latent variables: ω_{f_j} and ω_{t_h} respectively. Second, under the BPMF setting, the latent variables represent the computational coefficients for the participants' impression ratings R_{jh}. The coefficients are estimates based on two priors of $\omega_{f_j}, \omega_{t_h}$ following spherical Gaussian distributions (Gurkan and Suchow (2022)). In Formulas 1 and 2, the two-sided features are first condense to a consistent dimensionality within the latent spaces and then merged through a fusion process. The fusion products are processed via Gaussian likelihood functions. Subsequently, the resulting predictions, denoted as $\hat{R}_{jh}$, are derived through MCMC simulations.

$$p(\omega_{f_j}|\sigma) = \text{Normal}(\omega_{f_j}|\sigma^{-1}\mathbf{I}); \quad p(\omega_{t_h}|\sigma) = \text{Normal}(\omega_{t_h}|\theta^{-1}\mathbf{I}), \tag{1}$$

$$F_j = f_j \times \omega_{f_j}^T; \quad T_h = t_h \times \omega_{t_h}^T; \quad R_{jh}^* = F_j \times T_h^T + \varepsilon_{jh}; \quad \hat{R}_{jh} = \text{sigmoid}(R_{jh}^*) \times 100, \ (2)$$

where σ and θ are independent and Gamma distributed. f_j and t_h are row vectors and $R_{jh}^* \in (-\infty, \infty)$ is projected into predicted ratings as a continuous value $\hat{R}_{jh} \in (0, 100)$.

3.2 Actively Learning The Matrix Factorization Model

In the BPMF training data sampling process, we implement active learning to minimize training matrix entries, predict remaining responses, and achieve model convergence. Initially, a small training pool, $\boldsymbol{S^0}$ of size L, is randomly selected from the entire rated data pool $\boldsymbol{S}$ of size N. With a budget of Q opportunities, we query an oracle for an extra p data points (set $\boldsymbol{S^p}$) to be added in the training pool, guided by a certain active learning algorithm $A_{\boldsymbol{S}}$ as depicted in Appendix E or F. The algorithm updates parameters $\omega_{f_j}, \omega_{t_h}$ based on $\boldsymbol{S^{0*}} = \boldsymbol{S^0} \cup \boldsymbol{S^p}$, minimizing future expected learning loss: $\min_{\boldsymbol{S_q^p}} E_{(\mathcal{C}_1,\ldots,\mathcal{C}_N)\sim\boldsymbol{S}}[\mathcal{L}(\mathcal{C}_n; A_{\boldsymbol{S^0}\cup\ldots\boldsymbol{S_q^p}})]$. Diverging from traditional active learning, $A_{\boldsymbol{S}}$ queries a batch of p data points each round, as single data point additions have minimal impact in large datasets (Sener and Savarese (2017); Zhang et al. (2020)). We extend active learning to continuous prediction tasks by reformulating learning metrics and loss functions. Two strategies are explored: one on uncertainty-based sampling and another balancing sampling diversity and uncertainty, with performance comparisons conducted.

Uncertainty Sampling targets the least confident samples in each active iteration, with Sugiyama et al. (Sugiyama and Ridgeway (2006)) advocating for the selection of data points with maximum predictive distribution standard deviations. We adapt this strategy to BPMF due to its applicability in continuous prediction scenarios with a univariate Gaussian prediction distribution, similar to Gaussian linear regression. The iterative selection of the p most uncertain samples, as detailed in Algorithm 1 (Appendix E) and Formula 3 (Appendix B), enhances the training pool's informativeness. **k-Center Greedy sampling** is a pool-based active strategy where p points representing cluster centers of the entire data distribution are chosen from the unexplored data pool during each adaptive sampling cycle (Sener and Savarese (2017)). In our continuous target prediction scenario, the Bayesian model primarily manages generalization and training loss, while the active learning strategy aims to minimize core-set loss, comparing average empirical loss over known and unknown rating entries. Optimization specifics are elaborated in Appendix F.

4. Experiments and Results

We executed two experimental sets to evaluate the potential of integrating an active learning strategy for training the deep BPMF model, focusing on the optimal active learning strategy and batch size, and exploring if extending BPMF simulation steps in passive learning could surpass the active learning benefits. The experimental dataset is One Million Impressions dataset (Peterson et al. (2022)), which comprises over 1 million impression ratings on 34 traits for 1,000 machine-generated face images. Given that each image only receives ratings from a limited group of participants, the response matrix is relatively sparse. We assessed

the active learning method's effectiveness and efficiency using the test Root Mean Square Error (RMSE) averaged across three experimental repetitions.

4.1 Choices of Strategy and Batch Size

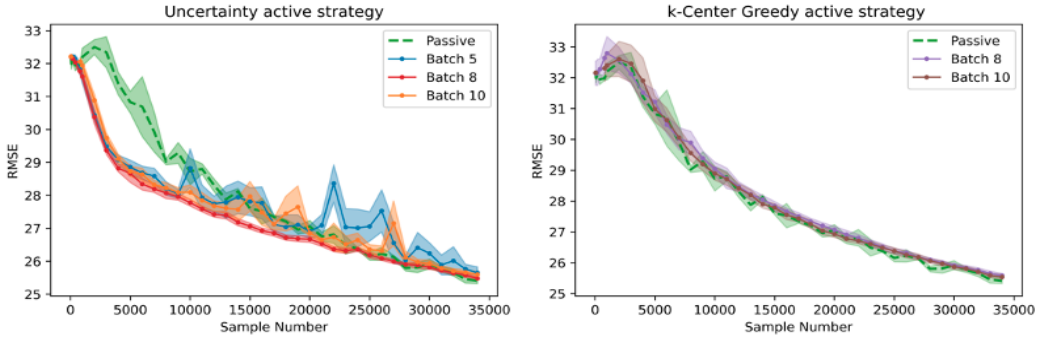


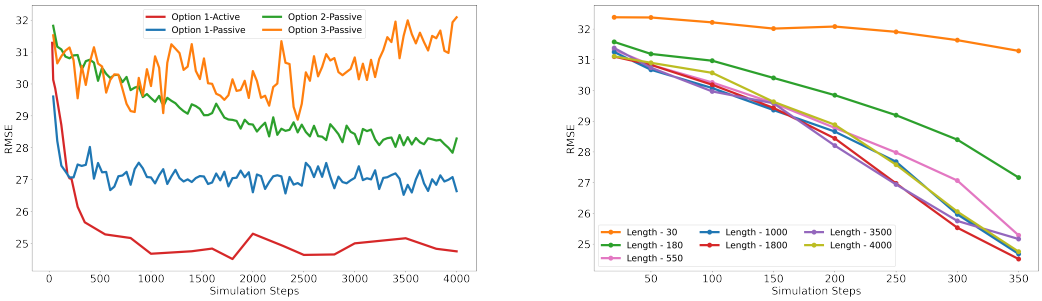
Figure 1: Test RMSEs over sample number with different batch sizes for uncertainty vs. k-Center Greedy active strategies with 95% confidence intervals.[1]

The first experiment investigates optimal batch size for large-scale active learning using uncertainty and k-Center Greedy strategies in deep BPMF for predicting impression ratings. Five distinct models with different active learning strategies and batch size combinations are run concurrently, contrasting with a passive baseline model. Figure 1 shows uncertainty strategy outperforming others. A batch size of 8 is optimal, showing the quickest RMSE reduction, notably beneficial under 10,000 samples, roughly 0.1% of the total dataset, where test RMSE drops from 32.1 to 27.8, about 1.0 lower than passive learning's test RMSE. Conversely, k-Center Greedy doesn't exceed the baseline in any scenario. After reaching 28,000 samples, batched uncertainty models converge with the baseline. This confirms the batched uncertainty strategy's effectiveness in reducing predictive RMSE with limited data, emphasizing the impact of batch size on prediction stability.

4.2 Impact of Simulation Chain Length

The second experiment aims to determine if extending the MCMC simulation chain in passive learning can match the efficiency gain of active learning. We tested three MCMC configurations: (1) incrementing both warm-ups and posterior samples in a 3:5 ratio by 8 samples; (2) extending posterior samples by 5 steps without warm-ups; and (3) increasing warm-ups by 5 steps with constant 5-step posterior samples. Using a 1000-sample subset from the same dataset comprising ten random responses from each combination of ten faces and ten traits, we compared predictive RMSE on 650 test data points to assess model performance. We employed the prior best strategy – uncertainty sampling, on the BPMF, adjusting the batch size to two due to the dataset size reduction.
First, testing three passive-sampling BPMF models revealed that Option (1) efficiently reduced predictive RMSE with increasing simulation steps (Figure 2(a)). Applying this optimal simulation chain to an active sampling BPMF model significantly improved perfor-

1. To smooth out fluctuations, each dot in Figures 1 is averaged from 200 samples of the adjacent epochs.



(a) Test RMSEs for different model options over the lengths of the simulation chain with 350 training samples

(b) Test RMSEs for the BPMF using batch-2 active strategy with different simulation chain lengths over training sample numbers.

Figure 2: Result summary of the second experiment. [2]

mance with fewer steps compared to passive models. By 220 steps, the active model's RMSE dropped to 27.024, outperforming the best performed passive model. The training time analysis (Table 1 in Appendix D) highlighted the sustained superiority of the active strategy, not offset by extending the simulation chain in passive models. Second, with 350 queried samples, the active strategy's performance peaked with a 1,800-step simulation chain, maintained throughout the adaptive sampling process (Figure 2(b)). The 1,800-step model's test RMSE remained low as the training set expanded from 50 to 350 samples, indicating consistent performance across varying training sizes.

5. Conclusion and Discussion

In this work, we propose an active learning method using BPMF with deep neural networks to predict human behavioral data, which selects informative stimulus-attribute pairs based on model parameter uncertainty. Our empirical tests demonstrate superior performance over passive learning, crucial for budget-limited crowdsourced studies and applicable across domains like online recommendations for social media or e-commerce. Key factors affecting performance include sampling strategy, active learning batch size, and simulation chain lengths. Our experiments highlight the effectiveness of the uncertainty sampling strategy over the k-Center Greedy strategy, especially in large behavioral datasets. Despite a smaller training pool, Bayesian inference efficiently learns from less confident data, reducing predictive error. However, k-Center Greedy sampling fails to capture feature diversity effectively. Our findings suggest that an appropriate active strategy and batch size, along with an optimized simulation chain length in the BPMF model, significantly enhance predictive performance, especially under data query budget and computational resource constraints.
The integration of deep Bayesian matrix factorization with the uncertainty active strategy is promising for impression prediction, using deep image and trait features to generate a 2D response matrix. Given Bayesian factorization's capability with 3D tensors (Xiong et al. (2010)), there's interest in exploring our approach's adaptability to more complex behavioral datasets with diverse features.

2. To smooth out fluctuations, the dots in Figures 2 are averaged from 5 adjacent epochs.

Appendix A.Queried data samples in k-Center greedy active learning strategy.

The following graph Figure 3 depicts p centers and their associated coverage radii $d_{\boldsymbol{S^0 \cup S^p}}$ during one batch selection of k-Center Greedy active sampling. This representation is utilized to illustrate one essential component of the upper bound of Core-set loss.

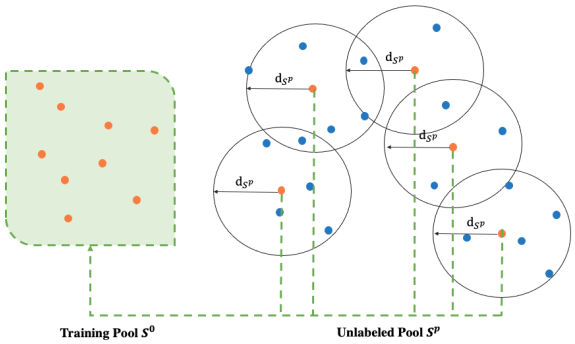


Figure 3: Queried data points and $d_{S^0 \cup S^p}$ in one batch.

Appendix B.Batched sample selection formula of uncertainty active learning strategy.

$$\boldsymbol{S^{p*}} = \underset{\mathcal{C}_n \in \boldsymbol{S} \setminus \boldsymbol{S^p}}{\operatorname{argmax}} \sum_{h=1}^{p} \sigma_{\omega^*_{\mathcal{U}_l q}}(R^* | \rho(R^*)) \tag{3}$$

, where the standard deviations of predictive distribution is $\sigma_{\omega^*_{\mathcal{U}_l}}$.

Appendix C.Extension of the mathematical details to formulate K-center Greedy active learning strategy.

Sener et al. (Sener and Savarese (2017)) regard active learning loss for classifications using CNN given a batch of p samples, $E_{(\mathcal{C}_1,\ldots,\mathcal{C}_N)\sim \boldsymbol{S}}(\mathcal{L}(\mathcal{C}_n; A_{\boldsymbol{S^p}}))$, as containing three components: generalization loss, training loss, and core-set loss. p denotes the number of central points in the unexplored data pool and also equates to the count of points chosen during each adaptive sampling cycle. In our scenario, which involves predicting continuous targets, the first two losses are primarily managed by the Bayesian model, while the active learning strategy concentrates on minimizing core-set loss. Core-set loss is defined as the distance between average empirical loss over the points with known ratings and that over the entire dataset, including entries with unknown ratings. Sener et. al shows its upper bound as $\mathcal{O}(d_{\boldsymbol{S^0 \cup S^p}}) + \mathcal{O}(\sqrt{\frac{1}{N}})$. Consequently, the optimization goal of loss can be converted to minimize the coverage radius from p centers ($d_{\boldsymbol{S^0 \cup S^p}}$), as illustrated in Figure 3 (in Appendix A).

The subsequent Formulas 4, 5, and 6 illustrate further deductions of the batch selection criteria, informed by the knowledge that the upper bound of the Core-set loss primarily consists of the component $\mathcal{O}(d_{\boldsymbol{S^0 \cup S^p}})$. The task of finding the upper bound for Formula 4

is equivalent to the k-Center problem presented in Formula 5, which is described as the min-max facility location problem in Wolf's work (Wolf (2011)).

$$\begin{aligned} &\frac{1}{N}\sum_{n\in \boldsymbol{S}} \mathcal{L}(\mathcal{C}_n; A_{\boldsymbol{S}^0\cup \boldsymbol{S}^p}) - \frac{1}{L}\sum_{l\in \boldsymbol{S}^p} \mathcal{L}(\mathcal{C}_l; A_{\boldsymbol{S}^0\cup \boldsymbol{S}^p}) \\ &\leq \mathcal{O}(d_{\boldsymbol{S}^0\cup \boldsymbol{S}^p}) + \mathcal{O}(\sqrt{\frac{1}{N}}) \end{aligned} \tag{4}$$

$$\min_{\boldsymbol{S}^p} d_{\boldsymbol{S}^0\cup \boldsymbol{S}^p} \simeq \min_{\boldsymbol{S}^p} \max_{\mathcal{C}_n\in S\setminus(\boldsymbol{S}^0\cup \boldsymbol{S}^p)} \min_{\mathcal{C}_l\in \boldsymbol{S}^0\cup \boldsymbol{S}^p} \triangle(\mathcal{U}_n, \mathcal{U}_l) \tag{5}$$

$$\boldsymbol{S^{p*}} = \sum_{l=1}^{p} \underset{\mathcal{C}_l\in \boldsymbol{S}\setminus \boldsymbol{S}^p}{\operatorname{argmax}} \min_{\mathcal{C}_L\in \boldsymbol{S}^p} \triangle(\mathcal{U}_l, \mathcal{U}_L) \tag{6}$$

The minimization of $d_{\boldsymbol{S^0}\cup \boldsymbol{S^p}}$ is further deduced as $\min\limits_{\boldsymbol{S^p}} \max\limits_{\mathcal{C}_n\in \boldsymbol{S}\setminus(\boldsymbol{S^0}\cup \boldsymbol{S^p})} \min\limits_{\mathcal{C}_l\in \boldsymbol{S^0}\cup \boldsymbol{S^p}} \triangle(\mathcal{U}_n, \mathcal{U}_l)$ when p data points are queried.This formula can be computed based on the distances of feature pairs $\mathcal{U}_n = (f_{jn}, t_{hn})$ and $\mathcal{U}_l = (f_{jl}, t_{hl})$ in a two-dimensional coordinate system. The $\min\limits_{\boldsymbol{S^p}} \boldsymbol{S^0}\cup \boldsymbol{S^p}$ is iteratively recomputed in Algorithm 2 to search for new data points to be queried.

Appendix D.Active learning vs. passive learning training time and test RMSE in the third experiment.

This training time analysis was run on a single NVIDIA GeForce RTX 4090 GPU with 24 GB memory.

MCMC chain length	Learning Type	Running time (minutes)	Test RMSE
220	Active	32.4633	27.024
280	Active	36.1294	26.1506
64,000	Passive	29.3237	27.4982
80,000	Passive	35.8137	27.7274
88,000	Passive	39.5863	27.1459

Table 1: Active learning vs. passive learning training time and test RMSE in the third experiment.

Appendix E.Pseudocode of Batched Uncertainty Sample Selection for Deep Bayesian Matrix Factorization.

Algorithm 1 Batched Uncertainty Sample Selection for Deep Bayesian Matrix Factorization

Input: The set of randomly selected d ratings as initial pool $\boldsymbol{S^0}$; The rating R_{il} in $\boldsymbol{S^0}$ is given by a participant i for a face image in terms of a certain trait. And budget $Q > 0$.

Output: Predicted ratings for all entries of the response matrix, $\hat{R}$

Extract deep features face images f_{jl} and trait t_{hl}

repeat Q times, initialize $q = 1$;

Update parameters $\omega^*_{f_j\,q}, \omega^*_{t_h\,q}$ via MCMC

Compute predicted $\hat{R}_q$, σ_q for the response matrix

Query $\boldsymbol{S^p_q}$ with argmax $\sum_{z=1}^{p}\ {\sigma_{zq}}^3$

$\boldsymbol{S^0_q} \leftarrow \boldsymbol{S^p_q} \cup \boldsymbol{S^0_{q-1}}$

q = q + 1

until $q = Q$

return $\hat{R}_Q$, $\boldsymbol{S} \setminus \boldsymbol{S^0_Q}$

Appendix F.Pseudocode of k-Center Greedy Sample Selection for Deep Bayesian Matrix Factorization.

Core-set loss is defined as the distance between average empirical loss over the points with known ratings and that over the entire dataset, including entries with unknown ratings. Sener et. al shows its upper bound as $\mathcal{O}(d_{\boldsymbol{S^0}\cup\boldsymbol{S^p}})+\mathcal{O}(\sqrt{\frac{1}{N}})$. Consequently, the optimization goal of loss can be converted to minimize the coverage radius from p centers ($d_{\boldsymbol{S^0}\cup\boldsymbol{S^p}}$), as illustrated in Figure 3 (in Appendices A and C). The minimization of $d_{\boldsymbol{S^0}\cup\boldsymbol{S^p}}$ is further deduced as $\min\limits_{\boldsymbol{S^p}} \max\limits_{\mathcal{C}_n\in\boldsymbol{S}\setminus(\boldsymbol{S^0}\cup\boldsymbol{S^p})} \min\limits_{\mathcal{C}_l\in\boldsymbol{S^0}\cup\boldsymbol{S^p}} \triangle(\mathcal{U}_n,\mathcal{U}_l)$ when p data points are queried.This formula can be computed based on the distances of feature pairs $\mathcal{U}_n = (f_{jn}, t_{hn})$ and $\mathcal{U}_l = (f_{jl}, t_{hl})$ in a two-dimensional coordinate system. The $\min\limits_{\boldsymbol{S^p}} \boldsymbol{S^0} \cup \boldsymbol{S^p}$ is iteratively recomputed in Algorithm 2 to search for new data points to be queried.

3. $\sigma_{z\,q}$ denotes the predicted standard deviation for the zth queried data point in a batch p samples during the qth iteration of the uncertainty strategy
4. $\boldsymbol{S}^p_{h\,q}$ denotes the kth queried data point in a batch p in the qth iteration of k-Center Greedy active strategy.

Algorithm 2 k-Center Greedy Sample Selection for Deep Bayesian Matrix Factorization

Input: The set of randomly selected d ratings as initial pool $\boldsymbol{S^0}$; the rating R_{il} in $\boldsymbol{S^0}$ is given by a participant i for a face image in terms of a certain trait. And budget $Q > 0$.

Output: Predicted ratings for all the entries of the response matrix, $\hat{R}$

Extract deep features face images f_{jl} and trait t_{hl}

repeat Q times, initialize $q = 1$;

Update parameters $\omega^*_{f_j q}, \omega^*_{t_h q}$ via MCMC

Set p learning centers

Choose p centers $(\mathcal{C}_{1q}, ..., \mathcal{C}_{p_q}) = S_q^p$ [4]by $\boldsymbol{S_q^p} = \underset{\mathcal{C}_{nz} \in S \setminus S_{zq}^p}{\text{argmax}} \min_{\mathcal{C}_{lz} \in S_{zq}^p} \triangle(\mathcal{U}_{nz}, \mathcal{U}_{lz})$, where set $\boldsymbol{S_q^p} \subseteq \boldsymbol{S} \setminus \boldsymbol{S_{q-1}^0}$

$\boldsymbol{S_q^0} \leftarrow \boldsymbol{S_{q-1}^0} \cup \boldsymbol{S_q^p}$

q = q + 1

until $q = Q$

return $\hat{R}_Q$, $\boldsymbol{S} \setminus \boldsymbol{S_Q^0}$

References

Ryan Prescott Adams, George E Dahl, and Iain Murray. Incorporating side information in probabilistic matrix factorization with gaussian processes. *arXiv preprint arXiv:1003.4944*, 2010.

Shayok Chakraborty, Jiayu Zhou, Vineeth Balasubramanian, Sethuraman Panchanathan, Ian Davidson, and Jieping Ye. Active matrix completion. In *2013 IEEE 13th International Conference on Data Mining*, pages 81–90, 2013a. doi: 10.1109/ICDM.2013.69.

Shayok Chakraborty, Jiayu Zhou, Vineeth Balasubramanian, Sethuraman Panchanathan, Ian Davidson, and Jieping Ye. Active matrix completion. In *2013 IEEE 13th International Conference on Data Mining*, pages 81–90, 2013b.

Mehdi Elahi, Francesco Ricci, and Neil Rubens. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20:29–50, 2016.

Thomas L Griffiths, Charles Kemp, and Joshua B Tenenbaum. Bayesian models of cognition. In Ron Sun, editor, *The Cambridge Handbook of Computational Psychology*. Cambridge University Press, New York, 2008.

Necdet Gurkan and Jordan Suchow. Cultural alignment of machine-vision representations. In *NeurIPS 2022 SVRHM Workshop*, 2022.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8107–8116, 2020.

Daniel Kersten, Pascal Mamassian, and Alan Yuille. Object perception as Bayesian inference. *Annual Review of Psychology*, 55:271–304, 2004.

Jon A Krosnick. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3):213–236, 1991.

Pascal Mamassian, Michael Landy, and Laurence T Maloney. Bayesian modelling of visual perception. *Probabilistic models of the brain: Perception and neural function*, pages 13–36, 2002.

Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, Boca Raton, FL, 2020.

Joshua C Peterson, Stefan Uddenberg, Thomas L Griffiths, Alexander Todorov, and Jordan W Suchow. Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences*, 119(17):e2115228119, 2022.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-Networks. arXiv preprint arXiv:1908.10084, 2019.

Ruslan Salakhutdinov and Andriy Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th international conference on Machine learning*, pages 880–887, 2008.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489, 2017.

Burr Settles. Active learning literature survey. Technical Report 1648, University of Wisconsin-Madison Department of Computer Sciences, 2009.

Masashi Sugiyama and Greg Ridgeway. Active learning in approximately linear regression based on conditional expectation of generalization error. *Journal of Machine Learning Research*, 7(1):141–166, 2006.

Edward Vul, Noah Goodman, Thomas L Griffiths, and Joshua B Tenenbaum. One and done? Optimal decisions from very few samples. *Cognitive Science*, 38(4):599–637, 2014.

Gert W Wolf. Facility location: concepts, models, algorithms and case studies. series: Contributions to management science: edited by zanjirani farahani, reza and hekmatfar, masoud, heidelberg, germany, physica-verlag, 2009, 549 pp.,€ 171.15, 219.00, £144.00, $isbn978-3-7908-2150-5(hardprint), 978-3-7908-2151-2(electronic)$, 2011.

Liang Xiong, Xi Chen, Tzu-Kuo Huang, Jeff Schneider, and Jaime G Carbonell. Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 211–222, 2010.

Chelsea Zhang, Sean J Taylor, Curtiss Cobb, and Jasjeet Sekhon. Active matrix factorization for surveys. *The Annals of Applied Statistics*, 14(3):1182–1206, 2020.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.