# About Project

**Objective:** The aim of the project is to understand the implementation of Variational Auto Encoder(VAE) for Collaborative filtering based on existing research paper and re-implement the algorithm with another dataset which is not used in earlier implementation.

**Research Paper**: [Enhancing VAEs for Collaborative Filtering: Flexible Priors & Gating Mechanisms](https://arxiv.org/ftp/arxiv/papers/1911/1911.00936.pdf)

**Implementation Source** Papers with code- [Github](https://github.com/psywaves/EVCF)

**Domain** Recommendation System, Collaborative Filtering

**Datasets Considered:**

* [Netflix](https://www.kaggle.com/netflix-inc/netflix-prize-data/downloads/netflix-prize-data.zip/1)
* [Amazon Movies data](https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/)
* [Food Recipe Review Data](https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions)
* [MUBI Platform Data](https://www.kaggle.com/datasets/clementmsika/mubi-sqlite-database-for-movie-lovers)

**Goal**: To understand VAE, Collaborative Filtering, Recommendation Engine Development.

**Problem Definition**:

Collaborative Filtering utilizes similarity among the user preferences to provide recommendations. Traditionally linear models having latent factor dominates in terms of usage. There is limitation due to modeling capacities and inability to model complex relationships.

**Problem Objective**:

The objective of implementing VAE for collaborative filtering is to generalize the linear latent factor models and explore the non-linear probabilistic relationships using neural networks. The considered paper provides insights to enhance the performance of neural network-based models for collaborative filtering by addressing two problematic characteristics of the current state-of-the-art variational autoencoder for CF: the too simplistic prior that restricts the model's ability to fully capture user preferences, and the model's inability to learn deeper representations with more than one hidden layer for each network.

# Research Paper Review

**1. What is the problem the selected article is trying to solve and why it is important to study that problem?**

Top of Form

The selected paper titled "Enhancing VAEs for Collaborative Filtering: Flexible Priors & Gating Mechanisms" aims to improve the accuracy of collaborative filtering. The paper addresses the limitations of traditional Variational Autoencoders (VAEs) that are used in collaborative filtering, such as the usage of simplistic prior for learning latent representation of user preference and model’s lack of ability to learn complex representation with more than one hidden layer. The authors propose a new model that introduces flexible priors and gating mechanisms to enhance the VAEs. The significance of the study is that it addresses the challenges of modeling user preferences and learning better representation.

**2. What is novel in this paper compared to what came before it. To accomplish that please select an article cited in the references and compare and contrast to it.**

The paper proposes a novel approach to enhance VAEs in collaborative filtering by introducing flexible priors and gating mechanisms. The authors compare their approach with previous work, such as the work of Liang et al. (2018) titled "Variational Autoencoders for Collaborative Filtering."

Both the papers proposed Variational Autoencoder for collaborative filtering in recommendation system providing a better alternative from traditional linear modelling techniques used for collaborative filtering.

The earlier paper introduced Variational Autoencoder for collaborative filtering along with implementation of multinomial likelihood distribution instead of logistic or Gaussian likelihood for loss optimization. They have considered adjusting the regularization and proposed a novel approach for hyper parameter tuning to achieve better results.

The selected paper has provided an extension over the earlier paper introducing flexible priors and gating mechanisms to enhance the VAEs. Their motivation was to learn richer latent representation and learn complex representation form user-item interaction history.

**3. What have been the impact and/ or limits in the scope of the application of the approach proposed in the selected article. Select an article that cited this article and describe it in context of that. You can use scholar.google.com to get a list of articles that cited the article.**

The article has been cited more than 30 times. Above is the graph of from the connected papers site showing the interdependence of paper with related papers. We can observe more research is happening in the domain with advanced neural network for recommendation. The authors have not mentioned any major limitations and have not discussed the scope for further research. The authors do discuss their observations of how adding more layers without gates did not improve and provided evidence that addition of layer with gates might further improve the accuracy. There is one paper with title “Adversarial and Contrastive Variational Autoencoder for Sequential Recommendation” explores another use case of the domain of recommender system. The paper attempts to tackle the common limitation that the representational ability of the obtained approximate posterior distribution is limited, resulting in lower quality of generated samples. By capturing temporal dependencies among items in user sequence and perform sequential recommendation Adversarial and Contrastive VAE-based models the authors have presented their novel approach.

Below image can be accessed by this link.

Source: <https://www.connectedpapers.com/main/d563427eba3545845bd60fea1a491c3be199bffa/Enhancing-VAEs-for-collaborative-filtering%3A-flexible-priors-%26-gating-mechanisms/graph>

Diagram

Description automatically generated

4. what is the broader implications of the work?

The proposed approach has broader implications beyond the field of recommender systems. Flexible priors and gating mechanisms can be applied to other domains, such as natural language processing, image recognition, and speech recognition. The interpretability of the model is also crucial for decision-making processes in various fields, such as finance and healthcare. Therefore, the proposed approach has the potential to impact various industries and domains.

# Basics of Recommendation Systems

1. **What is a Recommendation System?**

Recommendation System uses past behavior of an user as data to recommend things for future use of user. It can be used for selling relevant products to customers based on their needs and preferences.

1. **What type of data is considered relevant for Recommendation System?**

User Likes and dislikes for particular use case can be considered as relevant data.

1. **What are the different types of recommendations system?**

There are the various type of Recommendation system, some of the most prominent are mentioned below:

* Market Basket Analysis (Association Rule Mining Based)
* Content Based Filtering
* Collaborative Based Filtering
* Machine learning techniques: Clustering and Classification
* Deep Learning Techniques: Natural Language Processing

1. **Explain about the working of Content based filtering and Collaborative Based filtering techniques in recommendation system?**

**Content Based filtering**

Provides recommendation of items similar to the ones the user have previously selected or shown interest in. It depends on the content of the items such as a book’s inside content can determine can it be recommended to user or not.

It used Profile of user vector and Item vector containing user and item details respectively.

Similarity functions such as cosine similarity are used over the vectors to calculate score for items to recommend.

Cons of the content based filtering approach

The recommendation usually fall into similar categories and overtime becomes monotonous.

If the user is new user there would not be any accurate recommendation.

**Collaborative Based filtering**

Provides recommendations to user based on user-to-user similarity. It provides recommendation USER A based on the interests of similar user B.

User Item Matrix is created for individual user.

There are multiple ways of implementing collaborative filtering such as item-item collaborative filtering or user-user based collaborative filtering.

Cons of Collaborative filtering

Fails to perform well when there is lack of data to learn the relation b/w users & items.

1. **What is Auto Encoder?**

An autoencoder is a type of neural network that is used for unsupervised learning of efficient data encodings. It learns to compress the input data into a lower-dimensional representation, and then reconstruct the output data from that representation. Autoencoders consist of two main parts: an encoder and a decoder. The encoder takes the input data and maps it to a lower-dimensional representation, while the decoder takes that representation and maps it back to the original data space. The network is trained by minimizing the difference between the input and the reconstructed output using backpropagation. Autoencoders have a wide range of applications, such as data compression, anomaly detection, and image denoising.

1. **What is VAE and how it different compared to AE?**

VAE stands for Variational Autoencoder, which is a type of neural network architecture that is designed to generate new data samples similar to the ones it was trained on. It is a type of autoencoder that is based on probabilistic modeling, which means it learns the distribution of the data and then generates new samples from that distribution.

The key difference between VAE and a traditional autoencoder (AE) is that VAEs learn a probability distribution over the latent space, whereas AEs map input data to a fixed-dimensional encoding. In other words, VAEs are generative models that learn the underlying structure of the data and can create new samples from that structure, whereas AEs are usually used for encoding data and not for generating new samples.

The encoding and decoding process in a VAE is similar to that of an AE, but with the addition of a latent variable z. The encoder maps the input data x to a probability distribution over z, and the decoder maps a sample from that distribution back to the original input space. During training, the VAE minimizes the reconstruction error between the original input and the decoded output, as well as the Kullback-Leibler divergence between the learned distribution and a prior distribution. This encourages the VAE to learn a compact and smooth latent representation of the data that can be used for generative purposes.

In summary, the main difference between VAE and AE is that VAEs learn a probability distribution over the latent space, which enables them to generate new samples, whereas AEs simply learn an encoding of the input data.

1. **How VAE can be used for collaborative filtering?**

AE can be used for collaborative filtering by learning a low-dimensional representation of user-item interactions that captures the underlying preferences of users and the attributes of items. The idea is to train a VAE on a large dataset of user-item interactions to learn a mapping from the high-dimensional space of user-item interactions to a lower-dimensional space of latent variables. The learned latent variables can then be used to make personalized recommendations for each user by computing the similarity between their latent representation and the representations of the items in the dataset. By using VAEs for collaborative filtering, it is possible to model complex user-item interactions in a principled way, while also avoiding some of the common problems associated with traditional collaborative filtering methods, such as sparsity and cold-start.

# Analysis & Pre-Processing

Based on the existing source code. The data preparation code for Netflix dataset was rerun following are the initial data :

Netflix

The Data contains three column customer id, Movie Id and Rating.

Rating Range from 1-5

A screenshot of a computer

Description automatically generated with medium confidence

To focus on affinity of users only rating more than 3 were considered and “movie\_id” were considered as binary feature.

To filter out outliers and irrelevant data following two filters were implemented:

1. users who viewed at least one movie items.
2. Only keep those movies which has been viewed by at least 5 users.

After filtering, there are 9999011 watching events(entries) from 84154 users and 17631 movies (sparsity**: 0.674)**

Amazon Reviews

The dataset has columns as : 'overall', 'verified', 'reviewTime', 'reviewerID', 'asin', 'style', 'reviewerName', 'reviewText', 'summary', 'unixReviewTime', 'vote', 'image'

To maintain consistency only three columns were kept and mapped to new names as:

* Overall -> Rating
* reviewerID -> customerId
* asin -> MovieId

A screenshot of a computer

Description automatically generated with medium confidence

Based on the above distribution of rating, I inferred ratings were very skewed to higher rating indicating only higher rated items were given reviews and further analysis might hinder development of recommendation as the dataset is very homogenous.

Food Review Dataset

The dataset has columns: user\_id,recipe\_id,date,rating,u,i

To maintain consistency only three columns were kept and mapped to new names as:

* User\_id -> Customer\_id
* Recipe\_id -> Movie\_id
* Rating-> Rating

The distribution of rating is as follows:

A screenshot of a computer

Description automatically generated with medium confidence

Based on the above distribution of rating, I inferred ratings were very skewed to higher rating indicating only higher rated items were given reviews and further analysis might hinder development of recommendation as the dataset is very homogenous.

MUBI Dataset

The dataset has columns: movie\_id,rating\_id,rating\_url,rating\_score,rating\_timestamp\_utc,critic, critic\_likes,critic\_comments,user\_id,user\_trialist,user\_subscriber,user\_eligible\_for\_trial, user\_has\_payment\_method

To maintain consistency only three columns were kept and mapped to new names as:

* User\_id -> Customer\_id
* Movie\_id -> Movie\_id
* Rating\_Score-> Rating

The distribution of Rating is as follows:

. A screenshot of a computer

Description automatically generated with low confidence

To focus on affinity of users only rating more than 3 were considered and “movie\_id” were considered as binary feature.

To filter out outliers and irrelevant data following two filters were implemented:

1. users who viewed at least one movie items.
2. Only keep those movies which has been viewed by at least 5 users.

After filtering, there are 8412310 watching events from 130114 users and 94496 movies (sparsity: 0.068). Sparsity is very low compared to Netflix Database of 0.68.

The considered databases has the following data for their columns:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Database | Customers | Items | Rating entry | Item/Customer Ratio |
| Netflix | 17,770 | 480,189 | 100,480,507 | 27.02 |
| Amazon Movies | 182,032 | 3,826,085 | 8,765,568 | 21.01 |
| Food Review | 178,265 | 25,076 | 718,379 | 0.14 |
| MUBI | 142,698 | 451,757 | 15,519,989 | 3.16 |

We can clearly observe the data distribution of the databases is way more different.

# Methodology & Results

* The tabular present with three columns Movie\_Id, Customer\_Id and Ratings are reduced two columns by considering only the entries which have higher rating as 4and 5.
* The tabular structure is converted to matrix form with Customer\_ID as rows and Movie\_ID as binary feature. Each data cell with 1 value indicates Customers inclination of preferring that particular movie.
* The Matrix itself will be sparse matrix with high numbers of columns based on the count of movies considered.
* Traditionally the matrix is used to identify similarity between users(rows) by taking cosine similarity and similar user can be recommended with items taken by one and not taken by another.
* The referenced repository contains four different implementations of VAE as:
  + Mult-VAE (<https://arxiv.org/abs/1802.05814>)
  + Vamp
  + H+Vamp model with 1 hidden layer must use this model
  + H+Vamp model with 2 or more hidden layers use this model
* The common thing among the implementation is the presence of gated mechanism when the non-layer neural layer is defined with activation function like tanh are provide as arguments.

Model with following parameters is initialized for Netflix Dataset:

activation=None,

batch\_size=200,

cuda=False,

dataset\_name='netflix',

dynamic\_binarization=False,

early\_stopping\_epochs=50,

epochs=400, gated=False,

hidden\_size=600,

input\_size=[1, 1, 7738],

input\_type='binary',

lr=0.0005, max\_beta=1.0,

model\_name='vamp',

model\_signature='2023-05-12'

, no\_cuda=False,

num\_layers=1,

number\_components=1000,

pseudoinputs\_mean=0.05,

pseudoinputs\_std=0.01,

seed=14,

test\_batch\_size=1000,

Component/ Parameters of Models can be understood as below:

A picture containing text, screenshot, font, menu

Description automatically generated

Training and Validation Result

A screenshot of a computer screen

Description automatically generated with low confidence

We can observe clear decrease in the defined loss function. Loss here is used to understand the similarities/difference between the sampled data from the encoded inputs against the actual input provided to the Auto Encoder architecture. RE is another parameter which is used as metric to evaluate the reconstruction error during the sample decoding process. KL divergence another parameter use for understanding the difference between the reconstructed output and the actual output. Based on the above training log VAE model with beta values of 0.01 and 0.02 over 2 epochs. The model is being trained on a collaborative filtering task, where the goal is to recommend items to users based on their past interactions.

In each epoch, the model is trained on a training set and its performance is evaluated on a validation set. The train loss and validation loss are reported, along with their respective components: the reconstruction error (RE) and the Kullback-Leibler (KL) divergence.

The beta value controls the contribution of the KL divergence term to the loss function. In the first epoch, beta is set to 0.01 and the model achieves a train loss of 202.76 and a validation loss of 105.93. The best NDCG score achieved on the validation set is 0.13753.

In the second epoch, beta is increased to 0.02 and the model achieves a lower train loss of 114.04 and a lower validation loss of 94.55. The best NDCG score achieved on the validation set improves to 0.14973. The early stopping criterion is not met, meaning the training can continue.

Overall, the VAE is being trained to generate latent representations of user preferences, which can be used for personalized recommendations. The goal is to balance between accurately reconstructing the input and encouraging the model to learn diverse representations of user preferences.

# Discussion & Reflection

From a machine learning perspective, the use of a VAE for collaborative filtering presents several advantages. The VAE framework allows us to model the inherent uncertainty and diversity in user preferences by learning a distribution over latent representation. This captures the complexity of user-item interactions and helps overcome the limitations of traditional collaborative filtering methods.

The training log demonstrates the importance of striking a balance between reconstruction accuracy (RE) and regularization (KL divergence). The beta parameter plays a crucial role in controlling this trade-off. By adjusting the beta value, we can effectively control the level of emphasis on the regularization term and guide the model to learn more diverse and informative latent representations.

Furthermore, the improvement in NDCG scores indicates that the VAE model with a higher beta value produces more accurate and relevant recommendations. This aligns with the goal of collaborative filtering, where personalized and high-quality recommendations are desired.

However, it is important to note that the results may not be generalizable to all datasets and recommendation scenarios. The effectiveness of the VAE model heavily relies on the specific characteristics of the dataset, the quality of the input data, and the choice of hyperparameters.

In conclusion, the training log and the overall use of VAE for collaborative filtering highlight the potential of generative models in capturing complex user preferences and generating personalized recommendations. It showcases the importance of balancing reconstruction accuracy and regularization to achieve improved recommendation performance.

Gating mechanisms have gained significant attention in various neural network architectures, including Variational Autoencoders (VAEs) for collaborative filtering. Gating mechanisms aim to enhance the information flow and control the flow of data through the network by selectively weighting the input and output connections of the network.

Gating Mechanism

In the context of VAEs for collaborative filtering, gating mechanisms can be applied to improve the performance of the model in capturing user preferences and generating accurate recommendations. The use of gating mechanisms allows the VAE to learn higher-level interactions and dependencies between users and items, leading to more refined and nuanced representations.

One advantage of gating mechanisms is their ability to adaptively modulate the information flow based on the context or input. This flexibility enables the model to assign varying importance to different parts of the input and selectively focus on relevant information. This is particularly beneficial in collaborative filtering, where user-item interactions can vary in importance and relevance.

By incorporating gating mechanisms into VAEs, the model can effectively capture the dynamics and complexity of user-item interactions, leading to improved recommendation accuracy. Gating mechanisms enable the model to learn to attend to relevant features and suppress noise or irrelevant information, resulting in more discriminative and personalized recommendations.

Furthermore, the use of gating mechanisms in VAEs can help address the limitations of traditional VAEs, such as the model's inability to learn deeper representations with multiple hidden layers. Gating mechanisms facilitate the propagation of information through deeper models, enabling the VAE to learn more expressive and informative latent representations.

However, it is important to note that gating mechanisms also introduce additional complexity and the potential for overfitting. Careful design and regularization techniques are necessary to ensure that the gating mechanisms are effectively utilized without sacrificing the model's generalization performance.

In summary, gating mechanisms offer a powerful tool to enhance VAEs for collaborative filtering. They enable the model to capture intricate user-item interactions, learn higher-level dependencies, and improve the overall recommendation accuracy. By selectively modulating the flow of information, gating mechanisms provide a valuable mechanism for adaptive and context-aware learning in collaborative filtering applications.

# Model Card

**Model Details**:

- **Developed by**: Daeryong Kim and Bongwon Suh

- **Model date**: September 2019

- **Model version**: 1.0

- **Model type**: Collaborative Filtering using Variational Autoencoders (VAEs) with Flexible Priors and Gating Mechanism

- **Training algorithms**: Backpropagation with Adam optimizer

- **Fairness constraints**: The model is not optimized for any demographic or phenotypic group, and no demographic or phenotypic information is used in the training process.

- **Sources**:

1. Kim, Daeryong, and Bongwon Suh. "Enhancing VAEs for collaborative filtering: flexible priors & gating mechanisms." *Proceedings of the 13th ACM Conference on Recommender Systems*. 2019.

- **License**: Open source

- **Contacts**:

bongwon@snu.ac.kr

daeryong@snu.ac.kr

**Intended Use**:

- Primary intended uses: The VAE model is designed to generate personalized recommendations for users based on their historical interactions with items. This particular model can model user preferences and represent it using more complex deep neural network.

- Primary intended users: Recommender system developers and users who want to improve the accuracy and personalization of their recommendations.

- Out-of-scope use cases: The VAE model is not intended for use cases where user-item interaction data is not available, such as content-based recommendation systems.

**Factors**:

- Relevant factors: Implicit User Feedback, High rated items, minimum threshold for user selection,User-item interaction data including ratings.

- Evaluation factors: Model accuracy

**Metrics**:

- Model performance measured by recall@k, and Truncated normalized discounted cumulative gain (NDCG).

**Evaluation Data**:

- Datasets: The VAE model is evaluated on several benchmark datasets, including Movielens and Netflix price dataset,

- Benchmarks: The new model was compared against state of the art autoencoder models as baselines for comparison as WMF, SLIM, CDAE and Multi -VAE. For evaluation of flexible priors models such VampPrior, H+Vamp ,H+Vamp(Gated) and Multi-VAE(Gated) were considered.

- Motivation: The benchmark datasets provide a standard set of evaluation criteria for collaborative filtering models.

- Preprocessing: The datasets are preprocessed to remove items which were not related to users and users who are not frequent consumers of items, and the remaining data is split into training, validation, and test sets.

**Training Data**:

- The training data consists of user-item interaction data including majorly ratings and ratings were binarized by ignoring lower rated entries.

**Quantitative Analyses**:

- Unitary results: Multi-VAE is the strongest baseline while Vamp, H+Vamp, H+Vamp (Gated) shows sequentially improving performance. Vamp shows significantly better performance compared to Multi-VAE, indicating the benefit of changing the restrictive standard normal prior to a flexible VampPrior. Our final model H + Vamp (Gated) shows the best performance and significantly outperforms the strongest baseline Multi-VAE for both datasets on all metrics. The final model shows up to 6.87% relative increase in recall@20 for the Netflix dataset producing new state-of-the-art results.- Intersectional results: The VAE model is evaluated on subsets of the benchmark datasets based on demographic or phenotypic factors, such as age or gender, to ensure that the model is not biased towards any particular group.

Effects of Gating :

Models that lack gating mechanisms do not benefit from increased depth in terms of performance improvement, unlike gated models. This implies that gating assists the network in transmitting information through deeper models. However, we can observe significant performance improvements by simply adding gates without additional layers. This suggests that the higher-level interactions enabled by the self-attentive gates are also very beneficial in modeling user preferences. It may be argued that the gated model has more parameters, but it should be noted that ungated models cannot achieve similar performance by merely increasing the number of units.

**Caveats and Recommendations**:

- The VAE model is only as good as the quality of the input data and may perform poorly on noisy or incomplete datasets.

- The model is designed to generate personalized recommendations but may not be suitable for cases where users have very few or no interaction data.

# Data Ethics

1. The purpose of Data practice in development of Recommendation engine with VAE with flexible Prior and Gated mechanism:

Our project involves developing a recommendation system using VAEs with flexible priors and gating mechanisms for the data of user behavior of video streaming platform. The system will be able identify recommend movie and playlists to users based on their viewing history, preferences described by history of rated content.

2. Stakeholders:

The stakeholders involved in this project would be the video streaming platform, data professionals, video industry partners, and users. The platform has an interest in increasing user engagement and satisfaction, while data professionals aim to create an effective and efficient recommendation system. Video industry partners would like to see their content and playlist recommended to users, and users have an interest in discovering new content and having a personalized viewing experience.

3. Benefits and Risks:

The benefits of this project include increased user engagement and satisfaction, personalized recommendations, and potentially higher revenue for the platform and video industry partners. However, there are risks of potential harm, including loss of privacy, algorithmic bias, and the potential for users to be stuck in filter bubbles. Downstream impacts could include perpetuating existing inequalities in the video industry and contributing to a homogenization of user preferences.

4. Ethical Challenges:

The ethical challenges relevant to this project include privacy, transparency, accountability, fairness, and the potential for discrimination and bias in the recommendation system.

5. Ethical Obligations:

The data professionals working on this project have ethical obligations to be transparent about data collection and usage, ensure user privacy, and actively work to mitigate bias and discrimination in the recommendation system.

6. Disparate Impacts:

Disparate impacts could occur if certain user groups are underrepresented in the data used to train the recommendation system, leading to biased recommendations. This could disproportionately affect marginalized communities or individuals with niche music tastes.

7. Best and Worst Case Scenarios:

The best-case scenario for this project would be a personalized and fair recommendation system that increases user engagement and satisfaction, while the worst-case scenario would be a recommendation system that perpetuates inequalities in the music industry, limits user choice, and harms privacy.

8. Risk Reduction and Crisis Response:

To reduce the risk of the worst-case scenario, the data professionals could actively monitor and address potential bias and discrimination in the recommendation system, and ensure that user privacy is protected. In the event of a crisis, an effective crisis response could involve transparent communication with users, prompt remediation of any harms caused, and proactive measures to prevent similar harms in the future.

9. Ethical Proposals:

a) Include diverse perspectives in the development and testing of the recommendation system to mitigate bias and ensure fairness.

b) Allow users to easily opt out of data collection and personalized recommendations if desired.

c) Provide clear and accessible explanations of how the recommendation system works and how data is used to make recommendations, to promote transparency and user trust.

Resources:

[[Book]](https://link.springer.com/book/10.1007/978-1-4842-8954-9) Applied Recommender Systems with Python

[[Paper]](https://arxiv.org/ftp/arxiv/papers/1911/1911.00936.pdf) Enhancing VAEs for Collaborative Filtering: Flexible Priors & Gating Mechanisms

[[Paper]](https://arxiv.org/abs/1802.05814) Variational Autoencoders for Collaborative Filtering

[[Paper]](https://arxiv.org/pdf/2103.10693.pdf) Adversarial and Contrastive Variational Autoencoder for Sequential Recommendation