Evidential Stochastic Differential Equations for Time-Aware Sequential Recommendation

Krishna Prasad Neupane, Ervine Zheng, and Qi Yu

Rochester Institute of Technology

December, 2024

ENSDE

2024-11-11

Evidential Stochastic Differential Equations for Time-Aware Sequential Recommendation

Krishna Prasad Neupane, Ervine Zheng, and Qi Yu

Rochester Institute of Technology

December, 2024

Overview

Sequential recommender systems are designed to capture users' evolving interests over time to allow more personalized and timely suggestions. It has following key characteristics:

- **Time-Ordered Interactions:** Capture and analyze users' interactions in a specific sequence to reflect evolving interests.
- **User Behavior Prediction:** Predict the next item a user is likely to engage with, based on previous actions.
- **Dynamic Preference Modeling:** Track and adapt to users' changing preferences over time.
- **Context Awareness:** Often incorporate contextual factors, like time and location, for more relevant recommendations.
- **Personalization:** Tailor suggestions specifically to each user's unique behavior patterns.
- **Real-Time Adaptability:** Update recommendations in real-time as new user interactions occur.

Overview

Sequential recommender systems are designed to capture users' evolving interests over time to allow more personalized and timely suggestions. It has following key characteristics:

- **Time-Ordered Interactions:** Capture and analyze users' interactions in a specific sequence to reflect evolving interests.
- **User Behavior Prediction:** Predict the next item a user is likely to engage with, based on previous actions.
- **Dynamic Preference Modeling:** Track and adapt to users' changing preferences over time.
- **Context Awareness:** Often incorporate contextual factors, like time and location, for more relevant recommendations.
- **Personalization:** Tailor suggestions specifically to each user's unique behavior patterns.
- **Real-Time Adaptability:** Update recommendations in real-time as new user interactions occur.

Key Challenges and Motivation

- Existing sequential recommendation models (e.g., BERT4Rec, SASRec, etc.) often assume uniform time intervals between user interactions, which fails to capture real-world, non-uniform patterns.
- This oversight can lead to inaccurate predictions, as user preferences may change significantly over varying interaction gaps.
- As interaction intervals lengthen, the uncertainty in predicting user behavior grows, making it harder to provide accurate recommendations.

Table: Impact of interaction interval

Interaction	Interval (Seconds)	Ranking ↓	
		BERT4Rec	E-NSDE
$6 \to 7$	44	4	**4**
$13 \to 14$	623,591	24	**16**
$116 \to 117$	62	6	**3**
$150 \to 151$	896,291	56	**18**

- The unrealistic assumption of a uniform interaction interval could significantly impact the model's capability to capture users' continuously evolving behavior and subsequently hurt the recommendation performance as demonstrated in Table 1.

Key Challenges and Motivation
Existing sequential recommendation models (e.g., BERT4Rec, SASRec, etc.) often assume uniform time intervals between user interactions, which fails to capture real-world, non-uniform patterns.
This oversight can lead to inaccurate predictions, as user preferences may change significantly over varying interaction gaps.
As interaction intervals lengthen, the uncertainty in predicting user behavior grows, making it harder to provide accurate recommendations.
Table: Impact of interaction interval
The unrealistic assumption of a uniform interaction interval could significantly impact the model's capability to capture users' continuously evolving behavior and subsequently hurt the recommendation performance as demonstrated in Table 1.

Solution: E-NSDE Framework

- We propose an **Evidential Neural Stochastic Differential Equation (E-NSDE)** framework as shown in Figure below.

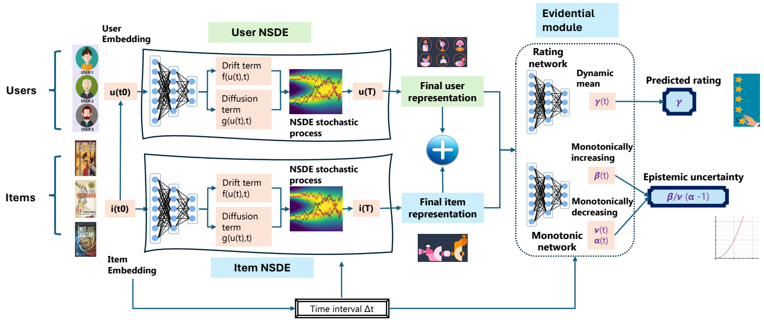
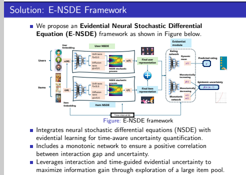


Figure: E-NSDE framework

- Integrates neural stochastic differential equations (NSDE) with evidential learning for time-aware uncertainty quantification.
- Includes a monotonic network to ensure a positive correlation between interaction gap and uncertainty.
- Leverages interaction and time-guided evidential uncertainty to maximize information gain through exploration of a large item pool.



E-NSDE Framework Contd.

- The NSDE module generates fine-grained user and item representations as follows:

$$u(T) = u(t_0) + \int_{t_0}^{T} f(u(t), t; \psi)dt + \int_{t_0}^{T} g(u(t), t; \omega)dB_t \quad (1)$$

  where $u(t_0)$ is the user's initial representation which aggregates a set of initially interacted items: $u(t_0) = \text{agg}(i_1..i_k)$ and T is the final time.

$$i(T) = i(t_0) + \int_{t_0}^{T} f(i(t), t; \psi)dt + \int_{t_0}^{T} g(i(t), t; \omega)dB_t \quad (2)$$

  where, $i(t_0)$ represents item's initial representation
- In Evidential module,
  - The rating network utilizes the fine-grained user u_t and item i_t representations from the NSDE and predicts the score $\gamma_{(u_t, i_t)}$.
  - The monotonic network is designed in such a way that the increase in time interval (Δt), increases the output, i.e., the variance of the predicted rating.

E-NSDE Framework Contd.
The NSDE module generates fine-grained user and item representations as follows:
$$u(T) = u(t_0) + \int_{t_0}^{T} f(u(t), t; \psi)dt + \int_{t_0}^{T} g(u(t), t; \omega)dB_t \quad (1)$$
where $u(t_0)$ is the user's initial representation which aggregates a set of initially interacted items: $u(t_0) = \text{agg}(i_1..i_k)$ and T is the final time.
$$i(T) = i(t_0) + \int_{t_0}^{T} f(i(t), t; \psi)dt + \int_{t_0}^{T} g(i(t), t; \omega)dB_t \quad (2)$$
where, $i(t_0)$ represents item's initial representation
In Evidential module,
The rating network utilizes the fine-grained user u_t and item i_t representations from the NSDE and predicts the score $\gamma_{(u_t, i_t)}$.
The monotonic network is designed in such a way that the increase in time interval (Δt), increases the output, i.e., the variance of the predicted rating.

Training: Evidential and Weighted Bayesian Personalized Ranking Loss

- The regularized EDL loss with negative log-likelihood ($\mathcal{L}^{NLL}[f_\Theta]$) to maximize the marginal likelihood and an evidential regularizer ($\mathcal{L}^{R}[f_\Theta]$) to impose a high penalty on the predicted error with low uncertainty for each sequential update is:

$$\mathcal{L}_{EDL}(u_t, i_t) = \mathcal{L}^{NLL}[f_\Theta(u_t, i_t)] + \lambda\mathcal{L}^{R}[f_\Theta(u_t, i_t)] \tag{3}$$

  where λ is a regularization coefficient.
- To leverage the effective exploration for the long-term, we formulate weighted BPR loss which is computed from non-interacted (negative) that are similar to the user's future interacted items.

$$\mathcal{L}_{\mathrm{WBPR}}(u_t, i_t) = \sum_{(u_t, i_t, j_t \in \mathcal{N}_t)} w_{(i_t, j_t)}\{-\ln[\sigma(\hat{r}_{(i_t, j_t)})]\} \tag{4}$$

  where $\hat{r}_{(i_t, j_t)} = \hat{r}_{(u_t, i_t)} - \hat{r}_{(u_t, j_t)}$, $\sigma(\cdot)$ is the sigmoid.
- The overall loss of the end-to-end model training is obtained by combining the EDL and WBPR loss:

$$\mathcal{L}(u_t, i_t) = \mathcal{L}_{\mathrm{EDL}}(u_t, i_t) + \zeta\mathcal{L}_{\mathrm{BPR}}(u_t, i_t) \tag{5}$$

Recommendation Performance Comparison

- Our experiment setting of sequential recommendation is based on next-item recommendation tasks.
- We split users by 70% into train and 30% in test.
- We leverage the actual time of interactions (in UNIX timestamp) to provide user preference evolution.

Table: Recommendation performance comparison

Category	Model	MovieLens-100K		Movielens-1M		Netflix		Amazon Book	
		P@5	nDCG@5	P@5	nDCG@5	P@5	nDCG@5	P@5	nDCG@5
Dynamic MF	timeSVD++	0.3842±0.015	0.3420±0.013	0.3917±0.016	0.3508±0.013	0.3756±0.016	0.2951±0.013	0.3601±0.014	0.3128±0.012
	CKF	0.3916±0.017	0.3620±0.013	0.3928±0.016	0.3552±0.015	0.3600 ±0.017	0.2986±0.014	0.3823±0.016	0.3214±0.015
Graph	NGCF	0.3859±0.014	0.3662±0.012	0.3978±0.016	0.3587 ±0.018	0.3574±0.015	0.3167±0.017	0.3574±0.012	0.3321±0.011
	LightGCN	0.4103±0.014	0.3702±0.013	0.4028±0.017	0.3632±0.015	0.3617±0.013	0.3204±0.016	0.3678±0.013	0.3382±0.012
Sequential	CASER	0.4096 ±0.012	0.3663±0.015	0.4021±0.014	0.3626±0.016	0.3658 ±0.013	0.3189±0.012	0.3722±0.012	0.3414±0.012
	SASRec	0.4105±0.013	0.3740 ±0.011	0.4112±0.015	0.3708±0.017	0.3746±0.012	0.3257±0.014	0.3812±0.014	0.3445±0.012
	BERT4Rec	0.4149±0.014	0.3781±0.011	0.4163±0.012	0.3754 ±0.013	0.3793±0.011	0.3295±0.013	0.3846±0.013	0.3463±0.013
	S^3-Rec	0.4124 ±0.012	0.3755±0.014	0.4134±0.013	0.3715 ±0.014	0.3786±0.016	0.3274±0.013	0.3725±0.014	0.3350±0.012
	CL4SRec	0.4210±0.016	0.3821±0.017	0.4205±0.013	0.3781 ±0.015	0.3814 ±0.016	0.3318±0.012	0.3858±0.014	0.3313±0.011
	SAR	0.4034±0.012	0.3741±0.012	0.4023±0.014	0.3747±0.014	0.3711±0.012	0.3224±0.013	0.3658±0.012	0.3320±0.014
	ResAct	0.4366±0.014	0.3948±0.012	0.4286±0.014	0.3814±0.011	0.3867±0.012	0.3360±0.011	0.3884±0.013	0.3472±0.013
ODE	LT-OCF	0.4267±0.013	0.3785±0.015	0.4141±0.016	0.3673 ±0.014	0.3848±0.012	0.3313±0.013	0.3841±0.014	0.3416±0.012
	GRU-ODE	0.4398±0.014	0.3902 ±0.017	0.4275±0.013	0.3792±0.013	0.3994±0.013	0.3417±0.015	0.3856±0.014	0.3455±0.012
Proposed	**E-NSDE**	**0.4711±0.015**	**0.4112±0.013**	**0.4551±0.011**	**0.3982±0.016**	**0.4194±0.013**	**0.3637±0.015**	**0.4021±0.014**	**0.3621±0.012**

Recommendation Performance Comparison
Our experiment setting of sequential recommendation is based on next-item recommendation tasks.
We split users by 70% into train and 30% in test.
We leverage the actual time of interactions (in UNIX timestamp) to provide user preference evolution.
Table: Recommendation performance comparison

Ablation Study

- **Uncertainty vs. Interaction gap:** We further investigate the impact of the user-item interaction gap and the corresponding uncertainty in providing important and diverse items in Table (a).

- Further, ablation study on key components of the E-NSDE model in Table (b).

Table: (a) Diverse recommendations by E-NSDE; (b) Ablation of key components

Model	Important Movies (Genre)	Future Movie's Genre
GRU-ODE	Dead Man Walking ('Drama') Richard III ('Drama') Mad Love ('Romance')	'Thriller' 'Drama' 'Mystery'
E-NSDE	GoldenEye ('Thriller') Taxi Driver ('Drama') Twelve Monkeys ('Sci_Fi')	'Crime' 'Sci_Fi'

(a)

NSDE	EDL	WBPR	Performance	
			P@5	nDCG@5
✓			0.4065	0.3677
✓	✓		0.4523	0.3962
✓		✓	0.4120	0.3715
✓	✓	✓	**0.4711**	**0.4112**

(b)

ENSDE
└Experimentation
└Ablation Study

2024-11-11

Ablation Study

- Uncertainty vs. Interaction gap: We further investigate the impact of the user-item interaction gap and the corresponding uncertainty in providing important and diverse items in Table (a).
- Further, ablation study on key components of the E-NSDE model in Table (b).

Table: (a) Diverse recommendations by E-NSDE; (b) Ablation of key components

Conclusions

- We propose a novel time-aware sequential recommendation model called E-NSDE.
- It includes NSDE module that enables continuous, time-aware modeling of user preferences, without requiring uniform time intervals.
- Further, monotonic network ensures the relationship between time intervals and uncertainty is properly reflected—longer time intervals increase model uncertainty.
- Provides SOTA performance in multiple real-world datasets.

Poster

More detailed information will be in the Poster with **ID: 96864 (Poster Session 2)**.

Conclusions

- We propose a novel time-aware sequential recommendation model called E-NSDE.
- It includes NSDE module that enables continuous, time-aware modeling of user preferences, without requiring uniform time intervals.
- Further, monotonic network ensures the relationship between time intervals and uncertainty is properly reflected—longer time intervals increase model uncertainty.
- Provides SOTA performance in multiple real-world datasets.

Poster

More detailed information will be in the Poster with **ID: 96864 (Poster Session 2)**.