Field-aware Probabilistic Embedding Neural Network for CTR Prediction

Weiwen Liu[1], Ruiming Tang[2], Jiajin Li[1], Jinkai Yu[2], Huifeng Guo[3], Xiuqiang He[4], Shengyu Zhang[1,5]
[1]The Chinese University of Hong Kong, [2]Huawei Noah's Ark Lab, [3]Harbin Institute of Technology (Shenzhen), [4]Tencent Data Service Center, [5]Tencent Quantum Lab

Motivation

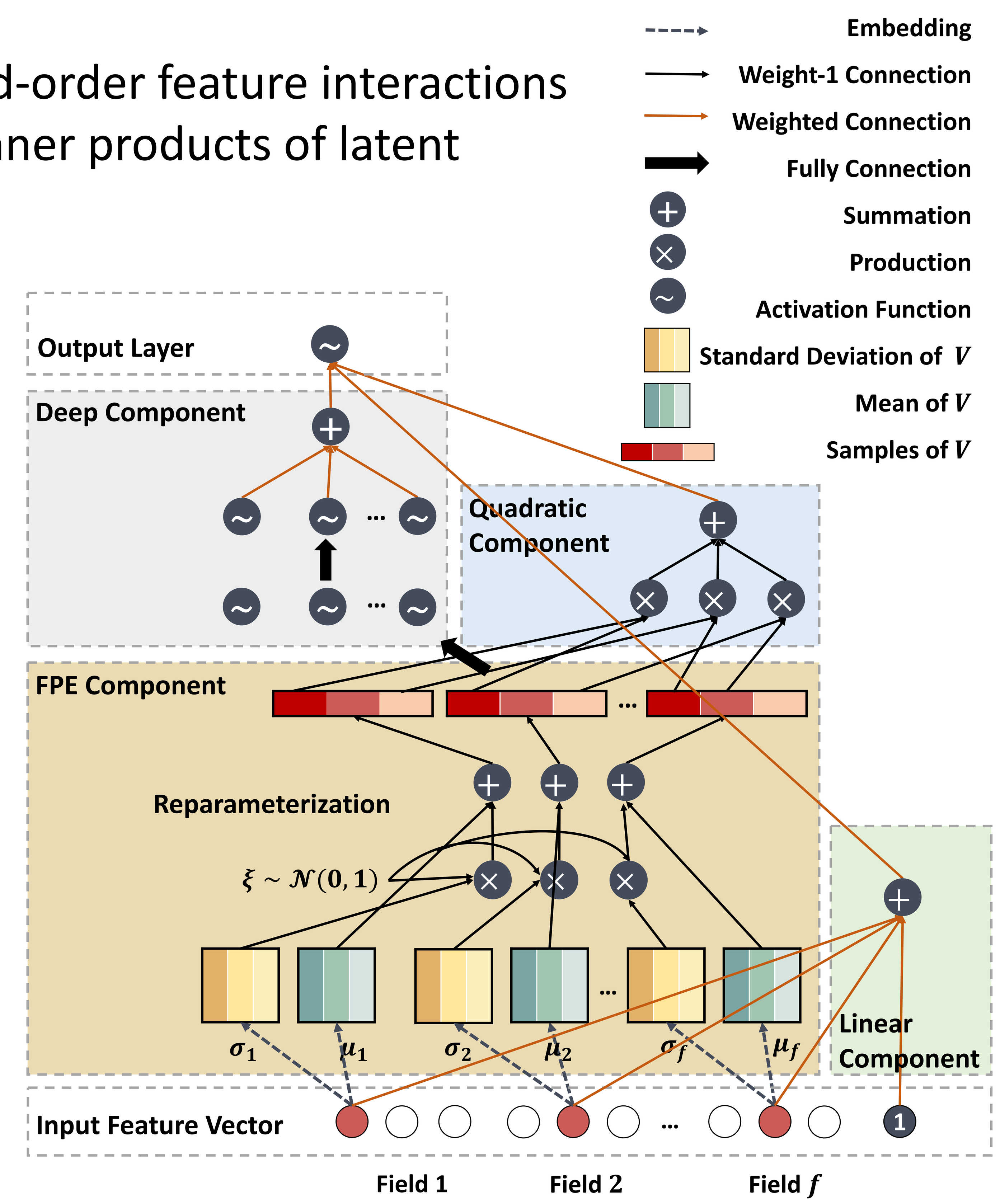
- The Click-Through Rate (CTR) prediction is a vital task for web search, recommender systems, online advertising, etc.
- Most of the existing deep models are biased to high-order interactions and fail to exploit the simple yet effective low-order feature interactions. Moreover, they did not consider the field information.
- Field-aware Factorization Machines (FFM) have exhibited great effectiveness in CTR prediction by considering field information. However, FFM suffers from the overfitting problem in many practical scenarios [1].
- We aim at designing a CTR prediction model that incorporates both low-order and high-order feature interactions, while preventing overfitting.

Contribution

- We propose a Field-award Probabilistic Embedding method (FPE) to estimate the probability distribution of the field-aware embedding instead of using the single point estimation. Various uncertainties for different feature interactions are are captured.
- The field information is incorporated into a unified end-to-end deep learning framework, FPENN, which combines the low-order and high-order feature interactions.
- We test our FPENN together with the state-of-the-art models (i.e., LR, FM, FFM, DeepFM) on two benchmark datasets to confirm the effectiveness of FPENN.

Model

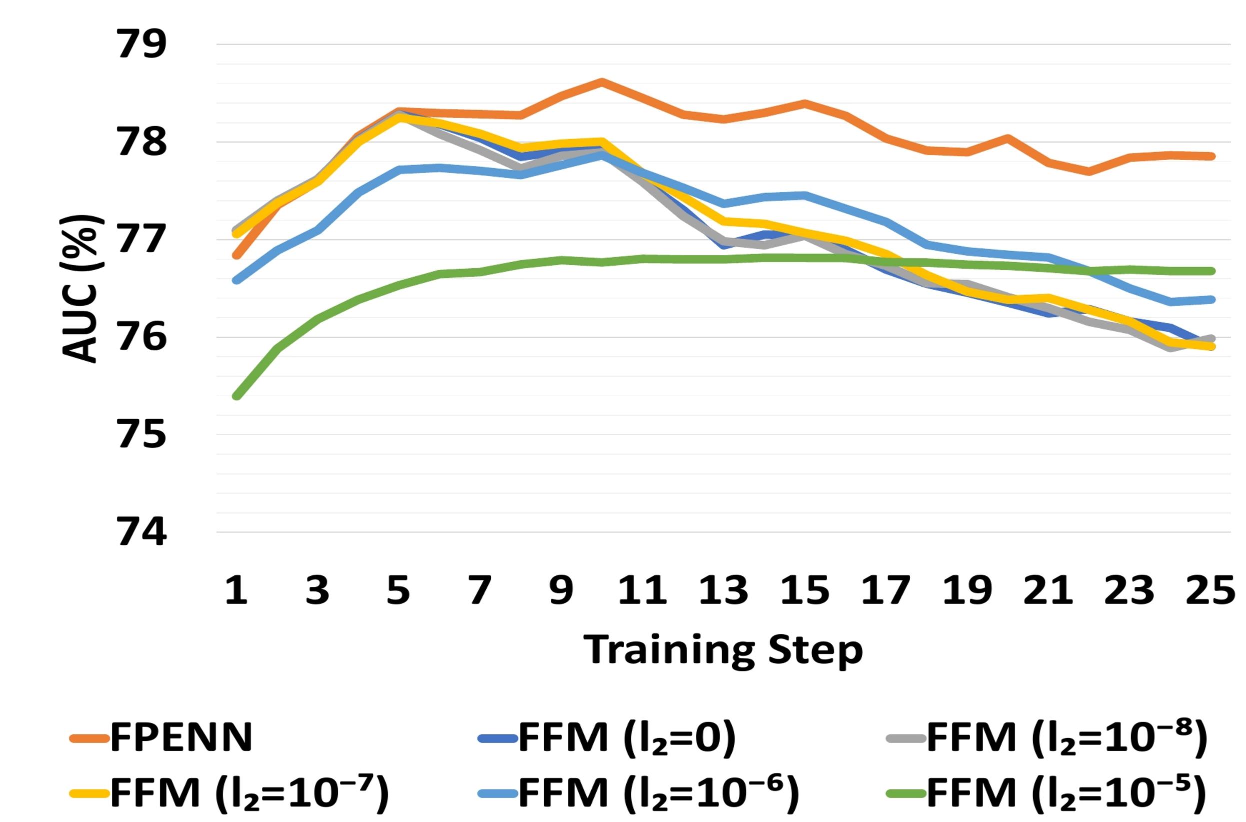
- Our proposed Field-aware Probabilistic Embedding Neural Network (FPENN) model consists of three components, i.e., FPE component, Quadratic component and Deep component.
- **FPE component:**
  - Embeds the sparse input feature vector to dense latent vectors to reduce the dimension.
  - A distribution is learned for each element in the latent vectors to enhance the robustness and effectiveness of the prediction.
  - **Training:** The probability distribution is learned using the reparametrization trick [2].
  - **Testing:** The mean and variance information are combined by the proposed UCB-strategy or TS-strategy.
- **Quadratic component:**
  - Captures the second-order feature interactions by computing the inner products of latent vectors.
- **Deep component:**
  - A neural network is deployed to learn high-order feature interactions, on the basis of feature embeddings generated by FPE component.
- **Overall architecture:**
  - The final output of FPENN is a weighted sum of the Quadratic component and the Deep component.

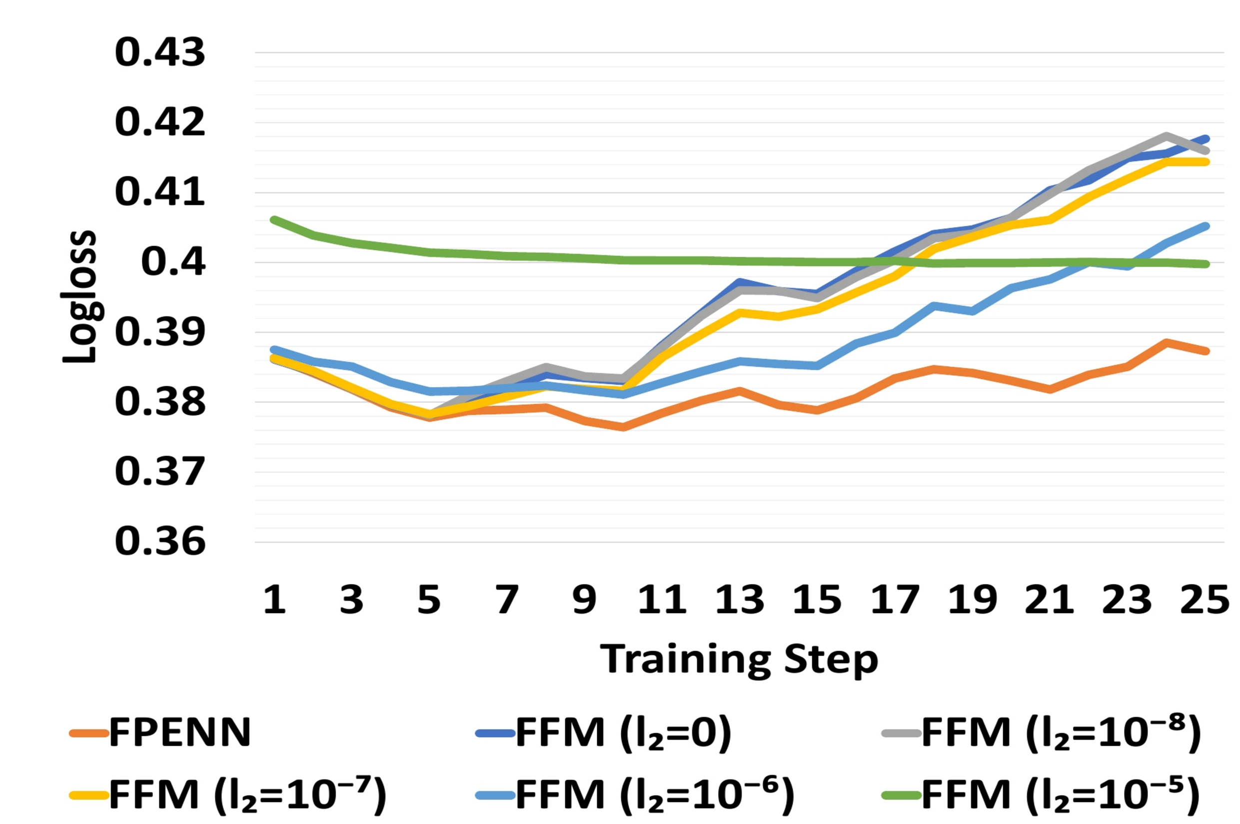


References:
[1] Juan, Y, et al. "Field-aware factorization machines for CTR prediction." *RecSys*, 2016.
[2] Ruiz, F. R., AUEB, M. T. R, and Blei, D. "The generalized reparameterization gradient." *NIPS,* 2016.

Experiments

- **Generalization ability:**
  - The performance of FPENN drops very little with training steps.





- **Accuracy**
  - FPENN performs better than other state-of-the-art models.

	Avazu		Criteo	
	AUC(%)	Logloss	AUC(%)	Logloss
LR	76.53	0.3883	78.00	0.5631
FM	77.97	0.3802	79.09	0.5500
FFM	78.25	0.3782	79.77	0.5440
DeepFM	78.36	0.3777	79.91	0.5423
FPENN	**78.61**	**0.3764**	**79.97**	**0.5417**