بسم الله الرحمن الرحیم

User Conversion Prediction in Display Advertisements

Mohammadreza Rezaei

Supervised by
Hamid R. Rabiee

Data Science & Machine Learning Lab (DML)
Sharif University of Technology

Winter 2020

Outline

1. Introduction
2. Related Works
3. Proposed Method
4. Experiments
5. Future Work

Introduction

Display Advertising

Ads

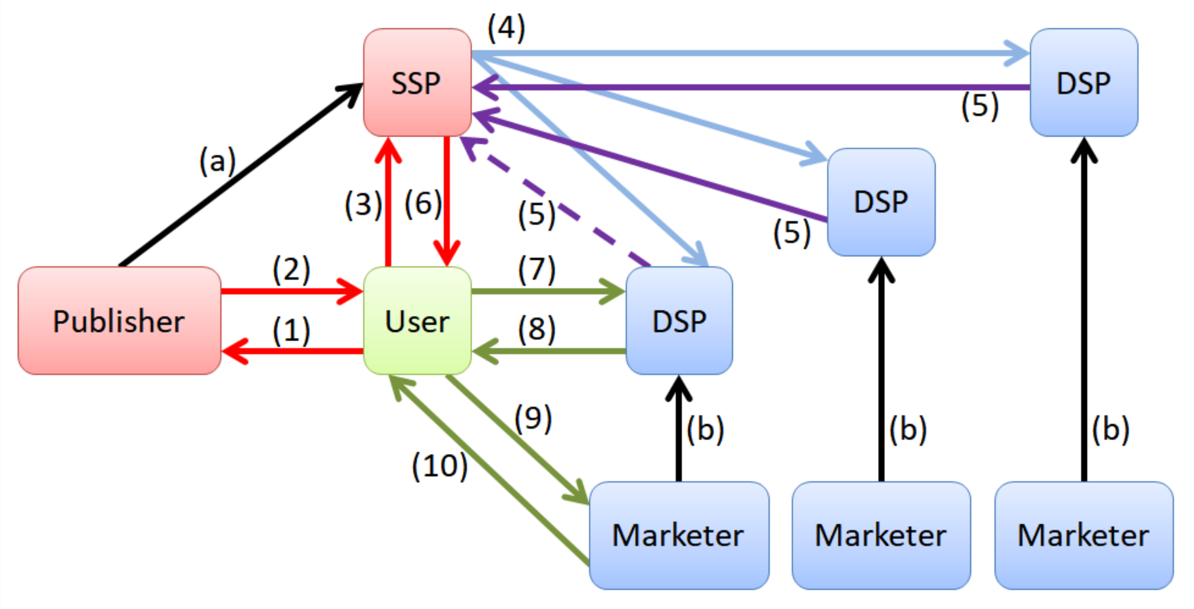
Google Ads

YAHOO! ADVERTISING

twitter Advertising

Real Time Bidding (RTB):

- A process that decides which ad will be shown to the user



Problem Definition

Click Through Rate (CTR) Prediction

- Predict probability of a click:
  - On a certain Ad
  - By a certain User
  - In a certain Page
- Sparse and High dimensional data:

Field 1						Field 2				Field 3			Field 4							
0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0

Challenges:

- Imbalanced data
- High dimensionality
- Sparsity
- Cold Start
  - Ad Cold Start
  - User Cold Start

Related Works

Related Work

- Classical approaches
- Factorization Machines
- Deep Methods
- Deep and Factorization Methods Combined

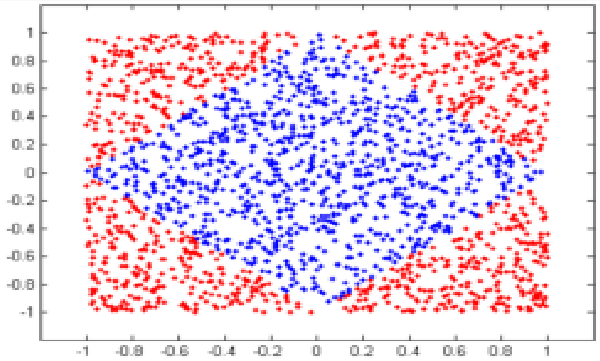
Related Work (Classical)

Classical approaches:

- SVM

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w'_{i,j} x_i x_j$$

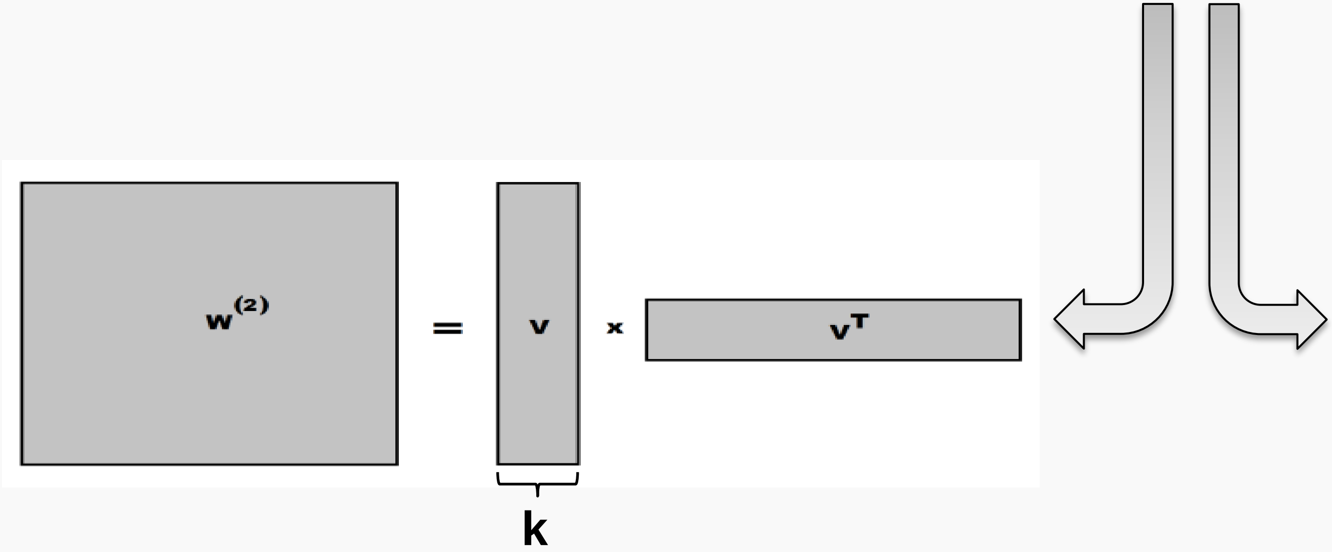
- $$y = g\left(\sum_{j=1}^{m} \sigma(u_j^T x)\eta(w_j^T x)\right)$$



Related Work (Factorization Machines)

Factorization Machines

$$\hat{y}(x) = w_{۰} + \sum_{i=۱}^{n} w_i x_i + \sum_{i=۱}^{n} \sum_{j=i+۱}^{n} w_{i,j}^{(۲)} \underbrace{x_i x_j}, \quad w_{i,j}^{(۲)} = \sum_{l=۱}^{k} v_{i,l} v_{j,l}, \; v_i \in \mathbb{R}^k$$



Cross features (interactions)

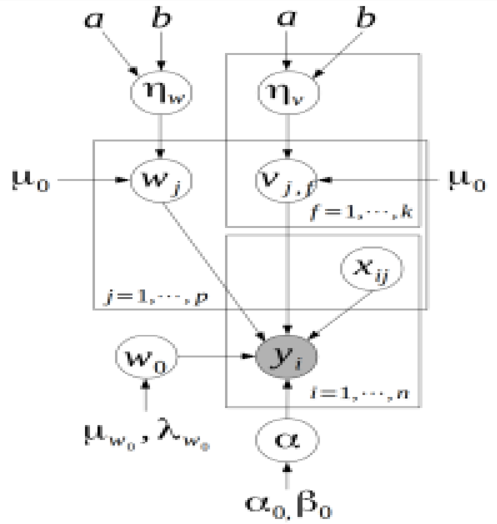
Field-Aware FM

$$\hat{y}_{FFM}(x) = w_. + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j < v_{i,F_j}, v_{j,F_i} >$$

Field-Weighted FM

$$\hat{y}_{FwFM}(x) = w_. + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j < v_i, v_j > r_{F_i,F_j}$$

Bayesian FM



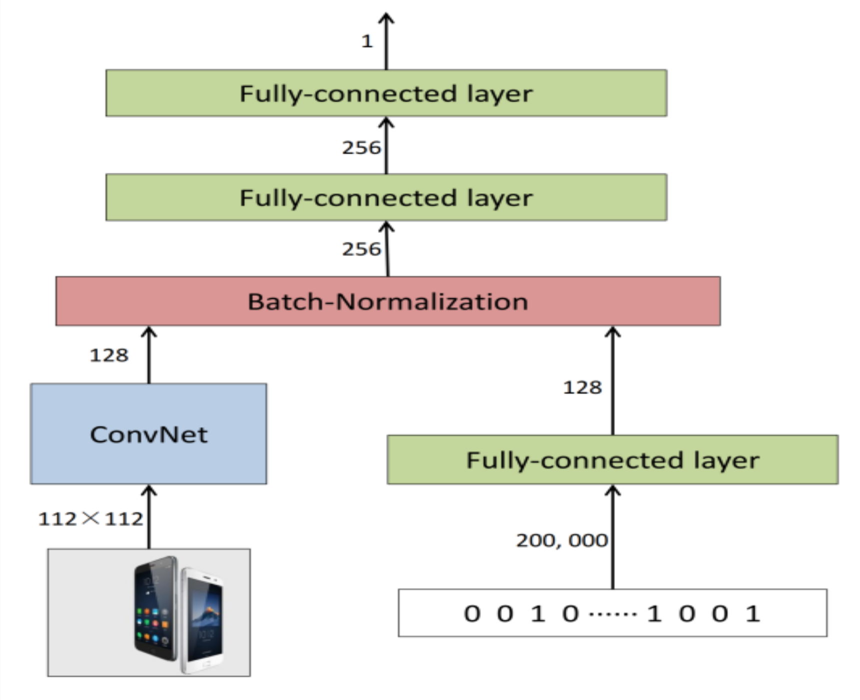
- Sparse FM

Attentional FM

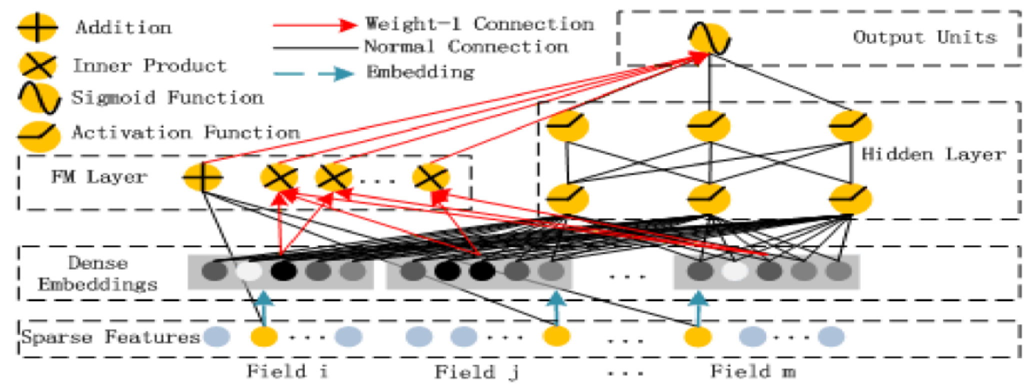
$$\mathcal{E}_{i,j} = (v_i \odot v_j) x_i x_j$$

$$\hat{y}_{AFM}(x) = w_. + \sum_{i=1}^{n} w_i x_i + \mathbf{P}^T \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{i,j} \mathcal{E}_{i,j}$$

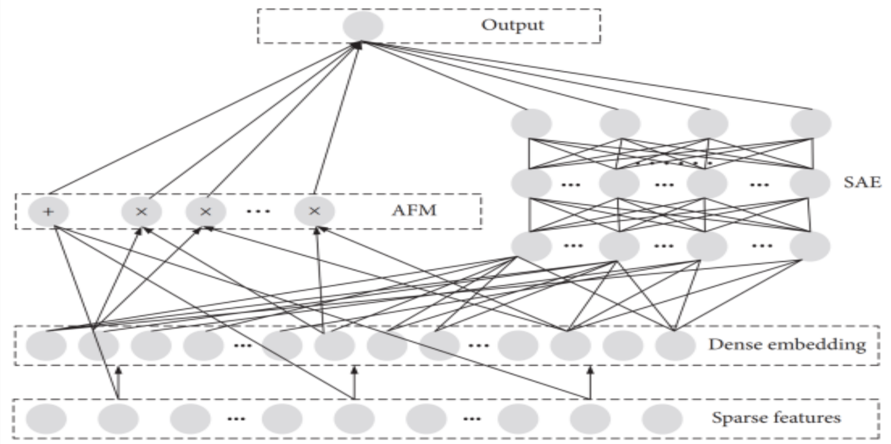
Deep CTR Prediction:



- Deep FM (Wide & Deep)



- ASAE (AFM + SAE)



Conclusion

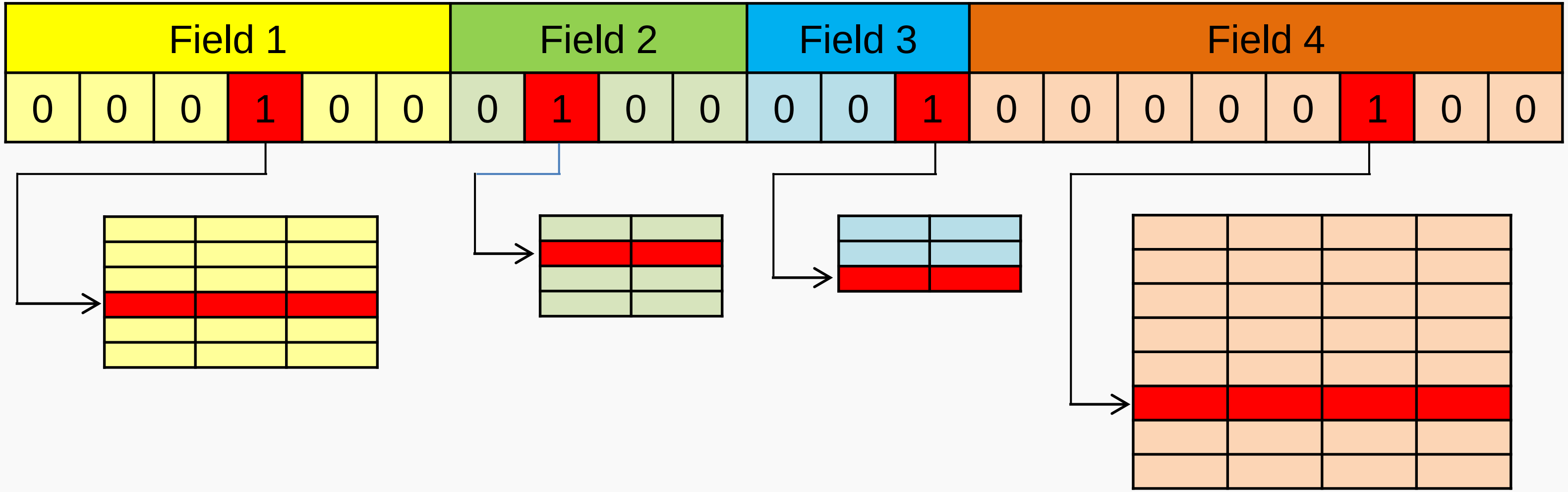
Model	Weaknesses	Strengths
SVM (with polynomial kernel)	•Too many parameters	•Fast training
Piece-wise linear model	•Too many parameters •Slow training •Difficult to tune hyper-parameters	•High flexibility •Sparse Parameters •Good interpretability
Factorization Machine	•Low complexity •Low interpretability	•Works well with sparse data
Field-Aware FM	•Lower complexity	•Faster training •More shared weights
Field-Weighted FM	•Lower complexity	•Less parameters

Conclusion

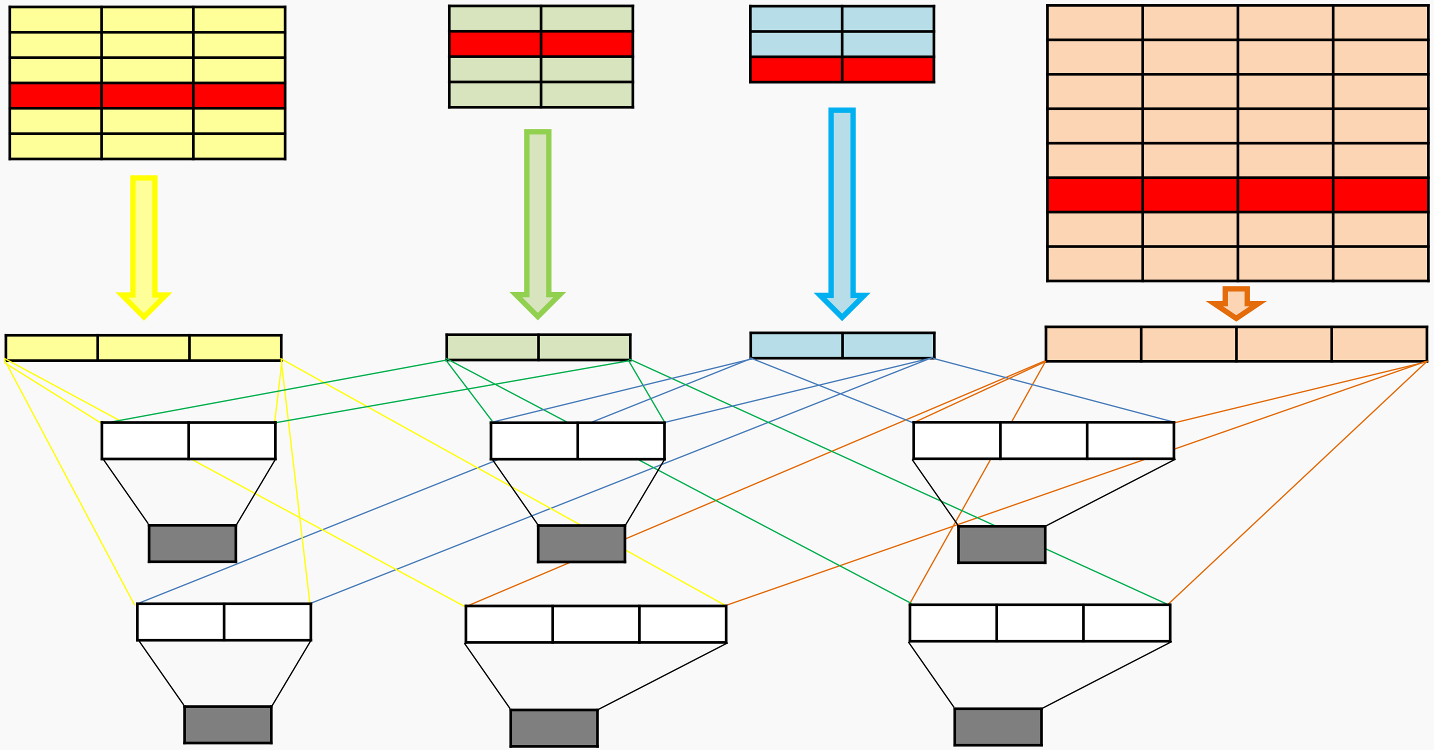
Model	Weaknesses	Strengths
Bayesian FM	•Intractable inference •Wrong normal assumption	•Balance of exploration and exploitation
Sparse FM	•Intractable inference •Approximate Laplace distribution	•More interpretability •More sparsity
Attentional FM	•Possible overfitting •Need to regularize	•More complexity •More interpretability
Deep CTR prediction model	•Need for Ad image •Possible overfitting	•Models higher degree interactions •More generalization when provided enough data
Deep FM (Wide & Deep)	•Too many hyper-parameters •Low interpretability	•Models higher degree interactions •No bias in high or low degree interactions

Proposed Method

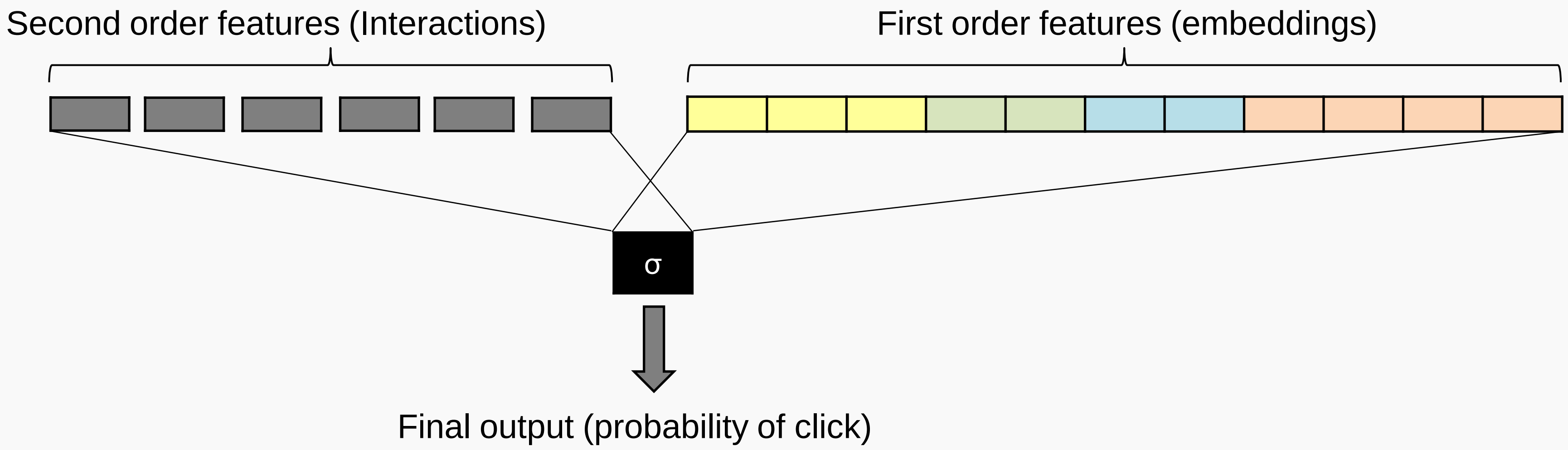
Different embedding sizes for different fields



Compute interactions via two layer neural networks



- Combine interactions and first order features using another layer of neural network



Proposed Method

Possible advantages:

- **Lower complexity & Lower number of parameters**
- **Interpretability (Embeddings, Interactions)**
- **High parameter sharing (helps facing cold start)**

Experiments

Dataset

Avazu dataset

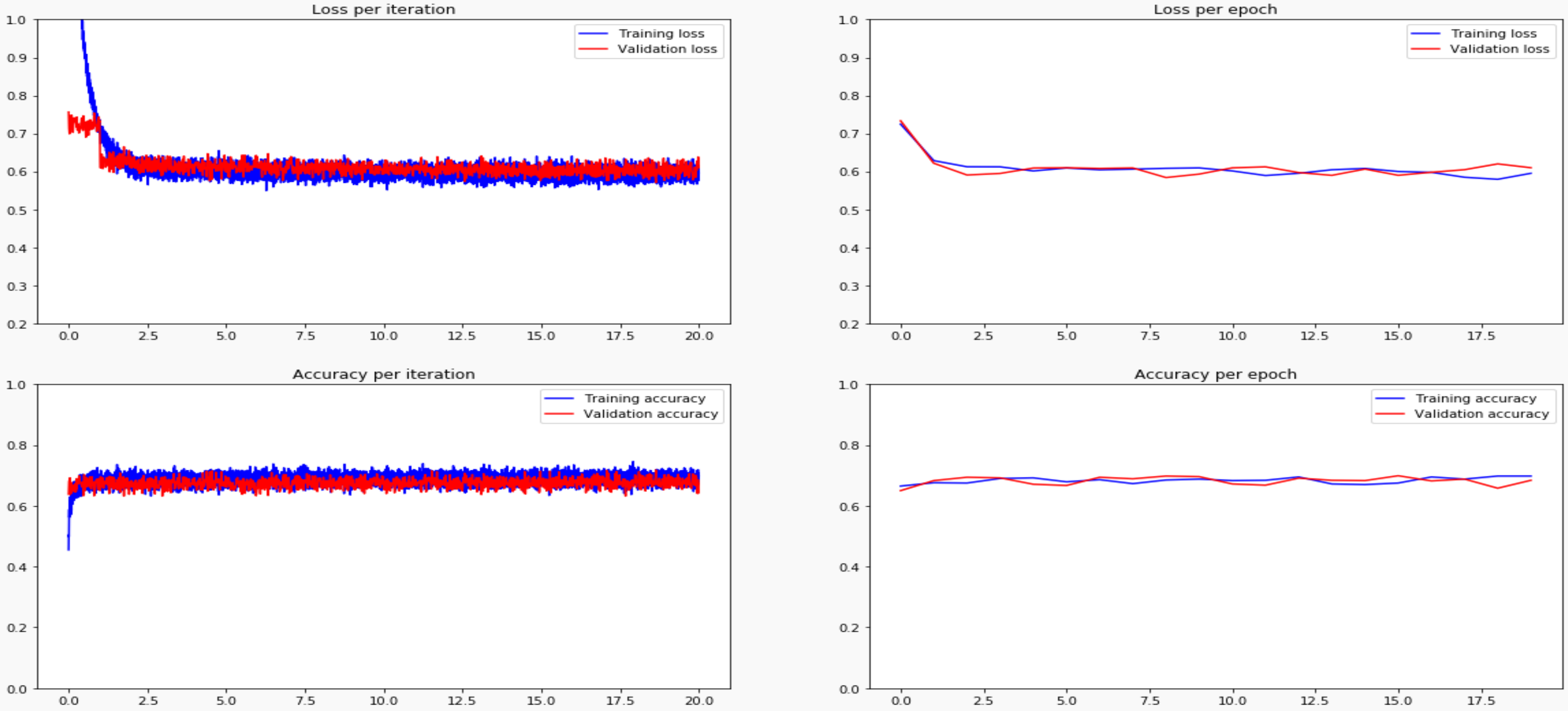
- +80M records
- Unbalanced (17% clicked)
- 22 Fields (+1.7M features)

Small Avazu

- +500k records
- Balanced (50% clicked)
- 19 Fields (+11K features)

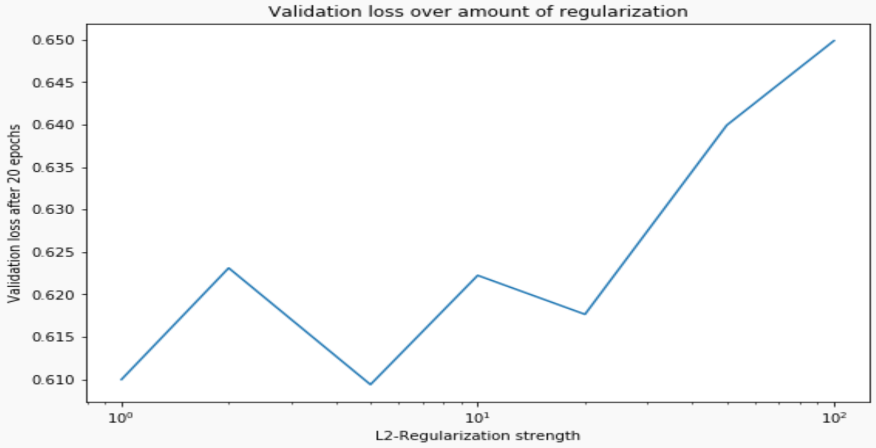
Results

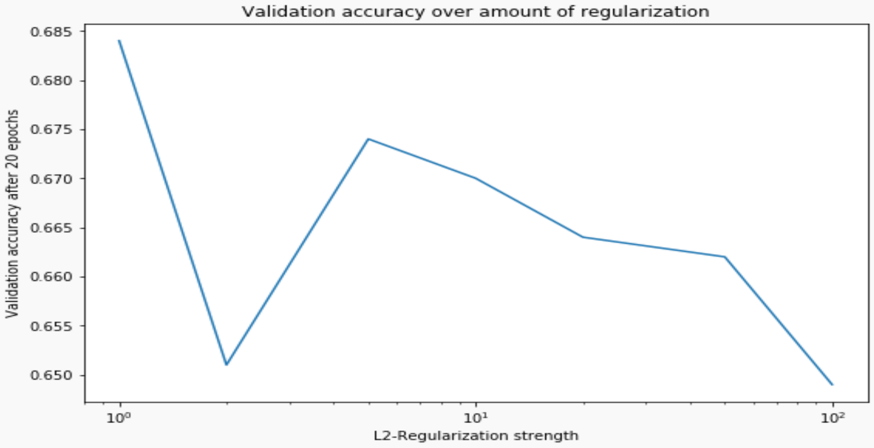
Cross-entropy and accuracy



Results

Effect of L2-Regularization (on embedding parameters)





Results

Final results:

Model	Accuracy(%)	Cross-entropy loss
FM (k = 5)	67.6	1.64
Proposed (Reg = 1)	**68.4**	**0.61**

Future Work

Future Work

Further work:

- Compare with Factorization Machines and other key models
- Try other regularizations (Dropout / L1 / Batch Normalization)
- Deploy the model in an online setting
- Work on larger datasets
- Visualize embeddings and find out if they are informative
- Check the most active interactions and try to interpret

Future Work

Title	Time needed	Progress	Finishing time
Reading related papers	5 months	80%	Feb. 2020
Developing proposed model	2 months	25%	Mar. 2020
Work with real data	2 months	0%	May. 2020
Writing paper	2 months	0%	Jul. 2020
Writing thesis	3 months	35%	Aug. 2020

Thank You!

Any Questions?

Refs

[1] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers,” p. 9, Jul. 1992.

[2] K. Gai, X. Zhu, H. Li, K. Liu, and Z. Wang, “Learning Piece-wise Linear Models from Large Scale Data for Ad Click Prediction,” arXiv:1704.05194 [cs, stat], Apr. 2017.

[3] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich, “Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft’s Bing Search Engine,” p. 8.

[4] S. Rendle, “Factorization Machines,” in 2010 IEEE International Conference on Data Mining, Sydney, Australia, 2010, pp. 995–1000, doi: 10.1109/ICDM.2010.127.

[5] Y. Juan, Y. Zhuang, W.-S. Chin, and C.-J. Lin, “Field-aware Factorization Machines for CTR Prediction,” in Proceedings of the 10th ACM Conference on Recommender Systems - RecSys ’16, Boston, Massachusetts, USA, 2016, pp. 43–50, doi: 10.1145/2959100.2959134.

[6] J. Pan et al., “Field-weighted Factorization Machines for Click-Through Rate Prediction in Display Advertising,” in Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW ’18, Lyon, France, 2018, pp. 1349–1357, doi:

Refs

[7] Z. Pan, E. Chen, Q. Liu, T. Xu, H. Ma, and H. Lin, “Sparse Factorization Machines for Click-through Rate Prediction,” in 2016 IEEE
16th International Conference on Data Mining (ICDM), Barcelona, Spain, 2016, pp. 400–409, doi: 10.1109/ICDM.2016.0051.

[8] C. Freudenthaler, L. Schmidt-Thieme, and S. Rendle, “Bayesian Factorization Machines,” p. 5.

[9] J. Xiao, H. Ye, X. He, H. Zhang, F. Wu, and T.-S. Chua, “Attentional Factorization Machines: Learning the Weight of Feature
Interactions via Attention Networks,” in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence,
Melbourne, Australia, 2017, pp. 3119–3125, doi: 10.24963/ijcai.2017/435.

[10] J. Chen, B. Sun, H. Li, H. Lu, and X.-S. Hua, “Deep CTR Prediction in Display Advertising,” in Proceedings of the 2016 ACM on
Multimedia Conference - MM ’16, Amsterdam, The Netherlands, 2016, pp. 811–820, doi: 10.1145/2964284.2964325.

[11] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, “DeepFM: A Factorization-Machine based Neural Network for CTR Prediction,”
arXiv:1703.04247 [cs], Mar. 2017.