Multi-view Machines

Bokai Cao, Hucheng Zhou, Guoqiang Li and Philip S. Yu

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

Microsoft Research

Multi-view Data

Multi-view data is prevalent and important in web applications.

- Web documents on Wikipedia
  - URL
  - Words on the page

- Web images on Instagram
  - Visual information
  - Textual tags

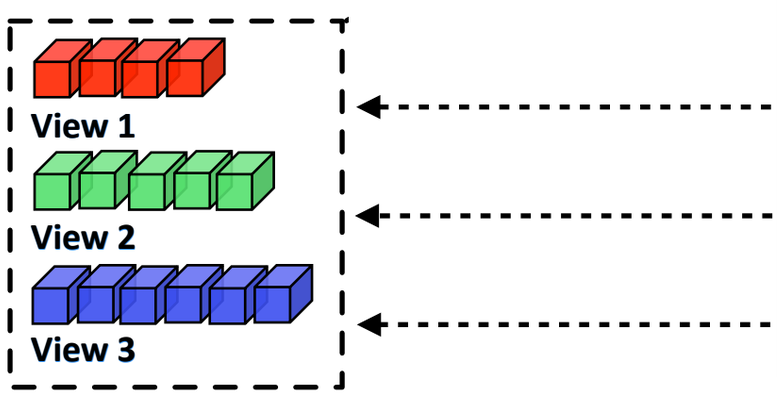
- Web search on Bing
  - User profile
  - Ad information
  - Query description

Multi-view Data

Multi-view data is prevalent and important in web applications.

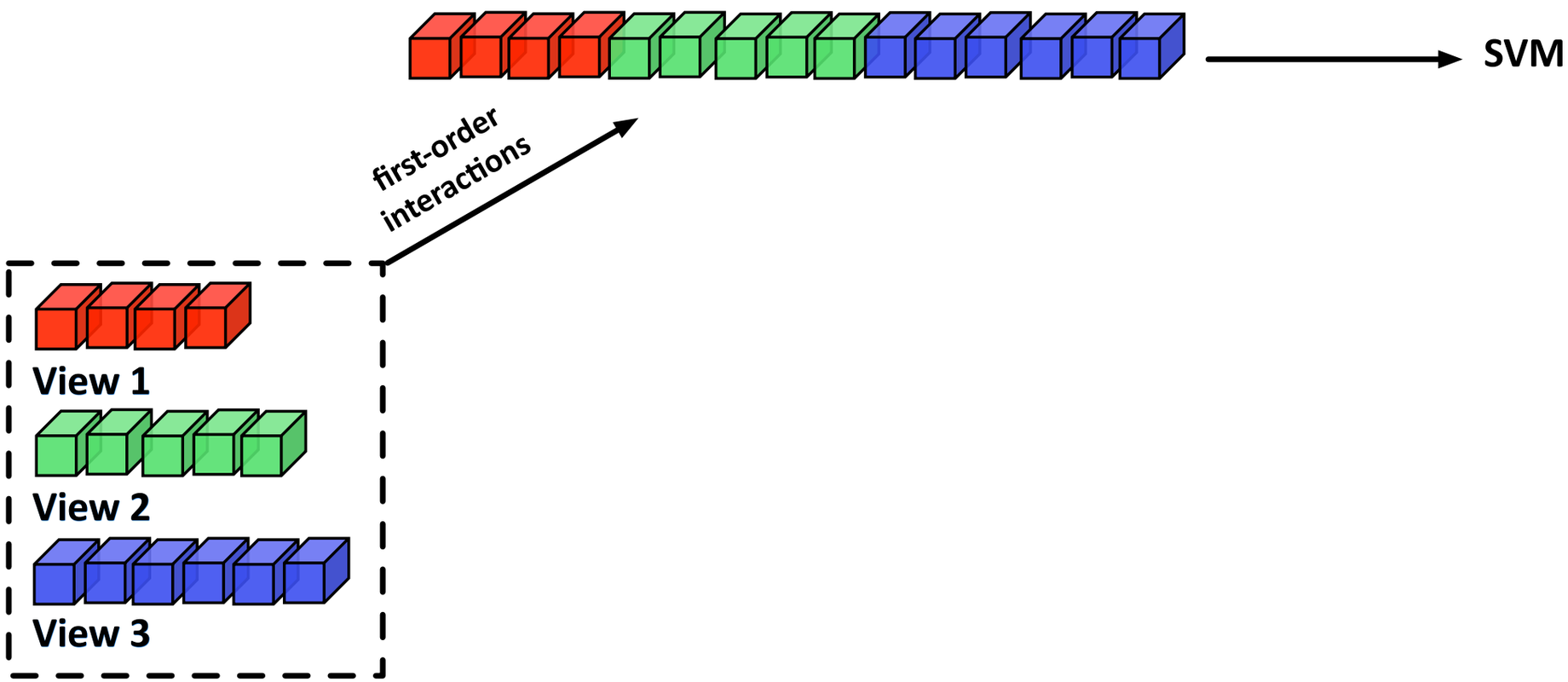
- Web documents on Wikipedia
  - URL
  - Words on the page

- Web images on Instagram
  - Visual information
  - Textual tags



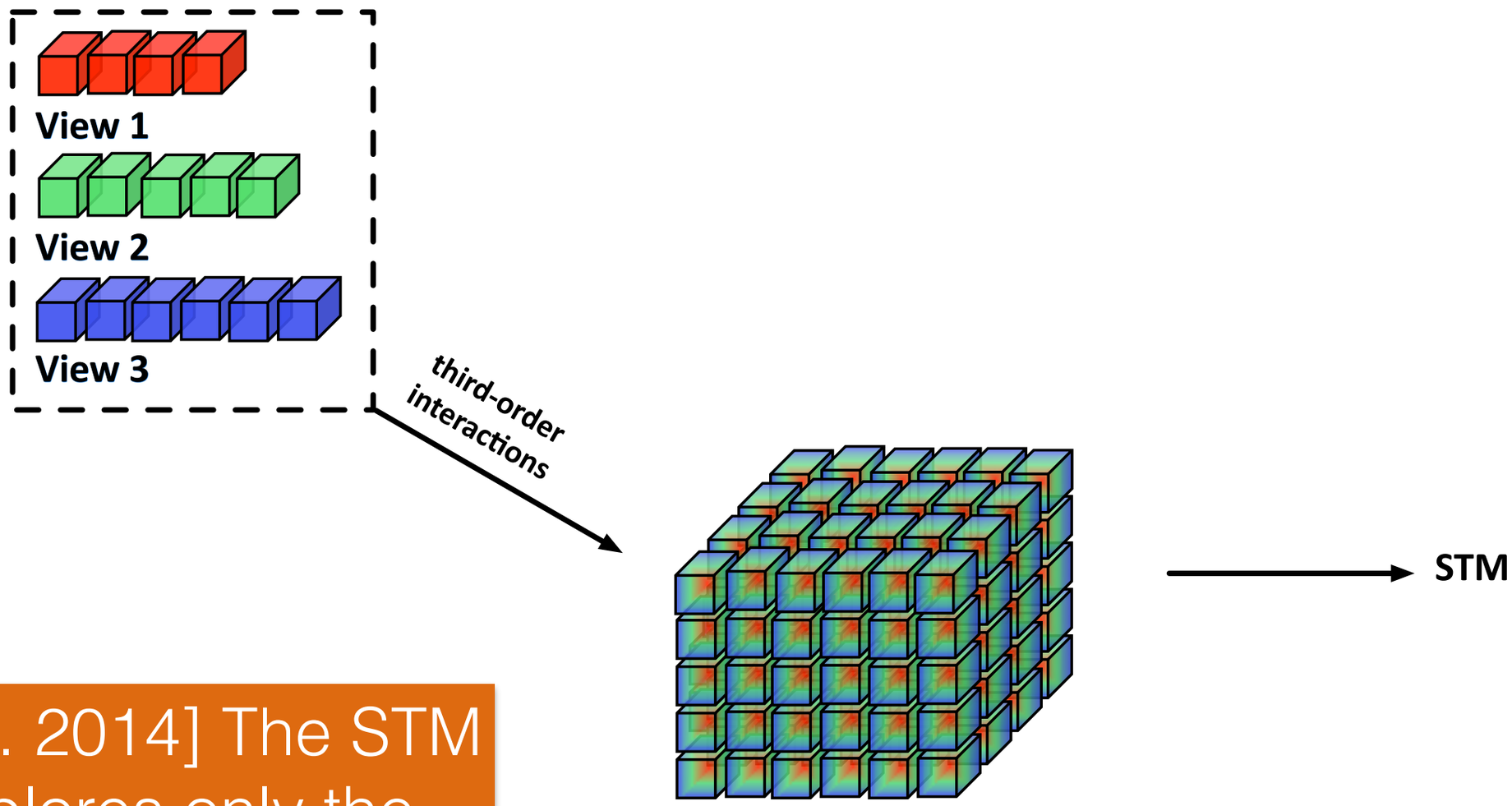
- Web search on Bing
  - User profile
  - Ad information
  - Query description

Support Vector Machines (SVMs)



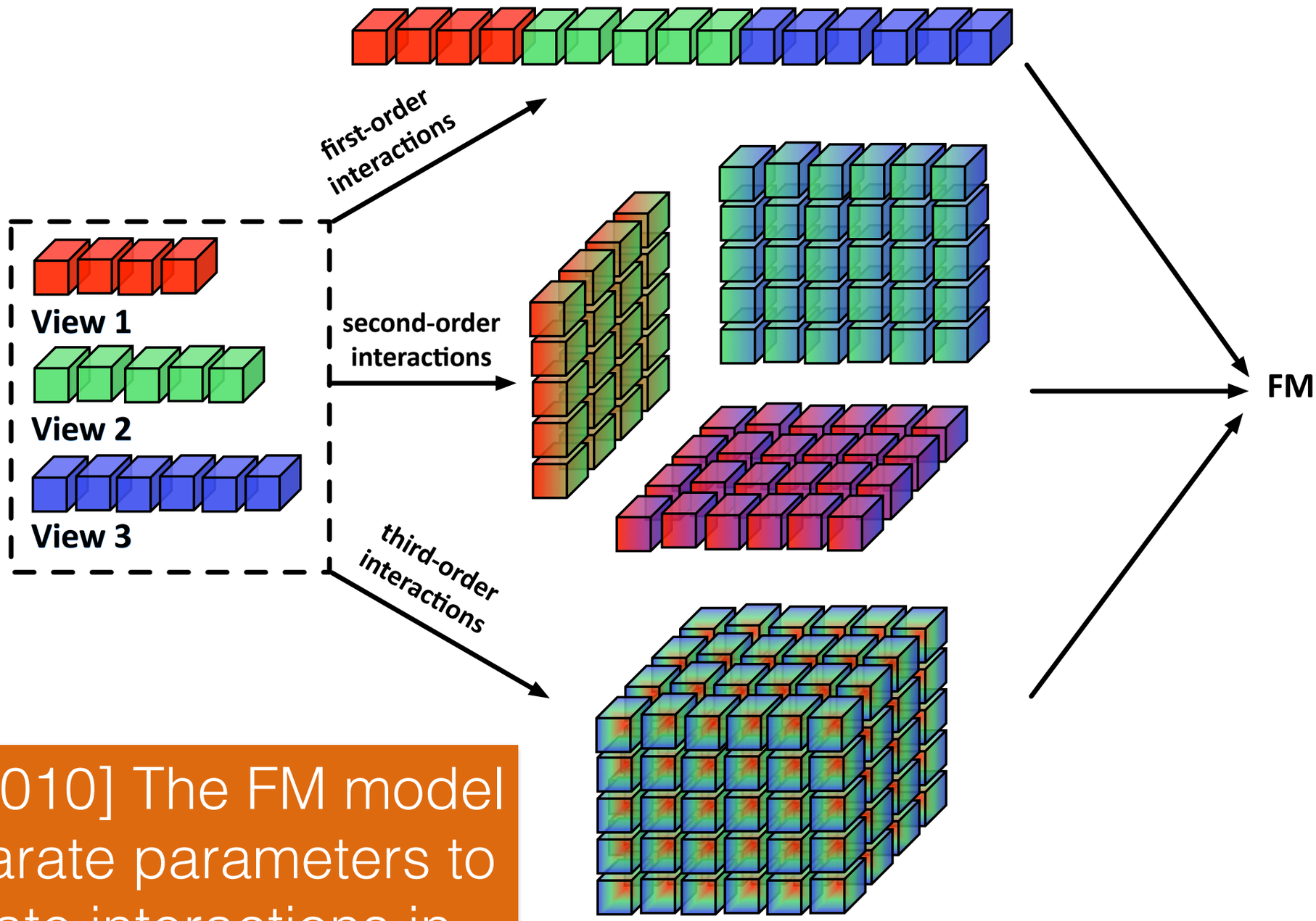
[Vapnik 1995] The linear SVM model (and LR etc.) is limited to the first-order interactions. Polynomial SVMs are likely to degenerate into linear SVMs on sparse datasets.

Support Tensor Machines (STMs)



[Cao et al. 2014] The STM model explores only the highest-order interactions.

Factorization Machines (FMs)



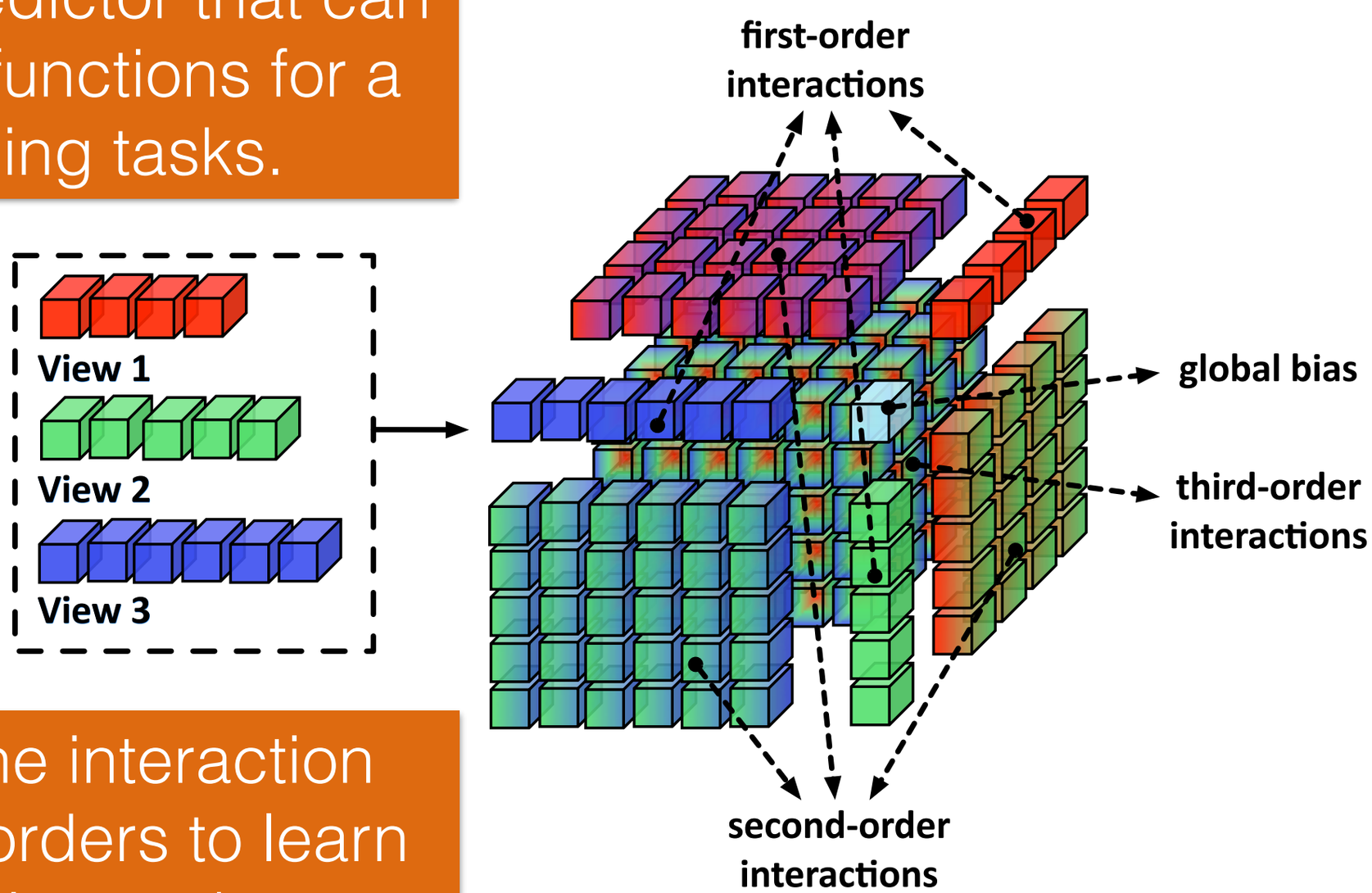
[Rendle 2010] The FM model uses separate parameters to approximate interactions in different orders.

Multi-view Machines (MVMs)

MVMs are a general predictor that can work with different loss functions for a variety of machine learning tasks.

MVMs include the full-order interactions between multi-view features in a compact model.

MVMs jointly factorize the interaction parameters in different orders to learn a consistent representation under sparsity.



Experiments

- Datasets
  - MovieLens: users (138,493), movies (27,278) and implicit feedback (27,278)
  - BingAds: queries (958,426), ads (1,935,510) and impressions (18)
- Baselines
  - Linear regression/logistic regression (LR)
  - Tensor factorization (TF)
  - Multi-view FMs (mvFM, mvFM-3d, mvFM-reg)
- Implementation
  - Gradient descent with AdaGrad
  - GraphX in Spark with iterative forward and backward steps
  - Code available at https://github.com/cloudml/zen/tree/mvm_opt/ml/src/main/scala/com/github/cloudml/zen/ml/recommendation

Experiments

Table 5: Prediction accuracy. ↓ indicates the smaller the value the better the performance; ↑ indicates the larger the value the better the performance.

Dataset	MovieLens (RMSE) ↓	BingAds (AUC) ↑
MVM	0.8376	0.7917
FM	0.8681	0.7872
mvFM	0.8447	0.7729
mvFM-3d	0.9060	0.7201
mvFM-reg	0.9807	0.6947
TF	0.8572	0.6645
LR	1.0017	0.7450

Acknowledgements

- WSDM Student Scholarship
- SIGIR Student Travel Grant
- We also thank Dinggang Shen for his discussions, Chunhui Zhang from the Bing Ads team for providing the dataset and Bo Zhao for helping with the experiments.