Hackathon @ UCSB

Hao-en Sung, Jeff Chang

Motivation

- One advertisement-related problem
- Retargeting makes great profit!!!

Problem

- Given the user’s behavioral logs such as views and orders history, we want to predict the possibility of the conversion.
- When one is attracted to click on advertisement, and also buy the item. It is called the conversion.

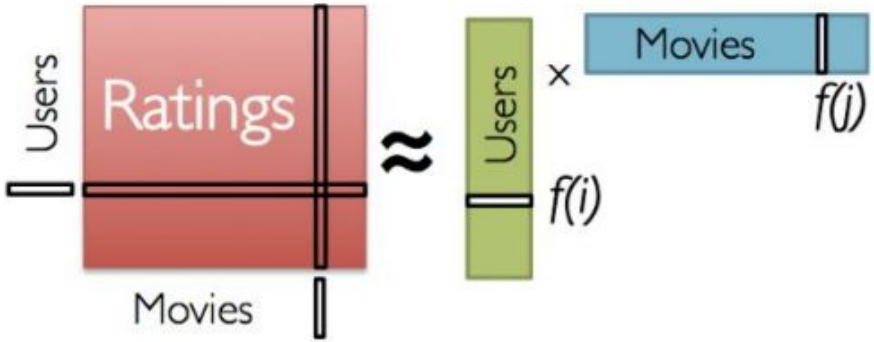
What we have?

- Train data (292784 Users)
- Test data (44094 users)
- View log (44165131 views)
- Order log (2568920 orders)
- … and many many features

Data Analysis

- Platform (iPhone/ Android/ iPad)
- Impression_datatime (timestamp)
- Network (3G/ 4G/ Wifi)
- Media code
- isClicked/ isConversion

Well-known Tools



- Recommendation Systems
  - Content-based model
  - Rating-based model, e.g. Matrix Factorization
- Machine Learning Models
  - Feature-based model, e.g. SVM, Random Forest, Logistic Regression

Image credit: http://www.diva-portal.org/smash/get/diva2:930575/FULLTEXT01.pdf

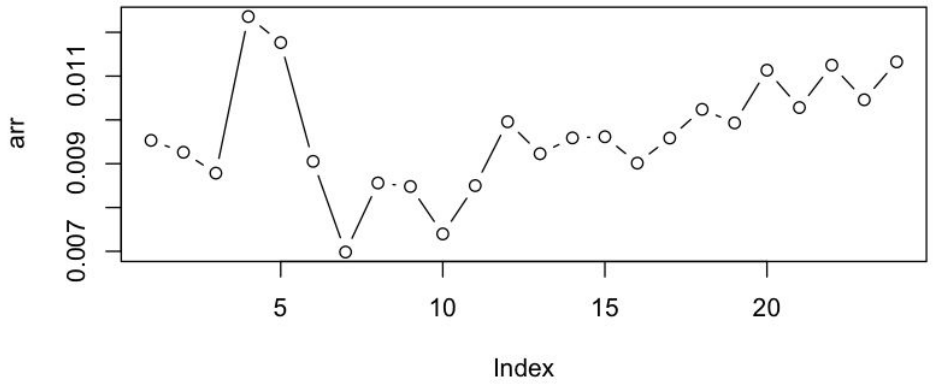
Evaluation Metric

- y = 1 if it is conversion, else y = 0
- p is our predicted possibility
- N is total number of dataset
- Loss function

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)\right)$$

What we once tried...

- Use categorical features, e.g. Platform and Network
- Use human heuristics
  - We usually browse webpages in the daytime, but only buy something when we are not so clear-minded.



Combination of Ideas

- Use Recommendation System to learn latent features.
- Use learned latent features as additional features to build general machine models.
  - however, our model only have little improvement
  - with only processed time feature (T: 0.05494466, V: 0.0546493)
  - with learned latent features (T: 0.05479516, V: 0.05467777)

Conclusion

- Machine learning problems usually take months or years to solve it.
- On top of model trials, more time is spent on finding and creating features.
- However, it is still interesting and exciting for attempts.