DiscoverTED

A TED Talk **Recommender**

Livia Chang
December 2016

Recommend talks to learn both *deeper* and *wider*

“Informative” Talks for users interested in “machine learning big data”

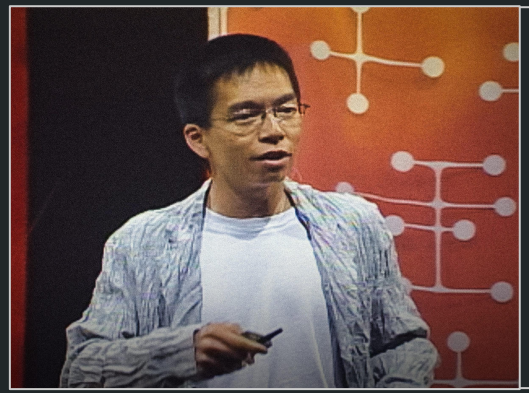
The jobs we'll lose to machines -- and the ones we won't

Anthony Goldbloom (TED2016)

How to fool a GPS

Todd Humphreys (TEDxAustin)

Talks for Deeper



My journey in design

John Maeda (Serious Play 2008)

The beauty of data visualization

David McCandless (TEDGlobal 2010)

Talks for Wider

Data

Talk Data

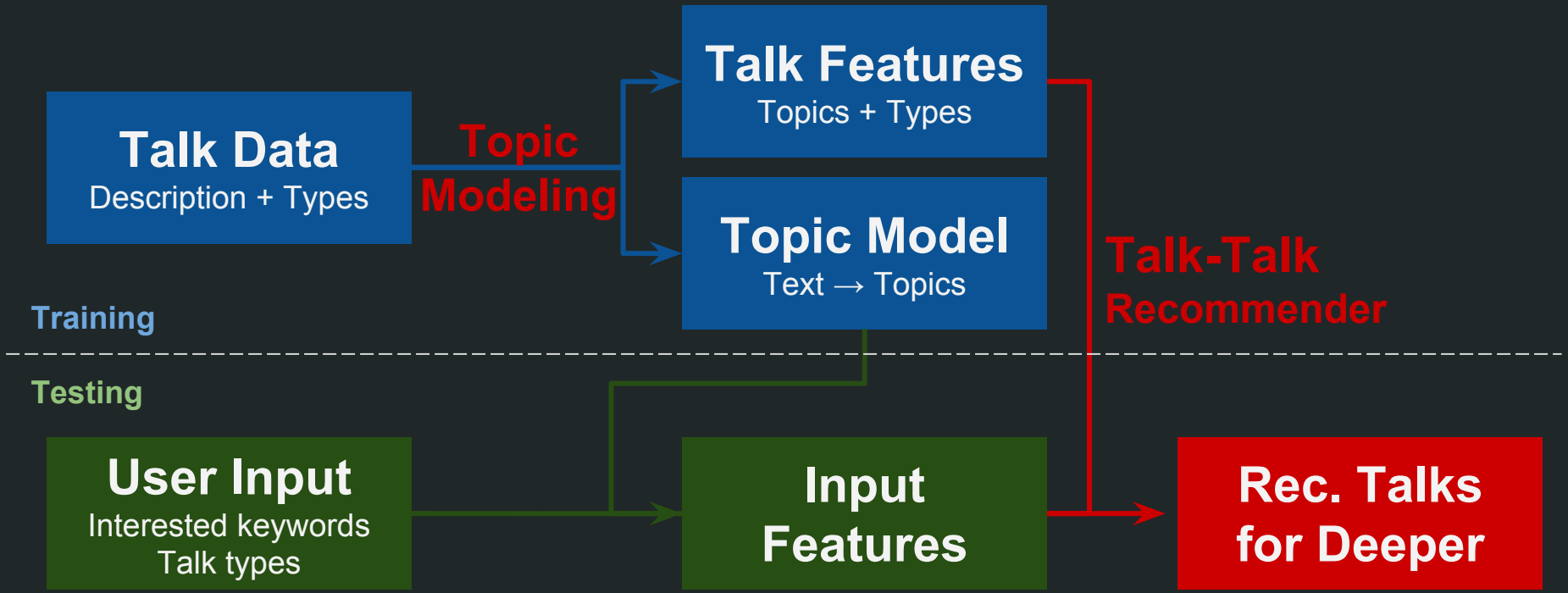
- Fields: titles, tags, description, talk types, ...
- Total 2,318 talks. 1,201 talks are “favorited”
- On average, there are 84.3 users per favorite talk
- Source: Scraped from **TED.com**

User-Talk Data

- Fields: users, favorite talks
- Total 12,401 users. 6,449 active users with 4+ favorite talks
- On average, there are 9.3 favorite talks per user
- Source: **IDIAP** from TED.com

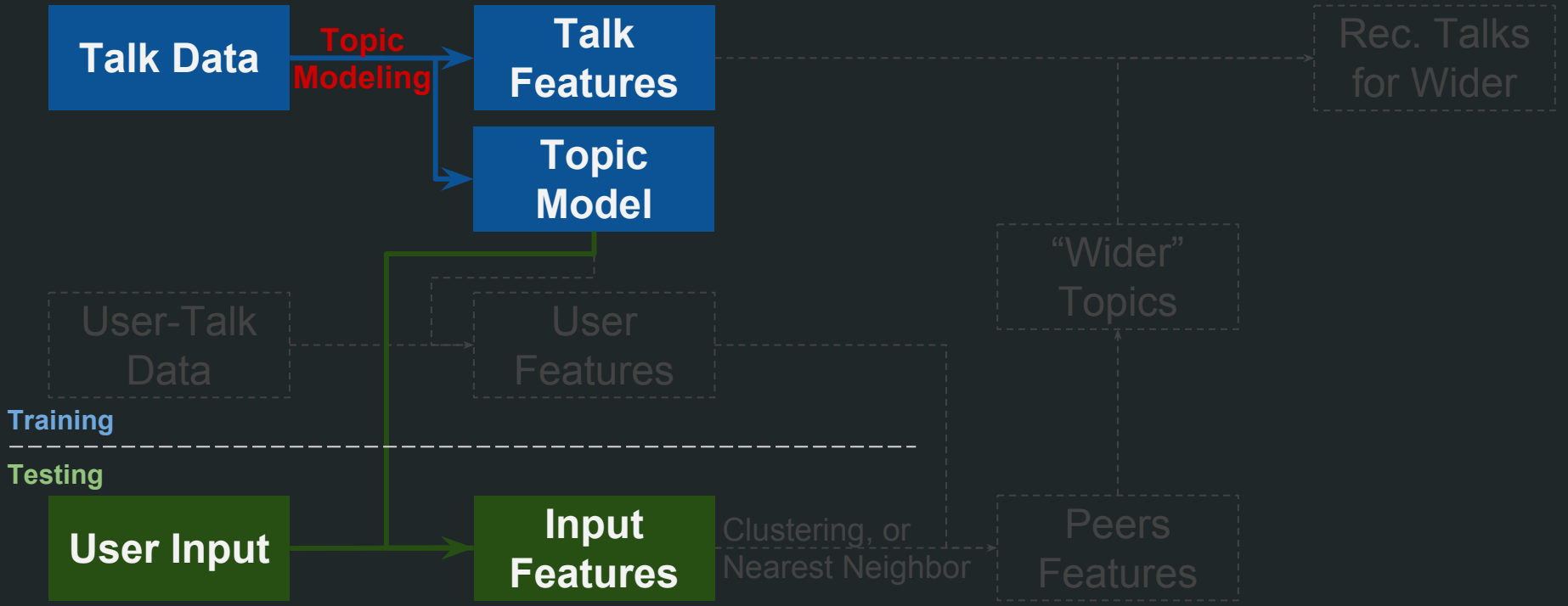
Learn Deeper: Talk-Talk Recommender

Recommend talks closest to a user's interested keywords and talk types



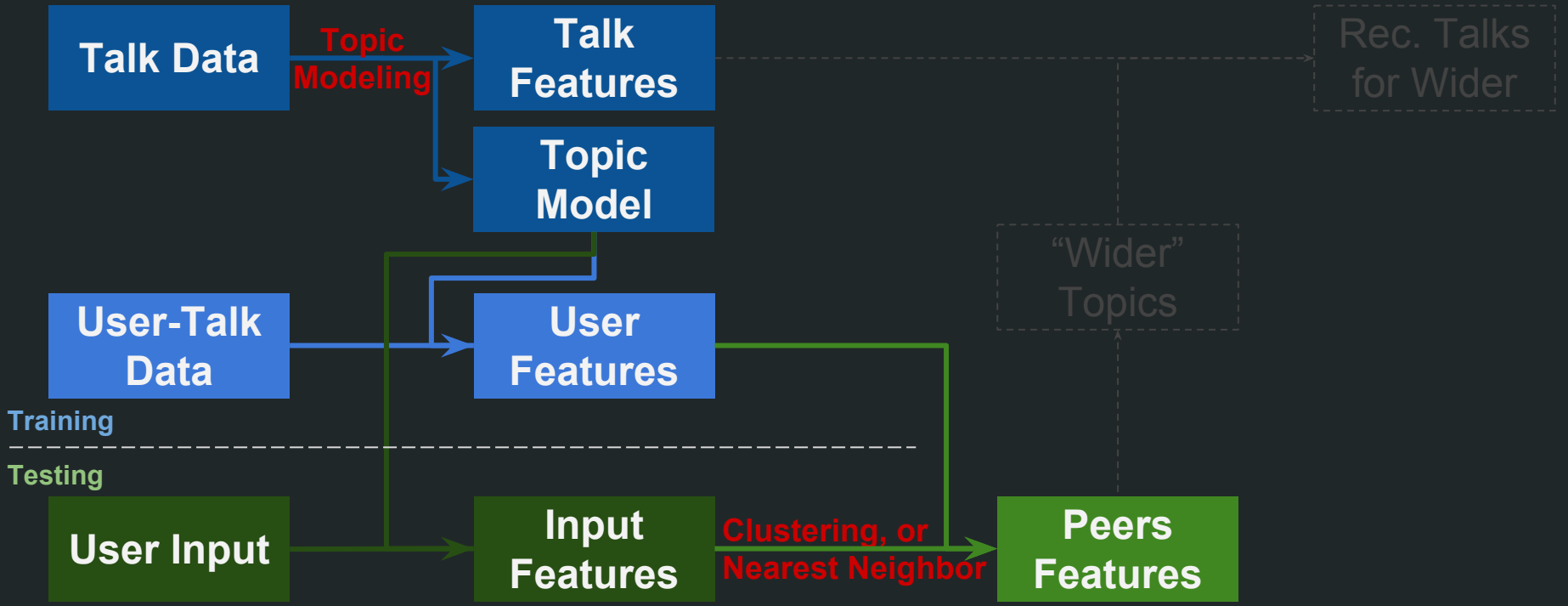
Learn Wider: User-User Recommender

Recommend talks in peers' next favorite topics and closest to a user's interests



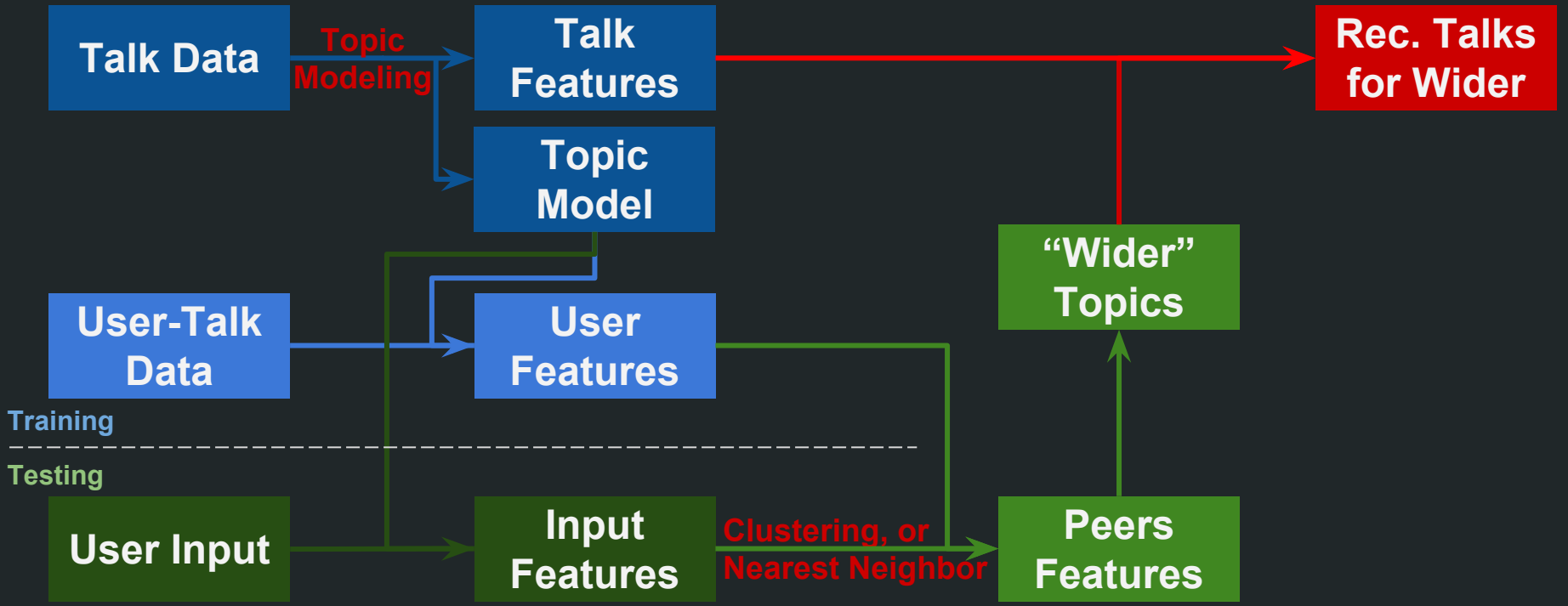
Learn Wider: User-User Recommender

Recommend talks in peers' next favorite topics and closest to a user's interests



Learn Wider: User-User Recommender

Recommend talks in peers' next favorite topics and closest to a user's interests



Model Selections

Natural Language Processing (NLP)
Latent Dirichlet Allocation (LDA)
Nearest Neighbor

Less preferred models:

- Non-negative matrix factorization (sparse data)
- Graphlab matrix factorization (sparse data)
- K-mean clustering (inter- vs. intra- distances)

Evaluation

Compared to random selections,
are recommended talks closer to a user's favorite talks?
➔ Yes!
Random: 1.01 | Deeper Only: 0.84 | Deeper+Wider: 0.89 (smaller distance = better recommendation)

Compared to "deeper" topics only,
do "wider" topics cover more favorite talks?
➔ Yes!
Deeper Only: 1.17 | Deeper+Wider: 1.11 (smaller distance = better recommendation)

Next Steps

Transcript is noisy but can be informative

Usage data for talks “viewed” can be helpful for better prediction

➜ “not like” v.s. “not visit”

➜ “1-minute” v.s. “full-length” watch

Acknowledge

Nikolaos Pappas, Andrei Popescu-Belis, "Combining Content with User Preferences for TED Lecture Recommendation", 11th International Workshop on Content Based Multimedia Indexing, Veszpré Hungary, IEEE, 2013 PDF Bibtex

Thank you & Happy Learning !

https://github.com/liviachang/DiscoverTED