

DiscoverTED: A TED Talk Recommender

Livia Chang

December 2016

TED

IDEAS WORTH **SPREADING**

Overview



Motivation

- TED.com provides good resources for learning
- How to learn more for interesting topics?
- How to discover new topics?

Difficulties

- Sparse Data → model selection
- Cold Start for New Users → required user input
- Evaluation

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



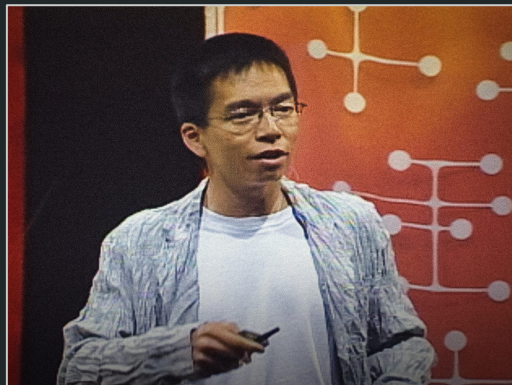
The jobs we'll lose
to machines -- and
the ones we won't

Anthony Goldbloom
(TED2016)

TALKS FOR DEEPER

How to fool a GPS

Todd Humphreys
(TEDxAustin)



My journey in
design

John Maeda
(Serious Play 2008)

TALKS FOR WIDER

The beauty of data
visualization

David McCandless
(TEDGlobal 2010)



Data

Talk Data

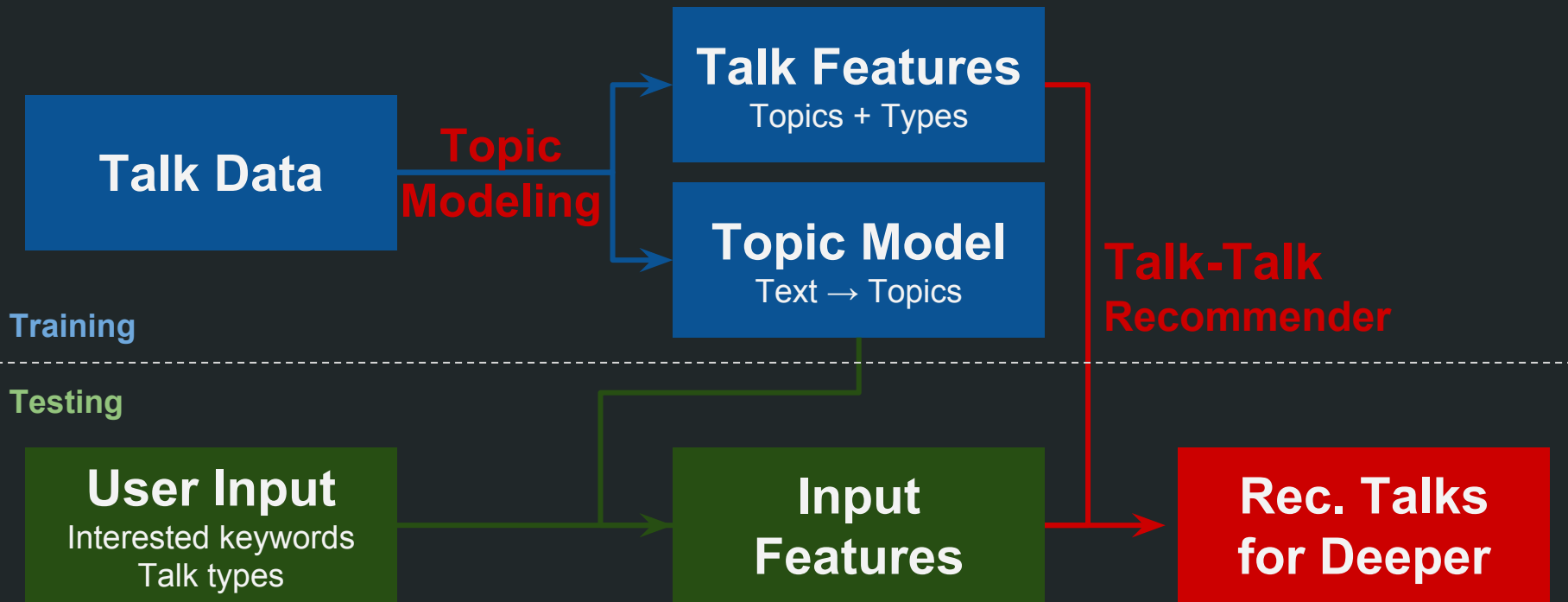
- Fields: titles, keywords, description, talk types, ...
- Total 2,318 talks (1,201 talks are favorited)
- On average, there are 84.3 users per favorited talk
- Source: Scraped from TED.com

User-Talk Data

- Fields: users, favorite talks
- Total 12,401 users. 6,449 active users with 4+ favorite talks (52% of total)
- On average, there are 9.3 favorite talks per user
- Source: IDIAP

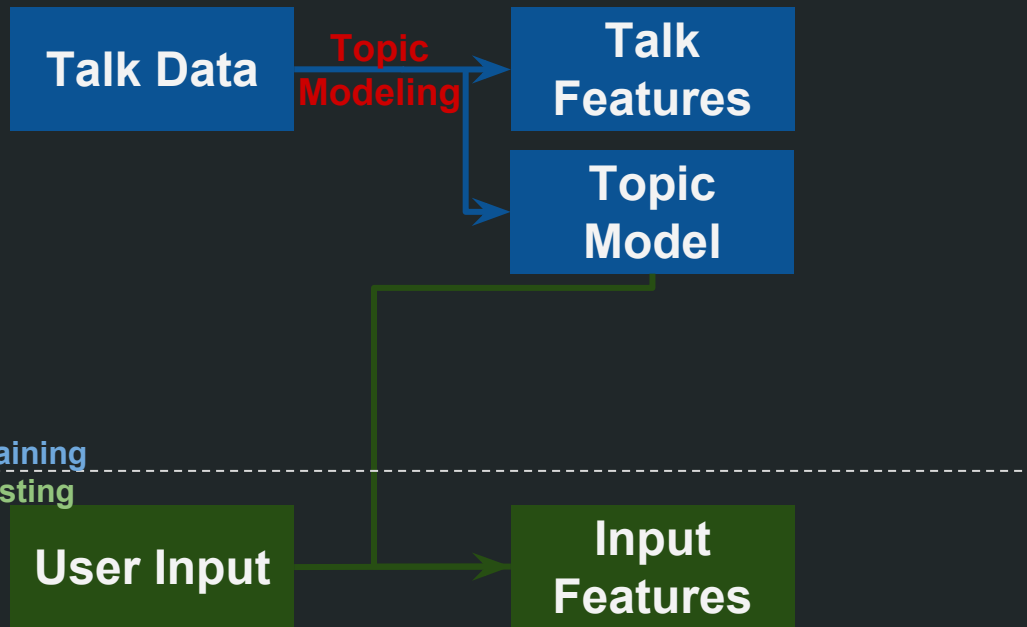
Learn Deeper: Talk-Talk Recommender

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



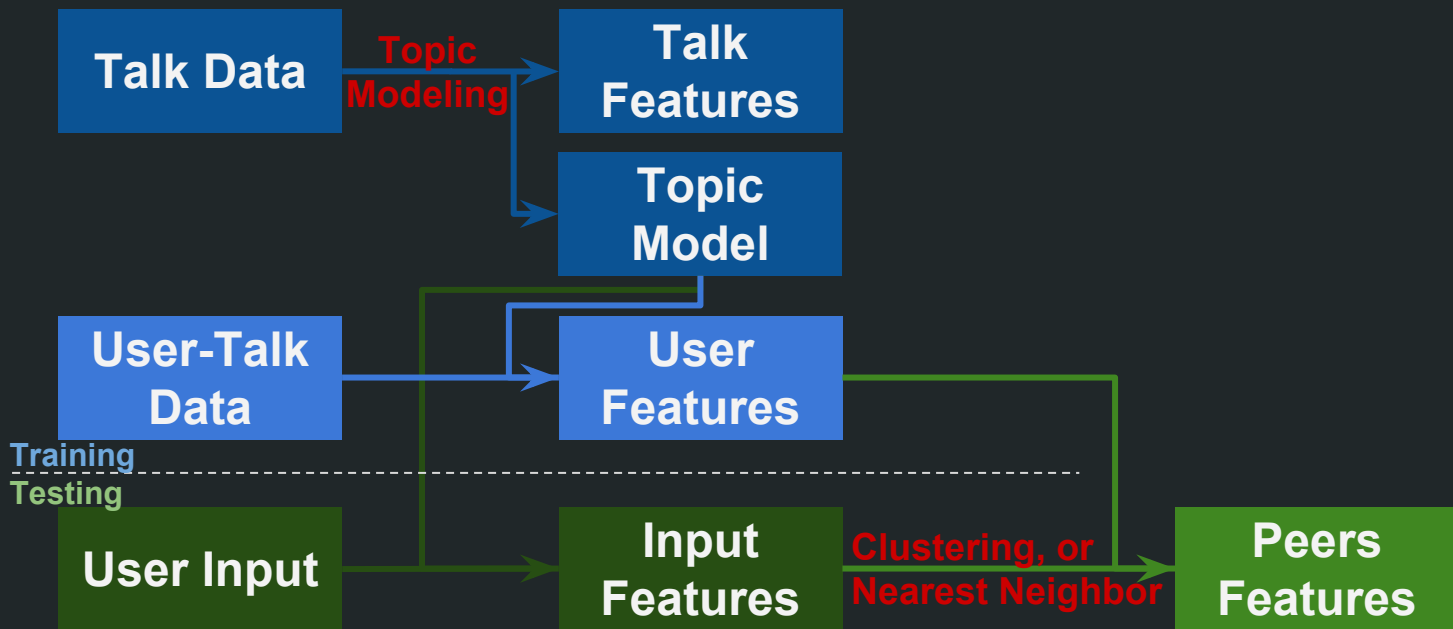
Learn Wider: User-User Recommender

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



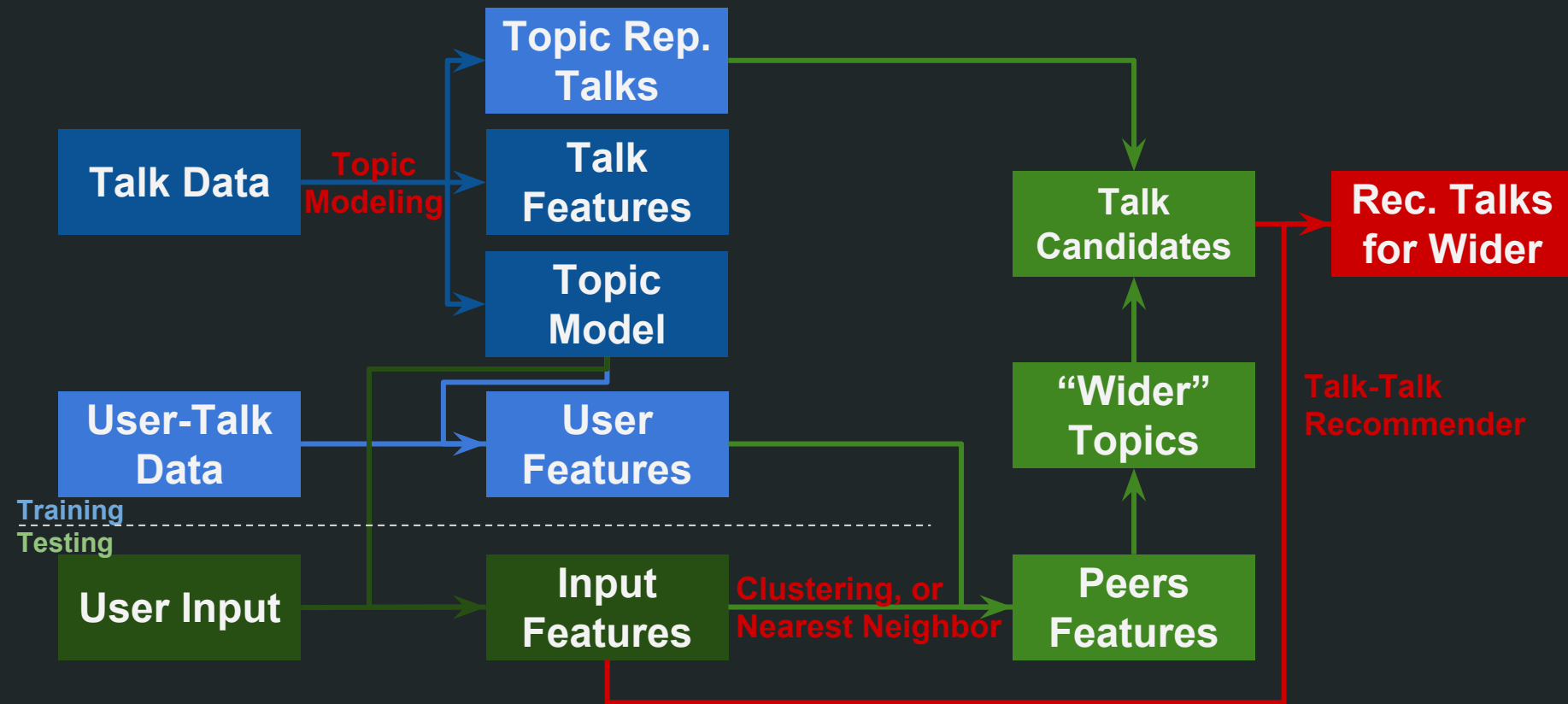
Learn Wider: User-User Recommender

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 and Evaluation

Topic Modeling: NLP + LDA

→ Matrix Factorization (sparse data)

User Modeling: Nearest Neighbor

→ Clustering (inter- vs. intra- distances), Matrix Factorization (sparse data)

Evaluation

- Split existing users into train vs. test users
- The *smaller* distance, the better
- Are recommended talks closer to a user's favorite talks?
 - Random: 1.01 | Talk-Talk Only: 0.84 | Ensemble: 0.89
- Do “wider” topics help cover a user's favorite talks?
 - Talk-Talk Only: 1.17 | Ensemble: 1.11

Future Work

Transcript is noisy but informative
More features to capture usage



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 :)