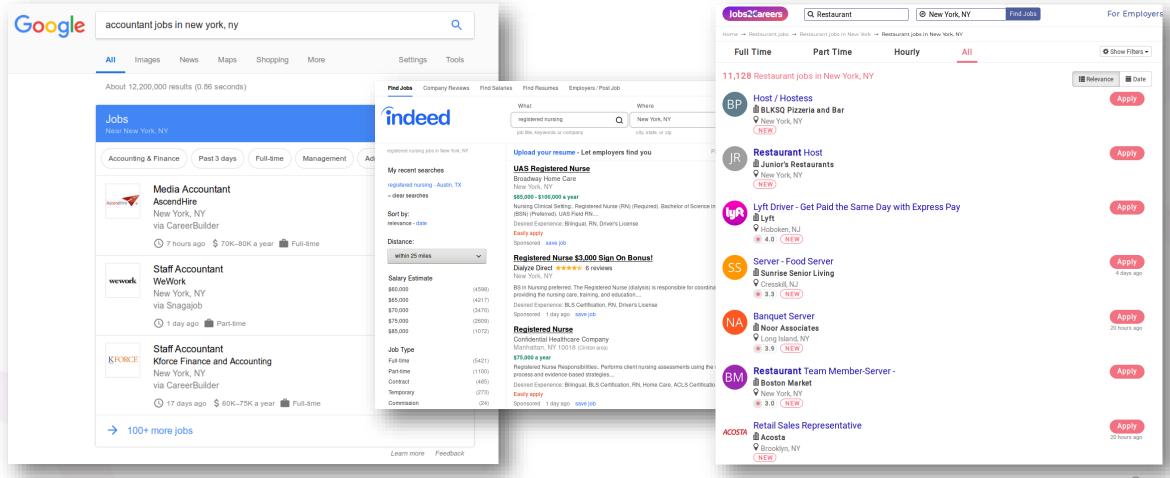
Talroo

Job recommendations leveraging Deep Learning using Analytics Zoo on Apache Spark and BigDL

Job Search

@ Talroo™



Speakers

® Talroo[™]



Wenjing Zhan
Data Scientist, Talroo



Guoqiong SongSoftware Engineer, Intel



Jacob Eisinger
Director of Data, Talroo

Agenda



- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Agenda



- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Where Talent and Recruiting Intersect

@ Talroo™

- Data-driven job ad network
- Self-funded growth
- Collaborative and social culture
- Austin, TX headquarters





Reach







SMS/text

Push notifications

Search engine

Niche job sites

Associations

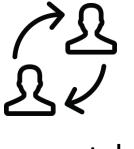
General job boards

Talroo reaches active candidates where they are searching.

Efficiency









Unique candidates

Better matching

 Talroo turns up what works, and turns down what doesn't, ultimately reducing cost-per-hire. Search





Jobseeker

Employer

We power search.

Talroo Monthly Scale



Billions of Queries

Tens of Millions of Interactions

Tens of Billions Impressions

Billions of Jobs Events

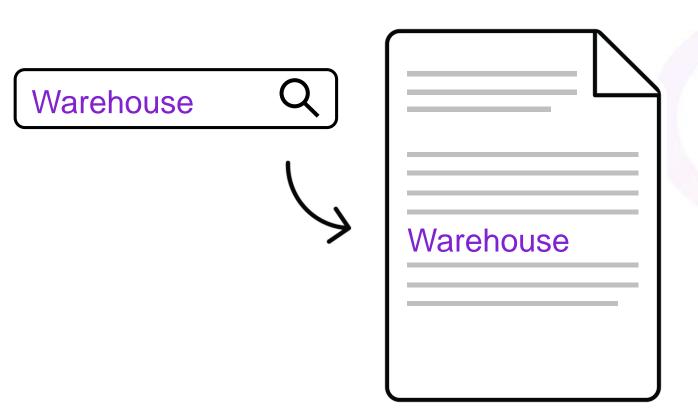
Agenda



- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Traditional Information Retrieval Sufferings @ Talroo

Solution challenges: stemming, synonyms, ontologies, sensitivity



Stemming



Accountant =!= Accounting

Stemming Solution



Accountant =!= Accounting

Accountant = Accounting

Stemming Sufferings



Accountant = Accounting = Account Representative

Synonyms



Synonyms Solution



Registered =!= RN Nurse

Registered Nurse

RN → registered nurse

Synonym Sufferings



Registered =!= RN Nurse

Registered Nurse

RN → registered nurse

Ontologies



Dishwasher =!= Back of House

Ontologies Solution



Dishwasher =!= Back of House

Dishwasher Back of House

Restaurant = Restaurant

Ontologies Sufferings



Dishwasher =!= Back of House

Dishwasher Back of House

Restaurant = Restaurant

• • •

Specificity Suffering

© Talroo[™]



Personalize Results Value



Find the right job faster



Resume

Employers

Find the right person



Job description

Agenda



- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Framing The Problem



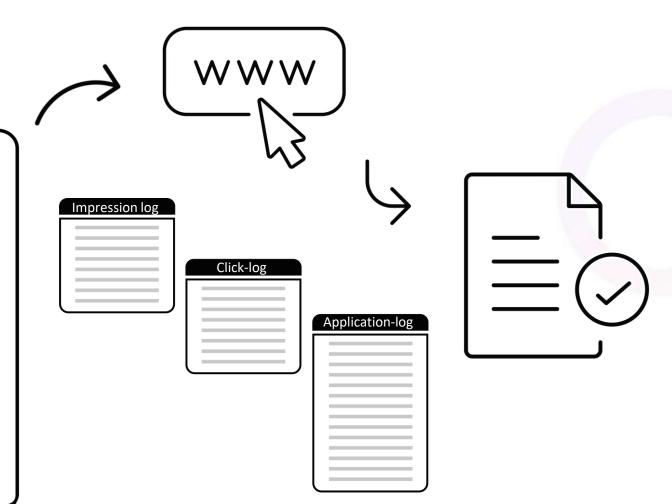
- What information is logged ?
- What do we do with the logged information?
- What did we find from logged information ?
- How to further tackle the findings?

What Information Do We Log

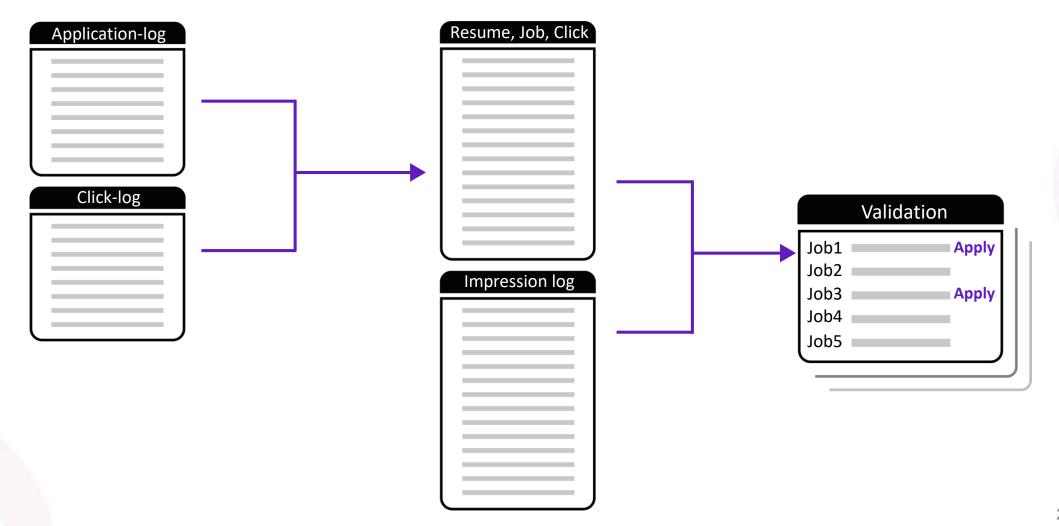
® Talroo™

Warehouse Q

- 1. Lead Warehouse Associate
- 2. Warehouse Workers Needed Immediately (\$17-\$38/Hour)
- 3. B2B Outbound Sales VI
- 4. Telesales Specialist Humana
- 5. Mail & Package Handler



Replaying the Job Application Process © Tolroo



® Talroo [™]

Warehouse Q

0

- 1. Lead Warehouse Associate
- 2. Warehouse Workers Needed Immediately (\$17-\$38/Hour)
- 3. B2B Outbound Sales VI
- 4. Telesales Specialist Humana
- 5. Mail & Package Handler

expect

- 1
- 2
- 3
- 4
- 5

® Talroo[™]

Warehouse Q

0

- 1. Lead Warehouse Associate
- 2. Warehouse Workers Needed Immediately (\$17-\$38/Hour)
- 3. B2B Outbound Sales VI
- 4. Telesales Specialist Humana
- 5. Mail & Package Handler

expect

reality

1

1

2

2

3

3

4

4

5

@ Talroo™

Bilingual International Sales Executive

09/2007 - 09/2017 Store Manager

03/2001 - 08/2007 International Sales Executive

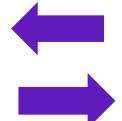
- 1. Lead Warehouse Associate
- Warehouse Workers Needed Immediately (\$17-\$38/Hour)
- 3. B2B Outbound Sales VI
- 4. Telesales Specialist Humana
- 5. Mail & Package Handler

™ Talroo

Bilingual International Sales Executive

09/2007 - 09/2017 Store Manager

03/2001 - 08/2007 International Sales Executive



2. Warehouse Workers Needed Immediately (\$17-\$38/Hour)

Lead Warehouse Associate

- 3. B2B Outbound Sales VI
- 4. Telesales Specialist Humana
- 5. Mail & Package Handler

Investigation Of The Findings

-- First Attempt

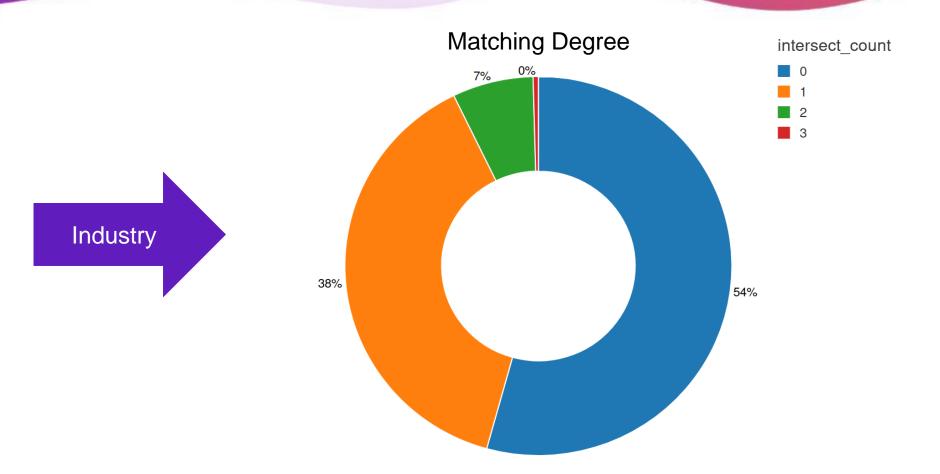


Bilingual International Sales Executive

09/2007 - 09/2017 Store Manager

03/2001 - 08/2007 International Sales Executive

3. B2B Outbound Sales VI

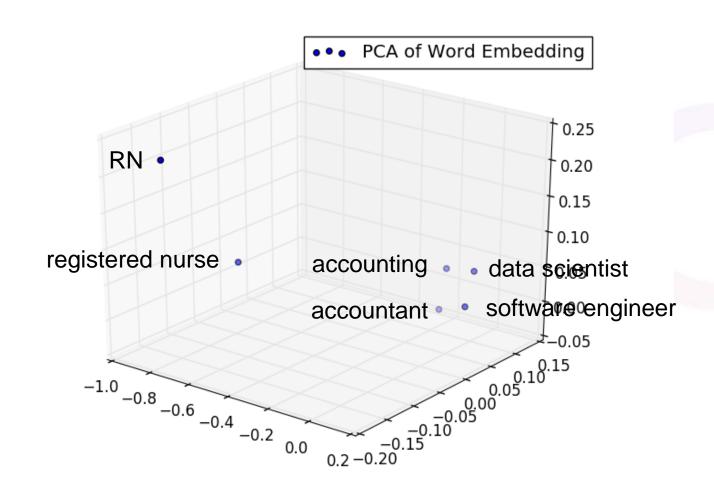


Investigation Of The Findings

-- Reframe The Solution



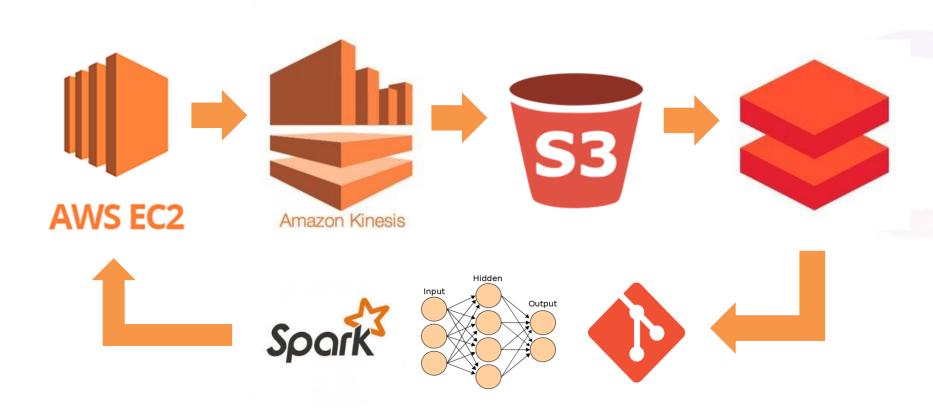
- Handle different terminology
- Identify and respect multiple industries
- Recognize other topics



Ideal Resume Search Requirements

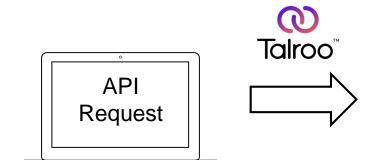
© Talroo[™]

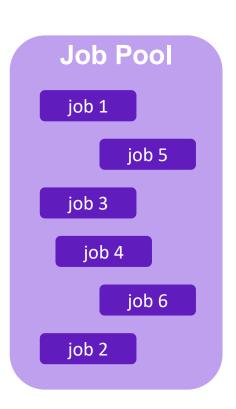


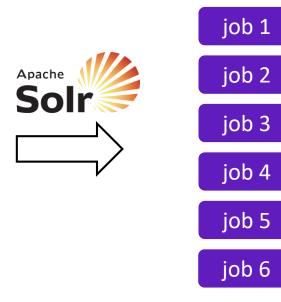


Scoring Before

© Talroo[™]

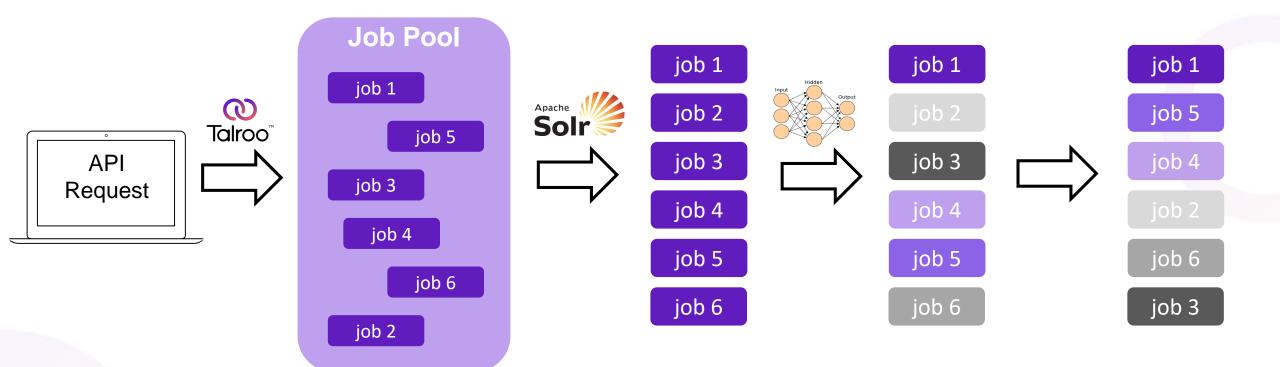






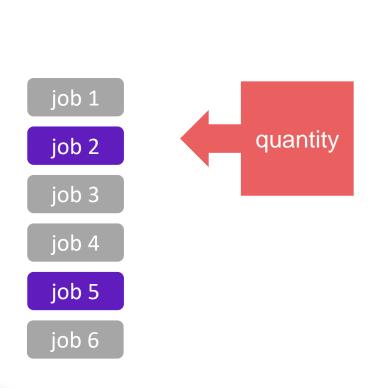
Scoring After - Reranker Solution

@ Talroo®



Evaluation Metric – Gold Standard

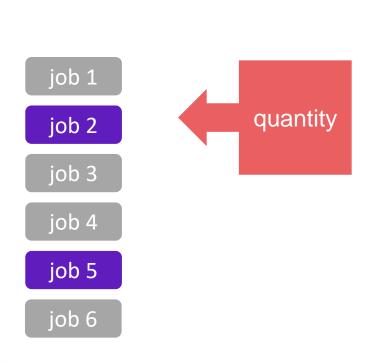




$$CTR = \frac{clicks}{impressions}$$

Evaluation Metric - Quantity

© Talroo[™]



Online

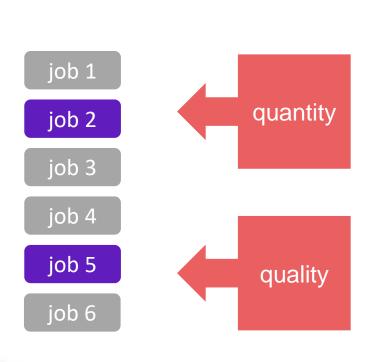
$$CTR = \frac{clicks}{impressions}$$

Offline

$$Precision = \frac{tp}{tp + fp}$$

Evaluation Metric - Quality

© Talroo[™]



Online

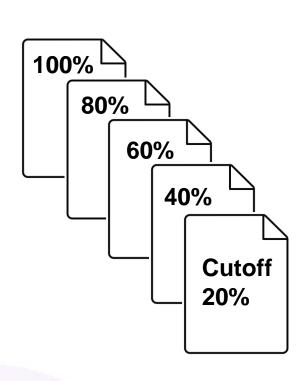
$$CTR = \frac{clicks}{impressions}$$

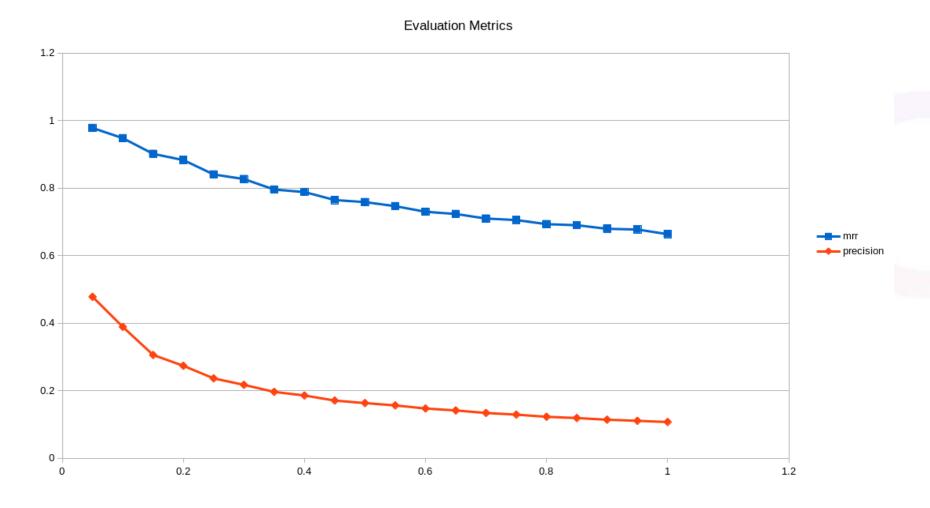
Offline

$$Precision = \frac{tp}{tp + fp}$$

$$MRR = \frac{1}{Q} \sum_{i=1}^{Q} \frac{1}{rank_i}$$

Evaluation Metric - Cutoffs





Resume Search Solution





Agenda

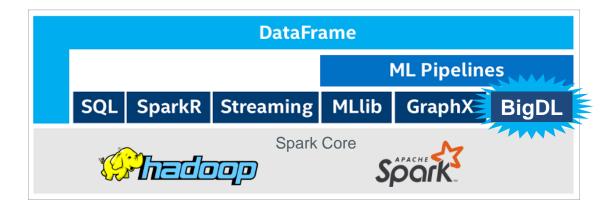


- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Intel BigDL



High Performance Deep Learning for Apache Spark* on CPU Infrastructure



BigDL is an **open-source** distributed deep learning library for Apache Spark* that can run directly on top of existing Spark or Apache Hadoop* clusters

Ideal for DL Models TRAINING and INFERENCE

Designed and Optimized for Intel® Xeon®

No need to deploy costly accelerators, duplicate data, or suffer through scaling headaches!







Lower TCO and improved ease of use *, with existing infrastructure



Deep Learning on Big
Data Platform,
Enabling Efficient
Scale-Out

43

Powered by Intel® MKL and multi-threaded programming

Intel Analytics Zoo



Build and Productionize Deep Learning Apps for Big Data at Scale

Reference Use Cases	 Anomaly detection Sentiment analysis Fraud detection Chatbot, sequence prediction, etc.
Built-In Deep Learning Models	 Image classification Object detection Text classification Recommendations Sequence-to-sequence, GAN, etc.
Feature Engineering	Feature transformations for Image, text, 3D imaging, time series, speech, etc.
High-Level Pipeline APIs	 Native deep learning support in Spark DataFrames and ML Pipelines Autograd, Keras and transfer learning APIs for model definition Support for model serving/inference pipelines
Backends	Spark, BigDL, TensorFlow, etc.

Agenda



- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Analytics Zoo solution



 Retrieve document vectors (resume and job) using Glove vectors

Analytics Zoo recommender model

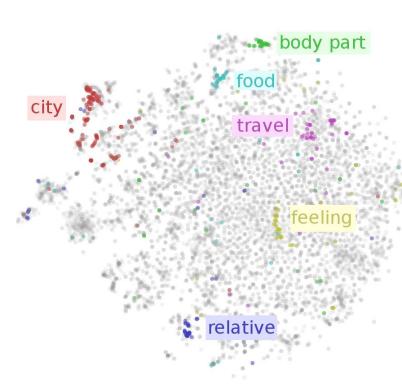
End-to-End flow(offline training)

Evaluation Results

Word Embeddings and GloVe Vectors

@ Talroo™

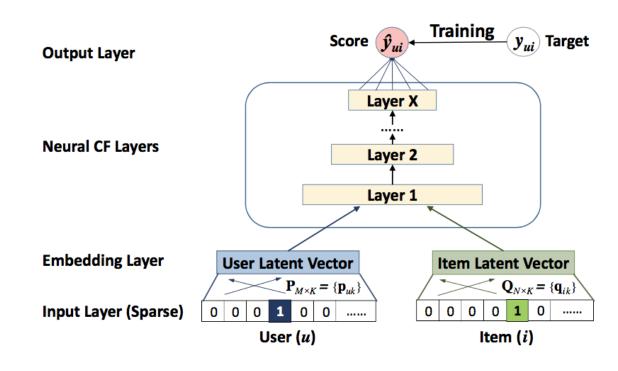
- Words or phrases from the vocabulary are mapped to vectors of real numbers.
- Global log-bilinear regression model for the unsupervised learning algorithm.
- Training is performed on aggregated global word-word co-occurrence statistics from a Wikipedia.
- Vector representations showcase meaningful linear substructures of the word vector space.



Analytics Zoo Recommender Model

@ Talroo®

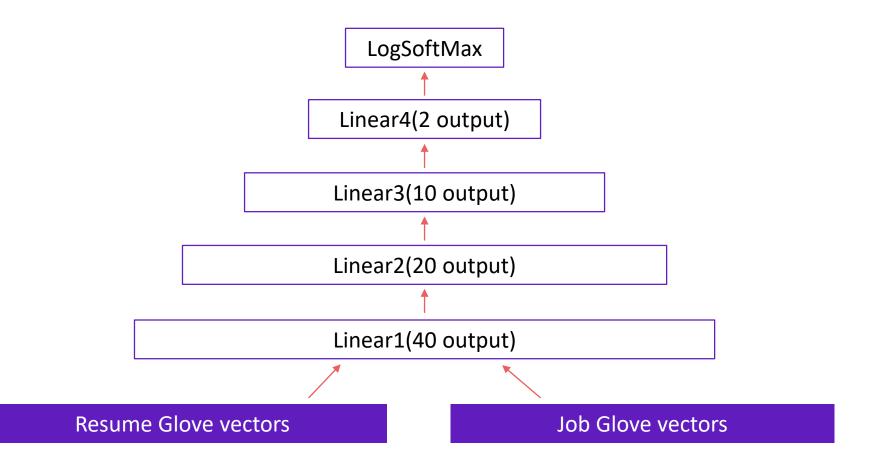
- Neural collaborative filtering,
 Wide and Deep
- Answer the question using classification methodologies
- Implicit feedback and explicit feedback
- APIs
 - recommendForUser
 - recommendForItem
 - predictUserItemPair



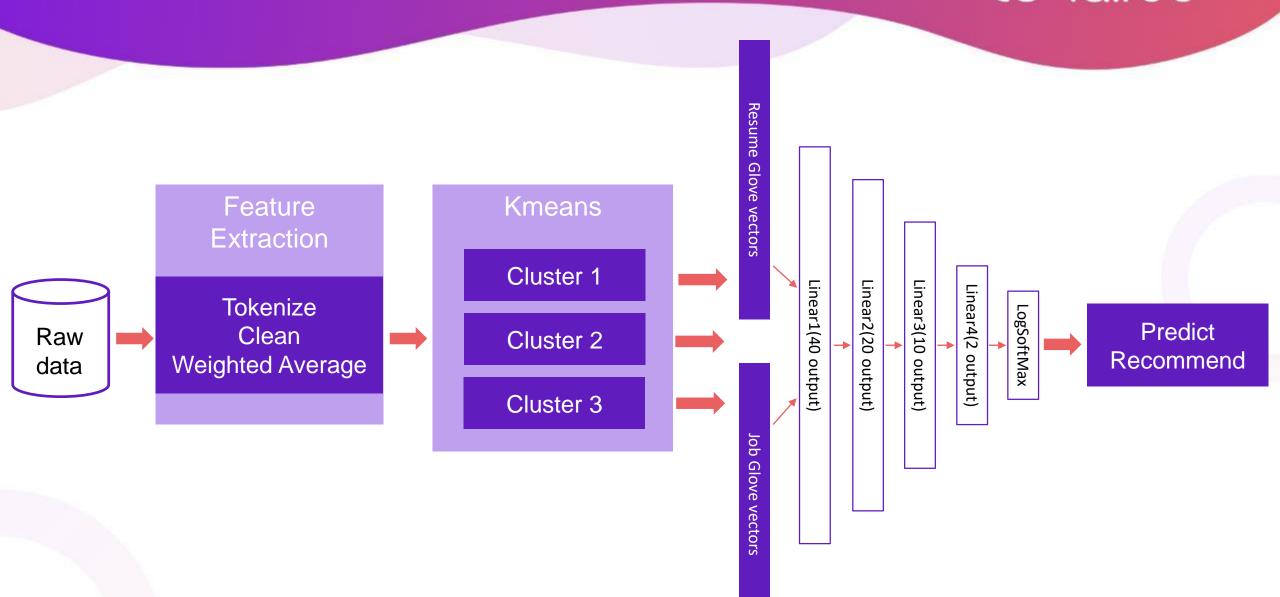
He, 2015

Recommender model



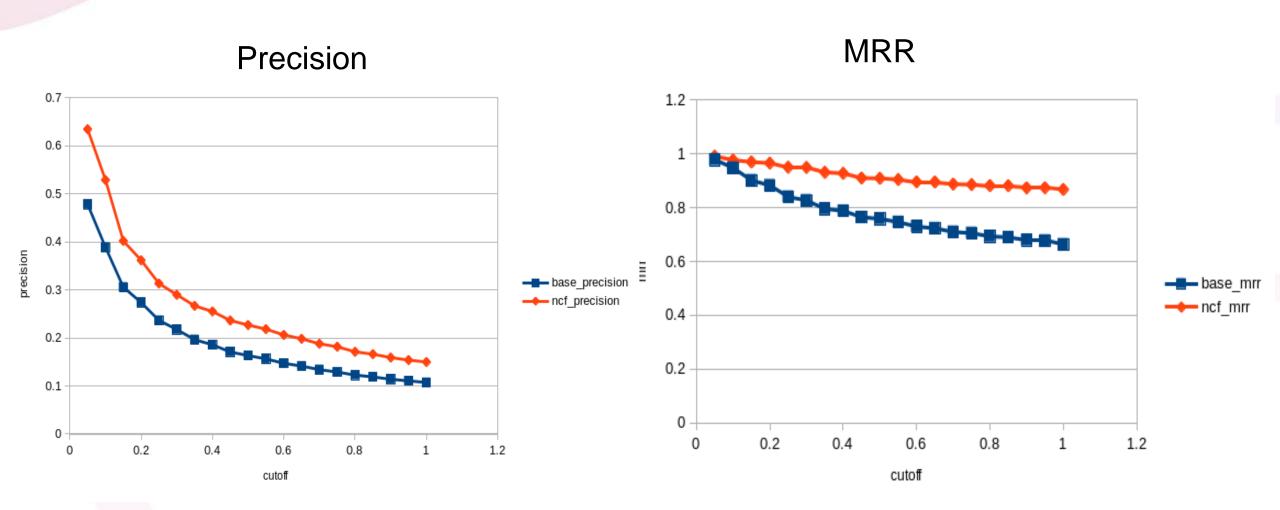


End to End Flow



Evaluation Results





Agenda



- Talroo Overview
- Job Search Challenges
- Resume Search Opportunity
- Analytics Zoo and BigDL Overview
- Resume Search Analytics Zoo Solution
- Lessons Learned

Lessons learned



- Analytics Zoo/BigDL integrates well into existing AWS Databricks
 Spark ETL and machine learning platform
- Analytics Zoo/BigDL scales with our data and business
- Jobs and resumes can be effectively modeled and processed through embeddings
- Ensembling multiple models and glove embedding feature embedding proved to be very effective for rich content
- More information available at https://analytics-zoo.github.io/

Legal Disclaimer



- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.
- No computer system can be absolutely secure.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Questions?



Wenjing Zhan
Data Scientist, Talroo
wenjing@talroo.com



Guoqiong SongSoftware Engineer, Intel guoqiong.song@intel.com



Jacob Eisinger
Director of Data, Talroo
jacob@talroo.com

End to End Flow

