Modeling Professional Similarity by Mining Professional Career Trajectories

Ye Xu, Zang Li, Abhishek Gupta
Ahmet Bugdayci, Anmol Bhasin
Outline

• Motivation
• Problem Statement
• Approach
• Evaluation
  ▪ Future Works
SimilarProfiles:
- Source profile -> return K target profiles with similar skills, titles, companies...
- keyword-based text matching

Temporal information in a LinkedIn profile
Software Engineer
LinkedIn
June 2012 – Present (1 year 3 months) | San Francisco Bay Area
SNA(Search, Network, and Analytics) team, working on Link Prediction, learning models, large scale data mining of big data etc.

Research Engineer
Cisco Systems
May 2011 – June 2012 (1 year 2 months) | San Francisco Bay Area
Data analysis and research team of STBU Security Application (machine learning and web security).

Research Assistant
The Pennsylvania State University
August 2007 – March 2011 (3 years 8 months) | State College,
Smart sensing lab. Doing research on access control, activity recognition, processing, semantic web, smart spaces, Internet of Things.

Engineer
China Mobile International Limited
June 2006 – August 2007 (1 year 3 months) | Guangzhou, China
Group of enterprise development and technology planning, technology and services (IMS, mobile positioning services, etc.).

Software Engineer
LinkedIn
August 2011 – Present (2 years 1 month) | San Francisco Bay Area
Application engineer

Research Assistant
Nokia Research Center
September 2010 – June 2011 (10 months) | Singapore
Data Mining, Semantic Web

Research Assistant
National University of Singapore
June 2008 – August 2010 (2 years 3 months) | Singapore
Data Mining, Statistical Analysis

Research Assistant
Duke University
September 2010 – July 2012 (1 year 11 months) | Durham, NC
Work on "Starfish," a self-tuning analytics system on MapReduce. Optimized Hadoop performance with Amazon EC2. Built cost model for MapReduce jobs and an optimizer. Implemented a visualizer to demonstrate the profiler. Applied optimizations to MapReduce workflows of data.

Software Engineer Intern in EC2 Team
Amazon.com
May 2011 – August 2011 (4 months) | Greater Seattle Area
Work in the High Performance Computing team at EC2. Designed and implemented the automatic regression test framework. 1 project

Software Engineer
Baidu, Inc.
January 2010 – August 2010 (8 months) | Beijing, China

People Similar to Ziang
1. Wilson Ngo
2. Zhiyong Zhang
3. Chang S.
4. Alek Kolcz
5. Jie Yang
6. Jimmy Chen
7. Fei Dong
Rerank Top K SimProfiles: Career trajectory Similarity
Outline

• Motivation
• Problem Statement
• Approach
• Evaluation
  ▪ Future Works
Keywords Profile Model

**Title:** Software Engineer, Research Engineer, Research Assistant

**Specialty:** Machine Learning, Data Analysis, Hadoop, Networks

**Company:** Cisco, Linkedin, Penn State

**Summary:** Software Engineer, Research Engineer, Machine Learning, Data Analysis, Networks, Research Assistant

Sequence Profile Model

- **Summary:** ML, Hadoop
  - **Company:** Linkedin
  - **Title:** Software Engineer
  - **Duration:** (2011.5-2013.3)

- **Summary:** Data Analysis
  - **Company:** Cisco
  - **Title:** Research Engineer
  - **Duration:** (2010.7-2011.4)

- **Summary:** Networks
  - **Company:** Penn State
  - **Title:** Research Assistant
  - **Duration:** (2006.9-2010.6)

How to model a profile with career trajectory?
Similar Profiles

- **Summary**: ML, Hadoop
- **Company**: Linkedin
- **Title**: Software Engineer
- **Duration**: (2011.5-2013.3)

- **Summary**: Data Analysis
- **Company**: Cisco
- **Title**: Research Engineer
- **Duration**: (2010.7-2011.4)

Similar Career Paths

- **Summary**: ML
- **Company**: Linkedin
- **Title**: Software Engineer
- **Duration**: (2012.7-2013.3)

- **Summary**: Mobile
- **Company**: Intel
- **Title**: Software Engineer
- **Duration**: (2010.5-2012.3)

- **Summary**: Hadoop, DM
- **Company**: Yahoo
- **Title**: Research Scientist
- **Duration**: (2008.8-2010.4)
Outline

• Motivation
• Problem Statement
• Approach
• Evaluation
  ▪ Future Works
Approach: Sequence Alignment

• Match and align positions from two career sequences according to the similarity between them.
  1. What is position-level similarity function?
  2. How to identify optimal match of pairs of positions?
Position Similarity Function

• Measure for position similarity
  – Build a Logistic regression model
  – Feature interaction
    • company, industry, title, seniority, job function, summary
    • E.g., title: whether the two positions share the same title name
  – Feature weight learning
    • Label: InMail via Similar Profiles
      – Use recruiter InMail to infer position similarity label.
      – Assumption: If recruiters send inMails to a target user from SimilarProfiles results of the source user, then the current positions (between source and target) are similar.
      – Amongst the rest of the recommended profiles, all the ones that were ranked higher but were not reached out by any recruiter are regarded as negative profile-pairs
Sequence Alignment

\[ \text{PathSim}[1:i, 1:j] = \max \left\{ \begin{array}{l}
\text{PathSim}[1:i-1, 1:j-1] + \text{PosSim}[i, j] \\
\text{PathSim}[1:i-1, 1:j] - \text{penalty} \\
\text{PathSim}[1:i, 1:j-1] - \text{penalty}
\end{array} \right\} \]

- **Match Pos i and Pos j**
- **Skip Pos i**
- **Skip Pos j**
Sequence of Positions

• Profile 1
  - **Summary:** ML, Hadoop
  - **Company:** Linkedin
  - **Title:** Software Engineer
  - **Duration:** (2011.5-2013.3)

  - **Summary:** Data Analysis
  - **Company:** Cisco
  - **Title:** Research Engineer
  - **Duration:** (2010.7-2011.4)

  - **Summary:** Networks
  - **Company:** Penn State
  - **Title:** Research Assistant
  - **Duration:** (2006.9-2010.6)

• Profile 2
  - **Summary:** ML
  - **Company:** Linkedin
  - **Title:** Software Engineer
  - **Duration:** (2012.7-2013.3)

  - **Summary:** Mobile
  - **Company:** Intel
  - **Title:** Software Engineer
  - **Duration:** (2010.5-2012.3)

  - **Summary:** Hadoop, DM
  - **Company:** Yahoo
  - **Title:** Research Scientist
  - **Duration:** (2008.8-2010.4)

  - **Summary:** Sensor
  - **Company:** Dartmouth
  - **Title:** Ph.D. Student
  - **Duration:** (2002.9-2008.7)
Sequence of Compositions

• Profile 1

- Summary: Data Analysis
- Company: Cisco
- Title: Research Engineer
- Duration: (2010.7-2011.4)

- Summary: Networks
- Company: Penn State
- Title: Research Assistant
- Duration: (2006.9-2010.6)

- SenChg: 7
- ComChg: YES
- Duration: 46

- SenChg: 2
- ComChg: YES
- Duration: 10

• Profile 2

- Summary: ML, Hadoop
- Company: LinkedIn
- Title: Software Engineer
- Duration: (2011.5-2013.3)

- Summary: ML
- Company: LinkedIn
- Title: Software Engineer
- Duration: (2012.7-2013.3)

- SenChg: 7
- ComChg: YES
- Duration: 21

- Summary: Mobile
- Company: Intel
- Title: Software Engineer
- Duration: (2010.5-2012.3)

- SenChg: 0
- ComChg: YES
- Duration: 4

- Summary: Hadoop, DM
- Company: Yahoo
- Title: Research Scientist
- Duration: (2008.8-2010.4)
Sequence of Compositions

• Similarity measure for composition node
  – **Position features**: company, industry, title, seniority, job function, summary
    • E.g., title: whether the two positions share the same title name.
  – **Transition features**: same company, same industry, title change, seniority change, year durations
    • E.g., same company: whether both of the two transitions happen within the same company or not.
Recency and Duration

- Incentivize matching more recent positions
- Incentivize matching longer duration positions

\[
S^{seq}(P_1[1 : i], P_2[1 : j]) = \max \begin{cases} 
S^{seq}(P_1[1 : i - 1], P_2[1 : j - 1]) + w(i, j)S^{mode}(X_i, Y_j) \\
S^{seq}(P_1[1 : i], P_2[1 : j]) - \lambda \\
S^{seq}(P_1[1 : i], P_2[1 : j - 1]) - \lambda
\end{cases}
\]

\[
w(i, j) = e^{-\tau_1 |r_i - r_j|} e^{-\tau_2 |d_i - d_j|} e^{-\tau_3 (r_i + r_j)} (1 - e^{-\tau_4 (d_i + d_j)})
\]
Similar Profiles score as a prior

• Sparse positions

• Bagging

\[ \text{score} = sp + \delta \times sc \]
Outline

• Motivation
• Problem Statement
• Approach
• Evaluation
  ▪ Future Works
Offline Evaluation: NDCG

- get the testing data from *Similar Profiles Recommender System* on LinkedIn Recruiter.
- we employ InMail sending and receiving to infer this ground truth.
Online Evaluation: A/B Test

\[
\text{score} = sp + \delta \ast sc
\]

- When giving more weight to the SimCareers score, we get a larger increase in both Profile Views and InMails sent.
- Capturing temporal information of career trajectory is crucial in computing profile similarity for hiring.
Rerank Top K Similar Profiles Recommendations
Outline

• Motivation
• Problem Statement
• Approach
• Evaluation
  ▪ Future Works
Future Works

- Skip VS. **merge**
  - Lateral move, promotion
  - Unaligned position does **not hurt**.
Thanks!