Translate Theory into Model

Decipher Job Mobility

Lun Li

Ph.D in Data Science
Stevens Institute of Technology

April 17, 2023

Overview

Two Main Theory

We attribute job mobility to the following two theories:

Theory I: Categorization Theory & Prototype Logic

An employee with high typicality in relation to a target title will be more likely to obtain that title.

Theory II: Social Influence Logic

If two employees have previously shared a title, and one of the employees switched to a different title, then the other employee will be more likely to obtain that title.

Hypothesis

The integration of Theory I and Theory II explains the major variability in job mobility.

Our Objective

Objective

The main objective is to mathematize theory I & II and incrementally build an explainable model to validate our hypothesis.

Approach

Build a probabilistic model for the conditional distribution of the target job,

$$\mathbb{P}(\textit{next job title} = t' \mid \textit{context } \mathcal{F}) = G(\textit{context}) \tag{1}$$

$\mathcal{F}$ encodes the information available before knowing the next job title. We make no assumption here, the advanced probability theory enforces the existence of $G(\cdot)$ as a function of context. Therefore, all we need is to propose a parametric form for G.

Translate Theory I (Markovian Setting)

Theorem I – Markovian Typicality

Assume $\mathcal{F} = \{current\ citle\}$, then the next job only depends on the current title, i.e.,

$$\mathbb{P}(next\ job\ title = t' \mid \mathcal{F}) = G(current\ title) \tag{2}$$

Proposals for $G(\cdot)$

Suppose each title has a lower-dimensional representation $z_\cdot$, the likelihood of getting job title t' while being currently titled t can be modeled as:

- Weighted Inner Product: $G_1(t) = [\langle z_t,\, z_{t_1}\rangle_w, \cdots \langle z_t,\, z_{t_m}\rangle_w]$;
- Generalized Linear Model: $G_2(t) = \mathtt{GLM}_w(z_t)$.

The parameter w adds flexibility. For instance, here, we make the assumption that the employee is fully representable by his/her title. If the employee has additional features, we need a parametric transformation to map it to the title space.

Translate Theory I (Non-Markovian Setting)

Theorem II – Past Title Dependent Typicality

Assume $\mathcal{F} = \{\textit{past job titles}\}$, then the next job only depends on the title history, i.e.,

$$\mathbb{P}(\textit{next job title} = t' \mid \mathcal{F}) = G(\textit{past job titles}) \tag{3}$$

Proposals for $G(\cdot)$

Denote the period i title as t_i and the representation (of employee) at period i by $h(i)$, we propose a path-dependent dynamic embedder,

$$h(i) = f(t_i) + h(i-1) \tag{4}$$

where $f(\cdot)$ encodes title t_i in lower-dimensional space. By self-substitution (and proper boundary condition), Eq (4) is equivalent to $h(i) = \sum_{j=1}^{i} f(t_j)$, the sum of all past title embeddings. Applying $G_1(\cdot)$ or $G_2(\cdot)$ on $h(i)$, we obtain the likelihood of getting t' as,

$$G_3(t) = G_x\big(h(i)\big), \; x = 1 \textit{ or } 2 \tag{5}$$

where $h(\cdot)$ follows Eq (4).

Translate Theory I (Non-Markovian Setting)

Theorem III – Past Experience Dependent Typicality

Assume $\mathcal{F} = \{\textit{past title and tenure}\}$, then the next job depends on the past title-tenure pairs, i.e.,

$$\mathbb{P}(\textit{next job title} = t' \mid \mathcal{F}) = G\big((\textit{title}_1, \textit{tenure}_1), \ldots, (\textit{title}_i, \textit{tenure}_i)\big) \tag{6}$$

Proposals for $G(\cdot)$

Keeping everything the same, we extend Eq (7) to a time-aware path-dependent dynamic embedder,

$$h(i) = f(t_i, \Delta_i) + \textit{df}(i-1, i) \cdot h(i-1) \tag{7}$$

where $f(\cdot)$ accounts for the tenure of each job, and discount factor $\textit{df}(\cdot)$ captures time decay of the past experience and satisfies $\textit{df}(i-k, i) = \prod_{j=i-k}^{i-1} \textit{df}(j, j+1)$. Self-substituting renders

$$h(i) = \sum \textit{df}(j, i) f(t_i, \Delta_i) \tag{8}$$

Essentially, each past title is re-weighted depending on its tenure and recency.

<u>Remark:</u> for instance, $\textit{df}(i-k, i) = e^{-r\sum_{j=i-k}^{i} \Delta_j}$ satisfies the above criterion.

Translate Theory II

Theorem IV – Social Influence

Assume $\mathcal{F} = \{\textit{all individuals' past titles}\}$, then the next job depends on the employee social network evolution, i.e.,

$$\mathbb{P}(\textit{next job title} = t' \mid \mathcal{F}) = G\big(\textit{social network evolution}\big) \tag{9}$$

Proposal for $G(\cdot)$

We break it down into 3 steps:

- Build a network to identify potential jobs through a referenced employee and the shared job title.
- For each individual, generate a lower-dimensional representation based on the network built above.
- Apply G_1 or G_2 to calculate the likelihood of getting the next job title.

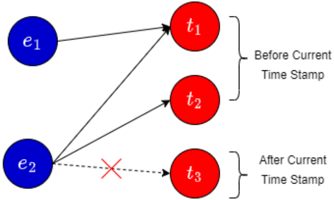
In the next, we detail the first 2 steps.

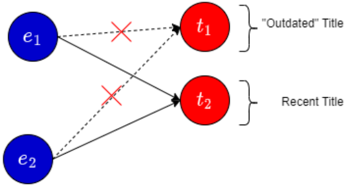
Translate Theory II – Build A Social Influence Network

Impose Constraints

When building the network at period i, we impose the following constraints:

- Non-Anticipativity: an employee cannot refer to the titles of another employee that occur in the future (left-hand-side).
- Limited-Memory: an employee shall not refer to the "outdated" titles of another employee (right-hand-side). That is to summarize social networks at a certain frequency (every k-year).





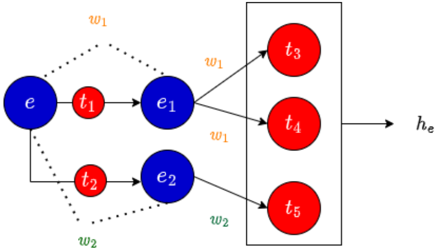
Translate Theory II – Representation Learning

Employee Weighted Node Semantics

Given a focal employee e, we denote his neighbor reference employees as $\mathcal{N}_{\text{emp}}(e)$ and neighbor potential titles (led by $\mathcal{N}(e)$) as $\mathcal{N}_{\texttt{title}}(e)$. The embedding of individual e is then

$$h = \sigma\left(\sum_{t \in \mathcal{N}_{\texttt{title}}(e)} w(e, e') \cdot z_t \right) \tag{10}$$

where e' is associated with title t, and $w(e, e')$ measures the similarity between focal employee e and reference employee e'.



Digression: Dynamic Title Embedding

Dynamic Title Embedding

To account for the evolving relationships among job titles, we can fine-tune the embedding with a title network. Fix a moderate sampling frequency, we periodically build a title graph $\mathcal{G}^{\texttt{title}}$, where the nodes are titles, and the edge weights are the number of transitions between connecting titles. Standard graph representation model can generate a modified embedding $\hat{z}_t$.

Concerns

- If we choose to use cosine similarity, the title embedding matrix changes over time, which makes learning unstable. That is, $z_e \cdot Z^T$ with a moving target Z^T.
- The edge weights of the title network can change dramatically, which makes title embedding very unstable.

Dynamic Embedding Smoothing

In objective function, using KL(Kullback–Leibler) Divergence, we penalize the modified embedding $\hat{z}_t$ deviates too far from the original embedding z_t.

Translate Integrated Theory

Theorem V – Prototype Logic & Social Influence

Assume $\mathcal{F} = \{\textit{all individuals' past titles including him/herself}\}$, then the next job depends on the employee's social network evolution, i.e.,

$$\mathbb{P}(\textit{next job title} = t' \mid \mathcal{F}) = G\big(\textit{social network evolution and self-evolution}\big) \tag{11}$$

Proposal for $G(\cdot)$

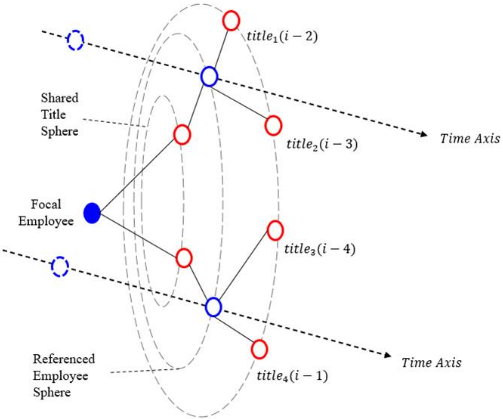
Given an individual e, at any period i,

- **(Optional*)** Generate dynamic title embedding z_t via title network;
- Generate time-aware employee embedding z_e via Eq (8) using dynamic title embedding z_t;
- Fuse social influence in generating final employee embedding $\hat{z}_e$ via Eq (10).
- Use GLM (G_2) to produce next title distribution;

Translate Integrated Theory – Measurable Tree

Measurable Tree

We propose a new structure measurable tree, which characterizes how a SINGLE agent interacts with a complicated environment, dynamic and has feedbacks). The proposed $G(\cdot)$ relies on this tree.



Engineering Considerations – Make Math AI Look

Add Buzz Word

- $G_1(\cdot)$ is called projection + cosine similarity;
- $G_2(\cdot)$ is called SoftMax;
- Eq (4) is the transition equation for RNN;
- Social influence based representation is called Meta-Path + HAN;

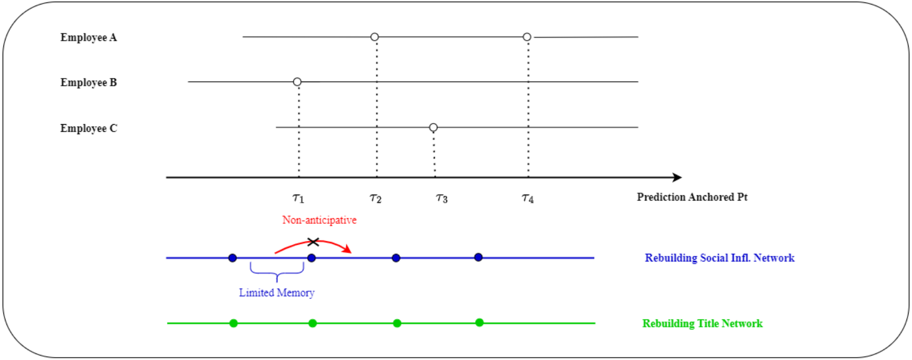
Our Additions So Far

- Eq (7) is inspired by Optimal Control and Financial Mathematics;
- Graphical Constraints is inspired by Branch and Bounding technique from Optimization;
- Employee Weighted Node Semantics generalized the original HAN node semantics aggregation;
- Using KL for stabilization is inspired by Reinforcement Learning for Language Model;

Engineering Considerations – Model Scheduler

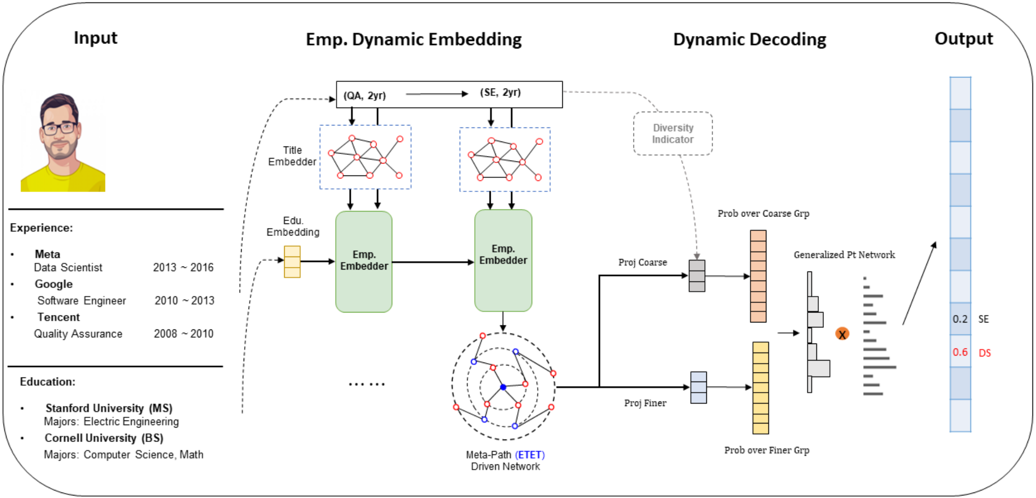
Tractable Scheduling

The prediction anchored points differ from employee to employee, while social infl. network and title network are rebuilt at a fixed frequency.



Model Scheduler

Engineering Considerations – Workflow



Engineering Considerations – Additional Tricks

Attention via Generalized Pointer Network

To derive the distribution over all titles, we leverage the coarse title grouping. Namely,

- Project employee embedding to a coarse embedding z_c and finer embedding z_f.
- Generate probability distribution over the finer titles using $G_2(\cdot)$, i.e., $\texttt{dist}(\cdot; z_f)$.
- Generate probability distribution over coarse title groups using $G_1(\cdot)$, based on z_c and personalized information, such as diversity score. That is, $\texttt{dist}(\cdot; z_c, \texttt{div})$;
- Compile $\texttt{dist}(\cdot; z_f)$ and $\texttt{dist}(\cdot; z_c, \texttt{div})$ into the ultimate distribution over title universe.

Remarks:

- Why use G_1 for coarse? Unlike evolving titles, we fix a static embedding for each title group. Therefore, we can use them as benchmarks to assess the likelihood of falling into each group.
- What's the intuition behind plugging diversity score into $\texttt{dist}(\cdot; Z_c, \texttt{div})$? Diversity controls the shape of the distribution. When diversity is high, the distribution becomes more even, when diversity is low, the distribution becomes more uni-modular.

Conclusion

Summary

In this paper, we build up a theory-driven AI framework to decipher job mobility. In particular,

- we capture prototype logic via a tenure-recency adjusted dynamic model.
- we capture social influence logic via a meta-path driven network with novel node semantic aggregation scheme.
- we capture evolving relationships among titles via a title graph and propose dynamic embedding smoothing to stabilize model learning.
- we integrate both theories via a novel Seq2Graph architecture.
- we capture personalized info, e.g., diversity, via entropy to control the likelihood in each title group (coarse title);
- we apply attention via a generalized pointer network to incorporate title group distribution in generating ultimate distribution over title universe.

The End