JAIST Repository
https://dspace.jaist.ac.jp/

Title	GCN と Few-Shot Learning を用いた少数サンプル職業推薦モデルのシミュレーションによる性能評価
Author(s)	趙, 泰銘
Citation	
Issue Date	2026-03
Type	Thesis or Dissertation
Text version	author
URL	https://hdl.handle.net/10119/20444
Rights	
Description	Supervisor:長谷川　忍, 先端科学技術研究科, 修士(情報科学)

Japan Advanced Institute of Science and Technology

Performance Evaluation of a Few-Sample Job Recommendation Model Using Graph Convolutional Networks and Few-Shot Learning in Simulation

2410114 Zhao Taiming

The contemporary labor market is undergoing a profound structural transformation driven by rapid digitalization, increasing technological specialization, and the emergence of highly differentiated industries. Along with these changes, large-scale vocational data have become widely available through online recruitment platforms, professional networking services, and corporate human resource systems. Although the volume of such data continues to grow, its distribution is far from uniform. In practice, occupational data commonly exhibit a power-law or long-tail distribution in which a small number of occupations account for the majority of available samples, while a large number of occupations are represented by only a limited amount of data.

In most real-world datasets, so-called "Head" occupations—such as general administrative roles, standard software engineering positions, or retail sales—dominate interaction logs and training samples. In contrast, a substantial number of occupations reside in the "Tail" of the distribution. These include highly specialized technical roles, region-specific professions, and emerging occupations associated with new societal demands, such as those related to the green economy, data governance, or AI ethics. For these occupations, the number of available samples often falls far below the threshold required for stable parameter estimation in conventional machine learning models, making reliable recommendation particularly challenging.

A central issue underlying this challenge is the systemic popularity bias inherent in many existing recommendation algorithms. Classical Collaborative Filtering methods and their neural extensions rely heavily on dense user–item interaction matrices. When an occupation has only a small number of interactions, the corresponding latent representation cannot be sufficiently optimized due to the lack of informative gradients. As a result, embedding quality deteriorates significantly for low-frequency occupations. During inference, models tend to favor popular occupations that have been well optimized during training, as these contribute more strongly to global performance metrics such as overall hit rate or accuracy.

This mechanism creates a self-reinforcing feedback loop in which popular occupations are recommended more frequently, generating even more interaction data, while rare but potentially relevant occupations remain underexposed. In the context of job recommendation, this phenomenon is not merely a technical limitation. It may restrict the career exploration space of job seekers, hinder labor mobility toward specialized or emerging sectors, and exacerbate mismatches between individual skill sets and occupational requirements. Consequently, mitigating popularity bias in long-tail occupations is a critical requirement for building fair and effective job recommendation systems.

To address the limitations of purely data-driven approaches in sparse and imbalanced environments, this research proposes a hybrid framework that shifts the learning paradigm from interaction-frequency dependence to knowledge-infused representation learning. Specifically, the proposed approach integrates Graph Convolutional Networks (GCN) for structural feature extraction with Prototypical Few-Shot Learning (FSL) for robust inference under limited supervision. By combining these two components, the framework aims to leverage both relational knowledge and metric-based generalization.

The foundation of the proposed method is the construction of a Job–Skill Knowledge Graph that explicitly models semantic relationships between occupations through their required skill sets. In this graph, nodes represent occupations and individual skills, while weighted edges encode the strength of association between an occupation and a skill. These weights are derived from large-scale job posting data using a weighted similarity measure that reflects both skill overlap and importance. By embedding domain knowledge into the graph structure, the model introduces a meaningful prior that complements sparse interaction data.

A GCN is then applied to this knowledge graph to perform iterative message passing. Through neighborhood aggregation, each occupation updates its representation by integrating information from semantically related occupations and shared skills. This process allows occupations with limited or even zero interaction data to acquire meaningful representations by leveraging their structural position in the graph. For example, a rare occupation such as a specialized cloud security role in the healthcare domain can inherit foundational semantic information from broader but related occupations, including cybersecurity or cloud infrastructure engineering. As a result, even tail occupations obtain representations grounded in skill semantics rather than historical popularity.

While GCN-based representation learning alleviates sparsity at the feature level, it does not fully address the challenge of decision making under extremely limited supervision. To tackle this issue, the proposed framework incorporates Prototypical Few-Shot Learning as a second stage. Few-shot learning methods are designed to generalize from a small number of labeled examples by learning a metric space in which similar instances are placed close to one another. In this study, an episodic training strategy is adopted to simulate the target low-data scenario during training.

In each training episode, a small support set is constructed by sampling only a few examples for each occupation, along with a corresponding query set. User profiles and occupation representations are projected into a shared metric space through a learnable projection function. Within this space, a prototype vector is computed for each occupation as the mean representation of its support samples. Recommendation is then formulated as a distance-based ranking problem, where users are matched to occupations based on their proximity to these prototypes. This formulation allows the model to generalize to new or rare occupations, as decision boundaries are determined by relative similarity rather than absolute frequency.

The effectiveness of the proposed GCN–FSL framework was evaluated using a large-scale real-world dataset consisting of over 150,000 vocational profiles, approximately 2,000 occupation categories, and a skill taxonomy containing more than 12,000 unique skills. To analyze performance across different data regimes, occupations were grouped into Head, Mid-frequency, and Tail segments according to their occurrence frequency. This experimental design enables a systematic investigation of model behavior in both data-abundant and data-scarce conditions.

Evaluation was conducted using a combination of accuracy-oriented and diversity-oriented metrics. Standard metrics such as Precision@N, Recall@N, and F1-score were used to assess recommendation correctness. In addition, to directly measure the mitigation of popularity bias, a TailShare@N metric was introduced, defined as the proportion of long-tail occupations appearing in the top-N recommendation lists across all users. Furthermore, Intra-List Diversity was employed to verify that the model provides a broad range of occupational options rather than shifting recommendations toward a narrow subset of niche roles.

Experimental results indicate that the proposed framework maintains competitive performance on Head occupations while demonstrating substantial improvements in the Tail segment. Compared with baseline neural collaborative filtering approaches, the proposed model exhibits significantly more stable performance when interaction data are scarce. The TailShare results further show that rare occupations appear more frequently in top-ranked recommendation lists, suggesting that the model effectively resists popularity bias. Visualization of the learned embedding space reveals that specialized occupations form tighter and more coherent clusters under the proposed approach, whereas traditional methods tend to collapse these occupations into diffuse regions dominated by popular roles.

In conclusion, this research demonstrates that the long-tail problem in job recommendation systems is not an inherent limitation but a challenge that can be addressed through the integration of structural knowledge and meta-learning techniques. By leveraging a Job–Skill Knowledge Graph and GCN-based representation learning, the proposed framework provides meaningful semantic embeddings even for occupations with little or no interaction data. The incorporation of Prototypical Few-Shot Learning further enables accurate and flexible matching in extreme cold-start scenarios, thereby mitigating popularity bias and promoting more diverse recommendations.

Future work will focus on incorporating temporal dynamics into the knowledge graph to model the evolving relevance of skills over time, allowing the system to adapt to technological change. In addition, large-scale human-in-the-loop evaluations will be conducted in collaboration with professional career counselors to qualitatively assess the usefulness and practical value of long-tail recommendations. Through these extensions, the proposed approach aims to contribute to the development of more equitable, adaptive, and human-centered career recommendation systems.