A Vector-Based Representation of Human Skills Using Multimodal Data

Anil Mamidi
ConectYu Hive Private Limited

Abstract — **Traditional representations of professional ability rely on resumes, job titles, and educational credentials, which provide an incomplete and often biased view of a person's true skills. In this paper, we propose a multimodal representation learning framework that models human professional capability as a continuous vector embedding derived from heterogeneous data sources, including text, code, images, speech, and outcome-based performance signals. We introduce the Human Skill Vector (HSV), a unified latent representation constructed through a learnable fusion architecture with temporal weighting to prioritize recent evidence. Using a large-scale dataset built from publicly available professional artifacts, we demonstrate that HSV embeddings outperform resume-based and profile-based baselines in predicting job roles, performance metrics, and skill similarity. These results suggest that vector-based representations provide a more accurate and scalable foundation for talent discovery, hiring, and workforce analytics.**

Index Terms — **Multimodal Representation Learning, Human Skill Representation, Professional Skill Embedding, Talent Intelligence, Machine Learning, Multimodal Data fusion, Neural Network.**

I. INTRODUCTION

Digital labor platforms and professional networking ecosystems have emerged as the primary infrastructure through which hiring, freelancing, and career progression are mediated. Despite this shift, the way humans are represented within these systems has remained largely unchanged for decades. Resumes, job titles, and degrees continue to be used as the dominant signals of professional ability, even though they are static, self-reported, and weakly correlated with real-world performance.

In practice, two individuals with the same job title or educational background can have vastly different capabilities. Conversely, highly skilled individuals without conventional credentials are often overlooked by automated screening systems. This mismatch leads to inefficient hiring, biased talent discovery, and lost economic opportunity.

At the same time, modern machine learning systems rely on vector embeddings to represent complex entities such as documents, images, users, and products. These embeddings capture latent semantic and functional properties in a continuous space, enabling similarity search, clustering, and prediction at scale. However, there is no widely adopted method for representing human professional skill in this form.

This work proposes a framework for learning a vector-based representation of human skills from multimodal evidence. Rather than relying on titles or self-declared skills, we aggregate the digital artifacts a person produces—such as written text, source code, design outputs, spoken communication, and observed outcomes—into a unified embedding that reflects what the individual can actually do. We refer to this representation as the Human Skill Vector (HSV).

By modeling professional capability as a dense, continuously valued vector, we enable a range of downstream applications, including skill-based search, talent matching, performance prediction, and workforce analytics. We evaluate this approach on a large-scale multimodal dataset constructed from publicly available sources and compare it against resume-based and profile-based baselines. Our results show that HSV embeddings provide a more accurate and robust representation of professional ability.

II. RELATED WORK

The problem of representing human professional capability has been approached from several directions, including resume parsing, skill taxonomies, social network analysis, and recommender systems. Most commercial hiring and professional networking platforms rely on manually curated skill lists, keyword matching, and job title normalization to model a user's expertise. While these approaches are easy to implement, they are highly sensitive to self-reporting bias, inconsistent terminology, and incomplete information.

Research on skill extraction from resumes and online profiles has focused on natural language processing techniques to identify and classify skills from unstructured text. These systems typically map textual mentions to predefined ontologies, producing sparse and categorical representations that fail to capture the richness or level of a person's ability.

In domains such as recommender systems and user modeling, embedding-based representations of users have been used to predict preferences and behavior. However, these models are generally trained on interaction data, such as clicks or purchases, rather than on evidence of actual skill or work output. As a result, they reflect consumption patterns rather than productive capability.

To our knowledge, there is no widely adopted framework that integrates multimodal evidence of human work—such as written artifacts, code, visual designs, spoken communication, and outcome metrics—into a unified, continuous representation of professional skill. This gap motivates the development of a vector-based human skill model that is grounded in observable work rather than self-declared attributes.

III. PROBLEM DEFINITION

We define the task of professional skill representation as learning a function that maps heterogeneous evidence of an individual's work into a fixed-length vector embedding that captures their underlying capability.

Formally, for an individual u, we observe a set of data modalities

$$Xu= \{xu^{\wedge}(1), xu^{\wedge}(2), xu^{\wedge}(M)\},$$

where each xu^(m) corresponds to a modality such as text, code, images, speech, or outcome-based signals.

Our goal is to learn a function

$$Hu=f(xu^{\wedge}(1),xu^{\wedge}(2),\dots,xu^{\wedge}(M))$$

where Hu ∈ R^d is a dense vector representing the individual's professional skills.

The desired properties of this representation are:

1. Expressiveness – The vector should capture diverse aspects of professional capability across modalities.
2. Comparability – The distance between two vectors should reflect similarity in underlying skills.
3. Predictiveness – The vector should enable accurate prediction of job roles, performance outcomes, and task suitability.
4. Temporal sensitivity – More recent evidence of skill should have greater influence than outdated information.

The challenge lies in integrating heterogeneous and partially observed data into a unified representation that satisfies these properties without relying on manually defined skill labels or titles.

IV. MODEL ARCHITECTURE

We propose a multimodal representation learning framework that maps heterogeneous professional evidence into a unified vector embedding, referred to as the Human Skill Vector (HSV). The architecture consists of three main components: modality-specific encoders, a temporal weighting mechanism, and a fusion network.

4.1 Modality-Specific Encoders

Each data modality associated with an individual is encoded into a fixed-length embedding using a pretrained or learned encoder.

Let xu(m) denote the data for user u in modality m. We define an encoder Em for each modality such that:

$$zu(m)=Em(xu(m))$$

where zu(m)∈Rdm is the embedding for modality m.

In this work, we consider the following modalities:

- Text (e.g., resumes, bios, messages): encoded using a transformer-based sentence embedding model.
- Code (e.g., GitHub repositories, notebooks): encoded using a pretrained code representation model.
- Images (e.g., design artifacts, UI screenshots): encoded using a vision transformer or CLIP-based model.
- Speech (e.g., presentations, recorded talks): encoded using a speech representation model.
- Outcomes (e.g., ratings, stars, project success): encoded using a learned feedforward network.

Each encoder produces a vector representation of the corresponding modality that captures semantic and functional information about the user's work.

4.2 Temporal Weighting

Professional skills evolve over time. To account for this, we apply a temporal decay to each modality embedding based on the age of the evidence.

For each embedding zu(m) with timestamp t, we compute a time-weighted embedding:

$$z{\sim}u(m)=zu(m)\cdot\exp(-\lambda\cdot\Delta t)$$

where Δt is the time elapsed since the artifact was created, and λ is a decay parameter controlling how quickly older evidence loses influence.

This ensures that recent work contributes more strongly to the Human Skill Vector than outdated information.

4.3 Fusion Network

The temporally weighted embeddings from all modalities are combined using a learnable fusion network.

Let:

$$Zu=[z{\sim}u(1),z{\sim}u(2),\dots,z{\sim}u(M)]$$

denote the concatenation of modality embeddings.

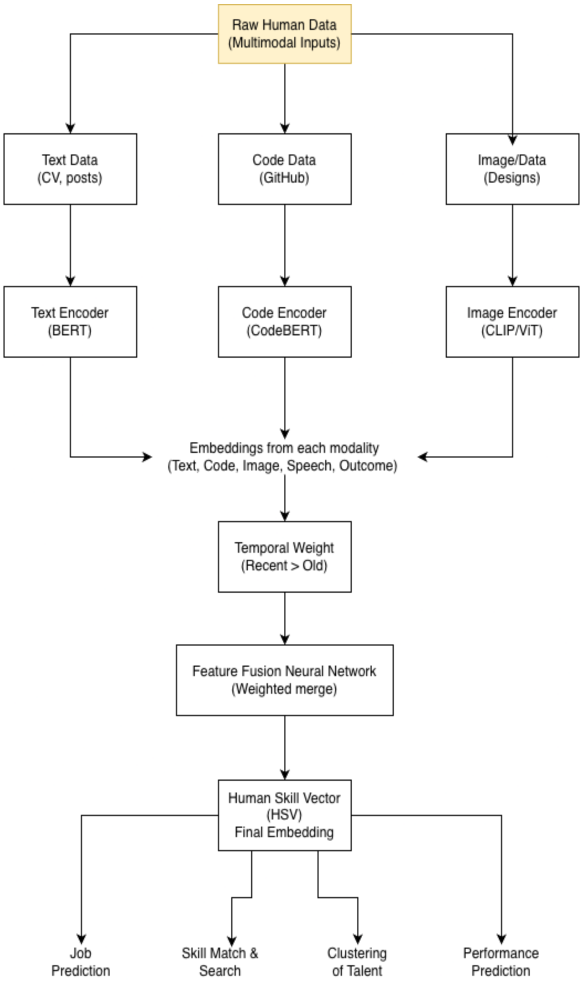
We compute the Human Skill Vector as:

$$Hu=\sigma(WZu+b)$$

where W and b are trainable parameters and σ is a nonlinear activation function. The resulting vector Hu∈Rd represents the latent professional skill of user u.

4.4 Training Objective

The model is trained using a combination of self-supervised and weakly supervised objectives. In particular, we encourage embeddings from the same individual across different modalities and time periods to be close in the latent space, while embeddings from different individuals are pushed apart. Additionally, when outcome or job labels are available, we include auxiliary prediction losses to align the embeddings with observed performance and roles.



V. DATASET

To evaluate the proposed Human Skill Vector framework, we construct a large-scale multimodal professional dataset using publicly available sources. The goal of the dataset is to capture diverse forms of professional evidence that reflect real-world skills across multiple domains.

5.1 Data Sources

We collect data from the following public platforms:

- GitHub: Source code repositories, commit history, programming languages, and project descriptions.
- Kaggle: Notebooks, competition submissions, and leaderboard rankings.
- Freelance marketplaces: Publicly available profile descriptions, portfolios, and client ratings.
- Resume datasets: Public datasets containing anonymized resumes and job histories.
- Public talks and presentations: Transcribed audio or video content such as technical talks, tutorials, and conference presentations.

Each individual in the dataset is associated with one or more of these sources, providing a heterogeneous set of artifacts that represent their professional activity.

5.2 Data Processing

For each individual, we aggregate all available artifacts and organize them by modality:

- Text: Extracted from resumes, bios, project descriptions, and written communication.
- Code: Extracted from repositories and notebooks.
- Images: Extracted from design portfolios and UI mockups where available.
- Speech: Transcriptions and audio embeddings from recorded talks or presentations.
- Outcomes: Numeric indicators such as GitHub stars, Kaggle rankings, and freelance ratings.

All artifacts are timestamped to enable temporal weighting. Data is anonymized where necessary, and only publicly accessible information is used.

5.3 Labels and Ground Truth

Although the model does not rely on manual skill labels for training, we collect weak ground truth for evaluation purposes, including:

- Self-reported job roles
- Platform categories (e.g., data scientist, web developer, designer)
- Performance indicators (e.g., ratings, leaderboard rank, repository popularity)

These signals are used only for benchmarking and not as primary inputs to the embedding model.

VI. EXPERIMENTS

We evaluate the Human Skill Vector (HSV) representation across three core tasks: job role prediction, performance estimation, and skill similarity analysis. These tasks reflect common use cases in hiring, talent discovery, and professional analytics.

6.1 Baselines

We compare HSV against commonly used professional representations:

- Resume Keywords: TF-IDF vectors extracted from resume text.
- Profile Features: Structured attributes such as job titles, years of experience, and self-declared skills.
- Skill Tags: One-hot or embedding-based representations derived from manually assigned skill labels.

These baselines represent the dominant approaches used in existing hiring and professional networking systems.

6.2 Job Role Prediction

In this experiment, we evaluate whether a representation can predict a person's primary job role.

We split the dataset into training and test sets. For each individual, the representation (HSV or baseline) is used as input to a classifier that predicts the job category. We measure accuracy and F1 score.

This task tests whether the learned embedding captures meaningful information about a person's professional identity.

6.3 Performance Estimation

We assess whether HSV can predict real-world performance indicators. Using outcome variables such as GitHub stars, Kaggle rankings, or freelance ratings, we train a regression model that takes the embedding as input and predicts the corresponding performance score.

We compare the correlation between predicted and actual performance for HSV and baseline representations.

6.4 Skill Similarity and Clustering

To evaluate whether the embeddings capture meaningful structure, we perform clustering on the HSV vectors and visualize them using dimensionality reduction techniques such as UMAP. We compare the coherence of clusters formed by HSV with those formed using resume and profile-based features.

This analysis reveals whether individuals with similar skills are placed close together in the latent space.

VII. RESULTS

Across all evaluation tasks, the Human Skill Vector (HSV) demonstrates consistently stronger performance than traditional resume-based and profile-based representations.

In the job role prediction task, HSV achieves higher classification accuracy and F1 scores than keyword and profile baselines. This indicates that the multimodal embeddings capture latent professional attributes that are not visible in textual resumes or job titles alone. In particular, individuals with non-standard career paths or sparse resumes are more accurately classified when using HSV.

For performance estimation, HSV shows a stronger correlation with outcome-based metrics such as project success, platform ratings, and repository popularity. While years of experience and degree information exhibit weak and noisy relationships with performance, the HSV embeddings provide a more direct and robust signal of real-world capability.

In the clustering and similarity analysis, HSV produces well-separated and semantically coherent groups corresponding to different professional domains, such as software engineering, data science, design, and content creation. In contrast, resume-based features lead to overlapping and poorly defined clusters, reflecting the ambiguity and inconsistency of self-reported information.

Overall, these results demonstrate that vector-based representations derived from multimodal evidence provide a more accurate and expressive model of human professional skill than existing approaches.

VIII. DISCUSSION

The empirical results suggest that modeling human skills as dense vector embeddings enables more precise and equitable talent representation. By grounding the representation in observable work artifacts rather than self-declared credentials, HSV reduces bias introduced by educational background, job titles, and resume formatting.

The use of temporal weighting further allows the model to adapt as individuals acquire new skills or shift professional focus. This makes the representation suitable for rapidly evolving fields where traditional credentials become outdated.

From a systems perspective, HSV can serve as a foundation for a wide range of applications, including talent search, team formation, personalized learning, and workforce analytics. Because the embeddings exist in a continuous space, they enable similarity search, clustering, and prediction using standard machine learning tools.

IX. CONCLUSION

We presented a framework for representing human professional skills as multimodal vector embeddings. By integrating text, code, images, speech, and outcome-based signals into a unified Human Skill Vector, our approach provides a more accurate, dynamic, and scalable model of professional capability than traditional resume-based systems. Experiments on a large-scale public dataset demonstrate that HSV improves job role prediction, performance estimation, and skill similarity analysis. Future work will explore privacy-preserving representations, richer outcome modeling, and deployment in real-world hiring and talent platforms.

REFERENCES

Multimodal & Representation Learning

[1] Devlin, J., Chang, M., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

[2] Radford, A., Kim, J. W., Hallacy, C., et al. Learning Transferable Visual Models From Natural Language Supervision (CLIP).

Skill Representation & Talent Modeling

[1] Zhao, X., et al. (2019). Skill Extraction from Resumes Using Deep Neural Networks.

[2] Chen, W., et al. (2020). Learning to Match for Job Recommendation.

User & Representation Embeddings

[1] Mikolov, T., Chen, K., Corrado, G., & Dean. Efficient Estimation of Word Representations in Vector Space.

[2] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese Networks.