Enhancing Job Salary Prediction with Disentangled Composition Effect Modeling: A Neural Prototyping Approach

Yang JI, Ying SUN, Hengshu ZHU

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50421-0

Problems & Ideas

- Problems Statements:
  - Existing salary prediction methods lack explainability in modeling skill composition effects
  - Traditional approaches fail to capture complex interactions between skill sets
- Ideas: Disentangled discrete skill subset selection, coupled with set-oriented prototypical learning

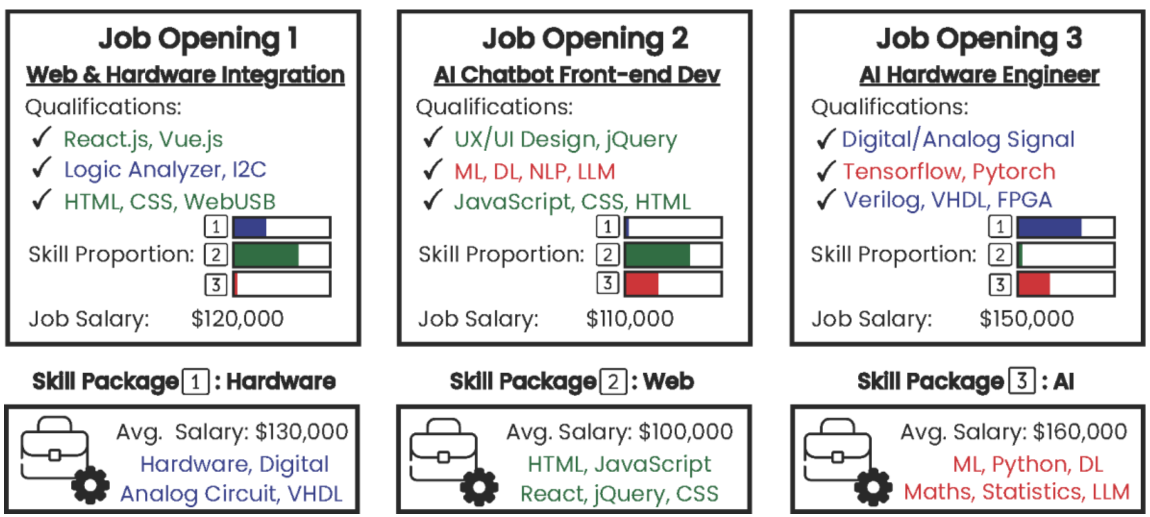

Image: Figure 1 - Illustrative example on skill-salary influence **Caption:** Traditional black-box models vs. our explainable approach showing how different skill packages (Hardware, Web, AI) command different average salaries with transparent reasoning.

Main Contributions

- Contributions:
  - **Superior Performance:** 10% improvement over state-of-the-art baselines across four real-world datasets
  - **Comprehensive Evaluation:** Validated on 400,000+ job postings across IT, Design, High-tech, and Financial sectors
  - **Explainability Achievement:** Successfully identifies semantically coherent skill subsets with quantifiable salary contributions

Methods	Skill-Salary Job Postings [4]				Job Salary Benchmarking [8]			
	IT		Designer		High-tech		Finance	
	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)
HSBMF	$5.291_{\pm 0.017}$	$3.939_{\pm 0.015}$	$4.612_{\pm 0.025}$	$3.371_{\pm 0.021}$	$3.844_{\pm 0.016}$	$2.872_{\pm 0.013}$	$4.299_{\pm 0.025}$	$3.150_{\pm 0.019}$
SSCN	$4.762_{\pm 0.063}$	$3.484_{\pm 0.052}$	$3.841_{\pm 0.143}$	$2.766_{\pm 0.102}$	$3.663_{\pm 0.162}$	$2.740_{\pm 0.106}$	$4.037_{\pm 0.134}$	$2.948_{\pm 0.096}$
NDP-JSB	$5.342_{\pm 0.021}$	$4.073_{\pm 0.021}$	$4.623_{\pm 0.044}$	$3.262_{\pm 0.023}$	$3.875_{\pm 0.086}$	$2.827_{\pm 0.067}$	$4.295_{\pm 0.075}$	$3.146_{\pm 0.071}$
BERT-JSB	$4.913_{\pm 0.032}$	$3.583_{\pm 0.014}$	$3.913_{\pm 0.046}$	$2.897_{\pm 0.018}$	$4.071_{\pm 0.042}$	$2.993_{\pm 0.034}$	$4.504_{\pm 0.045}$	$3.272_{\pm 0.036}$
Set-Tree	$5.552_{\pm 0.065}$	$4.367_{\pm 0.083}$	$4.339_{\pm 0.149}$	$3.269_{\pm 0.061}$	$3.953_{\pm 0.092}$	$2.998_{\pm 0.065}$	$4.406_{\pm 0.079}$	$3.206_{\pm 0.057}$
SESM	$4.489_{\pm 0.034}$	$3.532_{\pm 0.025}$	$3.812_{\pm 0.123}$	$2.767_{\pm 0.064}$	$3.623_{\pm 0.029}$	$2.817_{\pm 0.017}$	$3.967_{\pm 0.031}$	$2.940_{\pm 0.011}$
ProtoPNet	$4.503_{\pm 0.029}$	$3.423_{\pm 0.037}$	$3.794_{\pm 0.013}$	$2.772_{\pm 0.076}$	$3.602_{\pm 0.110}$	$2.786_{\pm 0.025}$	$4.116_{\pm 0.144}$	$3.013_{\pm 0.063}$
TesNet	$4.555_{\pm 0.102}$	$3.698_{\pm 0.037}$	$3.804_{\pm 0.063}$	$2.902_{\pm 0.012}$	$3.692_{\pm 0.096}$	$2.806_{\pm 0.081}$	$4.241_{\pm 0.108}$	$3.094_{\pm 0.071}$
ProtoConcepts	$4.468_{\pm 0.062}$	$3.412_{\pm 0.025}$	$3.792_{\pm 0.083}$	$2.774_{\pm 0.052}$	$3.504_{\pm 0.061}$	$2.776_{\pm 0.026}$	$4.023_{\pm 0.034}$	$2.961_{\pm 0.048}$
LGDESetNet	$\mathbf{4.162}_{\pm 0.012}$	$\mathbf{3.141}_{\pm 0.041}$	$\mathbf{3.473}_{\pm 0.058}$	$\mathbf{2.559}_{\pm 0.062}$	$\mathbf{3.327}_{\pm 0.038}$	$\mathbf{2.434}_{\pm 0.037}$	$\mathbf{3.775}_{\pm 0.046}$	$\mathbf{2.768}_{\pm 0.031}$

Image: Overall performance evaluation results **Caption:** Performance comparison showing LGDESetNet consistently outperforms existing methods across all datasets with significant improvements in both RMSE and MAE metrics.