Entity Recognition and Talent Profiles in Digital Industry Based on BERT and BiGRU

Ying Jin

12.04.2025

Introduction

Research Background

- Digital industry is a strategic emerging field, driving economic and social development.
- Talent profiles & knowledge graphs are critical for talent cultivation, curriculum optimization, and industry upgrading.

Limitations of Existing Studies

- Qualitative methods (expert evaluation, grounded theory): Subjective, lack data support.
- Quantitative methods: Descriptive stats (limited to structured data); LDA model (fails in deep semantic mining).
- NLP-based models (BERT-BiLSTM-CRF): Insufficient in highlighting key semantics; lack interpretable characterization.

Introduction

Research Goals & Core Contributions

- Goal: Propose BERT-BiGRU-Attention-CRF model to extract talent requirements and construct knowledge graphs/profiles.
- Contributions:

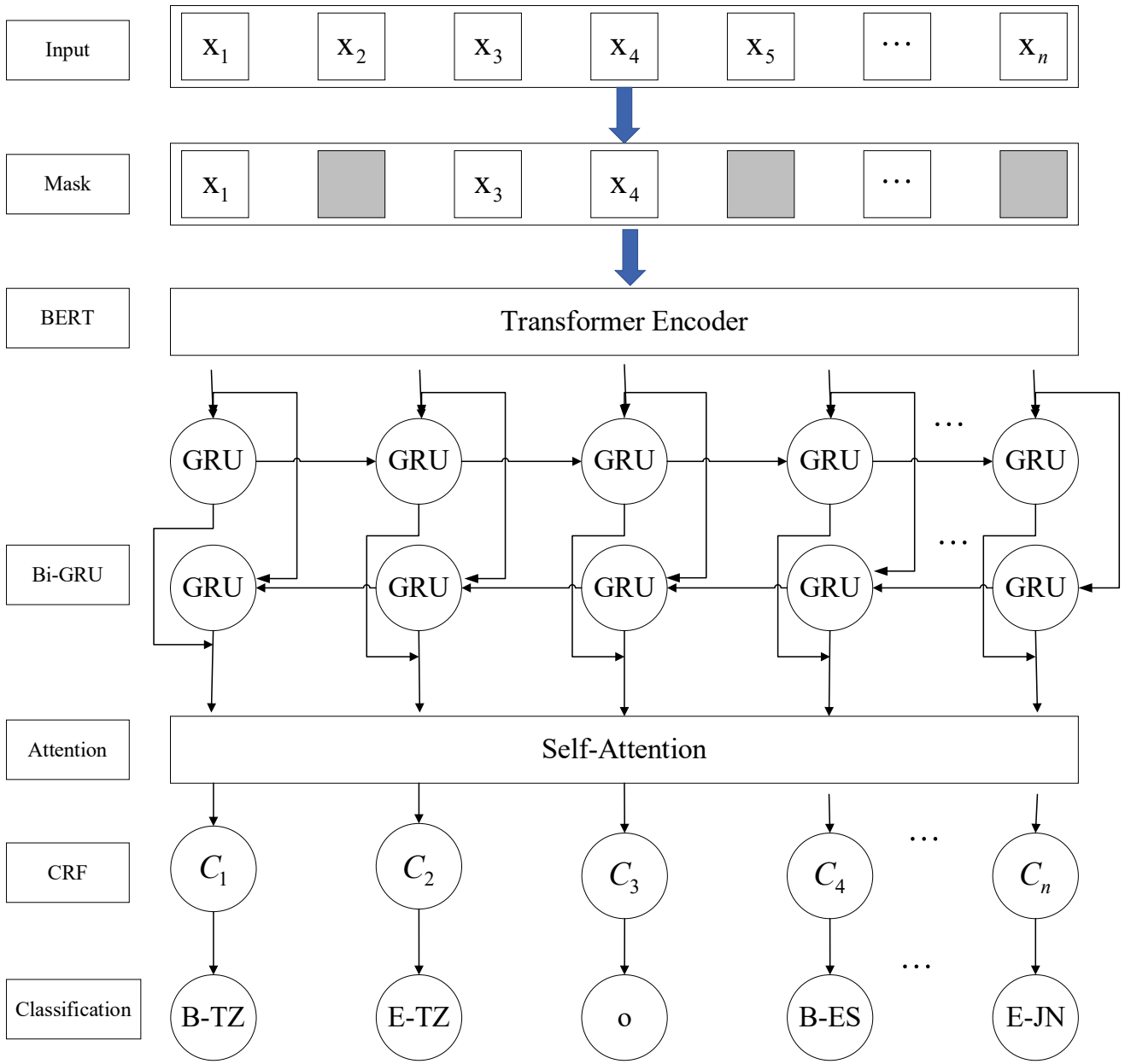
① Optimize model structure (BiGRU + Attention) for higher recognition accuracy.

② Build multi-dimensional talent profiles (knowledge, skills, traits, attitudes) and analyze position differences.

Model

BERT-BiGRU-Attention-CRF

- Text input layer
- Word embedding layer
- Sentence encoding layer
- Attention layer
- Text classification layer
- Knowledge graph layer



Model

BERT (Word Embedding)

- Based on Bidirectional Transformer, solves unidirectional text processing limitation
- Masked language model alleviates semantic bias
- Output: Comprehensive word embedding with contextual information

BiGRU (Sentence Encoding)

- Composed of forward & backward GRUs
- Reset/update gates control information flow, high computational efficiency
- Extracts semantic & grammatical features of words

Model

Multi-head Attention

- Parallel attention calculations, comprehensive context representation

CRF

- Captures tag dependencies, optimizes entity tagging order
- Loss function: Cross-entropy (predictions approach real tags)

Knowledge Graph Layer

- Store triples, build digital industry talent knowledge graph
- Dimensional profiles: Knowledge (K), Skills (S), Traits (T), Values (V)

Data Collection & Processing

Data Collection

- Sources: 3 mainstream Chinese recruitment platforms (Boss Zhipin, 51job, 58.com)
- Scope: 31 positions covering 10 digital industry directions (e.g., data analysis, product development)

Data Processing

- Deduplication → Extract talent demand info → Standardization (unify phrases/number formats) → Sentence segmentation → Truncation/padding (fixed sentence length)

Experiment Settings & Results

Evaluation Metrics

- Precision = TP/(TP+FP)
- Recall = TP/(TP+FN)
- F1=2×(Precision×Recall)/(Precision+Recall)

Model Comparison

- Baselines: BERT-CRF, BERT-IDCNN-CRF, BERT-BiLSTM-CRF
- Proposed model (BERT-BiGRU-Attention-CRF) outperforms all baselines in P/R/F1

Table 2 Comparison of model performance

Model	Precision	Recall	F1
BERT-CRF	0.7517	**0.8932**	0.8163
BERT-IDCNN-CRF	0.7683	0.8800	0.8203
BERT-BiLSTM-CRF	0.7848	0.8748	0.8273
BERT-BiGRU-Attention-CRF	**0.8017**	0.8825	**0.8401**

Experiment Settings & Results

Ablation Results

- BiGRU & Attention layers significantly improve performance

Table 3 Ablation study

Model	Precision	Recall	F1
BERT-BiGRU-CRF	0.7913	**0.8844**	0.8353
BERT-Attention-CRF	0.7872	0.8703	0.8267
BERT-BiGRU-Attention-CRF	**0.8017**	0.8825	**0.8401**

Digital Industry Talent Knowledge Graph

Talent Knowledge Graph Construction

- Extracted knowledge, skills, traits, attitudes from recruitment texts using BERT–BiGRU–Attention–CRF
- Formed triples such as (Big Data Consultant, Skill, Java)
- Built knowledge graph including:

  Knowledge: degree level, major, statistical/math foundations

  Skills: Java, Python, Linux, data governance, visualization

  Traits: communication, learning ability, logical thinking

  Attitudes/Values: responsibility, teamwork, service awareness

Relationships Between Jobs

Compute cosine similarity to identify job clusters

- Example cluster:

Data Algorithm Engineer

Data Mining Engineer

Data Analysis Engineer

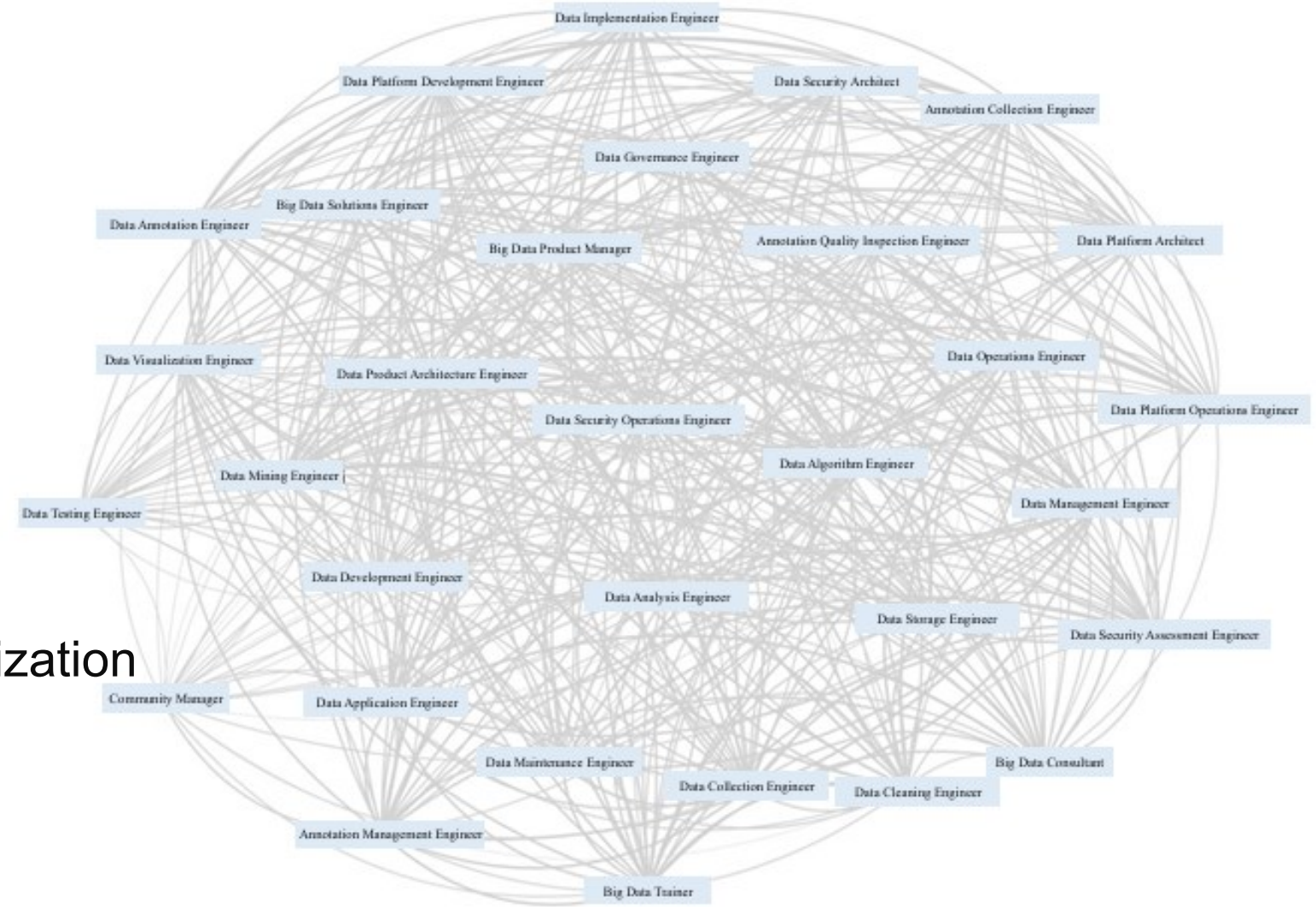
Data Visualization Engineer

- These positions share:

Python/R programming

Data cleaning, modeling, and visualization

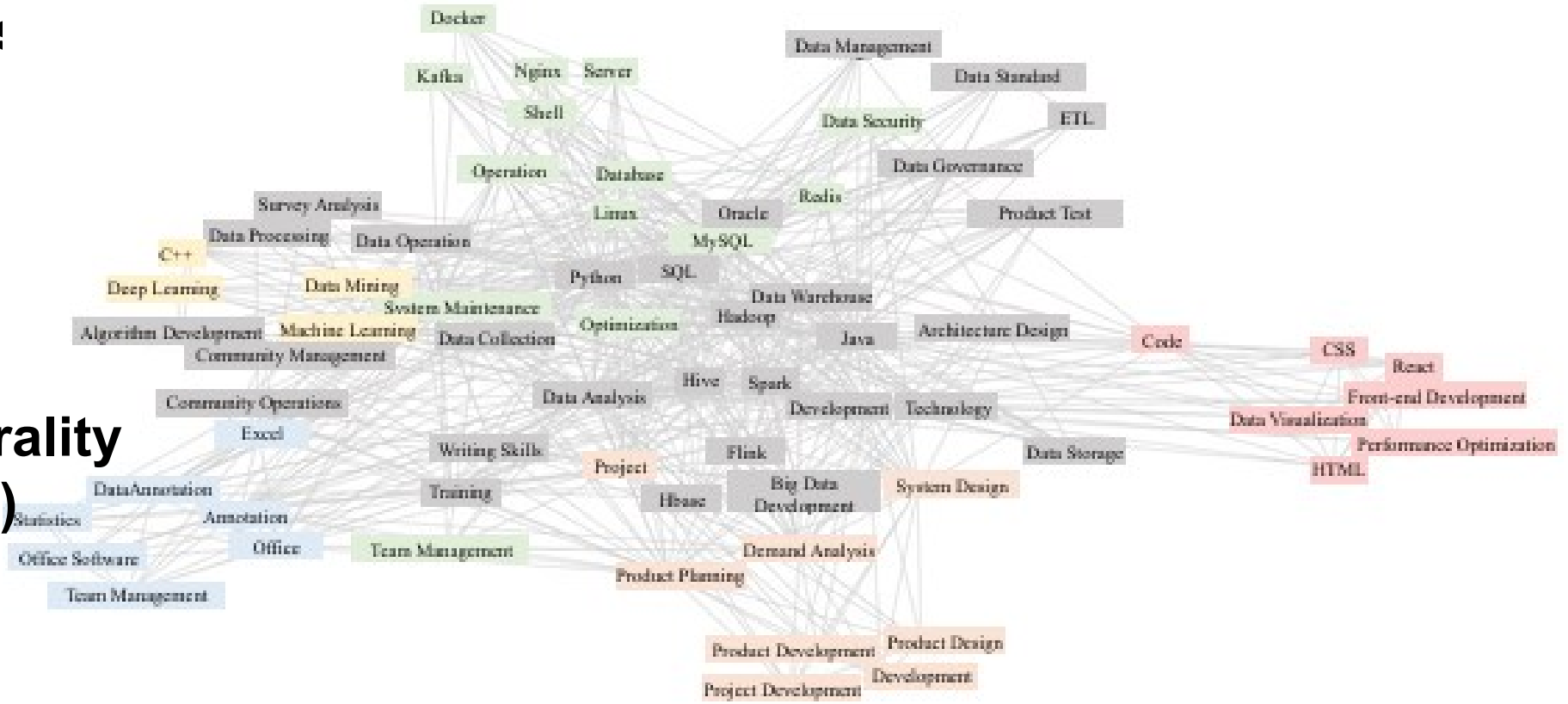
Algorithms & data processing



Relationships Between Skills

Degree Centrality (core foundational s

- Data Analysis
- Python
- Java



Betweenness Centrality (key bridging skills)

- Data Analysis
- Product Testing
- Project