

# 基于图匹配网络的人岗匹配

侯宇蓬 林子涵 刘浩宇



RUC AI Box

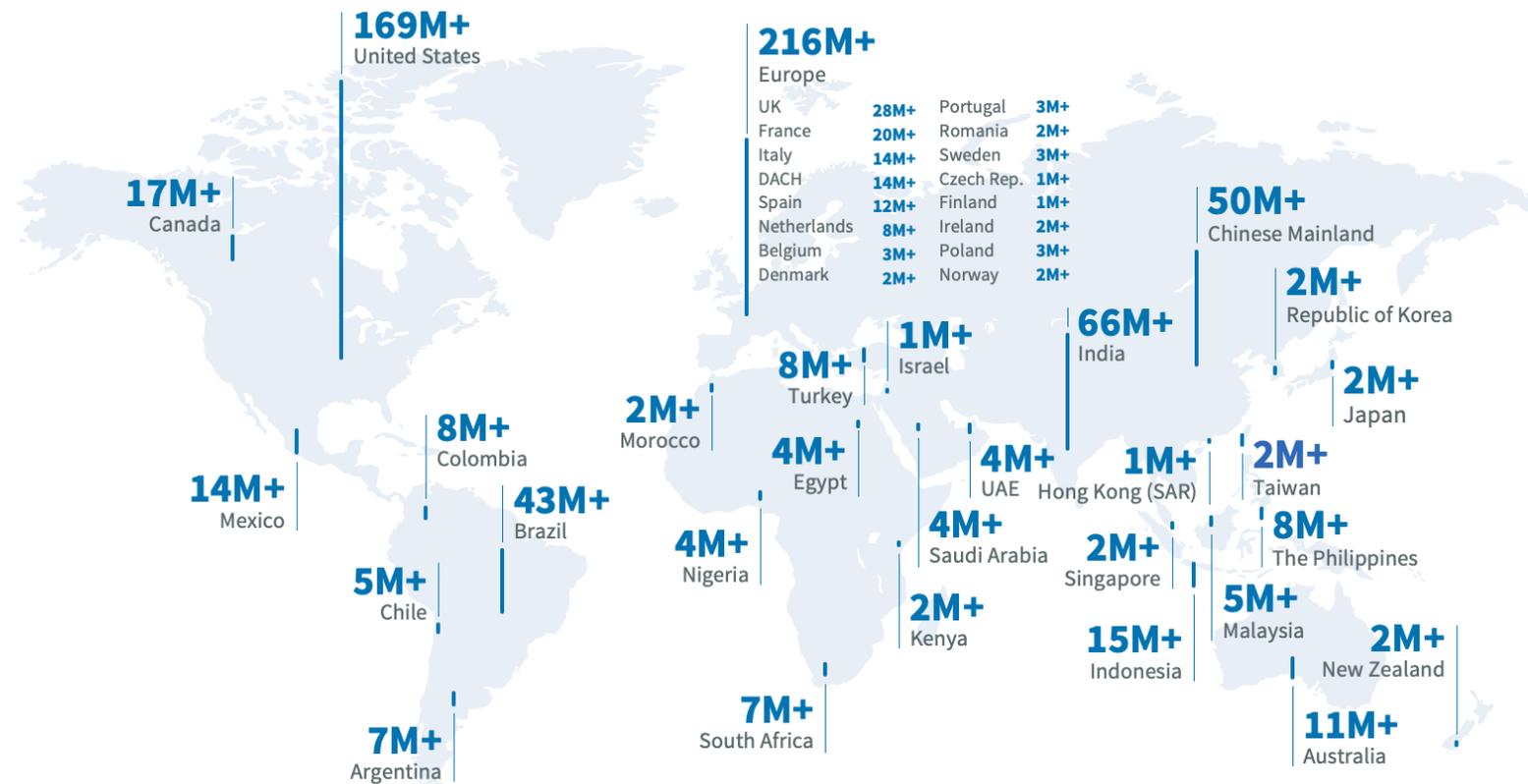
# Motivation

大量求职者与招聘方涌入互联网求职平台。

LinkedIn

6 亿 用户  
2000 万 岗位

BOSS直聘



# Motivation

通常求职者的能力会体现在**简历**文档上，而岗位的要求会写在**工作描述**文档上，所以人岗匹配工作可以建模为对简历-工作描述文档的匹配。

职业资格认证  人力资源管理师\* 专业技术资格 请选择

**专业特长** (200字以内)  
1. 理论知识水平。例如：具备系统的经济、金融知识。  
2. 专业技能水平。例如：具备熟练使用X线图进行数据分析的能力。  
3. 基础研究方法与工具。例如：熟练使用SPSS软件。

是否学生干部  是  否

**担任职务** (20字以内) 学生会主席

**奖学金** (200字以内)  
国家级：国家奖学金1次 2017年12月 全系仅2人  
省市级：省政府奖学金 2016年6月 全系仅2人；企业奖学金（与国家奖学金二选一） 全系仅1人  
校级：一等奖学金 本科3次，研究生1次

注意是填写奖学金哦，请将其他奖项填写至下面对话框

**其他奖励** (300字以内)  
学生工作类/志愿服务类/科研竞赛类/文体特长类  
具体的撰写方法，请参照上面【奖学金】的撰写方法

**计算机水平** (200字以内)  
计算机水平除了等级证书的证明，还应当包括基本的动手操作能力。例如：  
国家计算机二级证书（ACCESS）2016年  
熟练使用OFFICE办公软件，尤其擅长word文字排版、Excel表格制作及基本数据统计、Powerpoint汇报等

**其他情况说明** (200字以内)  
为了保持简历通篇的完整及精练，避免给人留下冗余的第一印象，此处应为【能力信息】未尽事宜的补充说明。  
例如：1. CPA在考，已经通过经济法、财务成本管理、审计、税法四门课程。2. 选修英语同声传译课程一学年，具备良好的英文听说能力。3. 性格阳光开朗，乐观自信，能够快速找到团队中的自我定位。

简历

## 5.4.7 工业大数据建模工程师

### a) 综合能力

- 熟悉工业大数据体系架构；
- 熟悉工业大数据应用场景与工业大数据相关业务流程；
- 熟悉数据建模的方法以及特征工程；
- 具备一定的业务问题分析能力，能围绕行业背景设计与构建机理模型；
- 具备良好的解决方案及技术文档编制能力；
- 具备良好的沟通表达及团队合作能力；

### b) 专业知识

- 掌握计算机基础、数据结构、数学建模及统计学相关知识；
- 掌握典型优化算法、机器学习、机理模型等知识；

### c) 技术技能

- 掌握Kettle、DataStage、Informatica等ETL工具应用，能进行ETL流程梳理、开发、调度、维护等工作；
- 掌握逻辑回归、决策树、神经网络等机器学习算法；
- 熟悉建模与优化算法设计；
- 熟悉MySQL、SQL Server、Oracle等常用数据库、数据库结构及其开发技术；
- 了解数据集市、仓库和ODS，了解数据仓库各类建模理论，以及数据仓库数据层级关系，了解3NF和多维数据模型设计；

### d) 工程实践

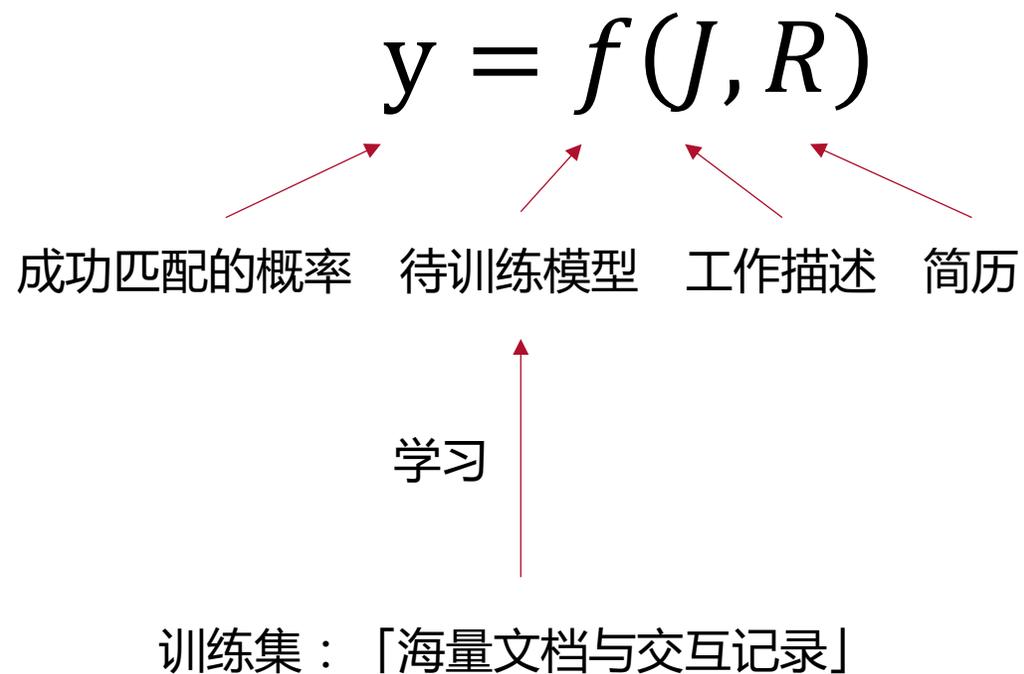
- 具备较强的统计分析、模型抽象和创新能力及基于工业大数据进行数据建模的能力；
- 具备数据模型上线、验证、监测、优化等项目经验。

工作描述



RUC AI Box

# Task



# Challenge

已有工作主要基于深度模型（如 CNN、LSTM 等）对文档进行建模：

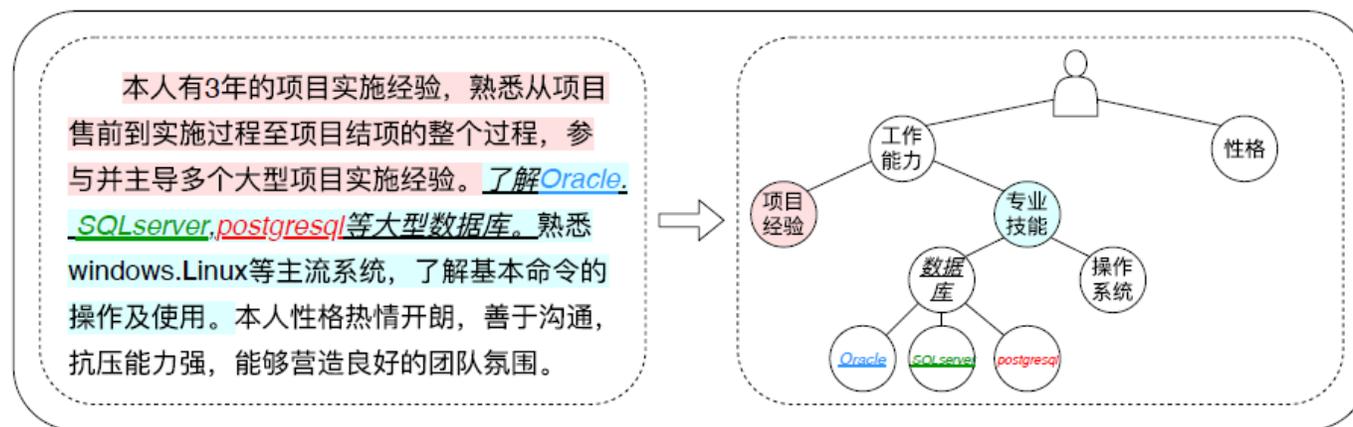
(1) 简历/工作描述文档是高度结构化的，现有工作没有对高度结构化的文档设计特有的归纳偏置；

(2) 已有模型缺乏可解释性。

# Idea

观察：简历/工作描述文档是高度结构化的

是否可以通过大量匹配数据，让模型自动学习文档的结构化形式？



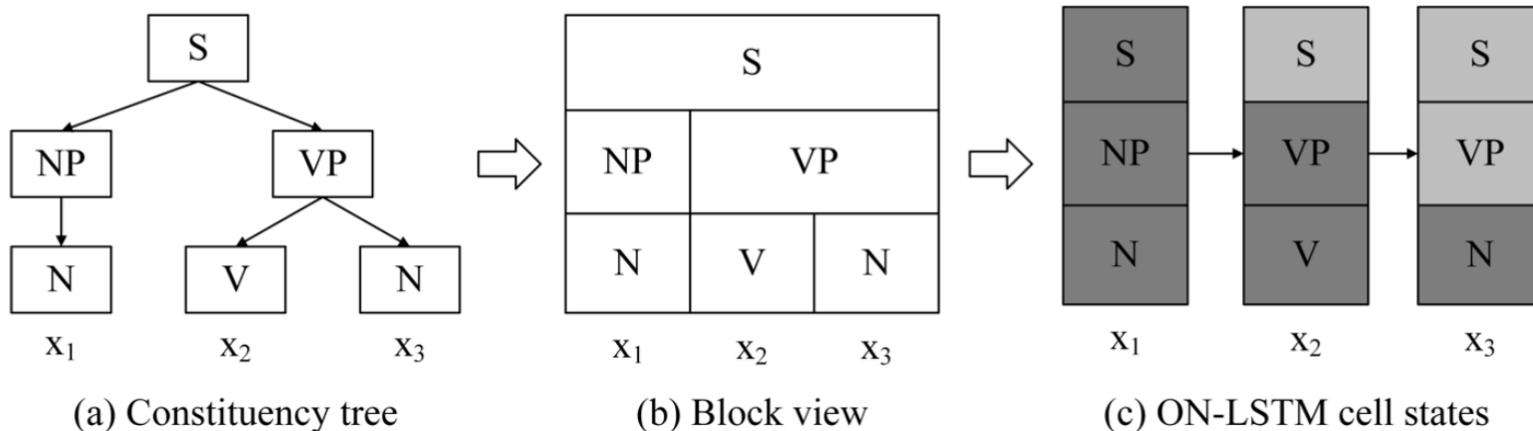
此过程模拟了人在阅读简历文档时构建的脑海中的技能逻辑梳理过程，在保证效果的同时具有很强的可解释性。

文档 -> 生成技能复合树 -> 树匹配



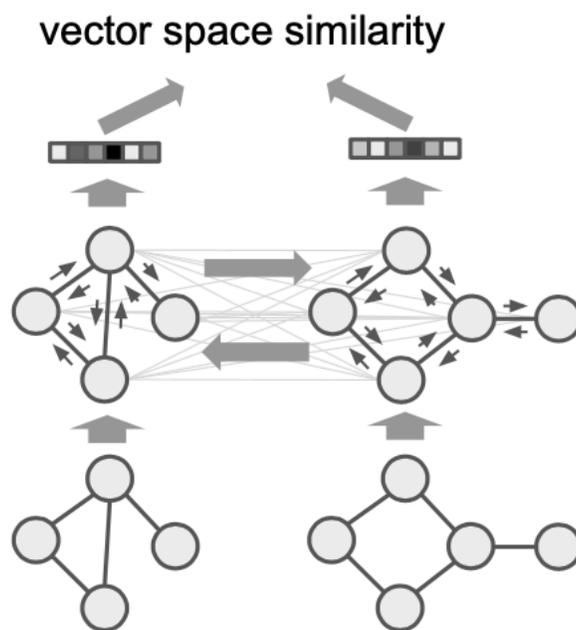
# Methodology

- 如何生成技能复合树？
- 人们倾向于连续书写逻辑上相关联的内容，使文本中尽可能少地出现  
语义转折 —— ON-LSTM

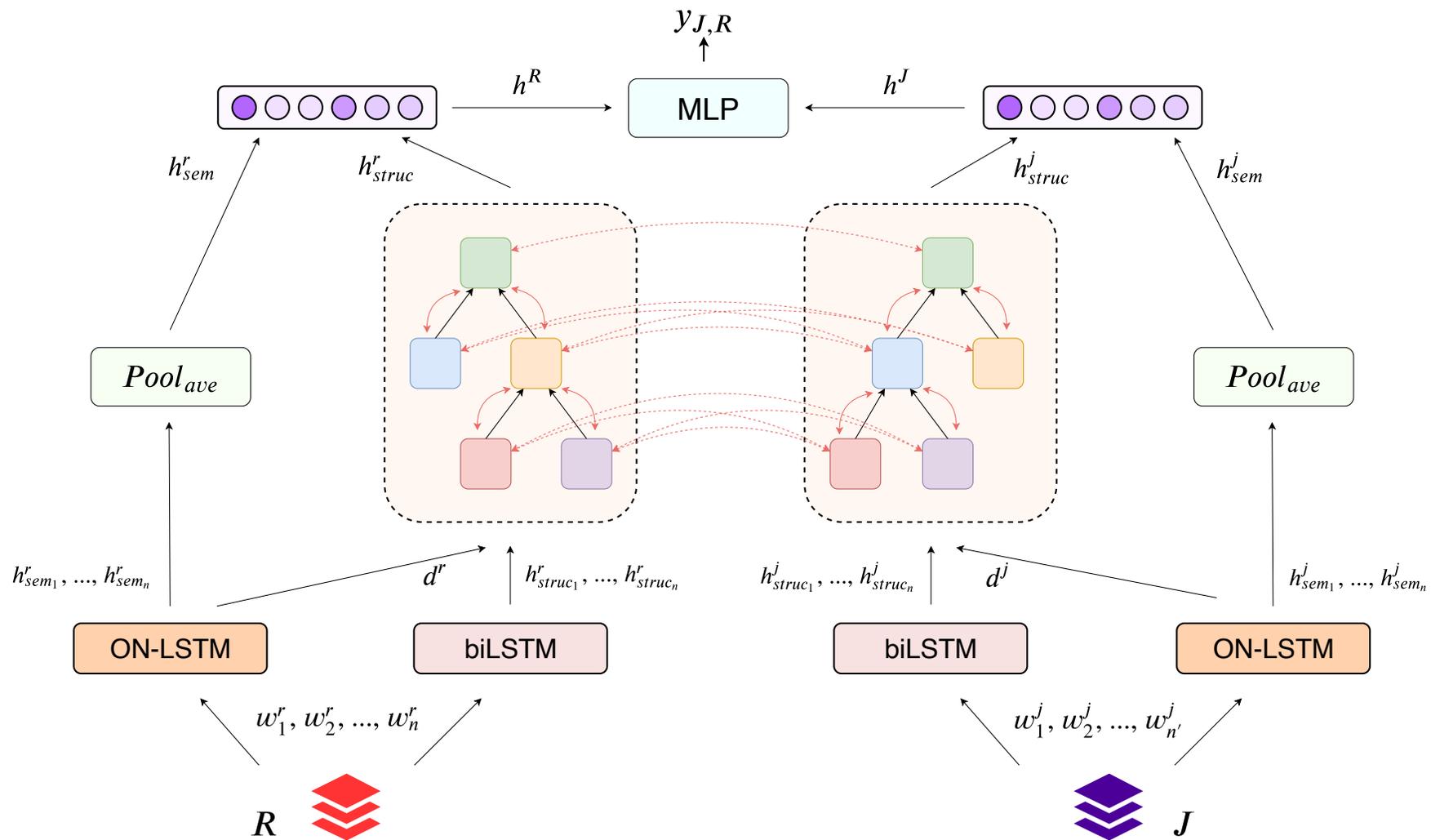


# Methodology

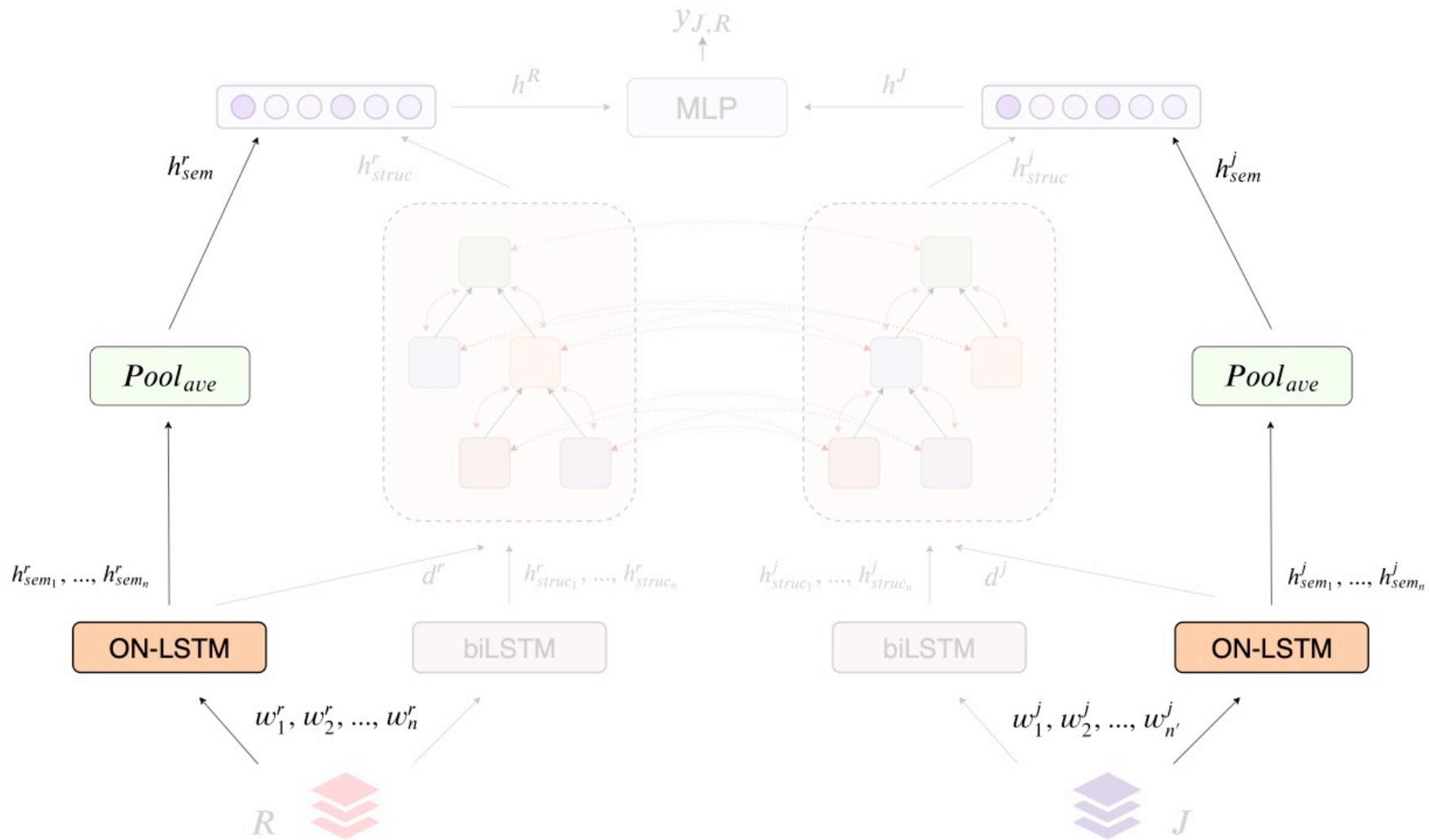
- 如何匹配两棵树？针对树结构改进的图匹配网络



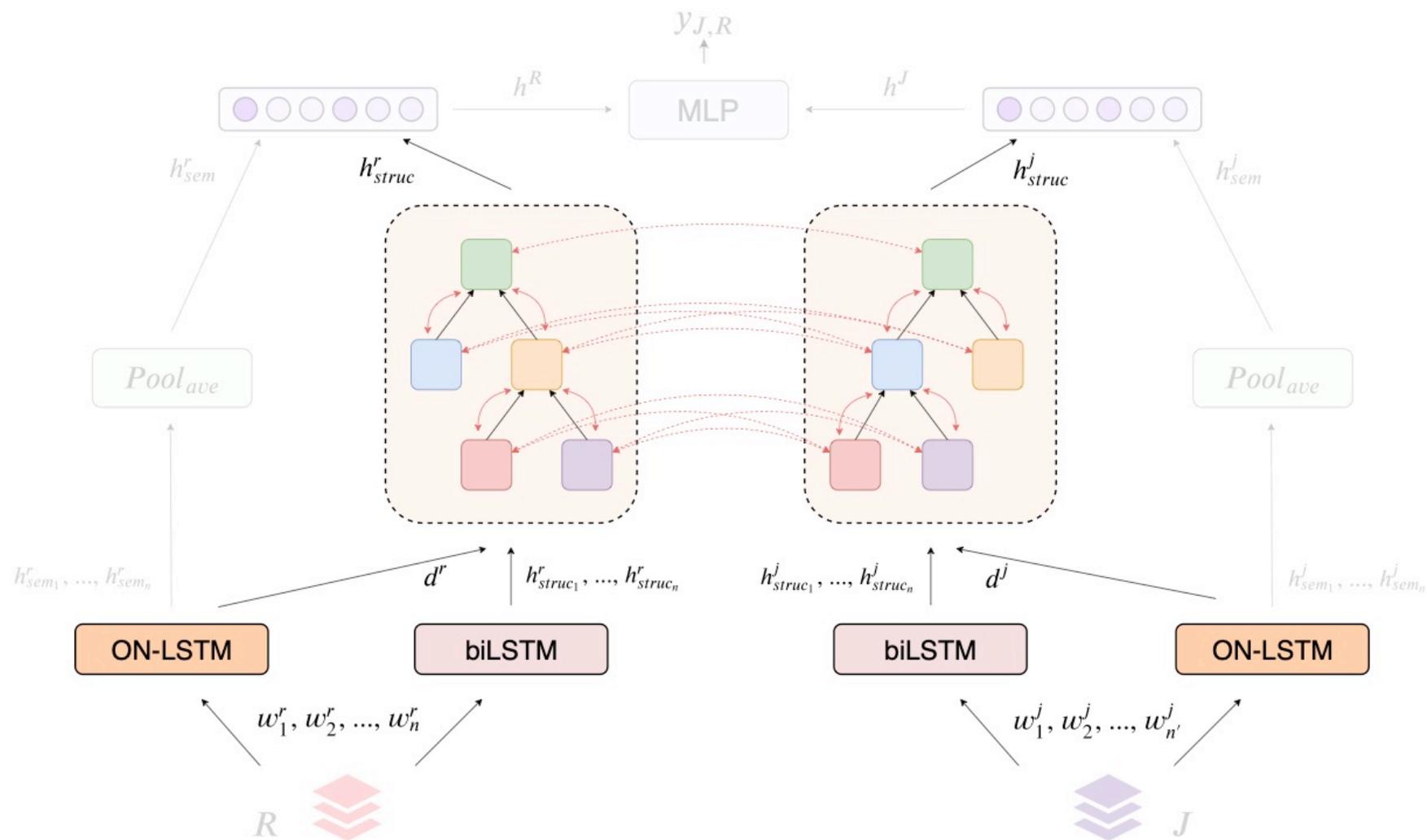
# Methodology - GMPJF



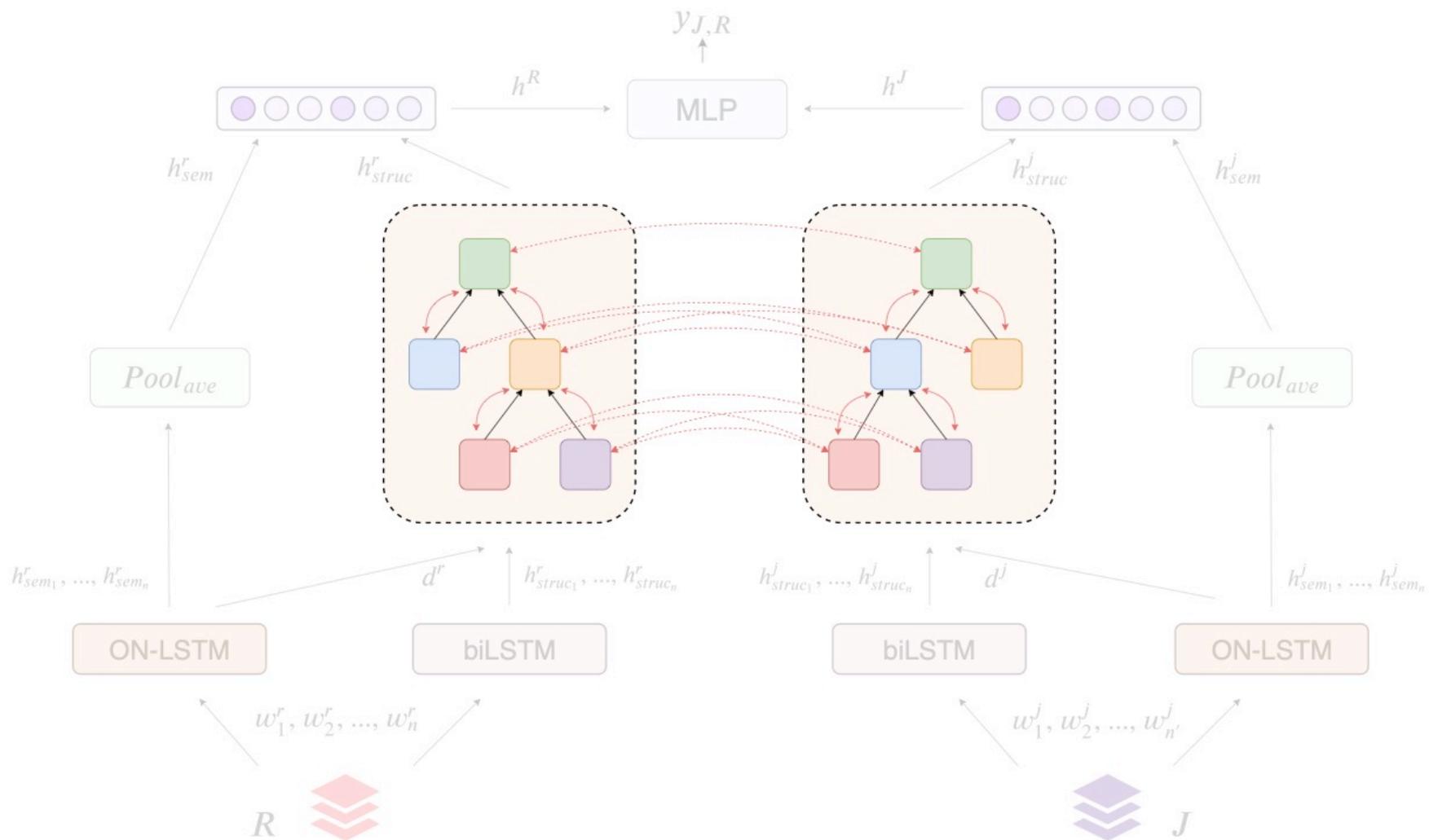
# Methodology – 语义编码



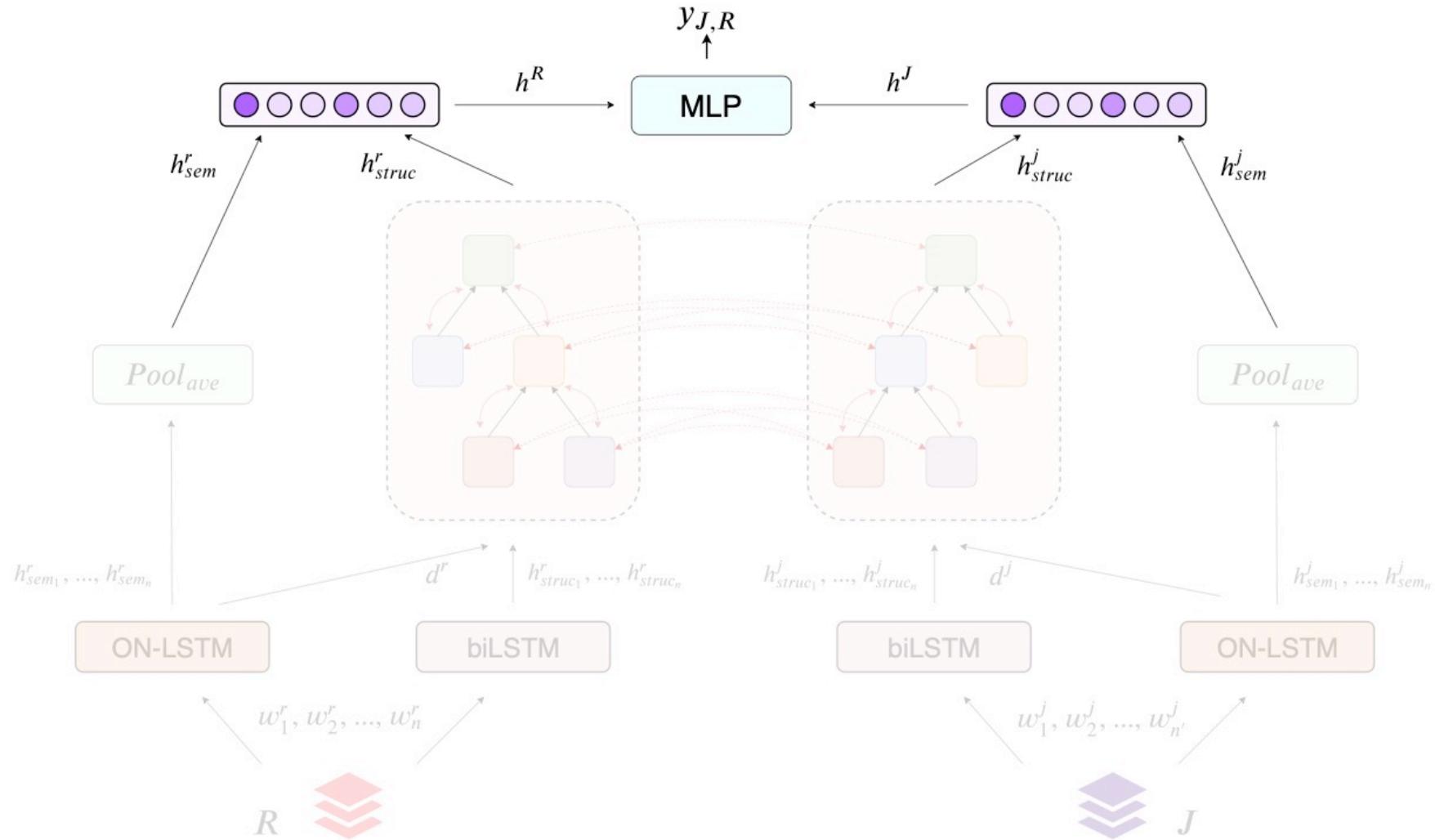
# Methodology – 结构编码



# Methodology – 图匹配



# Methodology – 概率预测



# Experiments

- BOSS 直聘某一线城市技术岗，六个月的真实线上数据

数据集合	训练集	验证集	测试集
数据组数	173,428	21,678	21,678

- 双方成功达成面试，被视作成功匹配。
- 正负样本 = 1:1
- Baselines : DSSM, PJFNN, BPJFNN, APJFNN, GMN
- Metrics : ROCAUC, ACC, Precision, Recall, F1 Score

# Results - Overall

模型	ROCAUC	ACC	Precision	Recall	F1 Score
DSSM	0.6775	0.6305	0.6423	0.602	0.6215
PJFNN	0.6781	0.6288	0.6341	0.6231	0.6286
BPJFNN	0.6936	0.6397	<b>0.6453</b>	0.6333	0.6392
APJFNN	0.6954	0.6403	0.6376	0.6636	0.6503
GMN	0.6876	0.6383	0.6351	0.6636	0.6490
GM-PJF	<b>0.6955</b>	<b>0.6449</b>	0.6395	<b>0.6773</b>	<b>0.6579</b>



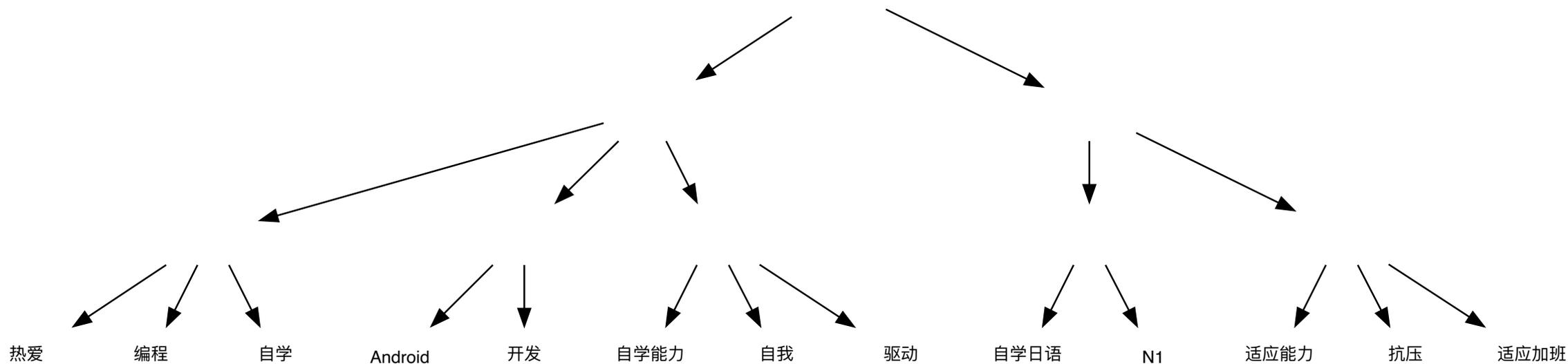
# Results – Ablation Study

模型	ROCAUC	ACC	Precision	Recall	F1 Score
BPJFNN	0.6936	0.6397	0.6453	0.6333	0.6392
GM-PJF - GM - Tree Struc (ON-LSTM)	0.6926	0.6396	0.6394	0.6535	0.6464
GM-PJF - GM	<b>0.6955</b>	0.6407	<b>0.6454</b>	0.6372	0.6413
GM-PJF	<b>0.6955</b>	<b>0.6449</b>	0.6395	<b>0.6773</b>	<b>0.6579</b>



# Case Study

## 技能复合树实例



# Demo

## Person-Job Fit

### 人岗匹配

个人简历

三年大数据开发经验，熟练使用spark框架对数据进行离线或实时计算；熟悉数仓架构，熟练使用hql处理数据。

Resume 树

2.

工作描述

1.负责开发streaming实时计算任务，服务于推荐、搜索和其他业务模块 2.负责大数据数据存储（如用户画像、物料画像）和计算平台的架构设计，提升离线计算和实时计算效率 3.开发部分在线后端服务，如引擎、rank服务  
岗位要求： 1.熟悉java、c++和python中的一种和多种 2.熟练使用至少一个实时计算框架(如jstorm、flink、storm) 3.熟悉大数据相关组件kafka、hbase、mr、hive等 4.掌握常用的rpc框架，对在>线服务系统性能优化有经验者优先 5.态度端正，做事踏实

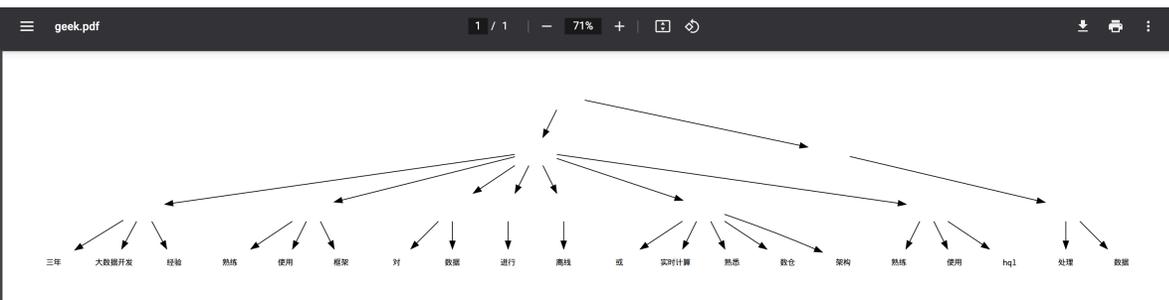
JD 树

匹配

1.

Score

[0.03872915357351303]



# Thanks & QA

问题定义：

$$y = f(J, R)$$

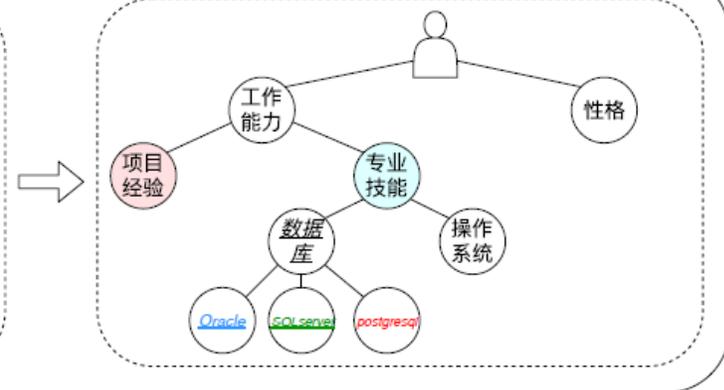
成功匹配的概率    待训练模型    工作描述    简历

学习

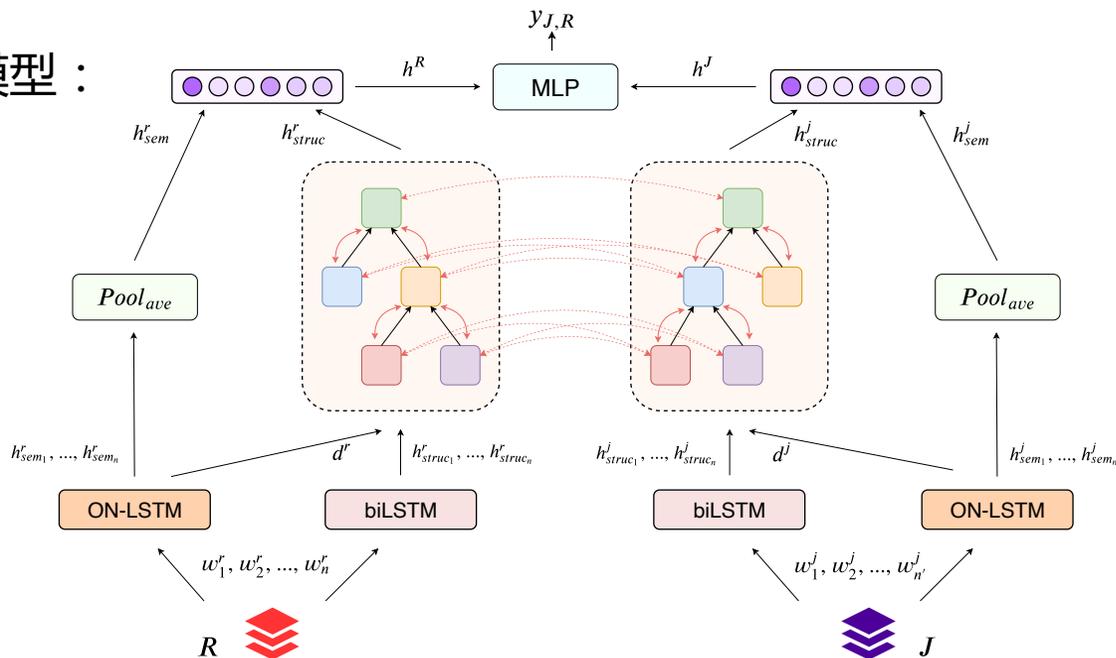
训练集：「海量文档与交互记录」

动机：

本人有3年的项目实施经验，熟悉从项目售前到实施过程至项目结项的整个过程，参与并主导多个大型项目实施经验。了解Oracle, SQLserver, postgresql等大型数据库。熟悉windows.Linux等主流系统，了解基本命令的操作及使用。本人性格热情开朗，善于沟通，抗压能力强，能够营造良好的团队氛围。



模型：



RUC AI Box