HUAeLOUD
---华/云/中/盛---

华云中盛法研杯方案

王旭鹏&姜焰

目录

1.任务介绍

本任务的主要目的是为了将案件描述中重要事实描述自动抽取出来，并根据领域专家设计的案情要素体系进行分类。本质上为短文本的multilabel分类问题.

案情要素抽取的结果可以用于案情摘要、可解释性的类案推送以及相关知识推荐等司法领域的实际业务需求中。

具体地，给定司法文书中的相关段落，系统需针对文书中每个句子进行判断，识别其中的关键案情要素。

本任务共涉及三个领域，包括婚姻家庭、劳动争议、借款合同等领域。

2.数据介绍

本任务所使用的数据集主要来自于“中国裁判文书网”公开的法律文书，每条训练数据由一份法律文书的案情描述片段构成，其中每个句子都被标记了对应的类别标签（需要特别注意的是，每个句子对应的类别标签个数不定），例如：

{"labels": ["DV9", "DV1", "DV2"], "sentence": "原告谢春佑诉称，原、被告因感情不和于2014年3月经衡阳县人民法院判决离婚，并判决婚生女孩周茵（2001年9月8日出生，现在衡阳县西渡镇蒸阳中学就读）由被告抚养。"}

{"labels": ["LB2", "LB3"],"sentence": "原告张倩向本院提出诉讼请求：1、判决被告补付原告工资7200元、经济补偿金2400元。“}

{"labels": ["LN2", "LN6", "LN4"]},"sentence": "本院认定事实如下：被告兄弟控股集团有限公司分别向中国工商银行股份有限公司永康支行借款3笔，本金共计人民币4000万元，分别为：“}

2.数据介绍

标签对应的具体内容:

Divorce

婚后有子女
限制行为能力子女抚养
有夫妻共同财产
支付抚养费
不动产分割
婚后分居
二次起诉离婚
按月给付抚养费
准予离婚
有夫妻共同债务
婚前个人财产
法定离婚
不履行家庭义务
存在非婚生子
适当帮助
不履行离婚协议
损害赔偿
感情不和分居满二年
子女随非抚养权人生活
婚后个人财产

Labor

解除劳动关系
支付工资
支付经济补偿金
未支付足额劳动报酬
存在劳动关系
未签订劳动合同
签订劳动合同
支付加班工资
支付未签订劳动合同二倍工资赔偿
支付工伤赔偿
劳动仲裁阶段未提起
不支付违法解除劳动关系赔偿金
经济性裁员
不支付奖金
违法向劳动者收取财物
特殊工种
支付工亡补助金|丧葬补助金|抚恤金
用人单位提前通知解除
法人资格已灭失
有调解协议

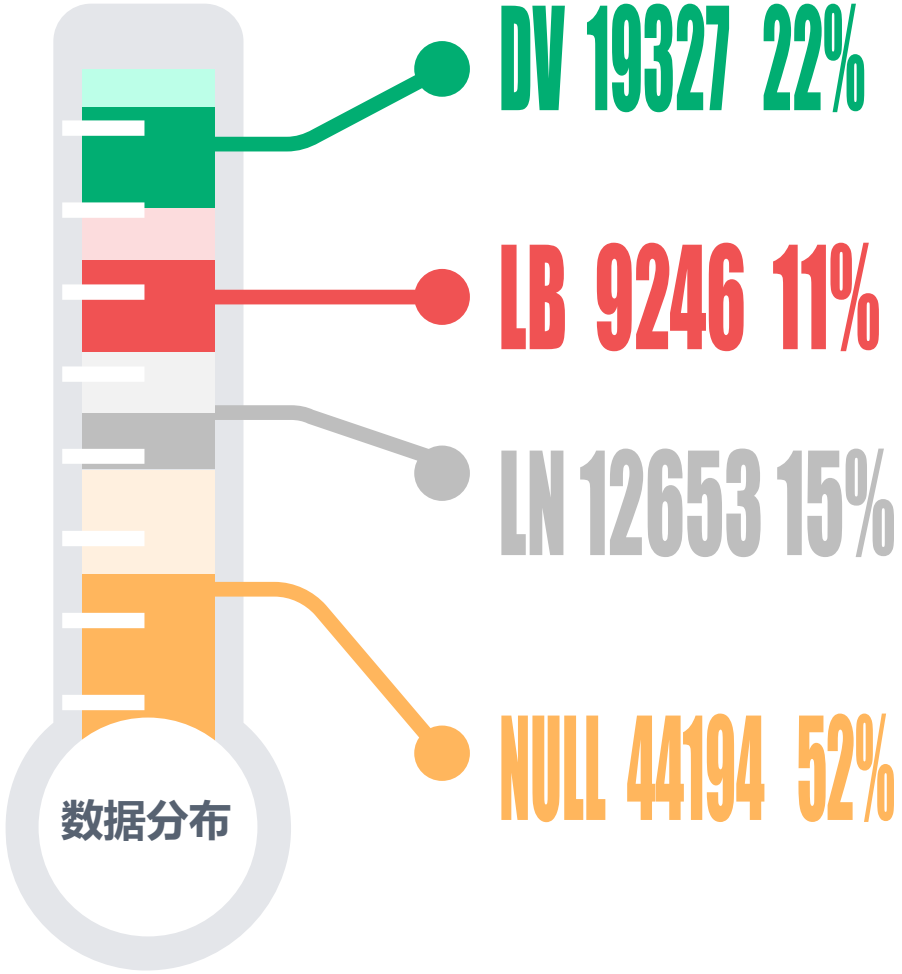
Loan

债权人转让债权
借款金额x万元
有借贷证明
贷款人系金融机构
返还借款
公司|单位|其他组织借款
连带保证
催告还款
支付利息
订立保证合同
有书面还款承诺
担保合同无效|撤销|解除
拒绝履行偿还
免除保证人保证责任
保证人不承担保证责任
质押人系公司
贷款人未按照约定的日期|数额提供借款
多人借款
债务人转让债务
约定利率不明

2.数据介绍

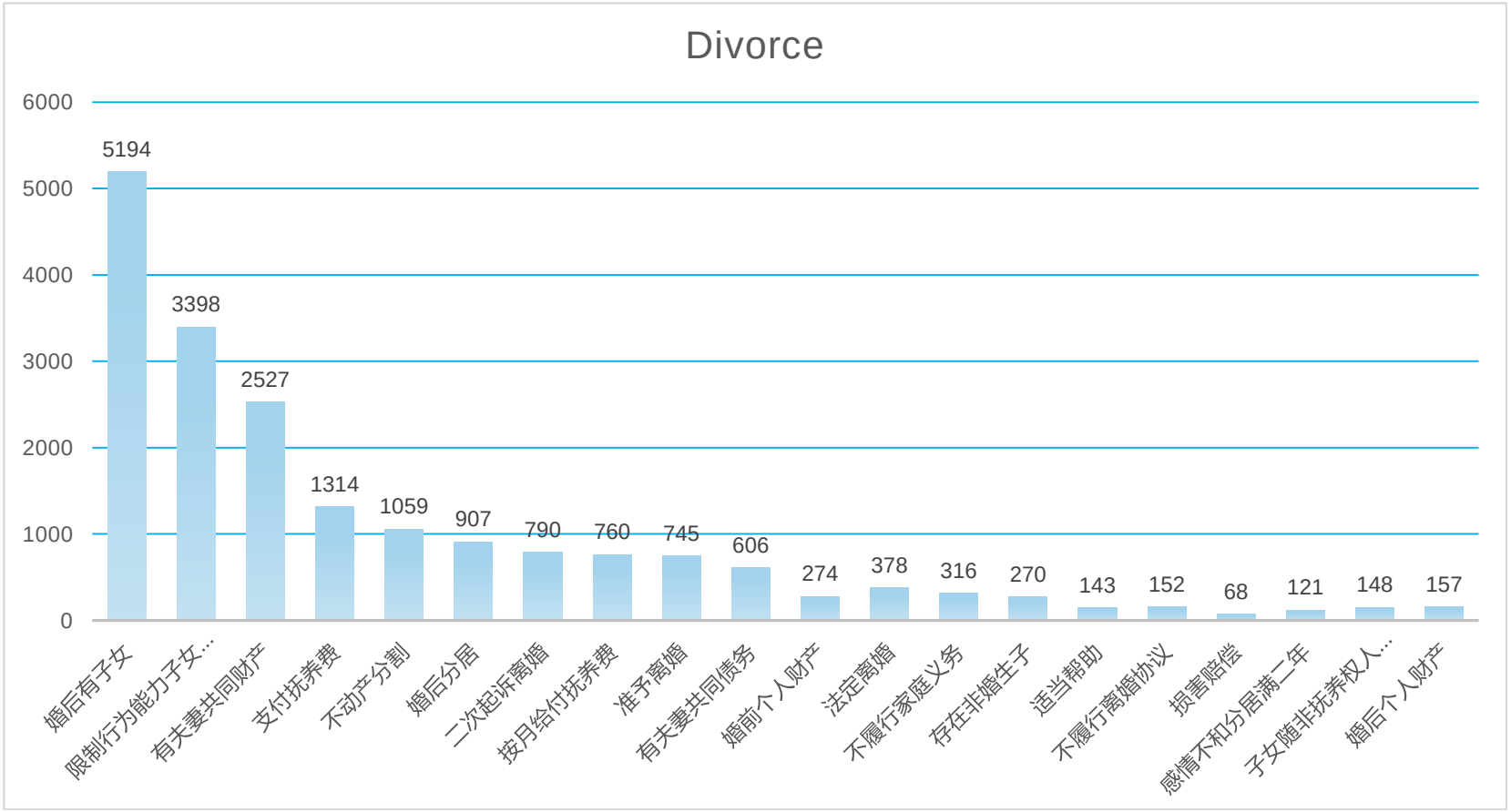
数据分布:

将多标签拆成单标签后统计



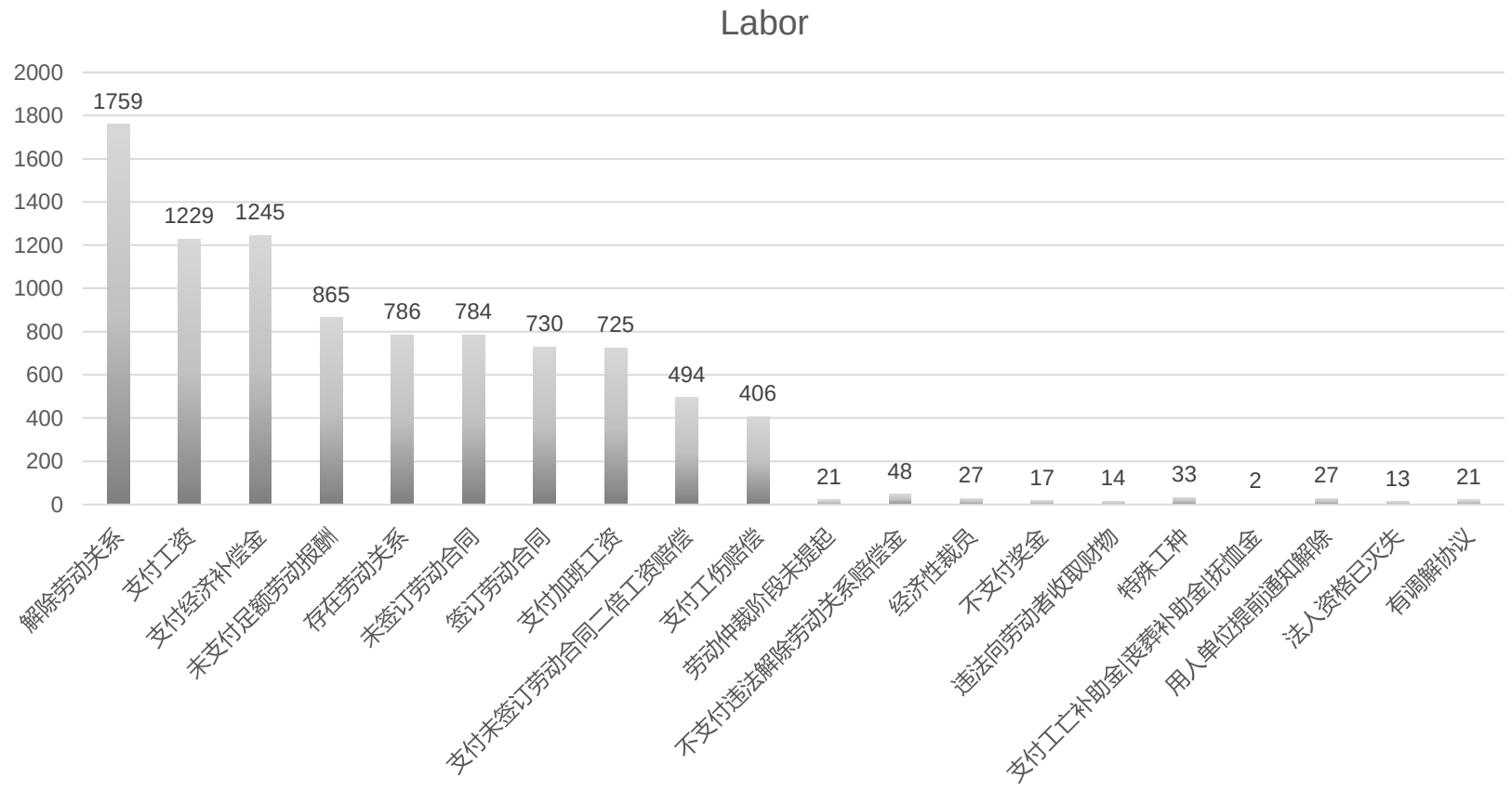
2.数据介绍

数据分布:


Divorce
6000
5000
4000
3000
2000
1000
0
5194
3398
2527
1314
1059
907
790
760
745
606
274
378
316
270
143
152
68
121
148
157
婚后有子女
限制行为能力子女...
有夫妻共同财产
支付抚养费
不动产分割
婚后分居
二次起诉离婚
按月给付抚养费
准予离婚
有夫妻共同债务
婚前个人财产
法定离婚
不履行家庭义务
存在非婚生子
适当帮助
不履行离婚协议
损害赔偿
感情不和分居满二年
子女随非抚养权人...
婚后个人财产

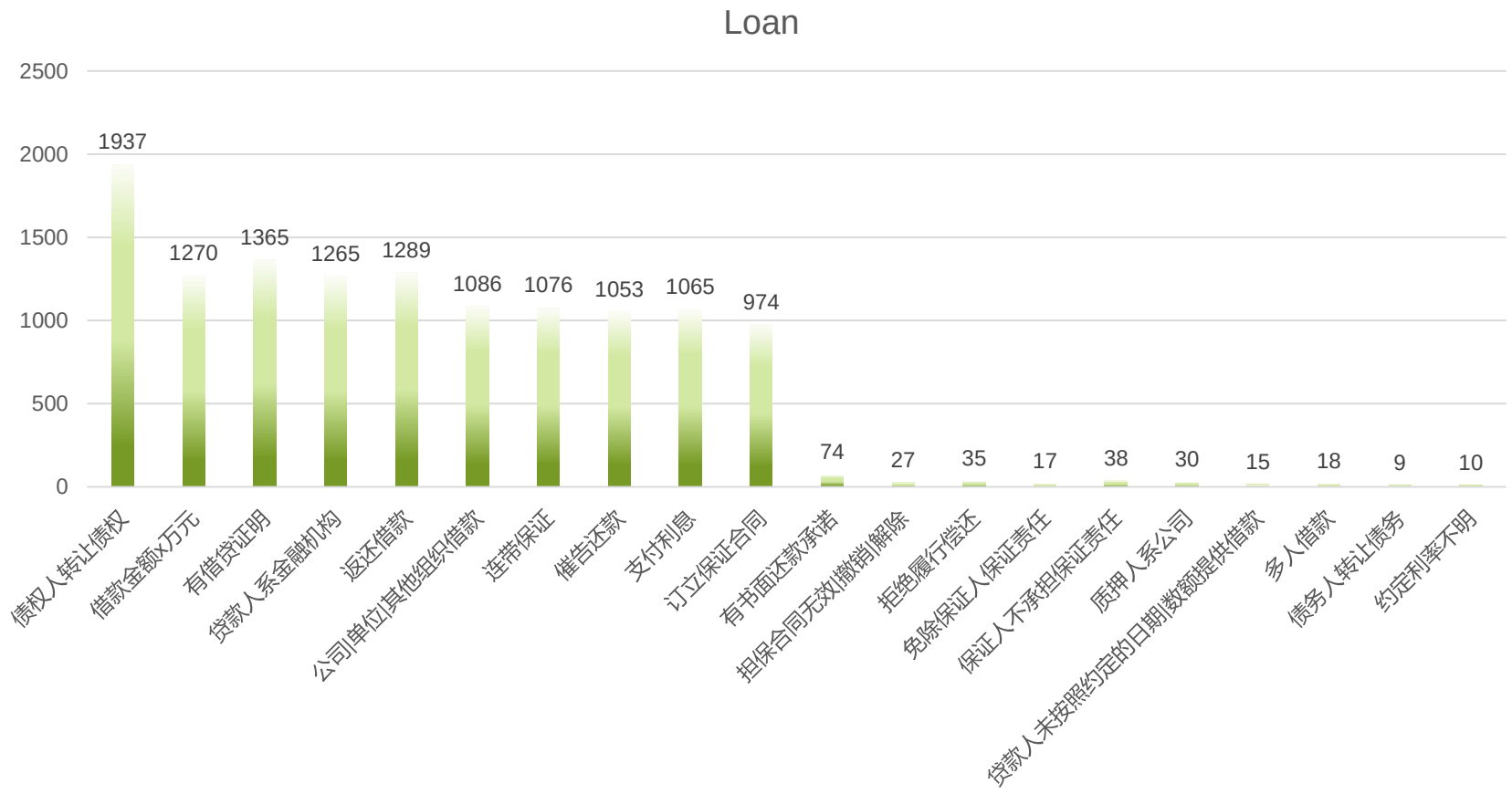
2.数据介绍

数据分布:


Labor
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1759
1229
1245
865
786
784
730
725
494
406
21
48
27
17
14
33
2
27
13
21
解除劳动关系
支付工资
支付经济补偿金
未支付足额劳动报酬
存在劳动关系
未签订劳动合同
签订劳动合同
支付加班工资
支付未签订劳动合同二倍工资赔偿
支付工伤赔偿
劳动仲裁阶段未提起
不支付违法解除劳动关系赔偿金
经济性裁员
不支付奖金
违法向劳动者收取财物
特殊工种
支付工亡补助金|丧葬补助金|抚恤金
用人单位提前通知解除
法人资格已灭失
有调解协议

2.数据介绍

数据分布:


Loan
2500
2000
1500
1000
500
0
1937
1270
1365
1265
1289
1086
1076
1053
1065
974
74
27
35
17
38
30
15
18
9
10
债权人转让债权
借款金额x万元
有借贷证明
贷款人系金融机构
返还借款
公司|单位其他组织借款
连带保证
催告还款
支付利息
订立保证合同
有书面还款承诺
担保合同无效|撤销|解除
拒绝履行偿还
免除保证人保证责任
保证人不承担保证责任
质押人系公司
贷款人未按照约定的日期数额提供借款
多人借款
债务人转让债务
约定利率不明

3.模型和tricks介绍

模型的输入:

在模型的输入中我们不仅仅只输入了文本,同时加入了对应类别的标签一起进入Bert(最终模型为XLnet),具体如下:

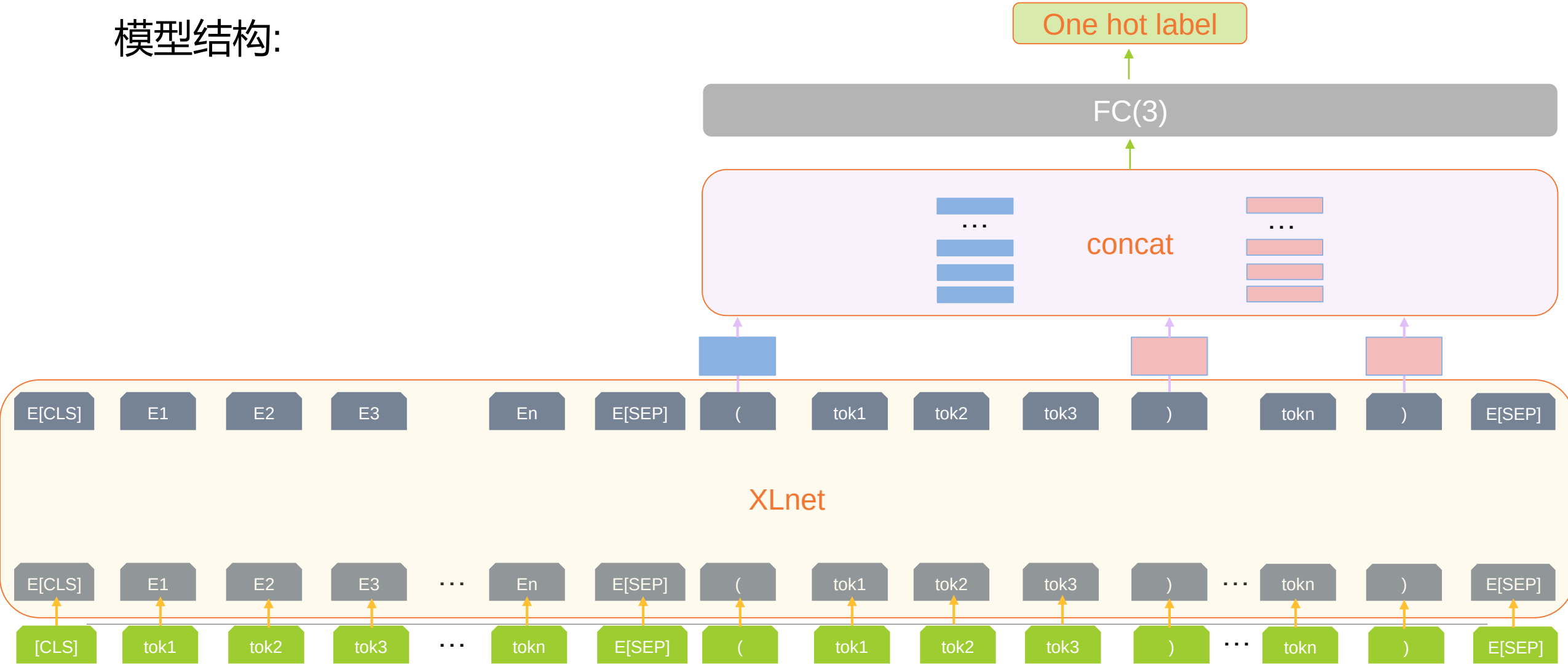
[CLS]原告谢春佑诉称，原、被告因感情不和于2014年3月经衡阳县人民法院判决离婚，并判决婚生女孩周茵（2001年9月8日出生，现在衡阳县西渡镇蒸阳中学就读）由被告抚养。[SEP] (婚后有子女)……(婚后个人财产)[SEP]

模型的标签:

对不同的标签做One Hot-encoding

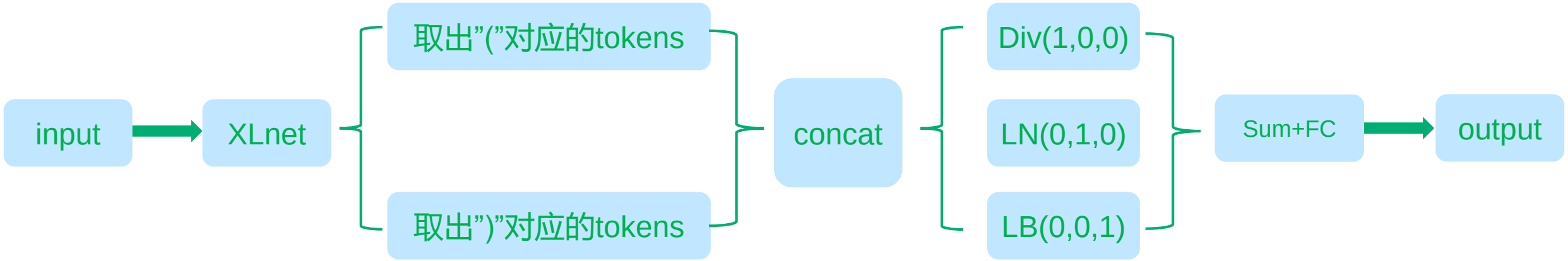
3.模型和tricks介绍

模型结构:



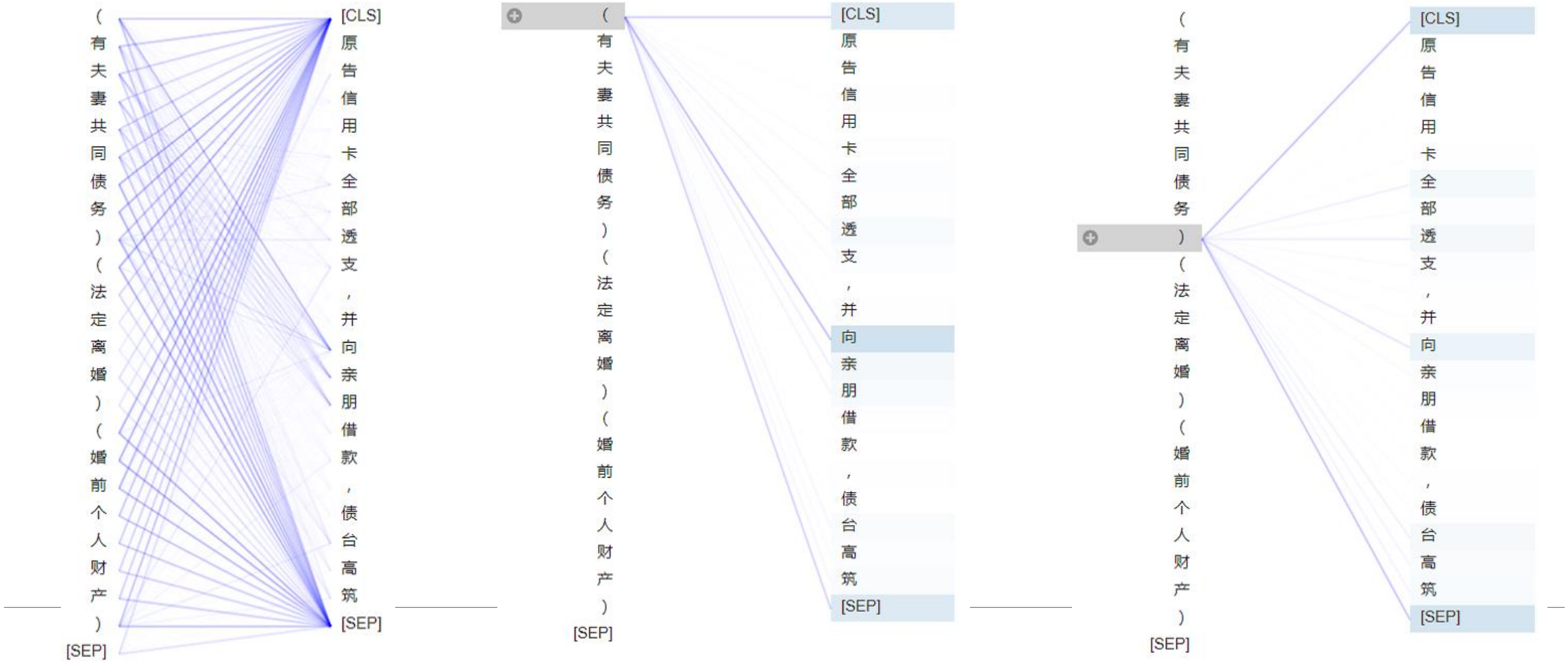
3.模型和tricks介绍

模型结构:



3.模型和tricks介绍

Why label embedding:



3.模型和tricks介绍

Why label embedding:

1.将标签的文字信息加入到模型学习，可以利用到标签的语义信息

2.通过Bert,可以增加标签与文本之间的联系

3.通过label embedding可以提升小样本类的效果

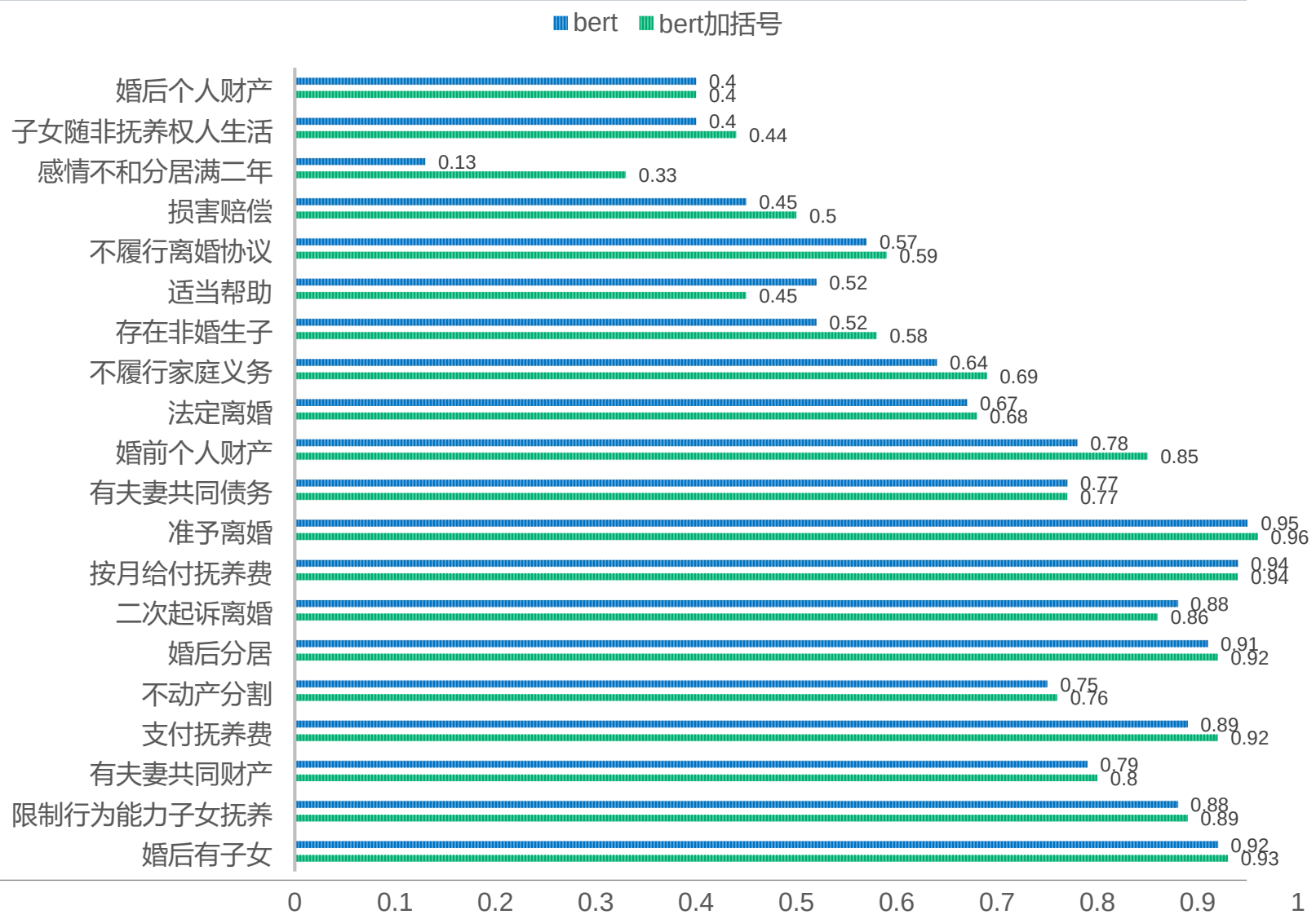
3.模型和tricks介绍

Why label embedding:

Label embedding增加了小数据类的f1值

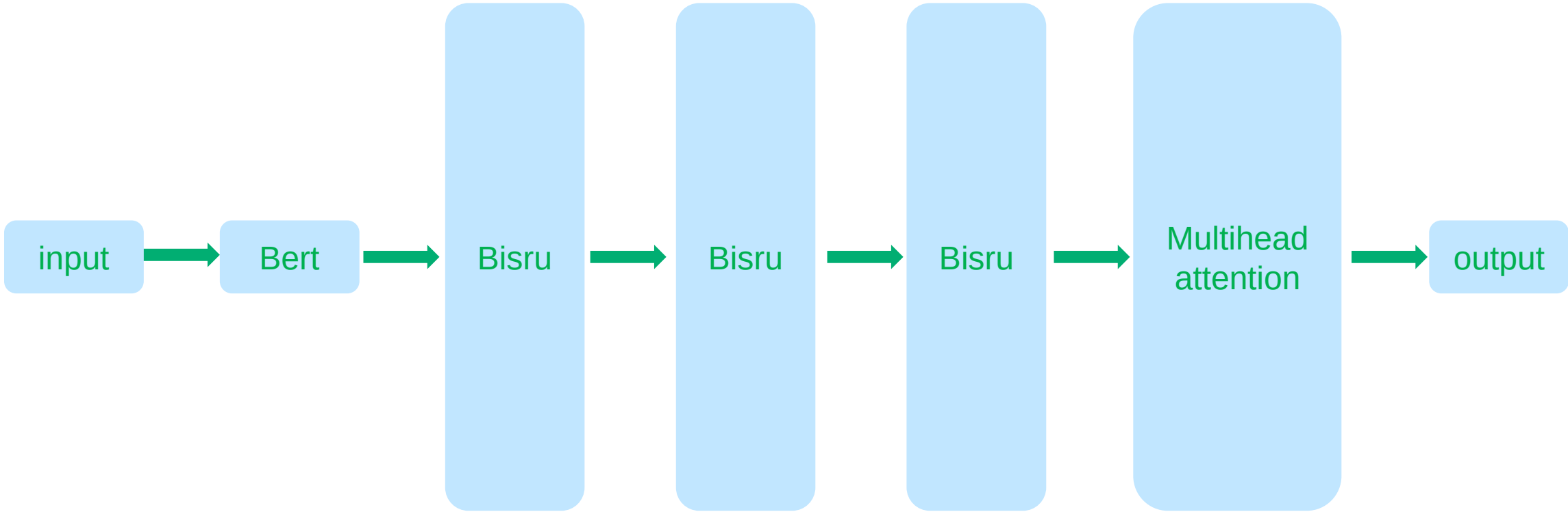
小

大



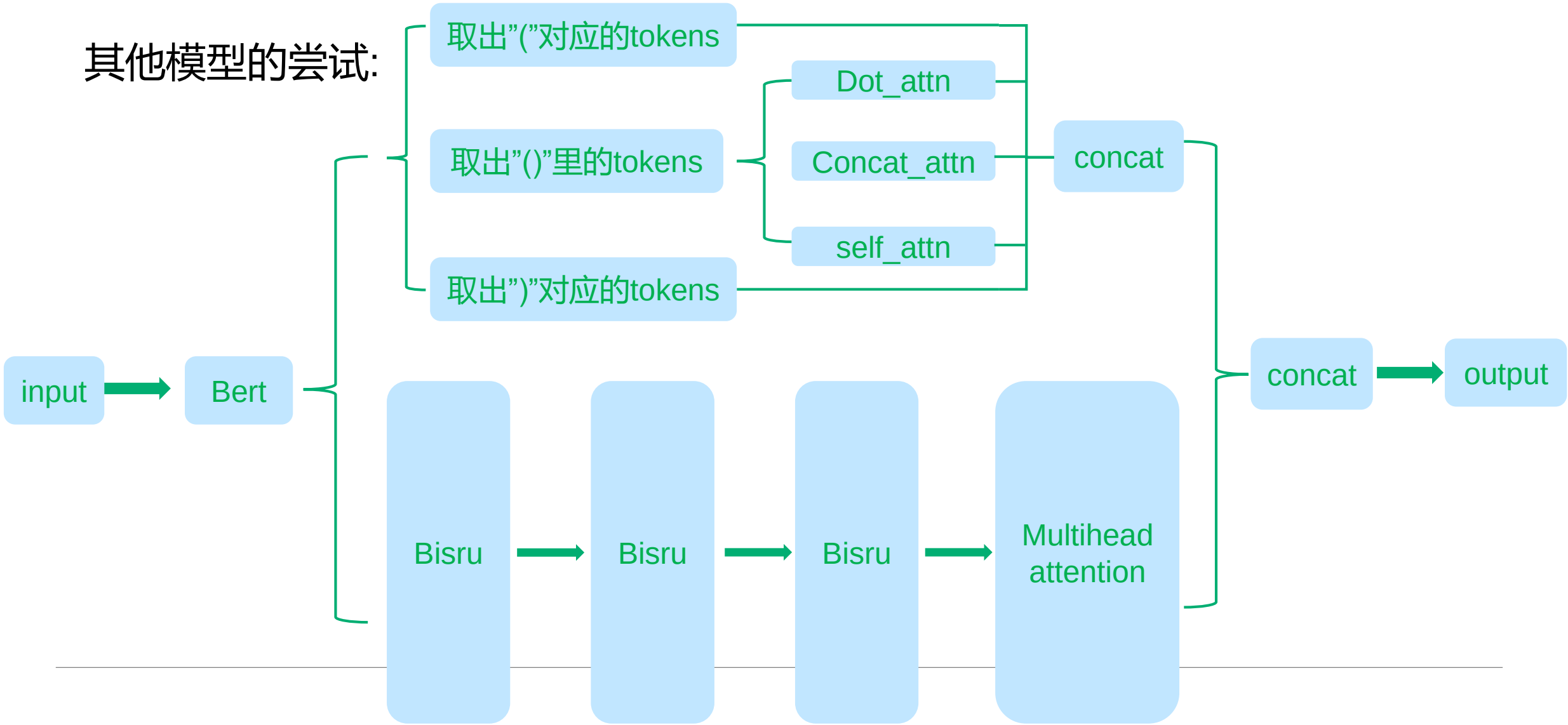
3.模型和tricks介绍

其他模型的尝试:



3.模型和tricks介绍

其他模型的尝试:



3.模型和tricks介绍

数据增强:

1.网上数据,手动标注—4200

2.百度翻译API– 8000

3.模型和tricks介绍

单模型线上对比:

3.模型和tricks介绍

最终提交:

$$[Xlnet + \text{括号}](\text{切分数据}) * 0.6 \ + \ [Xlnet + \text{括号}](\text{全量数据}) * 0.4 \ = \ 75.727$$