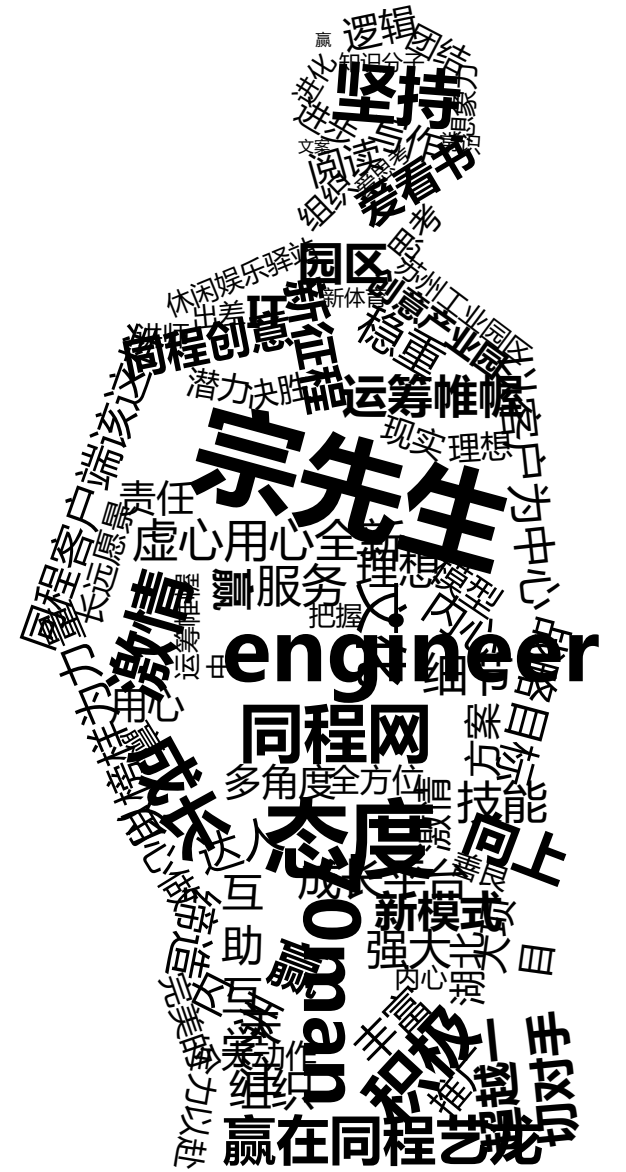
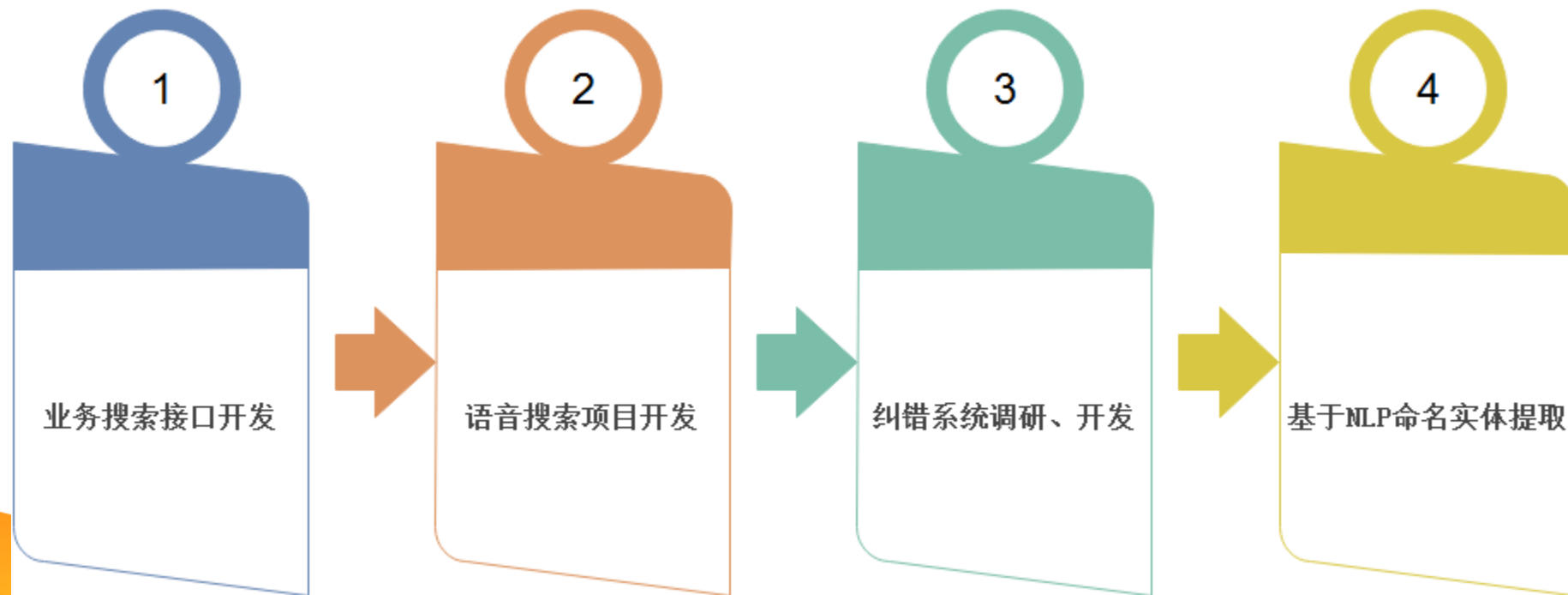


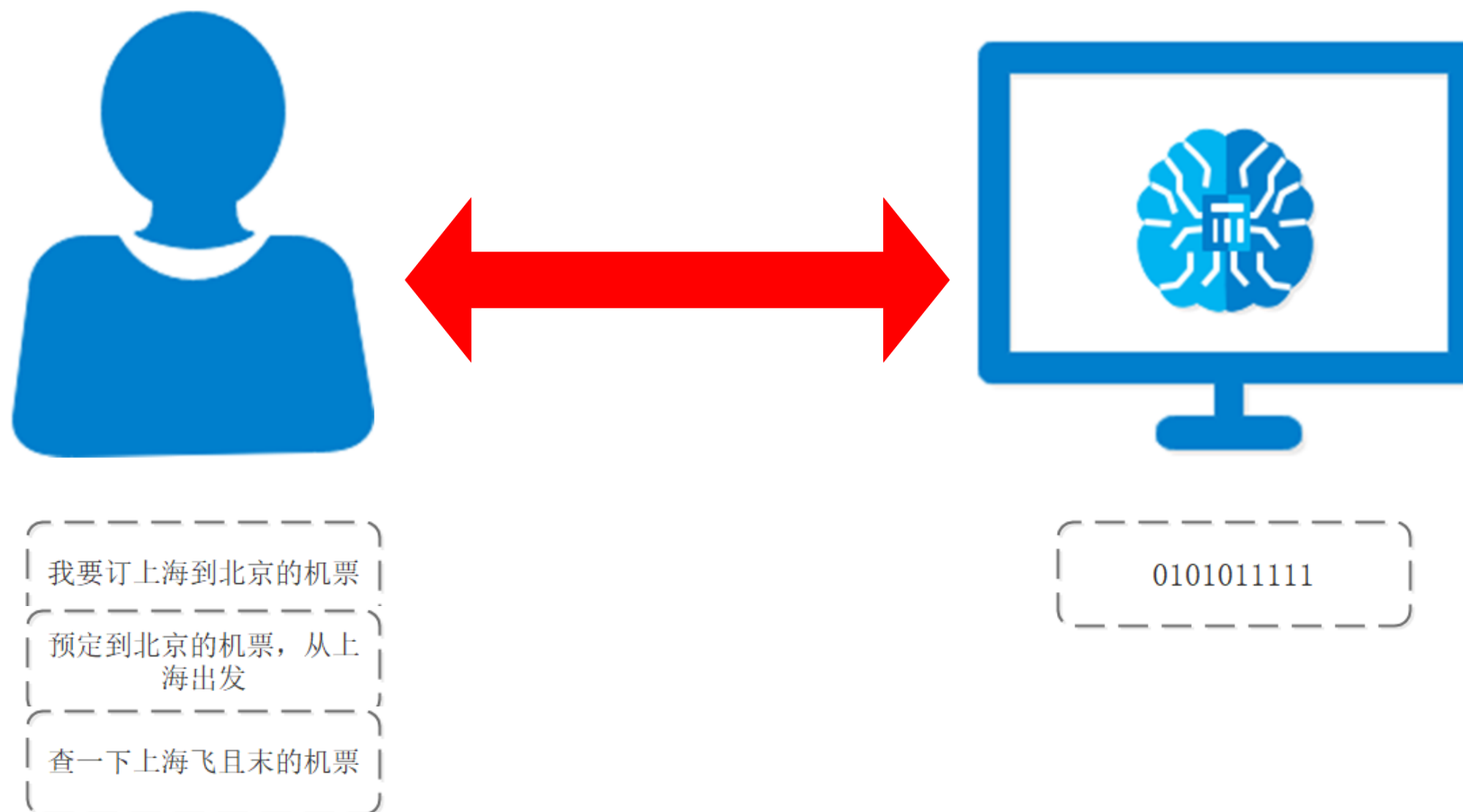
# 智能中文实体识别系统

智能搜索研发部--宗寿金

# 个人简介

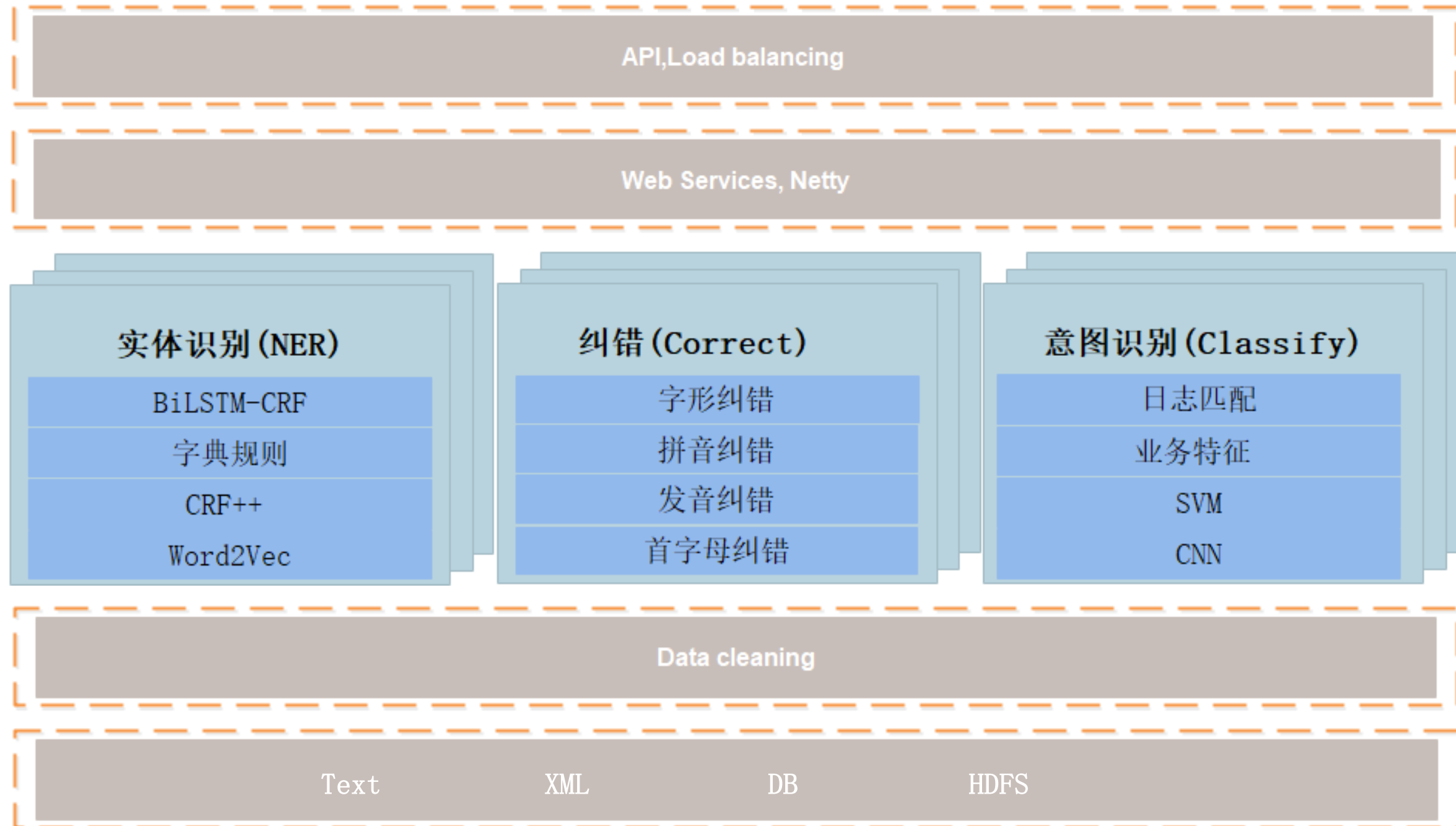
- 姓名:宗寿金
- 部门:智能搜索研发部

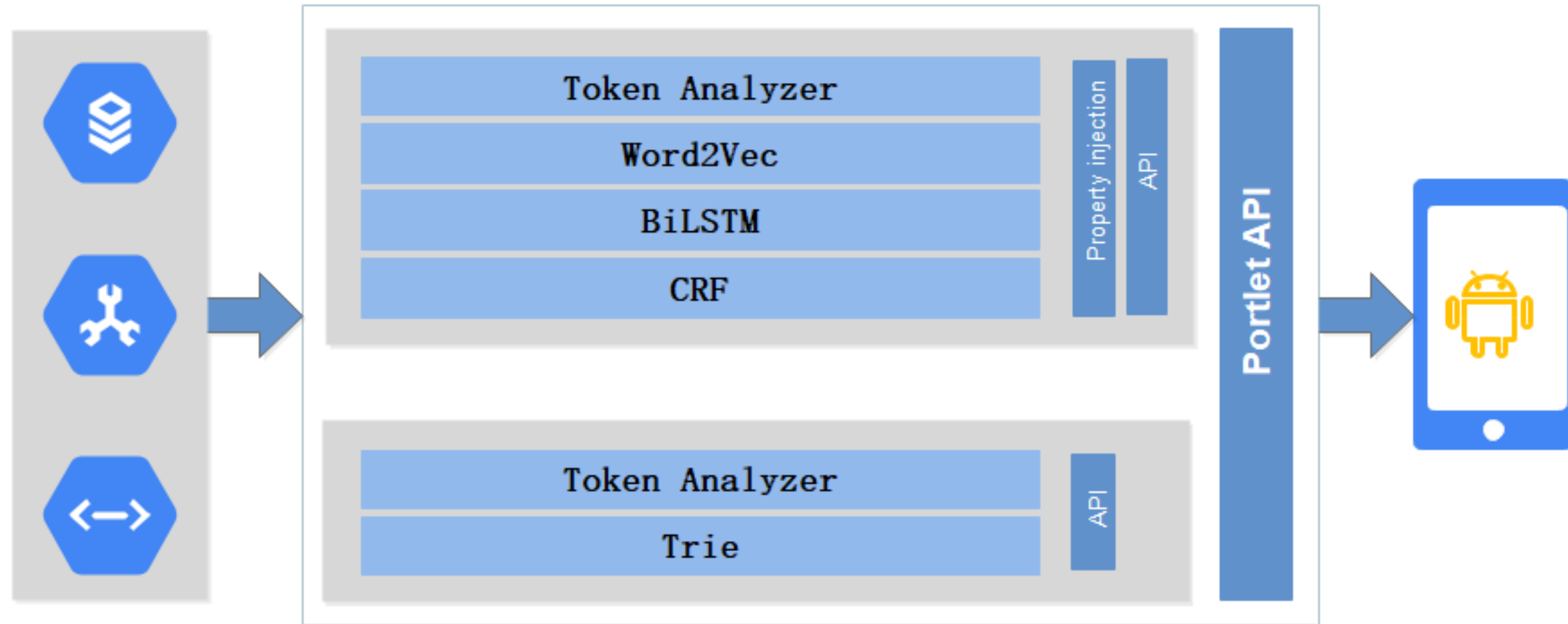




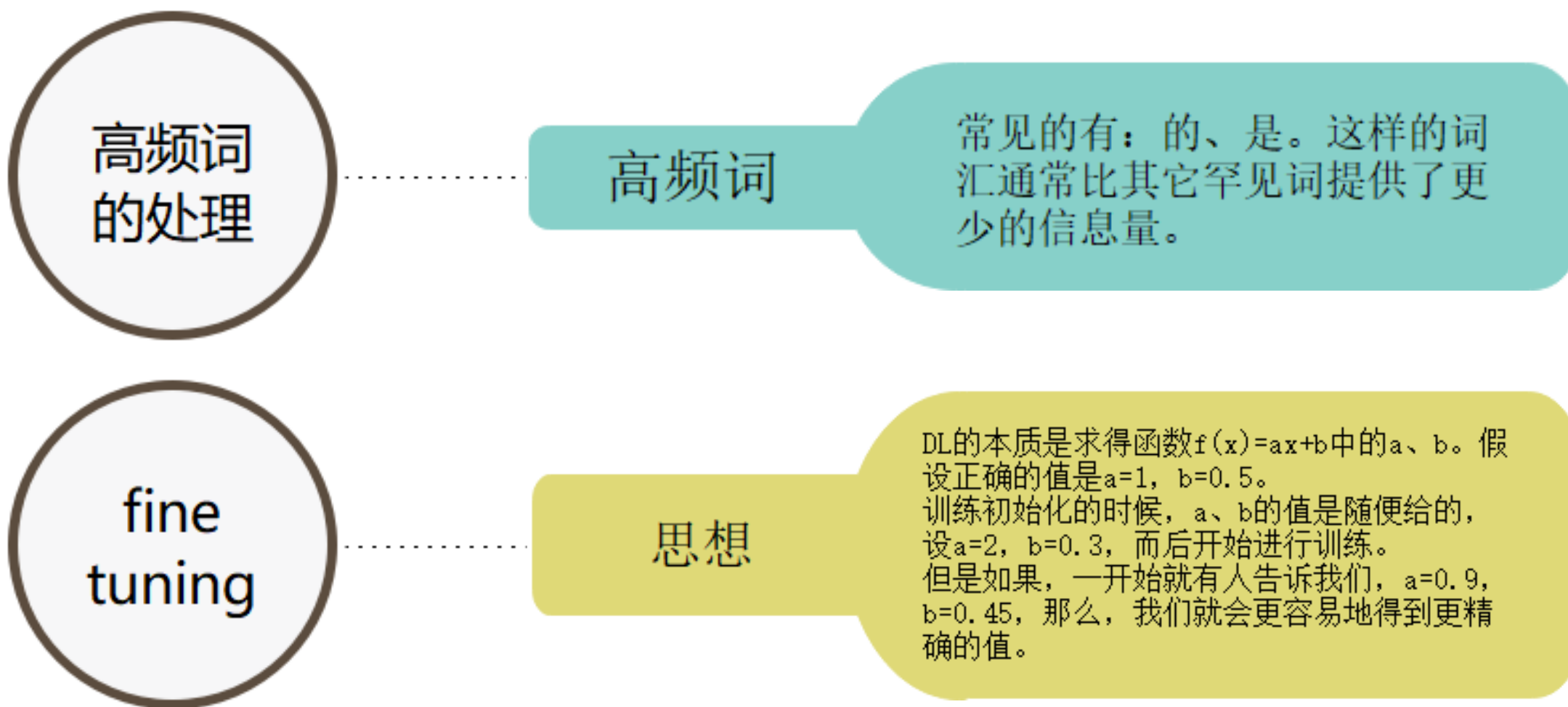
让机器理解人类语言是很难的，因为语言天生就是复杂的。业界目前在全文领域的处理是不够优秀的，但是在特定的垂直领域是有好的表现的。并且有了实际应用，像各类机器人产品。





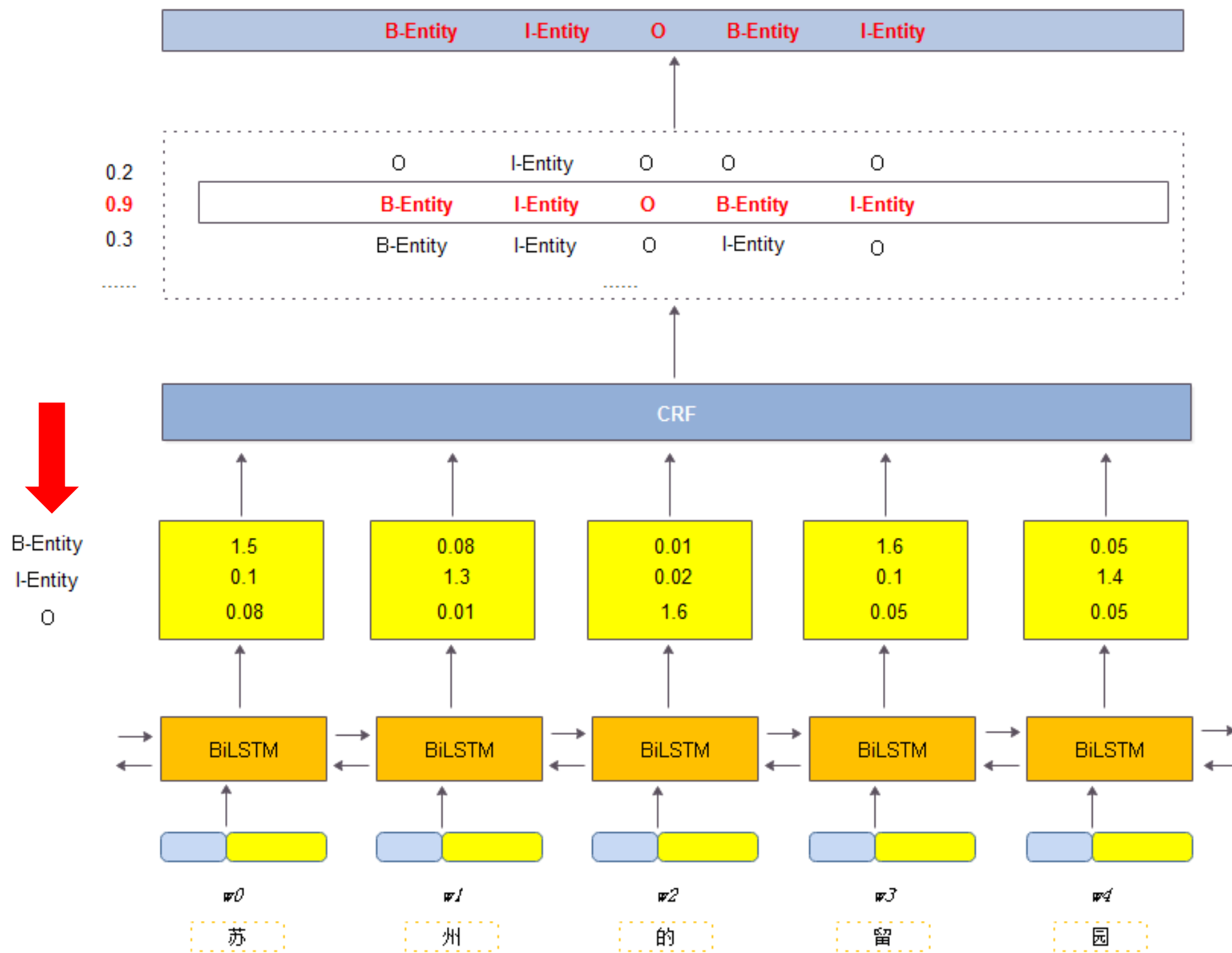


第一 0.046976987 0.009867641 -0.7570944 0.946452 0.54130673 -0.12159368 0.6204817 0.13075459 0.64436877 0.  
次 0.8586752 -0.0033830514 -0.31004173 0.5718351 0.5531515 0.0034371624 0.3425517 0.19923757 0.019328168 (  
来 0.044967934 0.4036975 0.39606753 0.1671985 0.42527685 0.3256366 0.80985934 0.14617704 0.26508832 -0.526  
宏村 0.038774718 -0.821415 -0.0627652 0.09608057 0.63518333 -0.22294678 0.0474165 0.34509218 0.39332274 0.  
。 1.0832804 -0.81887233 -0.06815572 -0.834906 -0.42175865 0.04209236 -0.8169991 -0.7914735 0.5055454 0.64  
感觉 -0.014481613 -0.6871397 -0.35476017 -0.029884212 0.9952648 -0.018171955 -0.06145633 -0.6315788 0.7571  
就 1.5487165 0.5153036 0.22363093 -0.82874006 0.50525165 -0.41960517 0.161687 -0.38740015 0.6599034 -0.464  
是 -0.7820301 0.44910857 0.5440042 0.15221055 0.7254627 0.7928974 0.4372032 -0.8809837 0.60538864 0.923375  
美 -0.43188626 -0.6928499 -0.31370726 0.17424013 0.83703595 0.2397026 0.13877438 -0.027584264 0.6411822 0.  
, 1.0589479 -0.511765 -0.30412745 -0.15296313 0.52246344 -0.3599321 -0.67532533 -1.1650944 1.0317103 1.029  
古老 0.3271905 0.15529023 0.08716673 0.27165768 0.48993617 0.10764594 0.47878364 0.16330142 0.52194715 0.5  
的 0.1948509 0.3968573 0.23971519 -0.6034616 0.05829314 0.46361104 -0.35814217 -0.39090428 0.17520912 0.57  
文化 0.16628847 -0.33042908 -0.4870814 -0.17614456 0.28886527 -0.20365526 0.22200266 0.3737608 -0.03528237  
值得 -0.8436707 -0.6734584 -0.56877106 0.6226293 0.6547742 -0.40323654 0.73107845 0.021554755 0.39337924 (  
一 0.5321799 0.14190404 -0.011887787 0.86443233 -0.6859676 -0.014601703 1.1109279 -0.53397816 0.5776924 0.  
去 0.21264657 0.5435075 -0.25712153 -0.27577817 0.02570348 0.017156696 0.58853567 -0.035688188 -0.07287674  
可以 1.1203761 0.002989558 -0.9685687 0.004069076 1.1697568 -0.05813827 0.24196921 0.16538809 -0.14073017  
不错 0.15558587 0.14841036 -0.26350752 0.28769186 1.1193488 0.4964584 0.10302042 0.77626383 0.97578144 -0.  
呦 0.23077351 -0.4871792 -0.27177277 0.40461916 0.17646761 -0.16412182 0.39146584 0.007292048 0.7026126 0.  
哈哈 -0.08473275 -0.49340478 -0.4431695 -0.0037185496 0.9926002 -0.101199254 0.42340153 -0.1260695 0.60201  
高速 0.07892663 -0.5783191 -0.39839178 0.33715907 0.64722365 0.004699737 0.0347277 -0.038368292 0.311369 (  
下来 0.13549761 -0.504659 -0.004344655 0.55801946 0.94196886 -0.40074784 0.47118944 -0.1705189 0.59298414

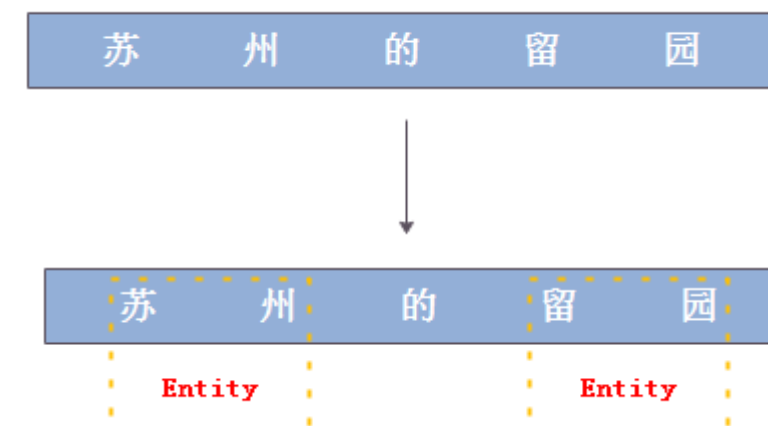




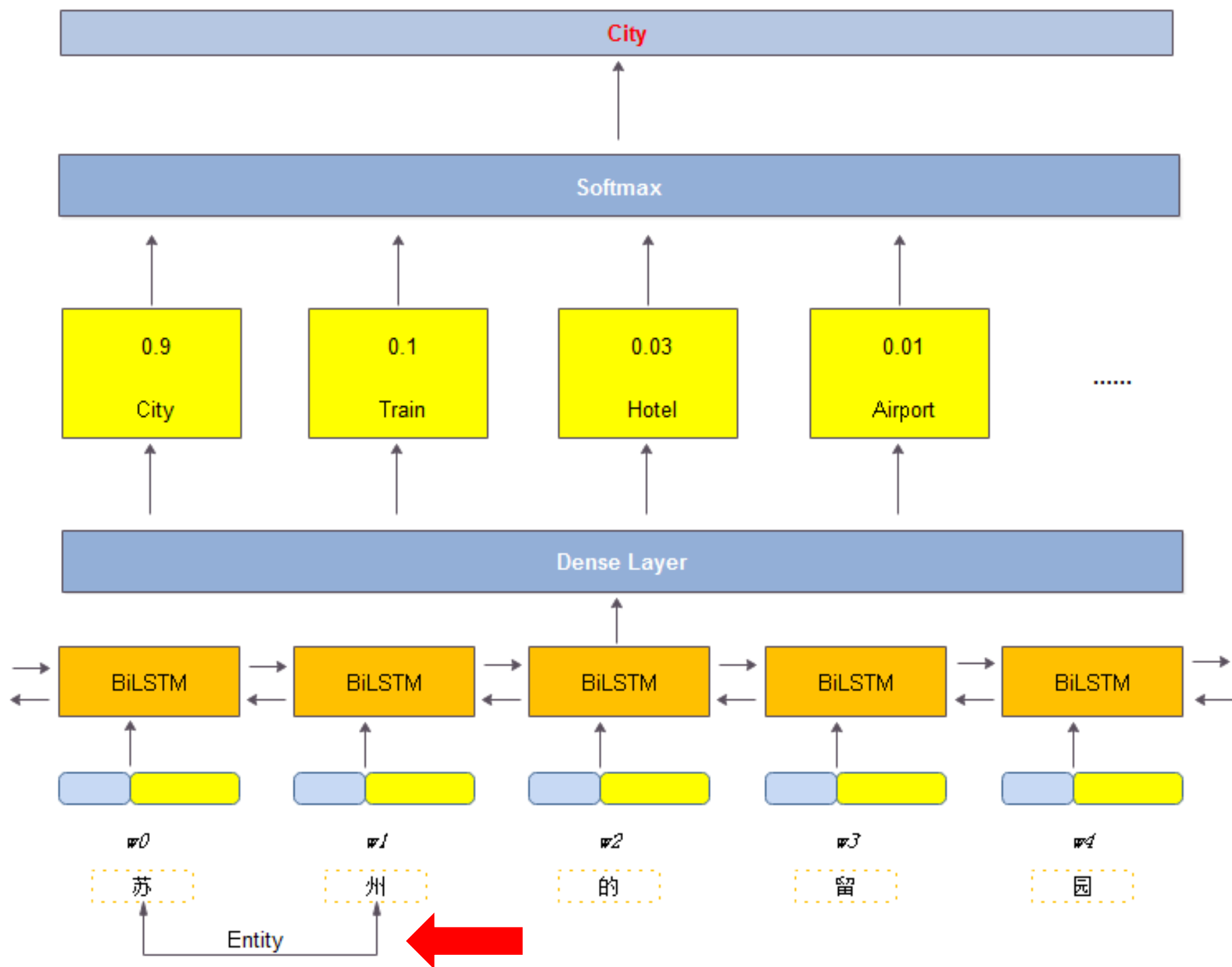
# 实体识别模块



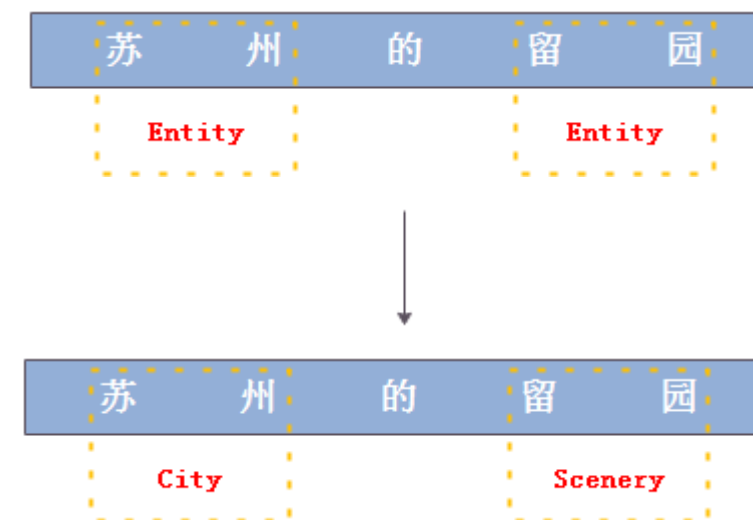
- 输入文字
- 转化词向量
- BiLSTM-CRF
- 高维实体



# 实体识别模块



- 输入高维实体
- 转化词向量
- 全链接神经网络
- 最终实体



16	到 <span>岐山</span> 的 <span>高铁票</span>
17	刚 到 <span>苏州</span>
18	<span>霸州</span> 到 <span>衡水</span>
19	<span>阜阳</span> 到 <span>武昌</span>
20	开 2336
21	祁阳 到 湛江 的 怎么说的 孕周到 湛江 的
22	永定路 汉庭 酒店
23	高铁 票 <span>虎门</span> 智慧 灯 的 <span>高铁票</span>
24	阴阳 战斗 <span>东莞</span> 司机
25	是 <span>榆树</span> 的 火车 <span>火车</span>
26	我在 飞机 的 价格 票价
27	我要 订 <span>丹东</span> 到 <span>大连</span> 的 14 号 的 票 <span>高铁票</span>
28	等 我 到 <span>桂林</span>
29	广州 南 支票 转 到 云南 富宁
30	兰州 到 金昌

欢迎使用 - 搜索数据运营平台 ^\_^

### Edit Annotation

16

17 虎门 [Link](#)

18 [Google](#), [Wikipedia](#)

19

20

21

22

23

24

25

26

27

28

29 广州 南 支票 转 到 云南 富宁

30 兰州 到 金昌

**Entity type**

- ☐ brand
- ☒ hoteltype
- ☒ hotel-star
- ☒ hotel-service
- ☒ landmark
- ☒ cbd
- ☒ common
- ☒ business suffix
- ☒ trip way
- ☒ around
- ☒ date
- ☒ flightno
- ☒ trainno

[Add Frag.](#) [Delete](#) [Move](#) [OK](#) [Cancel](#)

纠错前



同程



友商1



友商2

纠错后

19:46

上海到昆明

以下为您显示"上海到昆明"的相关搜索结果:

上海到昆明

周边跟团游跟团游自由行目的地参团定制旅行

昆明

昆明长水...昆明南站昆明站昆明火车站滇池昆明西部...翠湖公园东风广场

上海到昆明

火车票机票汽车票

上海到昆明

【国庆节】【亲子游】昆明+西双版纳+野象谷+中科院热带植物园+抚仙湖+洱海+大理+丽江双飞7日自由行航班自选昆明进丽江出昆明2晚大理2晚昆明出发跟团游

¥ 3784起

【中秋节】【亲子游】昆明+原始森林公园+西双版纳+抚仙湖+野象谷+大理+丽江双飞7日自由行航班自选昆明进丽江出昆明2晚大理2晚昆明出发跟团游

¥ 3727起

昆明+大理+丽江双飞7日自由行航班自选昆明进丽江出昆明2晚大理2晚昆明出发自由行

¥ 2760起

同程

19:46

上海到昆明

上海

上海欢乐谷上海一日游上海的酒店上海上海的景点上海上海的旅游产品上海上海温泉度假上海朋浩公寓上海上海到坎昆的旅游产品上海迪士尼乐园上海上海骏朋酒店上海上海卢宏朋公寓上海上海的福朋喜来登酒店上海上海到泰国的旅游产品上海到日本的旅游产品

友商1

19:46

上海到昆明

苏州到上海

上海到喀什市的机票上海到马拉喀什的机票上海·中国2015上海到喀什叶城全程骑行8700公里平安完成上海大观园上海欢乐谷上海玛雅海滩水公园【D21-D23】上海新疆33天自驾行记：喀什到叶城再到民丰，上天...上海到普陀山2日往返大巴票上海到普陀山巴士观光2日大巴票上海站上海上海海上铂悦酒店中国，上海浦东机场核心上海宾馆中国上海静安寺南京西路

友商2

n-gram模型

在中文错别字查错情景中，我们判断一个句子是否合法可以通过计算它的概率，假设一个句子 $S = \{w_1, w_2, \dots, w_n\}$ ，则可以转换成如下形式：

$$P(s) = P(w_1, w_2, \dots, w_n) = P(w_1) * P(w_2 | w_1) * \dots * P(w_n | w_1, w_2, \dots, w_{n-1})$$

Bigram

$$P(w_i | w_{i-1}) = count(w_i, w_{i-1}) / count(w_{i-1})$$

count (wi)

上	海	到	昆	明
2533	927	2417	746	158

count (wi, wi-1)

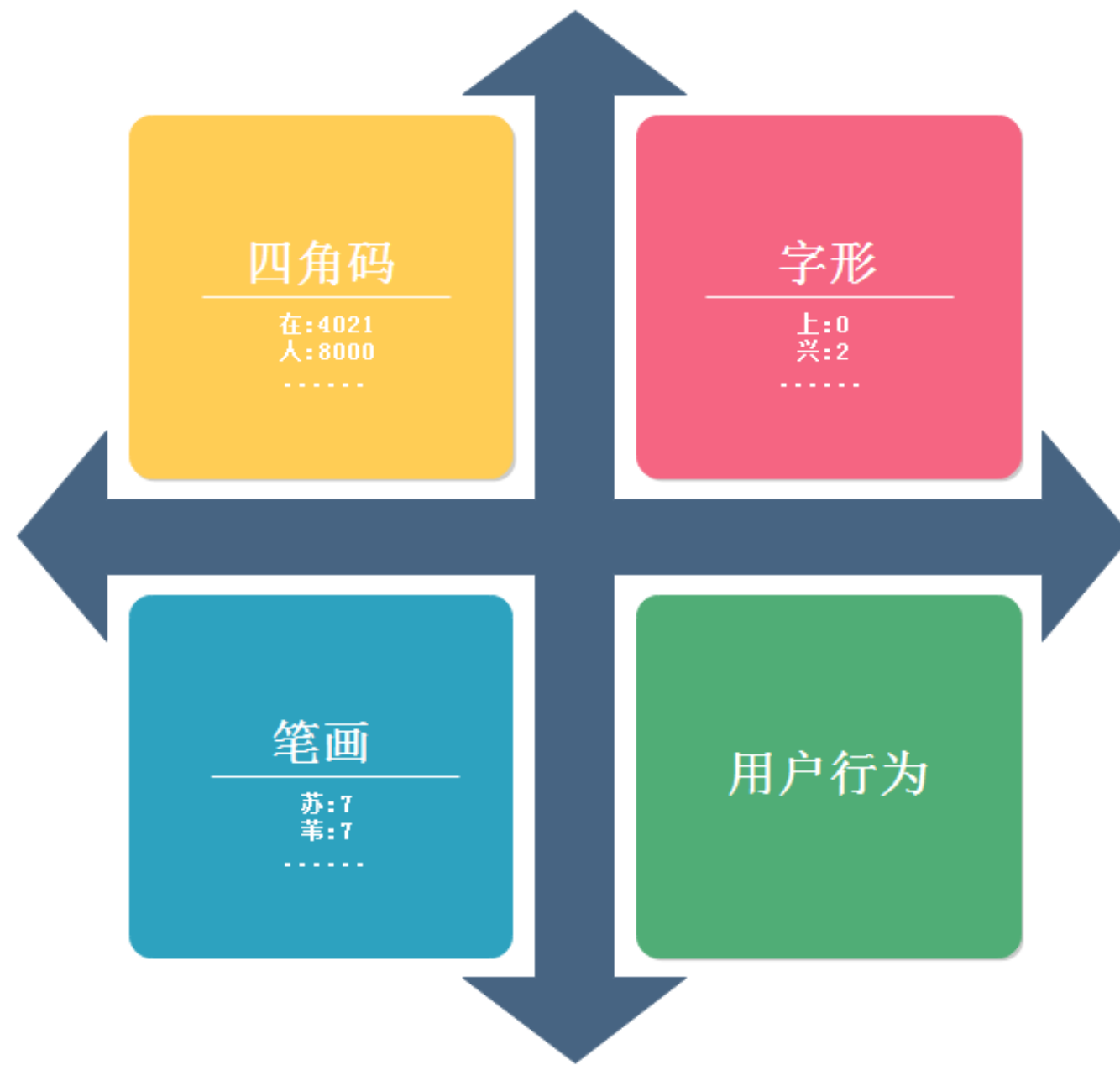
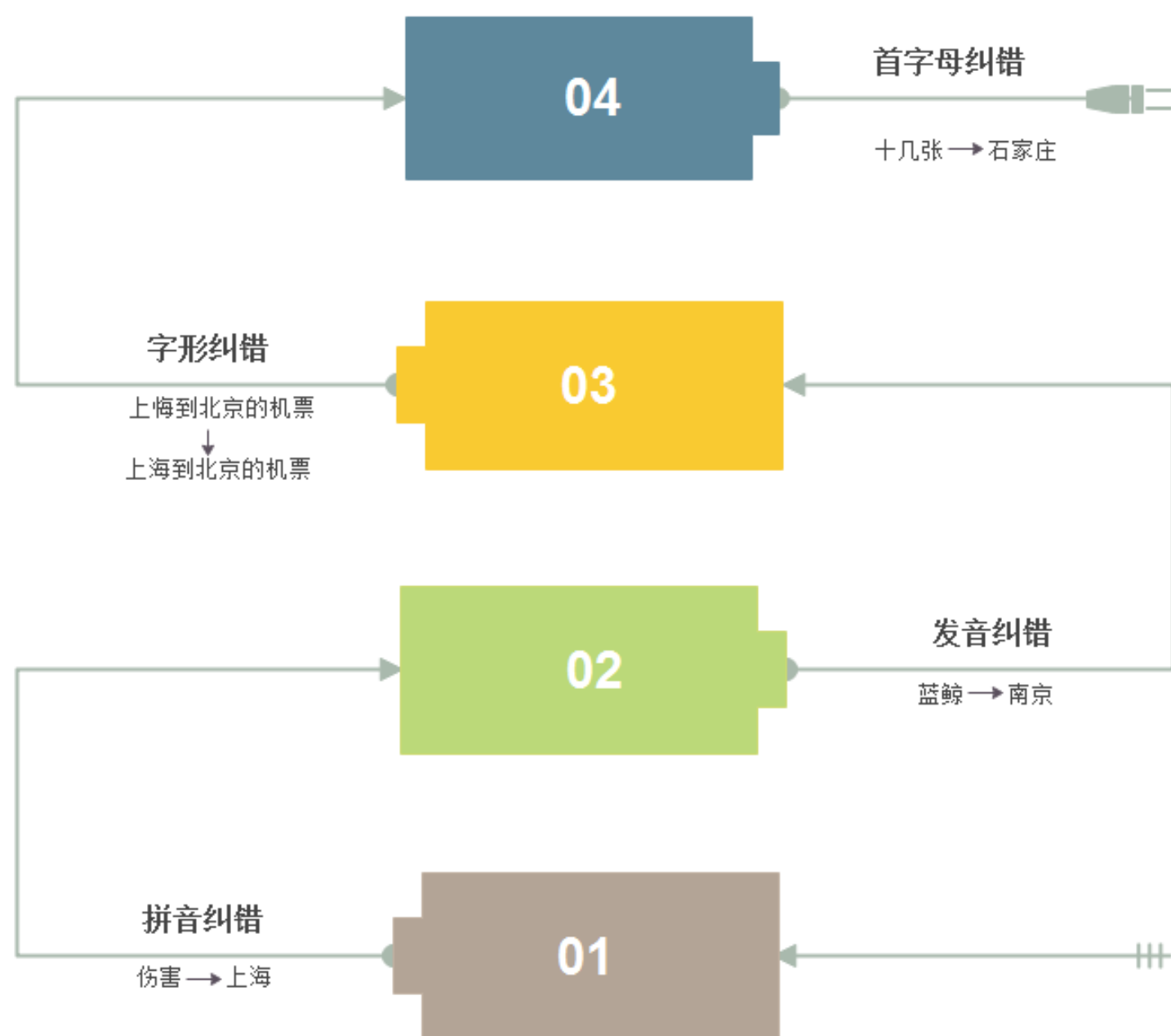
<div>wi-1 \ wi</div>	上	海	到	昆	明
上	5	827	0	9	0
海	2	0	608	1	6
到	2	0	4	686	2
昆	0	0	2	0	16
明	1	0	0	0	0

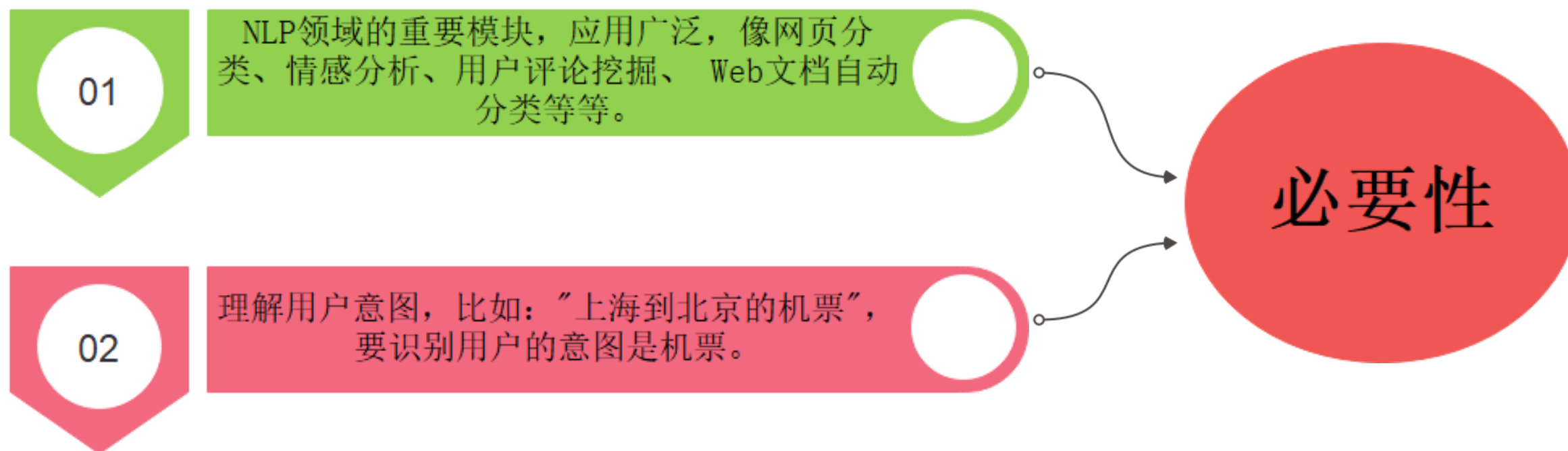
则*Bigram*概率矩阵计算如下：

<div>wi-1 \ wi</div>	上	海	到	昆	明
上	0.002	0.33	0	0.0036	0
海	0.0022	0	0.66	0.0011	0.0065
到	0.00083	0	0.0017	0.28	0.00083
昆	0	0	0.0027	0	0.021
明	0.0063	0	0	0	0

P(上海到昆明)=P(上)\*P(海 | 上)\*P(到 | 海)\*P(昆 | 到)\*P(明 | 昆)

=(2533/6781)\*0.33\*0.66\*0.28\*0.021

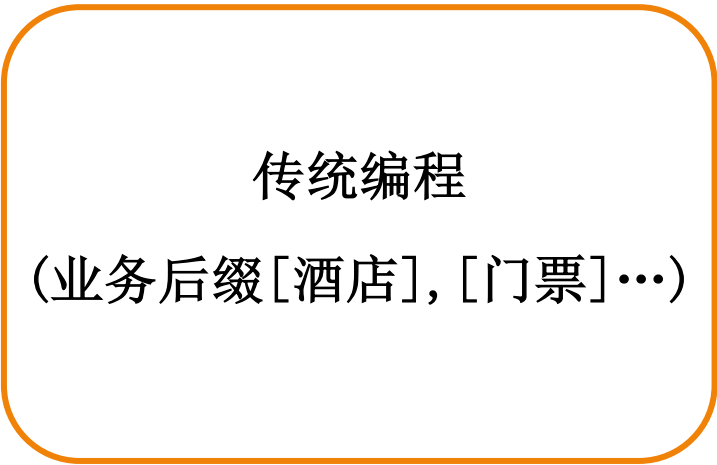






特征：X

分类规则



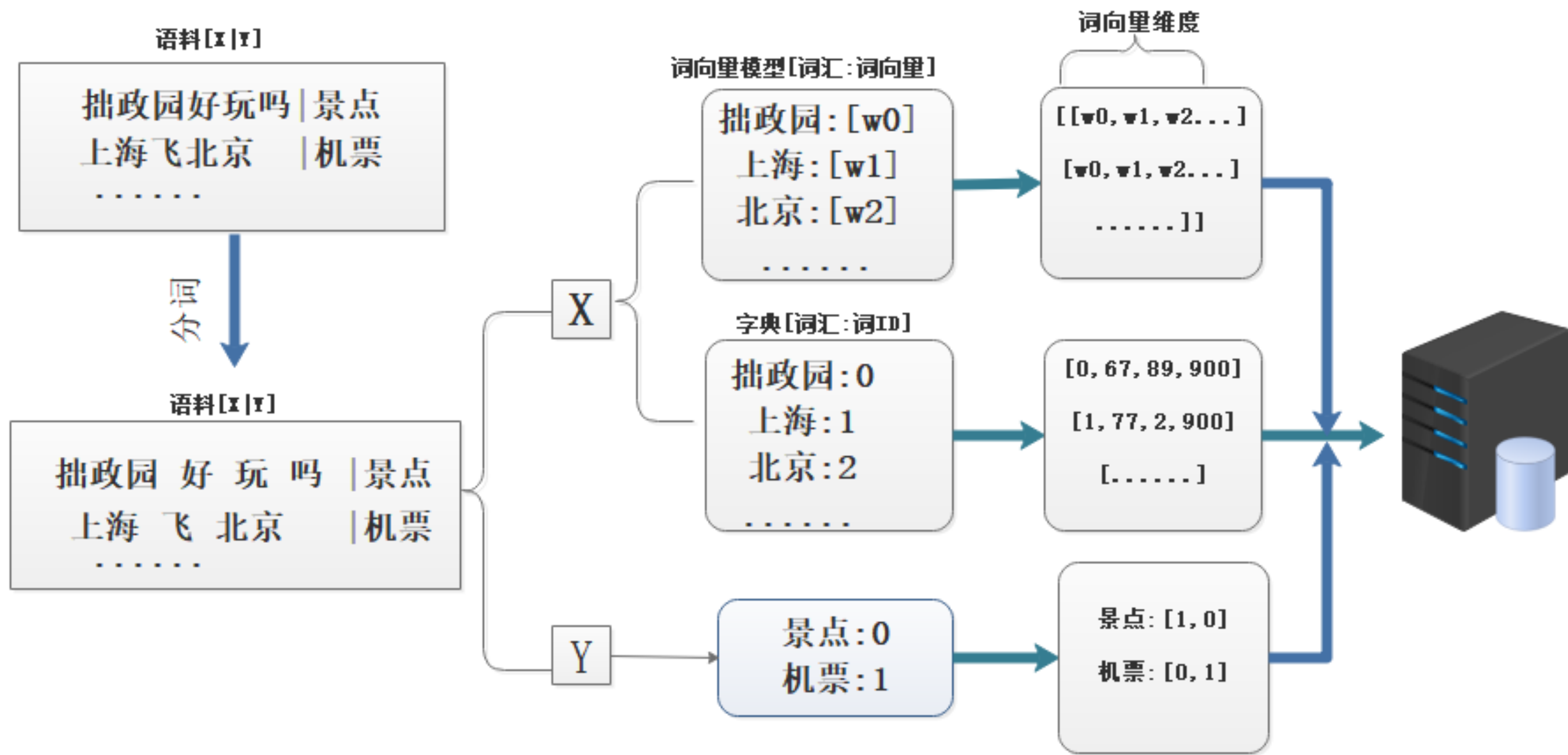
分类标签：Y

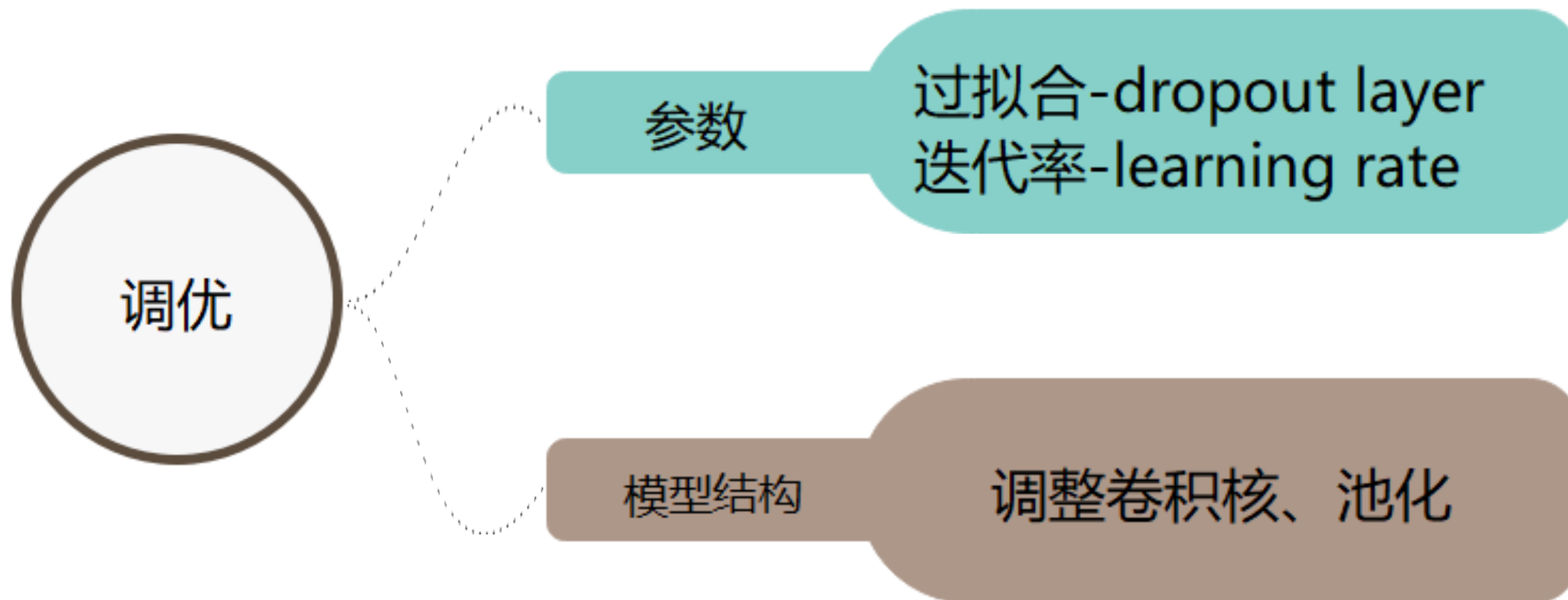
特征：X

分类标签：Y

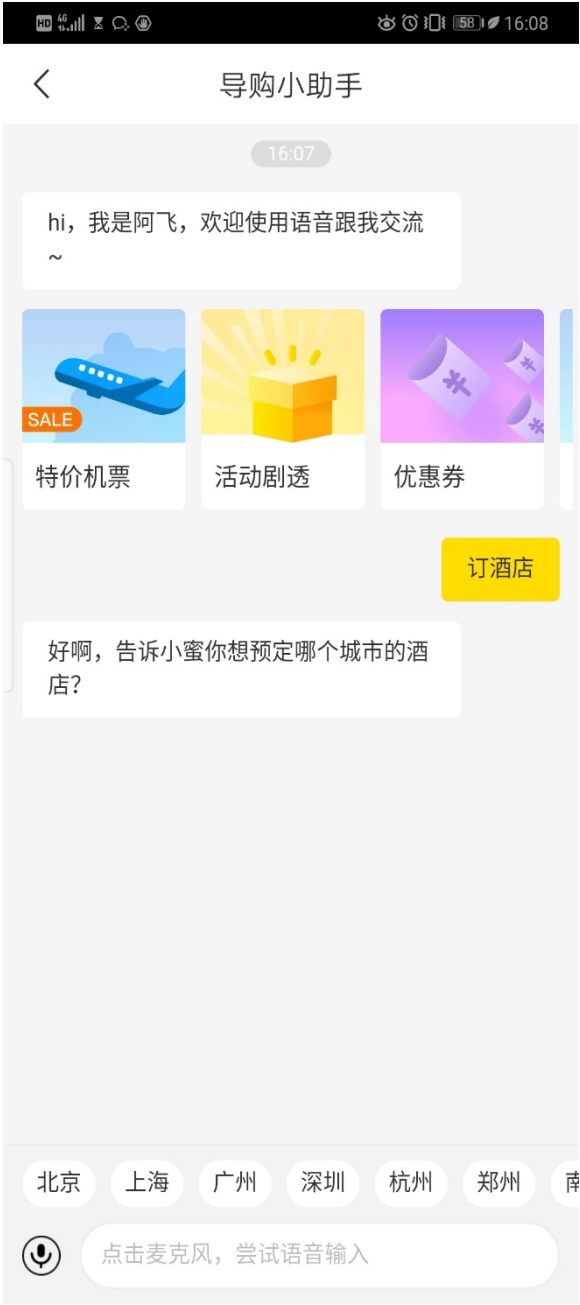


分类模型





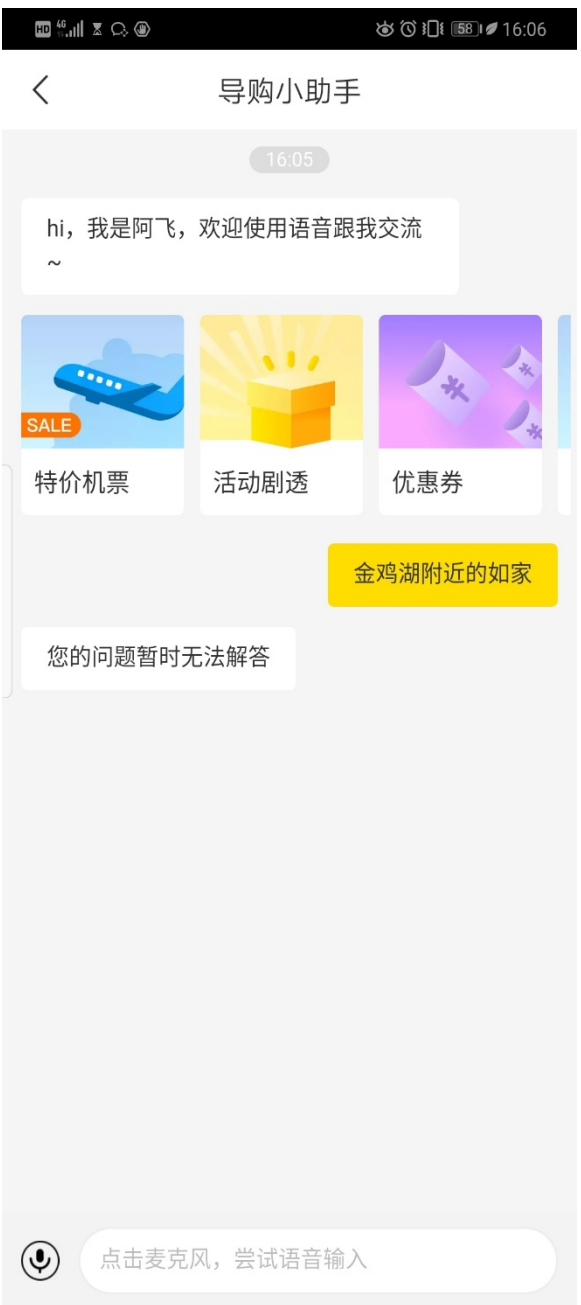
# 项目收益



友商



同程小程序

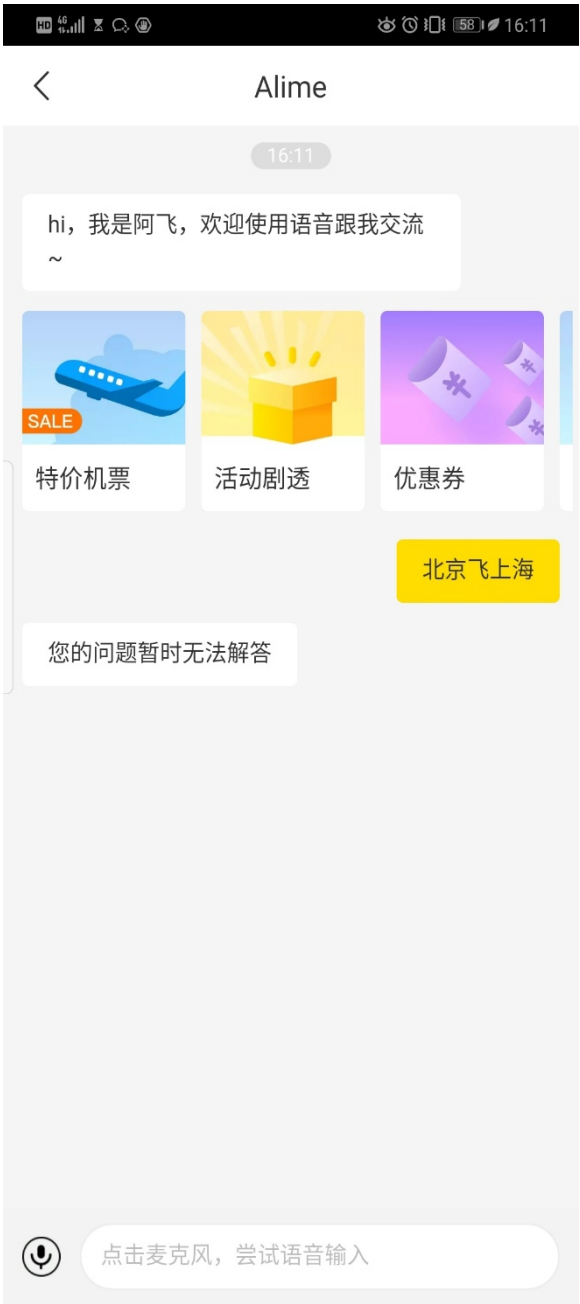


友商



同程小程序

# 项目收益



友商



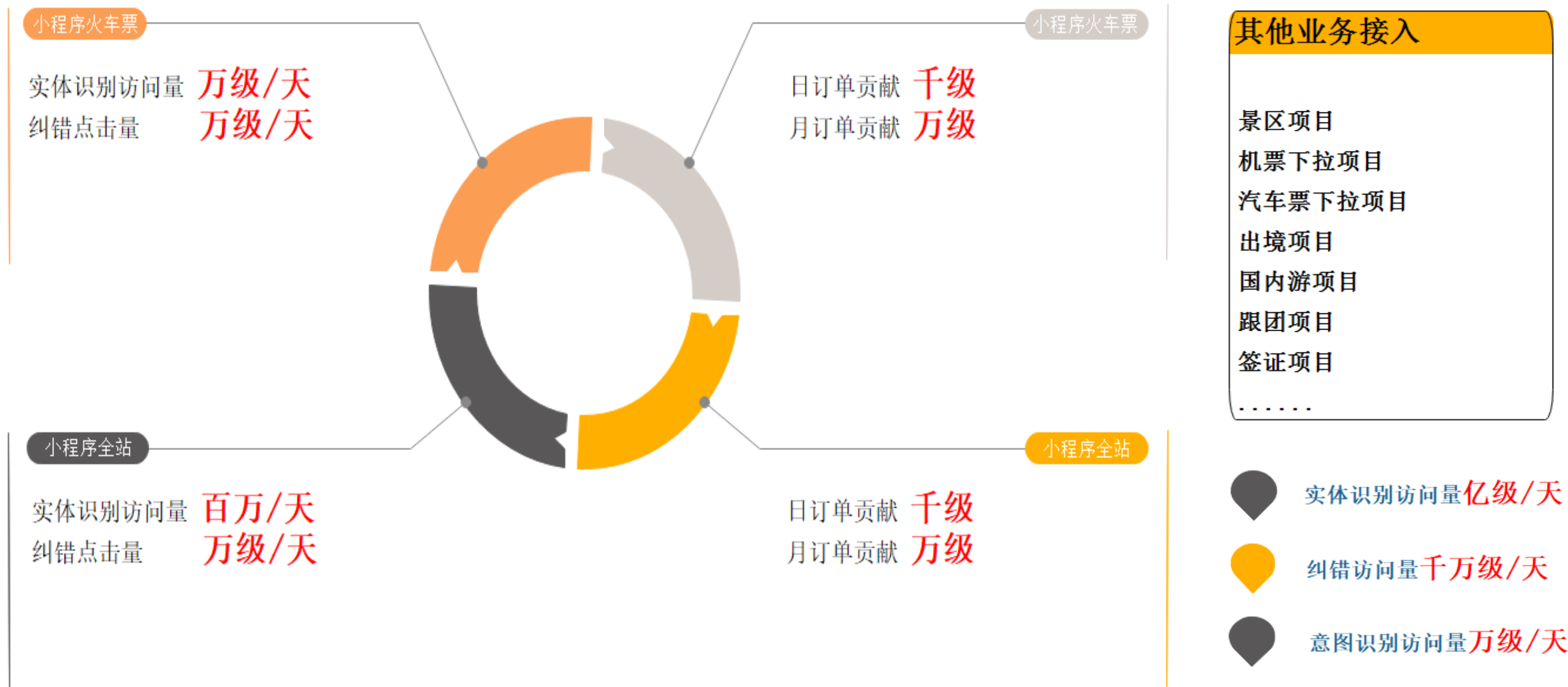
同程小程序



友商



同程小程序



01

海量用户咨询数据自动分类

02

推荐理由生成

03

商品评价的语义理解

01

提升NLP技术的掌握能力

02

功能模块的完善

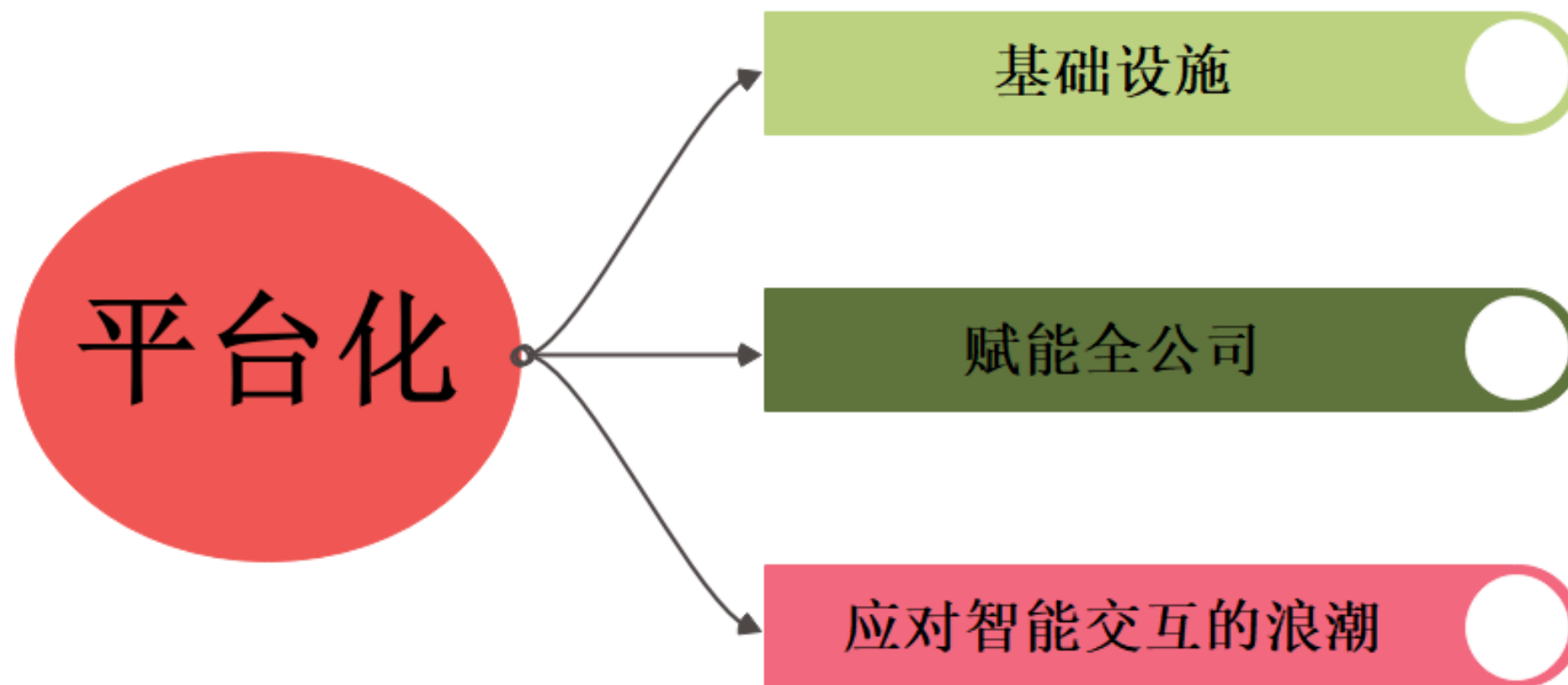
03

搜集更多真实的、有效的数据

04

NLP技术的升级迭代





在各种智能设备越来越盛行的趋势下，NLP技术是其人机对话最关键最核心的技术，各大互联网公司纷纷在储备NLP方面的能力，而作为公司NLP技术的先行者，该系统不使用第三方平台，自行研究储备NLP能力，很好的支持了语音搜索、搜索机器人等智能化产品的落地应用，并且可以支持公司所有需要使用NLP能力的业务，技术先行，助力公司实现从OTA向ITA的转变。

谢谢观赏