

意图识别

Query意图识别用于帮助理解用户在提交查询时的真正需求或目标，即将用户模糊或不完整的查询映射为系统能够理解的需求类型。意图识别通常依赖和服务于搜索相关产品，如：交易查询意图（如购买商品、预订服务等）、本地查询意图（如搜索附近的餐馆、电影院等）。本章将介绍常见的三种意图：

- **时效性意图**
 - 定义：用户的查询与当前时间或特定的时间范围密切相关，强调数据的实时性
 - 产品策略：在结果中优先展示最新内容
- **本地化意图**
 - 定义：用户希望获取与特定地理位置相关的信息或服务，侧重地理位置的关联性
 - 产品策略：在结果中优先展示距离最近的选项或提供地图视图
- **用户意图**
 - 定义：用户的查询直接涉及特定用户名、账号或社交媒体身份，关注个体身份或社交信息
 - 产品策略：在结果中展示相关用户信息内容

一、时效性意图

时效性意图通常有以下两个特点：

- 查询中通常包含时间相关的词语，例如“今天”、“最新”、“最近”等
- 涉及动态更新的信息，如新闻、事件、天气、股价、体育比分等

根据Query的时效性场景可以划分为三大类：

- **强时效性**
 - 定义
 - 突发事件：突发的新闻事件（重大事故、灾害、政治突发事件等）
 - 热门话题：短期内被广泛讨论的事件或话题（热门社交话题、体育赛事、娱乐事件等）
 - 语义时效性：周期性更新的信息，或者是希望获取最新版本的内容（天气预报、股价、节假日活动等）
 - 搜索结果策略：保证头部曝光结果中新Doc的占比
- **一般时效性**
 - 定义：查询与时间相关，但对实时性要求不高，允许在一定时间范围内满足需求
 - 搜索结果策略：提升搜索结果中新Doc的权重
- **无时效性**
 - 定义：查询与时间无关，不关注特定时间范围的信息

- 搜索结果策略：推荐权威或高质量的结果，结果中适当控制旧Doc的占比

时效性基础信号

通过建设时效性基础信号，识别查询Query和Doc的时效性，有效提升时效性意图搜索结果。对于突发热点事件需要建设热点事件识别流程以帮助判断Query是否指向热点事件。此外，对Query时效性区分通常借助语言模型和搜索后验进行识别判断。

热点事件识别

对于突发热点事件需要建立热点数据反馈机制，并据此判断检索词Query是否与热点内容相关：

1. 通过站内站外两个渠道来源搜集和挖掘热点数据
2. 将热点数据写入索引，构建实时更新的召回通道，并根据热点内容的时效衰减变化动态维护索引
3. 在线将Query和召回的热点内容进行相关性匹配计算，截断后综合热点发布时间排序
4. 根据排序后首条结果是否满足准出条件（与Query的相关性），判定Query是否为热点事件

Query时效性识别

Query时效性识别通常依赖基于语言模型的语义时效性打分以及搜索后验数据综合判断。

语义时效性打分

语义时效性分数采用BERT模型对Query进行打分，模型输入为Query，输出为时效性分数。

在训练样本的构建上，除了人工标注数据，可以引入自动样本进行扩量：

- **基于点击时新Doc的挖掘：** 挖掘给定时间周期内，点击高时效性Doc占比超过 $x\%$ 的Query作为正样本，反之作为负样本
- **基于共点Doc的挖掘：** 基于已标注的强置信的Query，通过Query-Doc点击二部图扩充该类别的Query

Query时效性分类

Query时效性分类模型的输入为语义时效性打分和搜索行为统计特征：

特征
语义时效性打分
关键词匹配
Query 曝光/点击量
Query 类别/类目
Query 点击 Doc 在不同发布时间下的占比（发布于 7 天/30 天/半年/1 年内）

时效性信号应用

对于识别为强时效性的Query，需要在搜索结果中优先展示新Doc，同时需要保证头部曝光结果中的新Doc占比。具体的，可以对满足时效性要求的Doc在保证召回（如：增加时效性Doc召回通路）的同时，在排序阶段对新Doc的权重进行提升。对于突发热点数据可以进一步通过干预的方式保证结果在结果首位展示。

二、本地化意图

本地化意图是指用户在搜索中希望获取与某一地理位置直接相关的信息，如：

- 查询附近的服务或设施（如“附近的咖啡馆”）
- 查询特定地点的活动或条件（如“北京明天天气”）
- 查找以地理位置为上下文的资讯或内容（如“上海有哪些好玩的地方”）

本地意图类别

用户可能明确或隐式表达位置需求：

- **显式位置意图**
 - 用户明确在Query中指出地理位置，通常包括地名、地址或地标
 - 示例：
 - “上海的五星级酒店”
 - “纽约时报广场附近的餐馆”
 - “东京迪士尼攻略”
- **隐式位置意图**
 - 用户未在Query中明确提及位置，但通过词语（如“附近”、“周围”）或上下文推断出位置相关性
 - 示例：
 - “附近的加油站”
 - “周边好吃的火锅店”
 - “离我最近的地铁站”

本地意图目标

本地意图需要识别出有位置需求的Query，并对Query位置解析，将其与地理位置信息进行关联，从而实现对特定位置的信息检索。

- **意图分类**
 - 定义：本地化意图可以进一步细分为精搜、泛搜等维度
 - **精搜意图**：用户明确地表达了具体的位置需求或服务目标，希望直接找到精确的答案或匹配的结果。需要重点提取具体的地理位置或服务类别，结合品牌关键词进行高精度匹配
 - **泛搜意图**：用户需求较为模糊，通常是为了探索某个区域或服务类别的信息，没有明确的具体目标。强调范围和多样化扩展，结合位置上下文和用户偏好优化结果
 - 示例：
 - 精搜意图：上海新天地

- 泛搜意图：上海美食

- 位置解析

- 定义：从Query中提取地理实体（城市、POI等），推断位置范围需求（指定区域范围限定）
- 示例：Query = “北京故宫附近美食”，抽取得到城市 = “北京”，POI = “故宫”

- 多义性处理

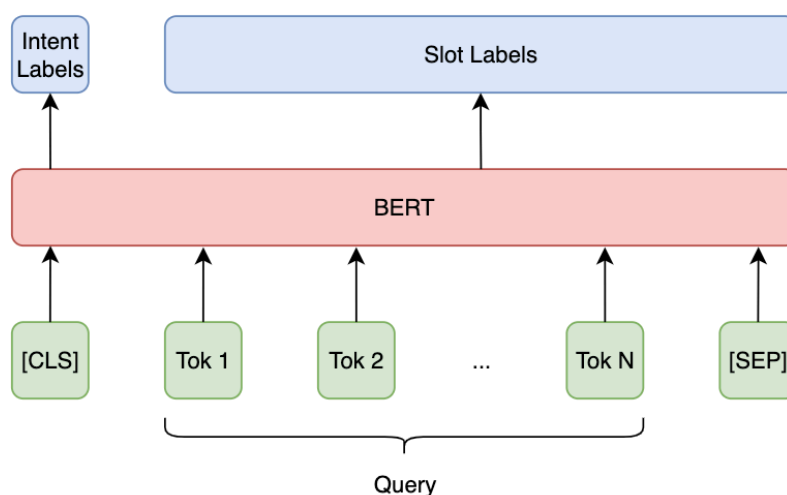
- 定义：一些地名可能存在歧义，需要结合用户实时地理位置和上下文解析
- 示例：Query = “人民公园地铁站”，有很多城市有人民公园，需要结合具体的用户位置信息进行解析

- 多意图处理

- 定义：在查询中可能同时隐含多个位置或需求层次，导致搜索系统无法直接确定用户的目标位置或意图方向
- 示例：在上海搜“北京烤鸭”
 - 改场景包含两种位置需求
 - 基于目标地点的内容需求（北京的烤鸭介绍或餐馆）
 - 基于当前地点的服务需求（上海提供北京烤鸭的餐馆）
 - 结合位置和上下文推测用户意图
 - 如果用户近期多次搜索“北京旅游攻略”，其意图更可能是查询北京本地信息
 - 如果用户最近搜索了“附近的餐厅”，则可能是希望找到上海的北京烤鸭店
 - 若查询发生在用餐时间，系统可倾向推测用户寻找上海的餐馆

本地意图模型

由于本地意图目标不仅需要实现意图识别，还需要完成位置解析，在模型可以选择基于BERT的意图识别和槽位填充模型的联合模型。



意图识别和槽位填充本质上分别为一个句子分类任务和一个序列标注任务，结合两者的loss即可实现多任务学习。给定Query文本，插入一个特殊的分类嵌入（[CLS]）作为第一个标记，并添加一个特殊的标记（[SEP]）作为最后一个标记。则输入序列为 $x = (x_1, \dots, x_T)$ ，BERT输出序列为 $H = (h_1, \dots, h_T)$ ：

- 意图分类

- 基于第一个特殊标记 ([CLS]) 的隐藏状态 h_1 , 学习目标函数为:

$$y^i = \text{softmax}(W^i h_1 + b^i) \quad (1)$$

- 槽位填充

- 基于隐藏状态 h_1, \dots, h_T , 学习目标函数为:

$$y_n^s = \text{softmax}(W^s h_n + b^s), n \in 1 \dots N \quad (2)$$

- 联合建模

- 目标函数为:

$$p(y^i, y^s | x) = p(y^i | x) \prod_{n=1}^N p(y_n^s | x) \quad (3)$$

- 学习目标是最大化条件概率:

$$p(y^i, y^s | x) \quad (4)$$

- 意图识别和槽位填充损失函数都采用交叉熵损失函数

三、用户意图

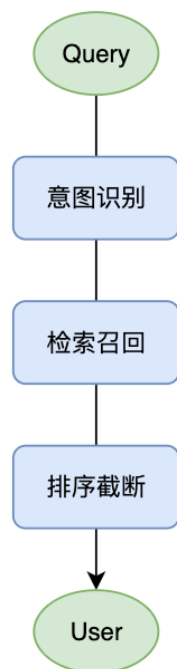
在一些有社交属性或UGC平台中, 用户通常有涉及用户名称、用户账号的检索需求, 检索内容一般为用户名、ID、昵称、标签等。这类需求往往旨在查找特定用户的信息、动态、或与之相关的内容, 以满足社交互动、内容消费、或平台探索的目的。

用户意图的产品形态是基于用户查询行为和需求设计的功能模块或交互界面, 通过解析用户意图并提供相关服务, 提升用户体验和检索效率。

除了单独给用户检索的搜索界面外, 一般在主界面的综合通搜页面内, 对于有明确用户意图的检索词也会触发如小窗口之类的交互界面以展示意图相关的用户信息。

用户意图识别链路

用户意图模块的目标是识别有用户意图的Query并返回对应用户信息, 常规的链路设计如下:



用户意图识别

由于用户意图Query在通用场景下流量占比通常较低，为了尽可能降低算力的消耗，在链路上一般会经由意图识别模块将有用户意图的Query筛选出来。而由于用户名、用户ID的文本构造通常有一定的规则，可以采用基于TextCNN、BERT等语言模型做分类判断。

训练数据的正样本的挖掘上可采用以下方案：

1. 百科人物数据
2. 站内用户名称/昵称（满足高活跃度、高粉丝量等）
3. 用户检索界面内历史高频检索词
4. 综合检索主界面历史触发交互窗口的检索词

用户多路召回

对于触发用户意图识别的Query需要进一步通过召回检索模块实现相关用户的候选召回，召回算法可以是基于向量表征的向量召回、基于倒排的文本召回、基于词表的KV召回等。

为了尽可能的保证召回率，通常会根据索引的内容不同而涉及多路召回通道：

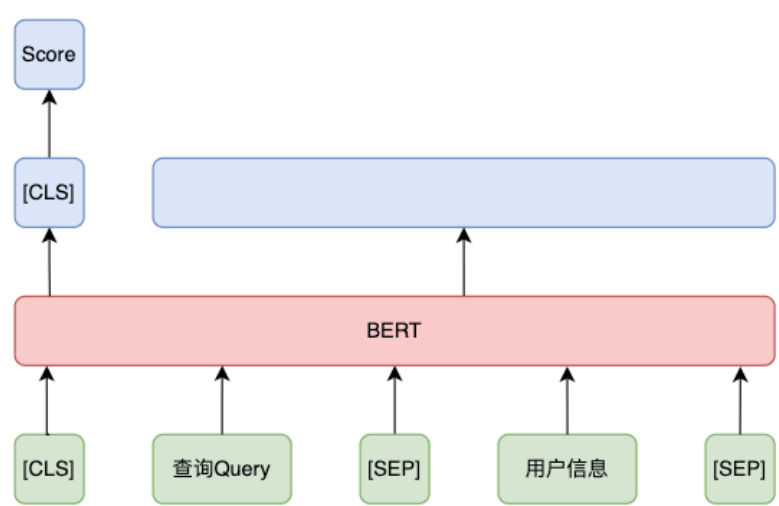
1. 用户名召回通道
 1. 昵称召回
 2. 拼音召回
 3. 改写召回
 4. 向量召回
2. 用户ID召回通道
3. 标签/领域召回通道
4. 关注列表召回通道

用户排序截断

对于召回的候选用户需要经过相关性判别和排序，去除低匹配相关用户和重名的低质用户（仿号、低活跃度用户等）。

相关性判别

采用BERT模型，输入为Query和候选用户信息（用户名/昵称/标签等信息），输出为相关性分数：



排序截断

对于满足相关性的候选用户，会通过结合后验行为特征的排序模型筛选出TopN作为最终检索结果展示。

排序模型常见特征如下：

特征	类型
Query 长度	Query 文本特征
Query 是否纯数字/英文	
Query 是否包含数字/英文/标点/表情符	
Query 包含的中文长度	
Query 意图/类目/NER	
Query 在综合搜索页面搜索曝光量/点击量	Query 统计特征
Query 在用户搜索页面搜索曝光量/点击量	
Query 历史触发用户交互窗口曝光量/点击量/点击率	
Query 在综合搜索页面触发用户交互窗口曝光量/曝光率	
Query 在综合搜索页面搜索后切换用户搜索的数量/比率	
Query-用户信息 相关性分数	Query-User 统计特征
Query-用户名 序列相似度 (LCS、Jaccard、SimHash、Levenshtein)	
Query-User 曝光量/点击量/点击占比	
用户信息（好友量/粉丝量/标签/平台认证等）	

总结

意图识别是用户检索体验中的核心环节，通过分析用户的查询内容和上下文，精准判断用户的真实需求，为后续的结果召回、排序和推荐提供有力支持。围绕不同场景，意图识别有多种维度和类型，其复杂性与丰富性决定了平台对用户需求的满足程度和交互质量。

意图识别是检索技术与用户体验的结合点，其复杂性与灵活性决定了平台在多样化需求中的竞争力。无论是时效性、本地化还是用户名意图，理解用户真实需求并精准满足，有助于提升用户对平台的信任感和依赖度，增强用户粘性。总之，理解并满足用户需求始终是产品优化和技术创新的核心方向。