# 基于查询表达式特征的时态意图识别研究*

桂思思[1,2]　陆　伟[3]　张晓娟[4]
[1](武汉大学信息管理学院　武汉 430072)
[2](武汉大学信息检索与知识挖掘研究所　武汉 430072)
[3](武汉大学信息资源研究中心　武汉 430072)
[4](西南大学计算机与信息科学学院　重庆 400715)

**摘要:** 【**目的**】针对时态意图识别问题, 探讨可抽取查询表达式特征的有效性及采用不同类别分类算法的识别准确度, 为后续相关研究提供一定的借鉴。【**方法**】按查询表达式特征与时间的关联性, 将其归类为时间无关特征、潜在时间特征、显式时间特征。在此基础上, 分别采用有监督分类算法及半监督分类算法, 探讨采用不同特征组合的有效性及不同分类算法的识别准确度。【**结果**】在抽取的三类查询表达式特征中, 仅使用显式时间特征的平均分类准确率最高, 且"查询是否包含年份"这一特征为强特征; 使用不同分类算法的识别准确度相差不大; 时态意图识别结果优于已有参与时态意图分类子任务(TQIC)测评的成果, 平均分类准确率为 81.14%。【**局限**】限于数据集的获取途径, 仅对 300 条查询的时态意图识别效果进行验证; 仅考虑已有的查询表达式特征, 未提出用于时态意图识别的新特征。【**结论**】查询表达式特征中与时间关联性高的特征能提高时态意图识别准确度, 而基于统计的特征(如查询词长度)对时态意图识别分类准确度的提升效果不明显。

**关键词:** 时态意图　有监督分类　半监督分类　特征抽取
**分类号:** G354
**DOI**: 10.11925/infotech.2096-3467.2018.0550

## 1　引　言

查询意图常被定义为用户通过查询表达式(即查询)而表达出的用户信息需求[1-2]。搜索引擎允许用户输入的关键词个数有限, 导致查询表达式不能完整描述用户的信息需求。因此, 理解查询意图、返回与用户信息需求相关的信息, 成为提高搜索引擎检索效率的主要途径之一。文献[3-5]表明, 时态是理解查询意图的一个重要维度。查询时态意图(简称时态意图)主要指用户通过查询语句所表达出来的对检索结果所属时间段的需求[4-5], 如查询"Olympics 2008(奥运会 2008)"表明用户想获得在 2008 年举行的奥运会的相关信息; 查询"Einstein early life(爱因斯坦 童年)"主要想获取爱因斯坦童年时期的故事、经历等。Campos 等[6]发现在 AOL 查询日志的样本中, 具有时态意图的查询占比约 8.21%, 因此具有时态意图的查询在一定程度上会影响搜索引擎的检索质量。在 2013 年–2016 年间, 信息检索领域的一些重要测评会议(如 SemEval、TREC 或 NTCIR 等)也对时态信息检索相关研究给予了高度重视, 然而在此之后, 时态信息检索研究成果较少, 且外文成果多于中文成果[7]。时态意图识别是时态信息检索研究中的一个基础问题, 旨在判断用户提交某查询后是否想获得某特定时间段的信息[7], 其结果有助于搜索引擎根据用户时态需求返回更精确的检索结果, 如在主题同等相关的条件下, 返回更新的文档(近因敏感排序, Recency-based Ranking)或为查询返回不同时间段的文档(时间依赖性排序, Time-dependent Ranking)。

通讯作者: 张晓娟, ORCID: 0000-0002-5889-5922, E-mail: zhangxiaojuan624@gmail.com。
*本文系国家社会科学基金青年项目"融合用户个性化与实时性意图的查询推荐模型研究"(项目编号: 15 CT Q019)的研究成果之一。

时态意图识别研究的主要方法是基于给定时态意图分类体系的查询分类研究[7], 即通过特征选择及模型训练, 将查询分类至给定分类体系中的某一个类别。常见的时态意图分类体系包括 Jones 等[8]提出的体系和 TQIC 体系[9]。大多数查询意图识别研究是基于 TQIC 体系展开的, 该体系由 NTCIR 测评会议中的时态意图分类子任务(Temporal Query Intent Classification, TQIC)①首次提出, 将时态意图分为 4 类[9]:

(1) 过去: 查询与过去相关的资源, 用户所需的查询结果不随查询时间而改变;

(2) 现在: 查询与现在相关的资源, 用户所需的查询结果及时更新, 随查询时间而改变;

(3) 未来: 查询与未来相关的资源, 如预测或预定的事件等;

(4) 与时间无关(简称为无关): 查询不具备上述时态意图, 用户所需的查询结果与时间无关。

基于该分类体系, TQIC 还发布测评数据, 为时态意图识别研究提供统一评测平台, 从而为时态意图查询的相关研究奠定基础。

基于 TQIC 体系的时态意图识别的分类特征主要来自查询表达式与伪相关文档集合。相比伪相关文档集合, 查询表达式是用户信息需求的一种表达形式[10], 且基于查询表达式的文本处理与特征抽取更为简易; 文献[9]表明, 在查询表达式特征的基础之上考虑伪相关文档集合特征(如 n-gram、发布时间、包含的时间信息等), 无论采用单一分类器, 或以投票思想聚合多个基分类器, 均难以提高仅使用查询表达式特征的时态意图识别准确度。由此可见, 查询表达式是时态意图识别的理想特征来源, 然而, 现有研究存在如下两个问题: 仅分别汇报各研究中所选取的查询表达特征及利用所选取特征实现时态意图识别的最终结果, 而未探讨所选取查询表达式特征对时态意图识别的有效性; 多数研究只采用有监督分类算法, 未对比分析采用相同特征时不同分类算法的识别准确度。

鉴于此, 本文对基于查询表达式的时态意图识别研究进行归纳总结, 将从查询表达式中抽取的特征按照其与时间的关联性归为与时间无关特征、潜在时间特征与显式时间特征三类; 在此基础上, 基于 TQIC 的时态意图分类体系(过去类、现在类、未来类、无关类), 采用有监督分类算法及半监督分类算法, 探讨不同类别的特征组合进行时态意图识别的有效性及采用不同分类算法的识别准确度。

## 2　相关研究

### 2.1　基于查询表达式特征的时态意图识别

该类研究只考虑查询表达式本身的特征, 主要分为与时间无关的特征(例如查询长度)和与时间相关的特征(例如动词时态与时间表达式等)。

Yu 等[11]抽取查询包含的时间与查询提交时间的时间差、核心动词的时态及实体特征三类查询表达式特征, 分别使用有监督分类算法(逻辑回归)与半监督分类算法(线性回归)实现时态意图识别。实验表明, 相比半监督分类算法, 使用全部特征的逻辑回归算法平均分类准确度在已有研究中最优。Shah 等[12]提出查询长度、查询中动词数量和查询中是否包含年份三个特征, 分别使用朴素贝叶斯、支持向量机及决策树算法进行时态意图识别。虽然最终的平均分类准确率不是最优, 但是“无关类”的分类准确率在已有研究中最优。Filannino 等[13]考虑 11 个与时间相关的特征及支持向量机、朴素贝叶斯、决策树和随机游走 4 种分类算法, 其实验表明仅使用 5 个特征的支持向量机平均分类准确率最高, 这 5 个特征为查询是否包含时间表达式、查询包含的时间与查询提交时间的时间差、查询中动词时态、查询中明显指示时态类别词汇(Triggers, 即时间关键词)的频率以及出现顺序。

上述研究为首次 TQIC 测评成果, 如何选取特征仍处于初期研究阶段, 故存在如下问题: 缺乏对可抽取特征的归纳总结; 缺乏对特征有效性的探讨。

### 2.2　融合查询表达式与伪相关文档集合特征的时态意图识别

该类研究同时考虑查询表达式特征及伪相关文档集合特征。其中, 抽取的查询表达式特征也主要包括与时间相关或时间无关的特征; 抽取的伪相关文档集合特征可以分为文档的时间特征(如发布时间、包含的时间信息等)和文档的一般特征(如 n-gram)。

Burghartz 等[14]抽取 n-gram、查询词项的时态类

①http://ntcirtemporalia.github.io/.

别、语言特征、主题特征 4 组查询表达式特征及文档发布时间和文档包含的时间信息两类伪相关文档集合特征，分别使用朴素贝叶斯与决策树识别时态意图。实验结果表明，朴素贝叶斯算法优于决策树算法，且同样采用朴素贝叶斯算法时，从 6 个特征组中人工选取的15 个特征优于利用模拟退火算法(Simulated Annealing)选取的特征组合。Hasanuzzaman 等[15]在查询表达式特征(查询包含的时间与查询提交时间的时间差、n-gram、查询词项的时态类别)的基础上，从伪相关文档集合抽取文档分类结果特征与文档时间信度值特征(Document Temporal Confidence Value)[16]，利用集成学习(Ensemble Learning)算法，以权重为各个基分类器分类准确率的加权投票思想聚合 8 个基分类器进行时态意图的自动识别。实验结果表明，“现在类”的分类准确率最高。在此基础上，Hasanuzzanman 等[17]提出聚合基分类器权重计算的其他三种方式，并将基分类器数量由 8 种扩展至 28 种，结果表明，其平均分类准确率优于 TQIC 任务的最优测评结果。与 Hasanuzzaman 等[15,17]的方法相似，Hou 等[18]通过两步实验进行时态意图识别：先采用PRISM算法进行时态意图识别，若该算法无法将查询归为某类，则采用投票思想聚合多个基分类器。基分类器训练的特征包括查询表达式特征(n-gram、实体特征、查询词项的时态类别、查询包含的时间信息、查询包含的时间与查询提交时间的时间差)及伪相关文档集合特征(n-gram)。实验结果表明，选择使用基于伪相关文档集合特征的基分类器优于基于其他特征的基分类器。

上述研究表明，在查询表达式特征的基础上考虑伪相关文档集合特征，无论采用单一分类器，或以投票思想聚合多个基分类器，均难以提高时态意图识别的平均分类准确率。

## 3 查询表达式特征归类与分类算法选择

### 3.1 查询表达式特征归类

将相关研究中涉及的查询表达式特征根据其与时间关联性的强弱，归为与时间无关的特征、显式时间特征及潜在时间特征三类。

(1) 与时间无关特征

与时间无关特征指无法体现“时间”的通用特征，主要包括：

①查询长度特征，即查询包含的词项个数，例如文献[12]。用户为表明对检索结果时间的需求，可能在查询表达式中加入与时间相关的限定词，导致查询长度的增加。例如查询“Martin Luther King Day(马丁·路德·金纪念日)”可返回关于该纪念日在多个时间段的相关信息，为明确查找 2013 年的相关信息，可在原查询的基础上添加年份信息“2013”，将查询表达式修改为“Martin Luther King Day 2013”，最终查询长度增加，查询意图更加明确。

②实体特征，即查询中是否包含人名、机构名、地址等实体信息，例如文献[11]。实体蕴含一定的时间信息，例如与实体相关事件的发生时间等。以人名实体“Neil Armstrong”为例，关于该人物的检索结果应多集中于他所生活的时间段。

(2) 显式时间特征

显式时间特征指查询中包含明显的与“时间”有关的特征，主要包括：

①年份信息特征[12]。年份信息限定查询结果对应信息的时间范围，例如查询“movies 2012(电影 2012)”只查找与 2012 年相关的电影。

②核心动词时态特征[11]。核心动词指主句中的动词。在英文中，动词的时态能表示行为发生的时间(过去、现在、未来)，然而从句使得一个句子中可能包含多个动词，例如“When did Neil Armstrong die(尼尔·阿姆斯特朗何时逝世)”中包含“did”与“die”两个动词，一个为过去时、一个为现在时，根据语法规则，只有主句中动词(did)的时态才能指示行为发生的时间(过去时)。因此，需先识别查询的句法结构，分清主句与从句，在此基础上识别核心动词(主句中动词)，并将该核心动词的时态作为特征。

③时间关键词(Dominant Keyword)特征，例如文献[13]。时间关键词指属于某一时间类别查询中反复出现的词项，可将查询所包含时间关键词的时间类别作为查询的时态意图类别。例如对于未来类的查询而言，包含时间关键词“will”、“forecast”、“shall”、“upcoming”或“next”的查询属于“未来类”的几率会很大。

(3) 潜在时间特征

潜在时间特征指该查询包含一些时间特征，但是需要借助一定的手段与方法分辨该特征的时间属性，主要包括：

①查询时间差特征，例如文献[11,13,15]。它指查询中的时间表达式所指代时间(查询包含的时间)与当前时间(查询提交至搜索引擎的时间)在时间测度上的差值。差值的正、零、负值可对应于时态意图分类体系中的“过去类”、“现在类”、“未来类”。例如，查询“Martin Luther King Day 2013(马丁·路德·金纪念日 2013)”中所包含的时间为 2013 年，假设查询提交的时间为 2015 年，则查询时间差为-2 年，在一定程度上可判断该查询的时态意图为“过去类”；假设查询提交的时间为 2013 年，则查询时间差为 0 年，在一定程度上可判

断该查询的时态意图为“现在类”。

②查询词项的时态特征[14-15]，即查询词项属于某一时态类别的概率值(获取方法参见 4.2 节)。查询词项的时态类别在一定程度上能反映该查询的时态，即查询的时态意图可由查询词所属时态类别体现，以查询“weather today(天气今天)”为例，查询词项“今天”与“天气”属于“现在类”的概率高，故该查询属于“现在类”的概率高。

### 3.2　查询自动分类算法选择

本文任务是依据查询的时态意图将查询分为过去类、现在类、未来类、与时间无关类 4 类。根据输入训练数据的标注程度，自动分类算法可分为有监督分类和半监督分类。两者均需区分训练数据与测试数据，并将依据某个分类体系标注过类别的数据作为训练数据，然后在训练数据上训练模型，最终利用该模型将未标记类别的测试数据自动分类至该分类体系中的某个类别；与有监督分类相比，半监督分类还可以将未标记数据作为训练数据。半监督分类中未标记的训练数据越多，分类器的泛化能力越强。

## 4　实验数据与实验构建

### 4.1　实验数据介绍及预处理

本实验采用以下两个数据集。

(1) TQIC 查询数据: TQIC 于 2014 年 5 月 9 日发布的 300 条英文查询①，示例如图 1 所示。

```
<query>
        <id>004</id>
        <query_string>price of samsung galaxy note</query_string>
        <query_issue_time>May 1, 2013 GMT+0</query_issue_time>
        <temporal_class>recency</temporal_class>
</query>
```

图 1　TQIC 查询数据示例

每一条数据记录以下信息：查询的唯一编号(id)；查询表达式(query_string)；查询提交至搜索引擎的时间(query_issue_time)；基于TQIC分类体系的人工标注时态意图类别(temporal_class)。

(2) AOL 日志数据②: 2006 年 3 月 1 日–2006 年 5 月 31 日连续三个月的查询日志，示例如图 2 所示。

```
AnonID  Query   QueryTime   ItemRank    ClickURL
53  mapquest    2006-03-01 15:18:21 1   http://www.mapquest.com
66  cajun candle    2006-03-01 13:20:18 1   http://www.cajuncandles.com
66  candle jars 2006-03-01 13:22:29 1   http://www.sks-bottle.com
```

图 2　AOL 日志数据的三条示例

每一条数据从左到右分别记录以下信息：用户 ID；查询表达式；查询提交至搜索引擎的时间；该 URL 在返回结果中的排名；用户点击的 URL。

本实验只需要考虑查询表达式、查询提交至搜索引擎的时间以及人工标注后的时态意图类别三类信息。其中，前两类信息均包含于两个数据集中，是有监督分类与无监督分类必须使用的信息；最后一类信息只包含于 TQIC 查询数据中，是有监督分类必须使用的信息，而非无监督分类必须使用的信息。

相较于文本数据，查询为短文本，需预处理的内容较少，针对以上两个数据集，将查询转化为小写，去除标点符号，最后以文本格式存储，以方便后续的特征抽取。

### 4.2　特征抽取

根据实验操作中特征的不同记录方式，3.1 节所述的三类 7 种查询表达式特征可具体为 19 个特征，为便于描述，将上述 19 个特征分为 A、B、C、D、E 与 F 这 6 个特征组，如表 1 所示。

表 1　本文所抽取的查询表达式特征

| 特征类别 | 分组编号 | 特征名 | 特征不同记录形式的编号 | 特征说明 |
|---|---|---|---|---|
| 与时间无关特征 | A | 实体 | 1 | 实体的数目 |
| | | 查询长度 | 10 | 词项的数目 |
| 显式时间特征 | B | 年份信息 | 11 | 是否包含年份信息 |
| | C | 核心动词的时态 | 6 | 动词的数目 |
| | | | 7 | 为过去时的核心动词数目 |
| | | | 8 | 为现在时的核心动词数目 |
| | | | 9 | 为未来时的核心动词数目 |
| | E | 时间关键词 | 12 | 过去类时间关键词的数目 |
| | | | 13 | 现在类时间关键词的数目 |
| | | | 14 | 未来类时间关键词的数目 |
| | | | 15 | 时间指代不明的词项数目 |
| 潜在时间特征 | D | 查询时间差 | 2 | 查询所表达的时间点数目 |
| | | | 3 | 指向过去的时间差数目 |
| | | | 4 | 指向现在的时间差数目 |
| | | | 5 | 指向未来的时间差数目 |
| | F | 查询词项的时态 | 16 | 属于过去类的词项数目 |
| | | | 17 | 属于现在类的词项数目 |
| | | | 18 | 属于未来类的词项数目 |
| | | | 19 | 属于与时间无关类的词项数目 |

(注：各个特征抽取的实验进度不同，因此表中编号未按照 3.1 节所述的顺序编号，而是按照实验抽取顺序编号。)

①http://research.nii.ac.jp/ntcir/permission/ntcir-11/perm-en-Temporalia.html.

②http://www.cim.mcgill.ca/~dudek/206/Logs/AOL-user-ct-collection/.

笔者使用 Stanford NLP 工具集①抽取表 1 中特征，涉及分词、实体抽取、句法分析、词项标注、时间表达式抽取等步骤。其中，特征组 E 包含的时间关键词特征借助文献[12]中的词典抽取；特征组 F 包含的查询词项的时态概率借助 TempoWordNet 词典(TWnH-1.0 版)②抽取，该词典是基于 WordNet[19]的词典，对收录的每一个词汇均标注了其属于“过去类”、“现在类”、“未来类”、“无关类”的概率。

### 4.3 实验构建

构建一个基准实验(Baseline)的特征组合以及多个对照组实验的特征组合，在相同特征组合上，分别使用 SVMlin③与 LIBSVM④实现半监督算法与有监督算法。对于 SVMlin，将 TQIC 中随机抽取的 80 条数据及 AOL 中随机抽取的 10 000 条数据作为训练数据，将 TQIC 剩余的 220 条数据作为测试数据；对于 LIBSVM，以 10 折交叉检验的方式训练参数，最终展示分类器 10 折交叉检验准确率的平均值。针对每类分类算法(SVMlin 或 LIBSVM)，均采用一对多的分类思想构建分类器，即针对每一个时态意图类别构造一个二值分类器，因此共构造 4 个二值分类器。与 TQIC 任务测评要求一致，分类器效果通过分类准确率(Accuracy)，即被正确分类的样本数除以所有样本数[20]测评，主要包括单类分类准确率和平均分类准确率两个指标，前者指单个二值分类器的准确率，后者指 4 个二值分类器的准确率平均值。

(1) 基准实验的特征组合构建

根据研究目的，基准实验的构建需满足两个原则：

①为探讨三类查询表达式特征的有效性，基准实验组需包含三类查询表达式特征；

②为探讨时态意图识别的准确度，基准实验组需为已有研究中最优结果。

在所有基于 TQIC 的测评研究中，文献[11]的平均分类准确率最高[9]，且其所抽取的三种特征(1、C、D)分别对应三类查询表达式特征，因此特征组合(1+C+D)可直接作为基准实验组。为减少对照实验组数，在特征组合(1+C+D)的基础上加入查询长度特征(编号 10)，从而构建一个新的特征组合(A+C+D)。随后基于 TQIC 分类体系，在以上两个特征组合下分别采用 SVMlin 与 LIBSVM 进行时态意图识别，结果如表 2 所示。

使用 SVMlin 时，采用(1+C+D)的准确度略高于采用(A+C+D)的准确度，但采用基于(A+C+D)的 LIBSVM 准确度在以上所有实验组中最高，因此本文使用特征组合(A+C+D)的实验结果作为基准实验。

(2) 对照组实验的特征组合构建

为验证不同类别查询表达特征的有效性，构建三个对照实验的特征组合，具体如表 3 所示。

表 2 时态意图识别初始实验结果

| 特征组合 | SVMlin | | | | | LIBSVM | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 过去 | 现在 | 未来 | 无关 | 平均 | 过去 | 现在 | 未来 | 无关 | 平均 |
| 1+C+D[11] | 0.764 | 0.755 | 0.777 | 0.746 | 0.760 | 0.840 | 0.777 | 0.803 | 0.767 | 0.797 |
| A+C+D | 0.818 | 0.736 | 0.741 | 0.741 | 0.759 | 0.840 | 0.777 | 0.803 | 0.770 | 0.798 |

(注：因特征处理的细节及抽样样本不同，故本实验重现的结果与原文略有不同。)

表 3 对照实验的特征组合

| 特征组合 | 说明 | 特征组合实例 |
|---|---|---|
| 显式时间特征组合 | 显式时间特征的 3 组特征(B、C、E)中任一两组的组合。 | B+C、B+E、C+E、B+C+E |
| 显式时间特征与潜在时间特征混合组合 | 在任一显式时间特征组合基础上，加入潜在时间特征(D、F)的组合。 | 以显式时间特征组合 B+C 为例，可行的组合为：B+C+D、B+C+F、B+C+D+F |
| 三类特征组合 | 分别选取上述两个类特征组合中的最优组合，在此基础上，加入与时间无关类特征组 A。 | / |

①http://nlp.stanford.edu/software/index.shtml.

②https://tempowordnet.greyc.fr/download_TWn.html.

③http://vikas.sindhwani.org/svmlin.html.

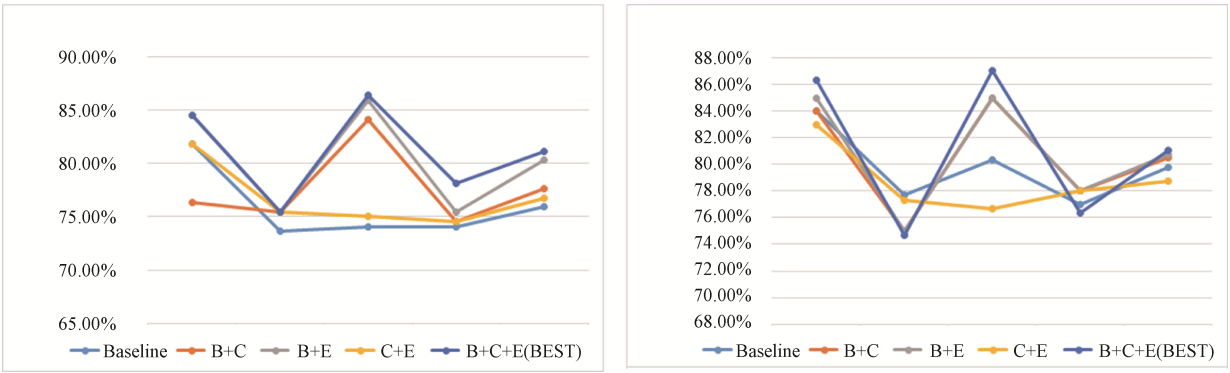
④https://www.csie.ntu.edu.tw/~cjlin/libsvm/.

## 5　实验结果分析

### 5.1　基于不同特征组合下的时态意图识别效果

(1) 显式时间特征组合的时态意图识别结果

4 个显式时间特征组合(B+C, B+E, C+E, B+C+E)下分别采用 SVMlin 及 LIBSVM 的时态意图识别结果如图 3 所示。可知, SVMlin 与 LIBSVM 的实验结果基本一致。对于平均分类准确率而言, Baseline 与 C+E 组较差, 其他三组(B+C, B+E, B+C+E)优于 Baseline; B+C+E 最优(81.14%), 且双尾 T 检验结果表明只有 B+C+E 组对于 Baseline 结果有显著提高($p$=0.048<0.05)。实验效果最优的特征组合 B+C+E 与 C+E 组相比, 增加表明查询是否包含年份信息的特征 B, 故可说明特征 B (是否包含年份信息)是时态意图识别的强特征; 对于单类的分类准确率, 所有实验组的“过去类”与“未来类”分类效果均明显优于“现在类”与“无关类”。因本实验中 B+C+E 的实验结果最优, 后文以“BEST”指代该组合。




(a) SVMlin　　(b) LIBSVM

图 3　显式时间特征组合的时态意图识别实验结果

(2) 显式时间特征与潜在时间特征组合的时态意图识别结果

测试潜在时间特征组合 D+F, 随后测试显式时间特征与潜在时间特征混合组合(见表 3), 采用 SVMlin 及 LIBSVM 的实验结果分别如图 4 和图 5 所示。

①潜在时间特征组合(D+F)与 Baseline 对比: 对于平均分类准确率, D+F 与 Baseline 相似, 但是采用 D+F“现在类”的分类准确率明显高于 Baseline, “过去类”的分类准确率低于 Baseline。

②显式时间特征与潜在时间特征混合组合与 Baseline 对比: 前者相对 Baseline 均有提升, 但因使用潜在时间特征的不同, 提升的效果也有差异。如图 4(a)、图 4(b)、图 4(d)所示, 在平均分类以及“过去类”的准确率提升上, 仅加入 F 的混合特征组合明显优于加入 D 或共同加入 D 和 F。

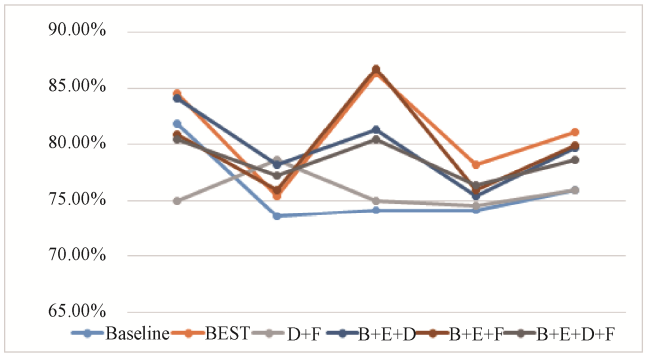
③显式时间特征与潜在时间特征混合组合与潜在时间组合(D+F)对比: 前者的平均分类准确率均高于潜在时间特征组合, 且在“过去类”、“未来类”、“无关类”的准确率上均有明显提高, 而在“过去类”的分类准确率不及 D+F 组合。

④显式时间特征与潜在时间特征混合组合与显式时间特征组合(BEST)对比: 前者对于 BEST 组而言, 平均分类准确率稍有下降, 且除了“过去类”的分类准确率均低于 BEST; 结合图 3 实验结果, 可知虽然显式时间特征与潜在时间特征的组合能提高仅采用潜在时间特征组合的平均分类准确率, 但不及仅采用显式时间特征组合的平均分类准确率。

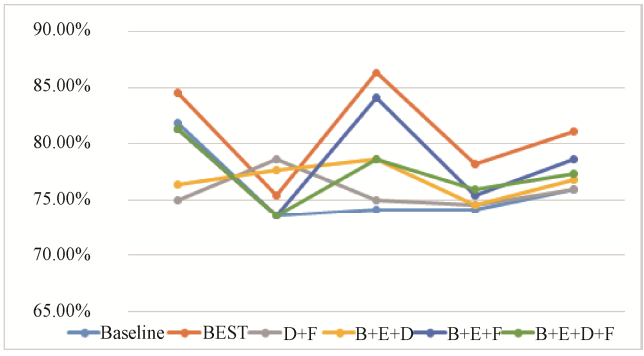
采用显式时间特征与潜在时间特征的混合实验组平均分类准确率高于 Baseline, 高于潜在时间特征组, 但不及采用显式时间特征组合; 在显式时间特征组合上加入不同的潜在时间特征对于准确度的提升有差异。图 5 与图 4 不同的结论为: 显式时间特征与潜在时间特征的混合组中 B+E+F 的平均分类准确率高于 BEST 组, 尤其是“无关类”的准确率高于 BEST 组的准确率, 然而双尾 T 检验表示 B+E+F 组、Baseline 以及 BEST 组之间并没有显著差异, 即识别效果提高有差异但不具显著性; 相对而言, 在显式时间特征组的基础上同时加入 D 与 F 的实验效果优于单独加入 D 或单独加入 F 的实验组。

(3) 三类特征组合的时态意图识别结果

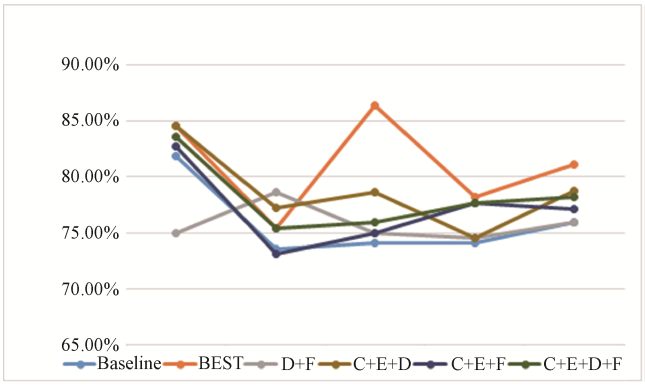
根据表 3, 三类特征组合时, 只考虑了不同分类算法下, 最优显式时间特征与潜在时间特征的组合加上与时态无关特征组(A)。例如, SVMlin 实验中组合 B+C+E+F 识别效果最优, LIBSVM 实验中组合 B+E+F 识别效果最优, 因此针对 SVMlin 与 LIBSVM, 本部分实验只考虑 A+B+C+E+F 和 A+B+E+F。时态意图识别的实验结果如图 6 所示。



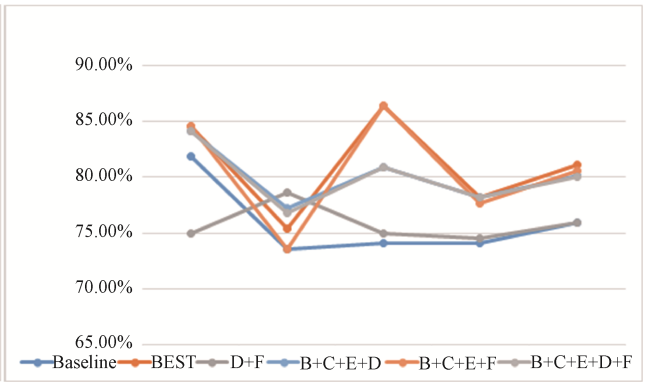

(a) B+E




(b) B+C

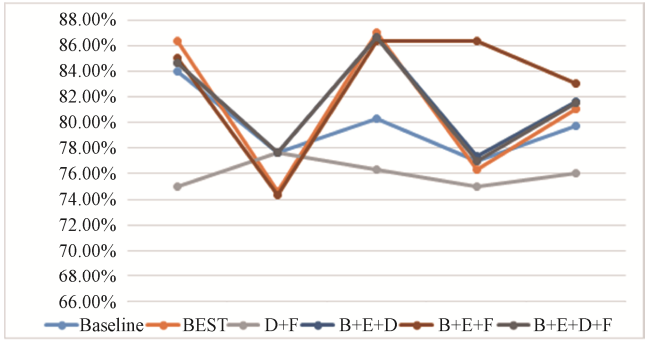



(c) C+E

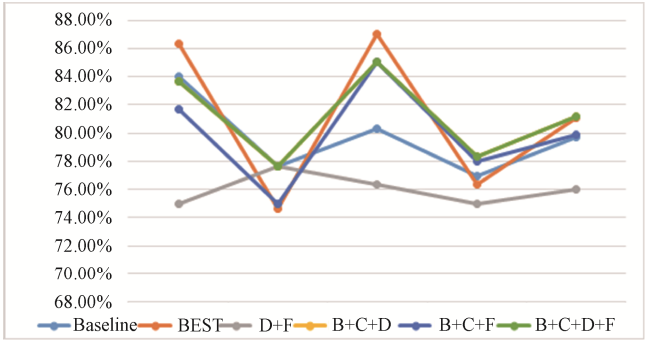



(d) B+C+E

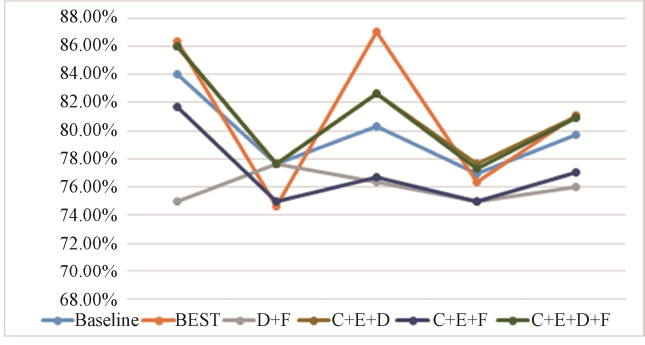
图 4　显式时间特征与潜在时间特征混合组合的时态意图 SVMlin 识别实验结果



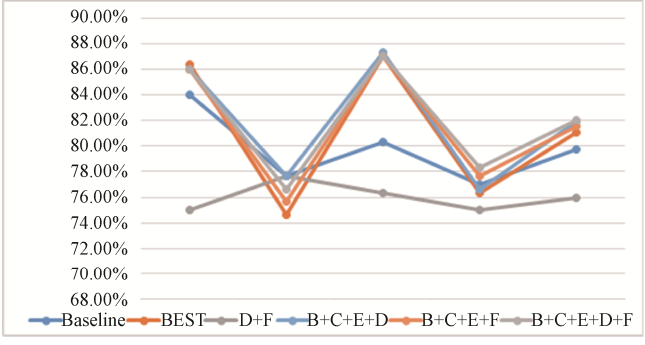

(a) B+E




(b) B+C

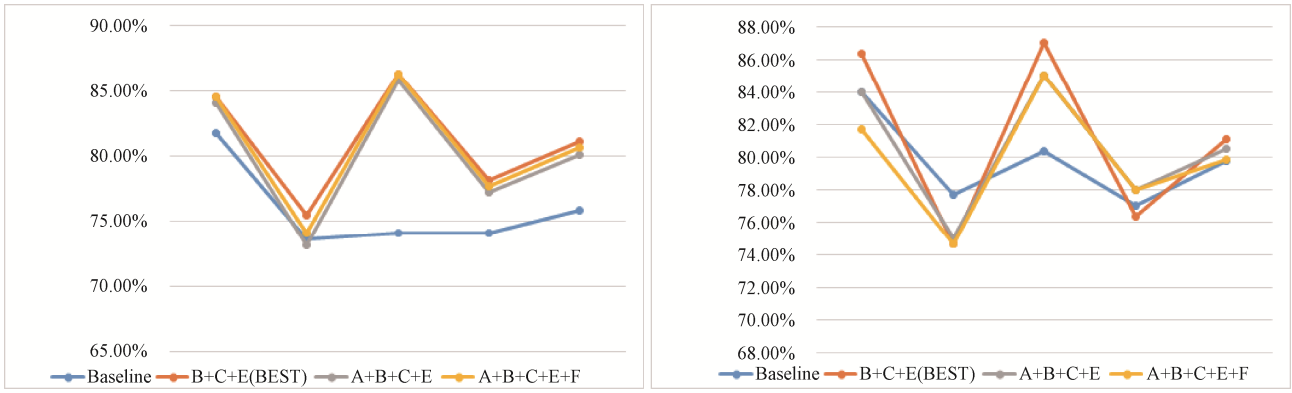



(c) C+E




(d) B+C+E

图 5　显式时间特征与潜在时间特征混合组合的时态意图 LIBSVM 识别实验结果




(a) SVMlin　　(b) LIBSVM

图 6　三类特征组合的时态意图识别实验结果

虽然其他实验组相对 Baseline 在平均分类准确率上均有提升，但是 BEST 组的平均分类准确率依然最优。就单类的分类准确率而言，使用 SVMlin 时，单类的分类准确率均有提高，尤其是“未来类”及“无关类”提升很明显；在使用 LIBSVM 时，只有“未来类”的分类准确率提升效果较为明显，BEST 组“现在类”及“无关类”的分类准确度反而较低。针对图 6 中实验结果的双尾检验 p 值如表 4 所示。

在两种分类器下，只有 BEST 组与 Baseline 组的 p 值小于 0.05，即具有显著差异。因此综合所有结果来看，仅考虑显式时间特征的组合相较于 Baseline 而言，具有显著提升，而加入潜在时间特征或者时间无关特征的实验结果虽然相对于 Baseline 有提升，但并非显著提升。

### 5.2　查询表达式特征统计分析

由上述实验结果可知，部分查询表达式特征组合用于时态意图识别的效果优于其他查询表达式特征组合，针对实验中涉及的 5 组查询表达式特征进行深入分析。各查询表达式特征统计结果如图 7 所示，虚线为 19 个特征的平均占比，为 26.79%。

表 4　三类特征组合的时态意图识别双尾检验结果

| 特征组合 | SVMlin | LIBSVM |
|---|---|---|
| B+C+E(BEST) | **0.048** | 0.456 |
| A+B+C+E | 0.110 | 0.559 |
| A+B+C+E+F | 0.076 | – |
| A+B+E+F | – | 0.954 |

(注：因 SVMlin 与 LIBSVM 只分别考虑 A+B+C+E+F 与 A+B+E+F，未对 SVMlin 下特征组合 A+B+E+F 的结果与 Baseline 结果进行双尾检验，也未对 LIBSVM 下特征组合 A+B+C+E+F 的结果与 Baseline 结果做双尾检验，故在表中以“–”表示结果。)

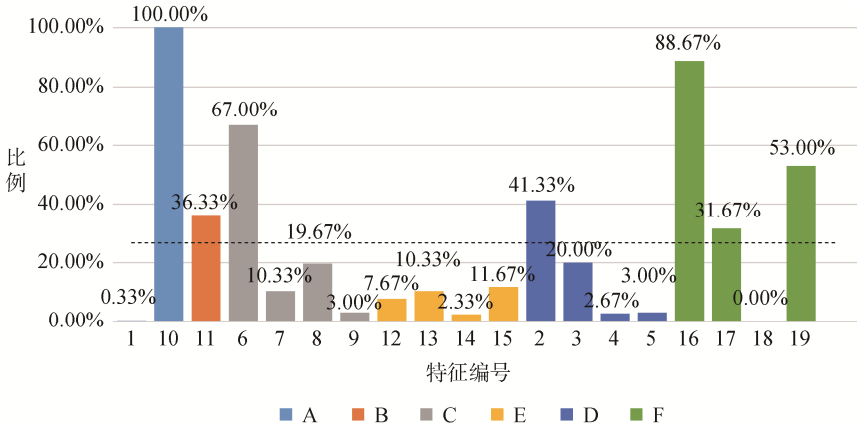



图 7　时态意图特征分析

部分特征占比偏高，例如特征10(100%)与特征16(88.67%)；部分特征占比几乎可以忽略不计，例如特征 1(0.33%)与特征 18(0.00%)。按平均准确率由高到低，分别基于上述三类时态意图识别的结果排序为：

显式时间特征>潜在时间特征>时间无关特征组，该排序与三类特征按照包含特征平均占比由低到高的排序相同：显式时间特征(B+C+E，18.70%)>潜在时间特征(D+F，30.04%)>时间无关特征组(A，50.17%)，即采用特征占比越低的特征组用于时态意图识别的准确度越高。笔者认为此现象的可能解释为：特征出现频率越高，说明该特征是普遍特征，因此区分度较低，在分类实验中有效性较差；但是若出现频率过低，则说明该特征几乎不存在于本实验所选用的具有时态意图的查询集合中，因此区分度也会较低，在分类实验中有效性较差。因此，较为理想的分类特征出现频率不能过高也不能过低。

## 6 结 语

针对时态意图识别问题，本文以查询表达式的特征为例，按照其与时间的关联性归为与时间无关特征、显式时间特征以及潜在时间特征三类，根据不同的特征组合分别使用有监督分类器和半监督分类器探讨上述三类特征的识别有效性，最终实现的时态意图识别结果优于同类任务测评的成果，平均分类准确率最高可达 81.14%。尽管如此，本文仍存在一些不足，这也将是需要进一步探讨的内容：

(1) 实验数据只有 300 条查询，可考虑自建一个较大的数据集，在此基础上验证结论的有效性；

(2) 主要考虑已有的查询表达式特征，未提出用于时态意图识别的新特征，后续研究需提出新的特征，进一步探讨如何提升时态意图识别的平均分类准确率；

(3) 采用本文方法准确识别时态意图的基础上，可考虑优化检索结果列表，实现近因敏感排序或时间依赖性排序。

## 参考文献:


[1] Broder A. A Taxonomy of Web Search[J]. SIGIR Forum, 2002, 36(2): 3-10.

[2] Sushmita S, Piwowarski B, Lalmas M. Dynamics of Genre and Domain Intents[C]// Proceedings of the 6th Asia Information Retrieval Societies Conference on Information Retrieval Technology. Springer, 2010: 399-409.

[3] Calderón-Benavides L, González-Caro C, Baeza-Yates R A. Towards a Deeper Understanding of the User's Query Intent[C]// Proceedings of the 2010 Workshop on Query Representation and Understanding. 2010: 21-24.

[4] Nguyen B V, Kan M. Functional Faceted Web Query Analysis[C]// Proceedings of the 16th International World Wide Web Conference. 2007.

[5] González-Caro C, Baeza-Yates R. A Multi-faceted Approach to Query Intent Classification[C]// Proceedings of the 18th International Conference on String Processing and Information Retrieval. 2011: 368-379.

[6] Campos R, Dias G, Jorge A M. What is the Temporal Value of Web Snippets?[C]// Proceedings of the 1st International Temporal Web Analytics Workshop. 2011: 9-16.

[7] 张晓娟，韩毅. 时态信息检索研究综述[J]. 数据分析与知识发现, 2017, 1(1): 3-15. (Zhang Xiaojuan, Han Yi. Reviews on Temporal Information Retrieval[J]. Data Analysis and Knowledge Discovery, 2017, 1(1): 3-15.)

[8] Jones R, Diaz F. Temporal Profiles of Queries[J]. ACM Transactions on Information Systems, 2007, 25(3): Article No.14.

[9] Joho H, Jatowt A, Blanco R, et al. Overview of NTCIR-11 Temporal Information Access (Temporalia) Task[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 217-224.

[10] Mizzaro S. How Many Relevances in Information Retrieval?[J]. Interacting with Computers, 1998, 10(3): 303-320.

[11] Yu H, Kang X, Ren F. TUTA1 at the NTCIR-11 Temporalia Task[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 461-467.

[12] Shah A, Shah D, Majumder P. Andd7@NTCIR-11 Temporal Information Access Task[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 456-460.

[13] Filannino M, Nenadic G. Using Machine Learning to Predict Temporal Orientation of Search Engines' Queries in the Temporalia Challenge[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 438-442.

[14] Burghartz R, Berberich K. MPI-INF at the NTCIR-11 Temporal Query Classification Task[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 443-450.

[15] Hasanuzzaman M, Dias G, Ferrari S. HULTECH at the NTCIR-11 Temporalia Task: Ensemble Learning for Temporal Query Intent Classification[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 478-482.

[16] Campos R, Dias G, Jorge A, et al. GTE: A Distributional Second-order Co-occurrence Approach to Improve the Identification of Top Relevant Dates in Web Snippets[C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management. 2012: 2035-2039.
[17] Hasanuzzaman M, Saha S, Dias G, et al. Understanding Temporal Query Intent[C]// Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015: 823-826.
[18] Hou Y, Tan C, Xu J, et al. HITSZ-ICRC at NTCIR-11 Temporalia Task[C]// Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies. 2014: 468-473.
[19] Miller G A. WordNet: A Lexical Database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
[20] Sokolova M, Lapalme G. A Systematic Analysis of Performance Measures for Classification Tasks[J]. Information Processing and Management, 2009, 45(4): 427-437.

## 作者贡献声明:

桂思思: 提出研究思路, 负责实验, 论文起草;
陆伟, 张晓娟: 论文修改及最终版本修订。

## 利益冲突声明:

所有作者声明不存在利益冲突关系。

## 支撑数据:

支撑数据由作者自存储, E-mail: sgui0229@whu.edu.cn。
[1] 桂思思. AOL_sample_10000.txt. AOL 美国在线查询日志中随机抽取的 10000 条数据.
[2] 桂思思. 300_query_features.txt. 300 条查询的查询表达式特征值.

收稿日期: 2018-05-17
收修改稿日期: 2018-06-21

# Temporal Intent Classification with Query Expression Feature

Gui Sisi[1,2] Lu Wei[3] Zhang Xiaojuan[4]
[1](School of Information Management, Wuhan University, Wuhan 430072, China)
[2](Institute for Information Retrieval and Knowledge Mining, Wuhan University, Wuhan 430072, China)
[3](Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China)
[4](School of Computer and Information Science, Southwest University, Chongqing 400715, China)

**Abstract: [Objective]** This paper investigates the effectiveness of query-based features and compares the performance of two types of classifiers in a query temporal intent classification task. **[Methods]** This paper first reviews all query-based features and then classifies those features into three types, according to their temporal relevance, namely, atemporal, implicit temporal and explicit temporal. Then, it tests accuracy of a temporal query intent classification task, using a supervised classifier and a semi-supervised classifier individually, with various combinations of query-based features of different types. **[Results]** Among all tested query-based features, using explicit temporal features achieves best accuracy, especially for the feature on whether a query contains a year; The performance hardly varies across classifiers; Our best macro average accuracy of 81.14% is higher than that in previous studies with the same experimental setups. **[Limitations]** Due to accessibility of dataset, our experiments are done on a limited size dataset. Only existing query-based features are studied and no new feature is proposed or tested. **[Conclusions]** Using highly temporal relevant features can improve accuracy in temporal query intent classification task, whereas using slightly temporal relevant features could hardly improve accuracy.

**Keywords:** Temporal Intent Supervised Classification Semi-supervised Classification Feature Engineering