

文章编号: 2095-2163(2023)12-0068-07

中图分类号: TP391.1

文献标志码: A

基于改进关联规则挖掘和距离聚类算法的知识融合

佟国香¹, 胡琪¹, 杨培威¹, 蒋伟¹, 岳继光²

(1 上海理工大学 光电信息与计算机工程学院, 上海 200093; 2 同济大学 电子与信息工程学院, 上海 200092)

摘要: 针对产品全生命周期管理(PLM)中来自于不同阶段和不同领域的设计文档, 本文提出一种基于实体抽取的改进关联规则挖掘与距离聚类相结合的知识获取与融合的算法。以汽车行业领域知识获取与融合为例, 从相关文档中抽取 8 183 组数据, 220 941 个实体, 将各组数据的 27 个特征两两交叉并与原来的特征规则共同作为候选集, 进行关联规则挖掘, 并构造初步的领域知识库。通过定义知识库中各实体之间的距离, 结合聚类算法减少冗余知识, 再根据融合后的知识构建知识库。实验表明, 该算法在一定程度上减少了知识模型的规模, 提高了领域知识库中知识的价值密度。

关键词: 实体抽取; 改进关联规则; 距离聚类; 产品全生命周期

Knowledge fusion based on improved association rule mining and distance clustering algorithm

TONG Guoxiang¹, HU Qi¹, YANG Peiwei¹, JIANG Wei¹, YUE Jiguang²

(1 School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2 School of Electronic and Information Engineering, Tongji University, Shanghai 200092, China)

Abstract: Aiming at the design documents from different generation stages and different fields in product lifecycle management (PLM), this paper proposed a knowledge acquisition and fusion algorithm based on improved association rule mining and distance clustering. Took the domain knowledge acquisition and fusion of automobile industry as an example. 8 183 groups of data and 220 941 entities were extracted from relevant documents. The 27 features of each group of data were crossed in pairs and the original feature rules were combined as candidate sets to association rules mining and constructed a preliminary domain knowledge base. The distance between entities in the knowledge base was defined, combined with the clustering algorithm to reduce redundant knowledge, and the knowledge base was constructed according to the merged knowledge. The experimental results show that the algorithm reduces the scale of the knowledge model to a certain extent and improves the value density of knowledge in the domain knowledge base.

Key words: entity extraction; improved association rules; distance clustering; product lifecycle management

0 引言

工业化与信息化的深度融合是实施制造强国战略的基本技术路线^[1]。随着智能制造和工业 4.0 等技术的发展, 产品全生命周期管理(Product Lifecycle Management, PLM) 系统逐渐被应用到很多大型机械制造企业中。根据国际标准 ISO/IEC 15288:2002 给出的设计定义流程、系统分析流程、实现流程、综合流程等 14 个技术流程, 分析全生命周期过程中可能出现的产品模型、过程模型以及知识模型, 探索模

型中信息的检索、重用以及模型间的互联互通等一系列的问题将更加复杂。因此, 如何获取构建产品全生命周期的模型所需知识、知识融合、知识模型的表达等一系列信息, 则成为工业智能化管理的研究热点^[2]。

在对知识获取和知识融合过程中, 针对知识过载和知识不完整等问题的研究备受关注。对本文文件中的数据进行挖掘和分析, 获取模型构建所需的知识, 首先要对这些本文文档中的数据进行抽取。罗辉等^[3]提出了一种基于 GATLexicon 网络的字符

基金项目: 国家重点研发计划项目(2018YFB1700902)。

作者简介: 佟国香(1968-), 女, 博士, 副教授, 主要研究方向: 嵌入式系统研究; 胡琪(1997-), 女, 硕士研究生, 主要研究方向: 自然语言处理; 杨培威(1998-), 男, 本科生, 主要研究方向: 问答系统; 蒋伟(1996-), 男, 硕士研究生, 主要研究方向: 知识图谱。

通讯作者: 岳继光(1961-), 男, 博士, 教授, 主要研究方向: 先进测控理论技术。Email: yuejiguang@tongji.edu.cn

收稿日期: 2022-11-16

信息融合部分实体信息的方法,将字符与其上下文信息进行部分融合,从而改善字符的表示,提升中文命名实体识别的性能;曾兰兰^[4]提出了一种基于 BERT 和联合学习的命名实体识别模型,有效提升裁判文书中各类实体的识别效果。由于初步构建的领域知识中,隐藏了大量潜在的知识,这些知识对领域知识模型的构建有很大的影响。融合领域相关信息,并进行分析和应用,构造更完善的知识模型,成为领域知识模型协同及数据管理的研究热点。因此,目前涌现出一些对领域知识融合的研究。在知识融合框架方面,周利琴等^[5]提出针对网络大数据的知识融合框架,在数据库融合更新方面取得了很好的成效,解决了数据实时更新产生的相关问题;A. Smirnov 等^[6]基于上下文的知识融合模式,对信息和系统结构进行自主性保存,为系统设计提供支持。在知识融合算法方面,文家富等^[7]结合贝叶斯和遗传算法,实现了分散关联知识元的融合;Ruta 等^[8]针对车载自组织网络的上下文信息共享问题,提出基于非单调、非标准推理服务的知识融合算法,实现车载网络节点不一致上下文注释的自动合并及协

调。

本文针对产品生命周期中各阶段知识模型构建过程中存在的知识冗余、知识不完整等方面的问题,提出一种基于实体抽取的改进关联规则挖掘与距离聚类相结合的算法,挖掘实体间关联规则,并对挖掘后的规则做二次处理,以得到更有价值且精简的知识。将本文算法应用到汽车领域某阶段信息管理,实现了部分领域知识的获取、融合以及知识库构建的全过程。

1 实体抽取

信息化背景下,大部分企业的管理体系都应用了信息技术,且在产品的全生命周期管理过程中形成了大量数据。要想对数据进行挖掘和分析,首先要对这些本文文档中的数据进行抽取。命名实体识别是自然语言处理中的一项基础任务^[9]。本文基于 Transformer 模型对汽车工业文档进行实体识别与抽取。本文提出改进的 Transformer 模型架构如图 1 所示,由输入表示层、上下文编码层和命名实体标签预测层构成。

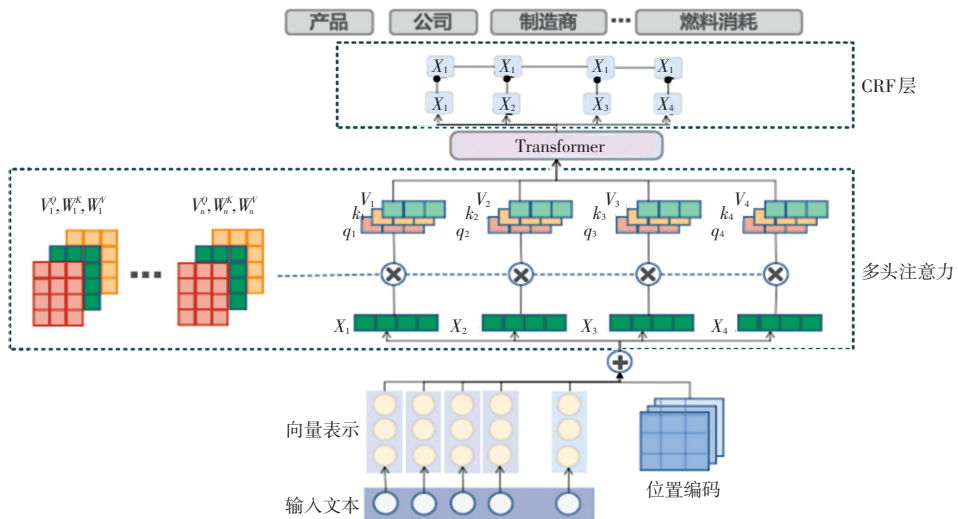


图 1 改进的 Transformer 模型

Fig. 1 Improved Transformer model

1.1 输入表示层

对文本信息进行向量化表示,作为模型的输入。传统的 Transformer 模型不同于循环神经网络,模型将文字转化为向量的过程中,并不考虑文字的上下文及文字的位置信息^[10]。为了弥补实体位置不同可能会产生不同语义的信息缺失问题,本文模型引入位置编码(Positional Encoding),用来表示实体之

间的序列关系。位置编码如式(1),式(2)所示:

$$PE_{(pos, 2i)} = \sin(pos / 10\,000^{2i/d_{\text{model}}}) \quad (1)$$

$$PE_{(pos, 2i+1)} = \cos(pos / 10\,000^{2i/d_{\text{model}}}) \quad (2)$$

其中, pos 为每个词向量在句子中所处的时间步; i 为词向量内每个元素所处的向量维度; d_{model} 为词嵌入的维度。

对于同一个编码的 pos , 偶数位置使用 \sin 函

数,奇数位置使用 \cos 函数。通过代入正余弦函数,将词嵌入数值的大小控制在 $-1 \sim 1$ 之间,以加速梯度计算。由三角函数性质式(3)对式(1),式(2)进行变形得到式(4):

$$\begin{cases} \sin(\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta \\ \cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta \end{cases} \quad (3)$$

$$\begin{cases} PE(pos + k, 2i) = PE(pos, 2i) \times PE(k, 2i + 1) + \\ PE(pos, 2i + 1) \times PE(k, 2i) \\ PE(pos + k, 2i + 1) = PE(pos, 2i + 1) \times \\ PE(k, 2i + 1) - PE(pos, 2i) \times PE(k, 2i) \end{cases} \quad (4)$$

通过 pos 与 k 的位置向量的 $2i$ 与 $2i + 1$ 维的线性组合,可以将词向量之间的相对位置信息加入到向量绝对位置信息中。得到位置编码后,将字向量与位置编码相加,得到一个最终的编码作为整个序列建模的输入。

1.2 上下文编码层

上下文编码层对输入表示层输出的序列向量进行特征提取。将词嵌入向量进行切分,输入到多头注意力机制模块中,使输入的词向量根据其处于不同上下文的位置时,词义能够具有更多元的表达。为了防止多重计算时参数出现过大或过小的情况,对多头注意力机制的输出进行加权和归一化处理;同时,为了防止注意力机制可能出现的拟合程度不够的问题,将归一化处理后的向量传入具有两层线性的全连接前馈神经网络,再对其进行加权和归一化处理。上下文编码层如图2所示, q 表示输入的文本信息, k 表示文本信息中的关键词, v 表示经过不断学习得到的文本信息提取出的关键词。

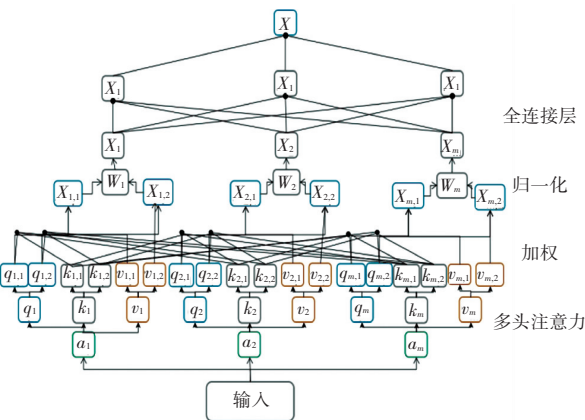


图2 上下文编码层

Fig. 2 Context coding layer

1.3 命名实体标签预测层

命名实体标签预测层以上下文编码层提取的特

征为输入。在Transformer模型中,通过解码器输出的结果,没有考虑序列之间的相关关联,可能会出现无效标注的情况。因此,在解码层的输出加入条件随机场模型,通过分析序列标签之间的相邻依赖关系,对最优标签序列进行求解。其最优标签得分计算公式如式(5)所示:

$$Score(X, x) = \sum_{i=1}^n Z_{x_i, x_{i+1}} + \sum_{i=1}^n M_{i, x_i} \quad (5)$$

其中, $Z_{x_i, x_{i+1}}$ 表示标签 X_i 转移到标签 X_{i+1} 的概率得分, M_{i, x_i} 表示第 i 个字符被标记为标签 X_i 的概率得分。

2 知识融合

由于初步构建的汽车领域知识库中存在语义相近、重复甚至错误的知识。因此,本文基于改进关联规则的方法对汽车领域数据进行关联规则挖掘,并结合距离聚类算法,对语义相近以及重复的知识进行剔除。

传统的关联规则挖掘算法存在两个缺陷:

- (1) 根据支持度和置信度两个指标得到的挖掘结果可能是错误的;
- (2) 挖掘过程中对频繁集的计算复杂度非常高。

改进的关联规则算法中,将每个项集集合 $S = \{t_1, t_2, \dots, t_n\}$ 的项目 t_i 进行两两交叉,得到长度为2的项集,原始项集和两两交叉后的项集作为候选集集合 $S_2 = \{t_1, t_2, \dots, t_n \& t_1 t_2 \& t_1 t_3 \& \dots \& t_{n-1} t_n\}$ 。只计算长度为1和2的规则,加入提升度(lift)指标,生成关联规则。改进关联规则挖掘算法流程如算法1所示。

算法1 改进关联规则挖掘算法

输入 实体数据集,支持度 sup,置信度 conf,提升度 lift,项集对应的项目的序号 T_{id}

输出 所有关联规则

一次扫描得到所有频繁1项集的集合 S

$E(S)$

两两交换项集中的所有项目

For (all $t_i \in S$) Do

 For (all $t_j \in S, j < i$) Do

 通过交集操作,产生长度为2的频繁项集

$S_2 = t_i \cup t_j$

 Count(S_2) = | $T_{id}(t_i) \cup T_{id}(t_j)$ |

 根据支持度 sup:

 sup($t_i \Rightarrow t_j$) = $P(t_i \cup t_j)$

If($|\text{count}(S_2)| \geq \text{sup}$) Then

$S = S \cup S_2, T_1 = T_1 \cup S_2$, 其中 T_i 初始为空

If($T_i \neq \emptyset$) 调用函数 $E(T_i)$

return H

根据置信度 conf 以及提升度 lift , 产生符合条件的规则, 同时保留支持度很低, 但是同样有价值的关联规则。

$$\text{conf}(t_i \Rightarrow t_j) = P(t_i | t_j)$$

$$\text{lift} = \frac{\text{conf}(t_i \Rightarrow t_j)}{\text{sup}(t_j)}$$

为了减少冗余知识, 采用基于距离的聚类算法对算法 1 获取的初步的领域知识库进行知识聚类, 以提高知识的价值密度, 如算法 2 所示。其中, 定义领域知识库中任意两项 t_i, t_j 之间的距离, $\text{Dist}_{\text{item}}(t_i, t_j)$ 表示 t_i, t_j 项共同出现与所有包含 t_i 或 t_j 项的比率如公式(6)所示:

$$\text{Dist}_{\text{item}}(t_i, t_j) = \frac{S(t_i \cup t_j)}{S(X) + S(Y) - S(t_i \cup t_j)} \quad (6)$$

其中, $S(X)$ 表示所有包含 t_i 的项集; $S(Y)$ 表示所有包含 t_j 的项集; $S(t_i \cup t_j)$ 表示出现 t_i 与 t_j 的集合。

$\text{Dist}_{\text{item}}(t_i, t_j)$ 的取值范围为 $[0, 1]$, $\text{Dist}_{\text{item}}(t_i, t_j) = 1$, 表示 $t_i = t_j$, $\text{Dist}_{\text{item}}(t_i, t_j) = 0$, 表示 t_i 与 t_j 总是存在于不同的项集中。

定义项集间的距离, 如式(7)所示:

$$\text{Dist}_{\text{set}}(X, Y) = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \text{Dist}_{t_i, t_j} \quad (7)$$

知识间的距离定义: 设知识 $r_1: X_1 \Rightarrow Y_1, r_2: X_2 \Rightarrow Y_2$, 知识 r_1, r_2 之间的 Dist 定义如式(8)所示:

$$\text{Dist}_{\text{rule}}(r_1, r_2) = \alpha \times \text{Dist}_{\text{set}}(X_1 \cup Y_1, X_2 \cup Y_2) + \beta \times \text{Dist}_{\text{set}}(X_1, X_2) + \gamma \times \text{Dist}_{\text{set}}(Y_1, Y_2) \quad (8)$$

其中, $\text{Dist}_{\text{set}}(X_1 \cup Y_1, X_2 \cup Y_2)$ 表示两个知识前后项并集后的距离; $\text{Dist}_{\text{set}}(X_1, X_2)$ 表示两个知识前项并集后的距离; $\text{Dist}_{\text{set}}(Y_1, Y_2)$ 表示两个知识后项并集后的距离; α, β, γ 为权重系数, 且 $\alpha + \beta + \gamma = 1$, 具体取值由用户决定。

算法 2 基于距离的聚类算法

输入 初步领域知识库; 知识半径 ε ; 知识密度阈值 limitkdp

输出 基于距离聚类的簇的集合 R

将所有待聚类知识存储在有序队列 Q 中, 并标记为未聚类

随机选择有序队列 Q 中的一个知识 k

DBSCAN(k)

标记 k 为已聚类

If k 的 ε 领域内至少包含 limitkdp 个知识

创建一个新簇 C , 并把 k 添加到 C
将 C 存储在 R 中

令 K 为 k 的 ε 领域中的知识集合

For K 中每个知识 k

If k 是未聚类

调用函数 DBSCAN(k)

End For

Else 标记 k 为噪声

迭代 Q 中的知识, 直到没有未聚类的知识

3 实验结果与分析

3.1 改进的 Transformer 模型对比实验

为说明改进的 Transformer 实体抽取模型的有效性, 本文在不同领域数据集上进行验证。以复杂的医学领域 EMR 数据集为例, 将改进的 Transformer 模型与基于传统机器学习方法的 CRF 模型、基于深度学习方法的 Transformer 模型、集成模型 LSTM-CRF 模型和 Lattic+LSTM 模型进行实体识别效果对比。采用精度、召回率和 F_1 值作为评价指标, 实验结果如图 3 所示。改进的 Transformer 模型在不同的指标上均有较高的分数, 分别为 89.67%, 87.54% 和 88.59%。与最优的对比模型相比, 改进的 Transformer 模型在精度、召回率和 F_1 值 3 个评价指标上分别提升了 6.1%, 7.65% 和 6.94%。实验结果表明改进的 Transformer 模型在实体抽取方面具有较高的准确率。

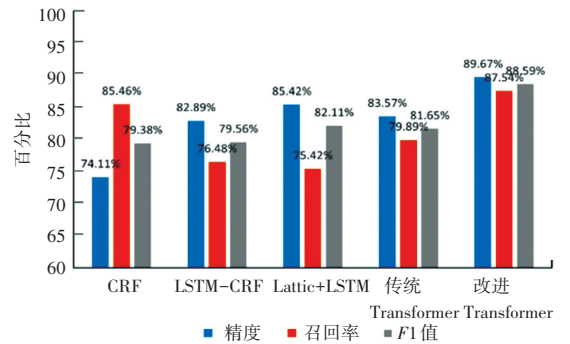


图 3 各模型在 EMR 数据集上的识别效果对比图

Fig. 3 Comparison diagram of recognition effect of each model on EMR dataset

3.2 汽车领域实验数据

本文基于汽车企业生命周期管理中部分项目文档信息, 通过改进 Transformer 模型对文档进行实体识别得出 220 941 个实体, 并根据识别结果进行分类得出 31 个实体类别, 两者共同作为汽车领域数据

集见表1。将每个句子中同时出现的实体都视作潜在的关系实例,实体项集见表2。根据实体抽取结果进行分类处理的结果如图4所示,形成了初步汽车领域知识库。

表1 汽车领域数据集属性示例

Table 1 Example of properties for the automotive domain dataset

编号	实体类	实体名
1	产品名	奥迪 A3、宝马 X1...
2	能源类型	汽油、纯电动...
3	制造商	宝马公司、德国大众汽车...
4	上市时间	2021、2019...
5	环保标准	国 VI、欧 V...
6	级别	中型车、跑车...
7	油耗	5.8、6.1...
8	进气形式	涡轮增压、自然吸气...
9	燃油标号	95号、98号...
...

表2 汽车领域数据集中的项集示例

Table 2 Example of item set in data set of automobile field

项集编号	项集
Item1	'奥迪 A4', '1.8T'
Item2	'宝马 M2', '汽油'
Item 3	'奥迪 A4', '国 VI', '汽油'
Item 4	'奥迪 A4', '承载式'
Item 5	'宝马 M2', '前置后驱'
Item 6	'奥迪 A4', '前置前驱', '涡轮增压'
Item 7	'宝马 M2', '宝马公司', '紧凑型车'
Item 8	宝马 M2', '8 挡手自一体'
Item 9	'宝马 M2', '直喷'
Item 10	'宝马 M2', '1.3T'
Item 11	'奥迪 A4', '铝合金'
...	...

奥迪A6L中央空调和空调面板的触摸阻尼和响应速度非常好。车型为中大型轿车,发动机马力2.0T,环保标准为国家六级标准,油耗为6.8L/100km。



图4 实体抽取及分类结果

Fig. 4 Entity extraction and classification results

本文将31个实体类型进一步拆分成产品的业务需求、设计约束、系统功能、性能评估、结构布局、

参数说明等共8张表格。为了便于知识的聚类,通过设置不同的最小置信度,对挖掘到的关联规则进行筛选,使得每个表中知识的数量接近。各表格信息见表3。若置信度小于最小置信度的阈值,将不会被纳入知识库中,依据表格实际规模和内容的差异,设定了不同的阈值。这样,将有助于基于距离的算法对知识进行更好地聚类。

表3 各表格信息

Table 3 Table information

表格的序列号	最小置信度的阈值	知识的数量
1	0.35	301
2	0.20	240
3	0.04	274
4	0.15	241
5	0.10	302
6	0.05	298
7	0.08	342
8	0.05	308

3.3 知识融合实验分析

为对比改进后的关联规则算法的高效性,将改进算法分别与Apriori算法和Eclat算法在运行时间上进行了对比分析,如图5所示。在8个表格中,改进关联规则算法的运行时间都小于Apriori算法和Eclat算法,整体上比Apriori算法提升了36.16%,比Eclat算法提升了54.81%。由实验结果可知,基于特征交叉的关联规则挖掘算法避免了大量的频繁项集的计算,降低了算法的时间复杂度,可以有效地减少算法的运行时间。

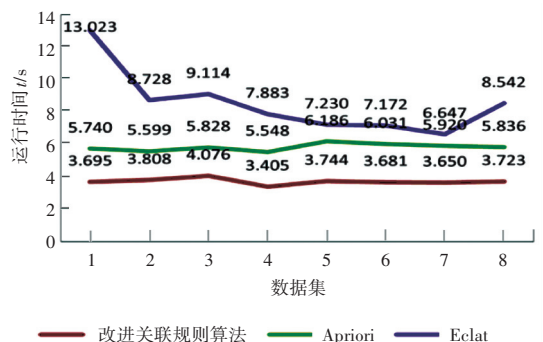


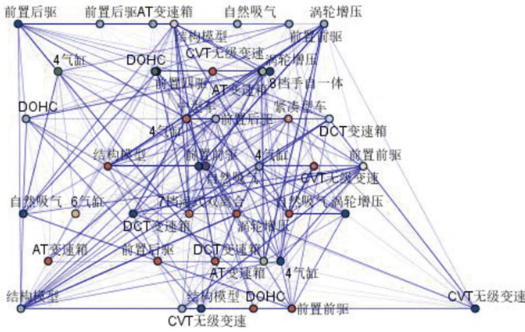
图5 不同关联规则在各数据集上的运行时间比较

Fig. 5 Comparison of running times of different association rules on each dataset

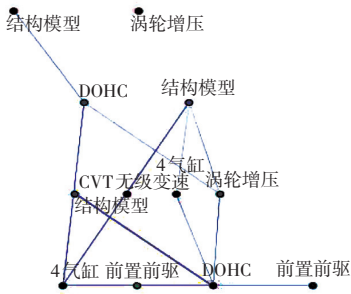
为了更清晰地观察各实体之间的关系,通过统计产品与服务解决方案(Statistical Product Service Solutions, SPSS)软件实现各表格实体间强弱关联关系,如图6所示。图6中强弱关联划分以百分之

五十置信度为界, 强弱关联规则反映了各表格实体之间的潜在关联程度。由此可知, 本文提出的改进关联规则挖掘方法可以有效地挖掘出数据实体之间的隐藏关系。

与融合, 从而提高了知识的价值密度。



(a) 弱关联规则图



(b) 强关联规则图

图 6 基于改进算法挖掘后的强弱关联规则图

Fig. 6 Strong and weak association rule graph mining based on improved algorithm

对实体进行共现网络分析, 判断实体之间是否有语义关系。根据两个实体共同出现的频率判断两个实体之间关联强度, 宝马系列车的实体共现网络图如图 7 所示, 其三维数据分析结果如图 8 所示。图 7 中圆圈大小代表该实体出现频次的多少, 颜色越深代表中心性越高, 联系紧密的实体会相对形成一个小团体, 通过对每个小团体的实体进行归纳总结, 实现聚类分析。图 8 反映了各实体之间的相似度, 由点与点之间的距离表示, 通过三维图中的各部分空间可以发现潜在的性质相近集群。由图 7、图 8 可知, “德国大众汽车”、“慕尼黑工厂”和“宝马公司”出现频率最高, 文档中心为“德国”“一汽”“奥迪”。

基于上述实体间关系的分析, 定义实体间的距离, 根据距离对实体进行聚类的过程, 对初步的汽车领域知识库进行整理, 得出聚类后的新汽车领域部分知识库, 见表 4。

由表 2 和表 4 对比可以看出, 经过改进关联规则和距离聚类处理后的领域知识包含表 2 中多个项集, 减少了知识库中的冗余知识, 实现了知识的获取

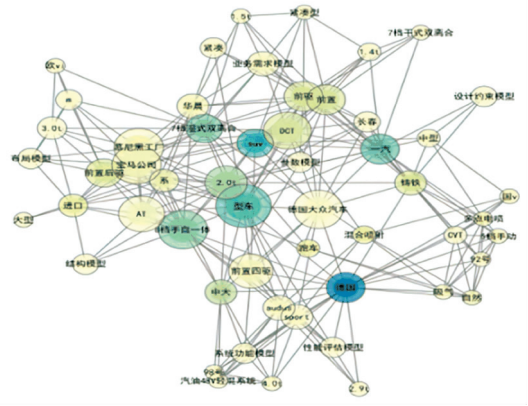


图 7 实体共现网络图

Fig. 7 Co-occurrence network of entities diagram

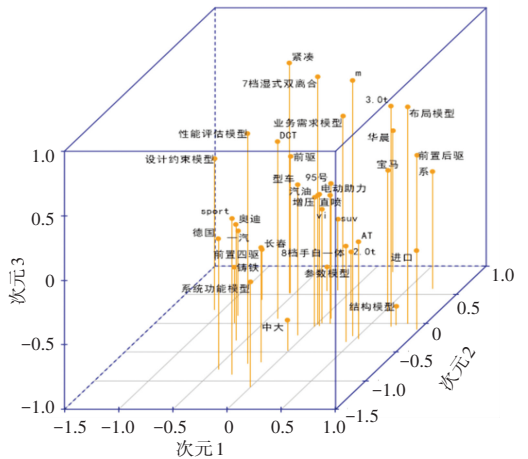


图 8 三维数据分析结果图

Fig. 8 3D data analysis results

表 4 聚类后领域知识库

Table 4 Domain knowledge base after clustering

表格的序列号	聚类后的领域知识
1,2	'奥迪 A4', '1.8T', '国 VI', '汽油', '涡轮增压', '一汽奥迪', '直喷', '前置四驱'
2,3	'冠道', '1.3T', '前置前驱', '涡轮增压', '92 号', '直喷'
3,4	'宝马 M2', '3.0T', '1332', '国 VI', '汽油', '铝合金', '7 档湿式双离合'
4,5	'思域', 'CVT 无级变速', '紧凑型车', '涡轮增压', '东方本田公司', '92 号'
5,6	'汉兰达', '中型 SUV', '国 V', 'DOHC', '前置后驱', '6 档自一体'...
...	...

4 结束语

针对产品生命周期管理系统中的数据模型化处理,本文首先引入位置编码和多头注意力等来改进 Transformer 模型,实验证明改进的 Transformer 模型在实体抽取能力上有较好的表现,精度、召回率和 F_1 值分别达到了 89.67%, 87.54% 和 88.59%。通过改进关联规则的方法对汽车领域数据进行关联规则挖掘,并结合距离聚类算法,对语义相近以及重复的知识进行剔除,进一步减少了冗余知识,对汽车行业领域相关文档的模型化处理结果验证了本文方法的有效性。基于此实现工业产品全生命周期模型知识获取与知识模型构建,将有助于提高管理效率,为实现模型的智能化管理提供技术参考。本文提出的方法仅适用于文本文档,而复杂产品全生命周期管理中的知识获取可能涉及音频、视频、图像等多种资源,因此,未来工作将侧重于:

(1) 针对多模态的复杂产品知识库统一构建方法展开进一步研究;

(2) 研究基于图的方法解决上下文语境导致的实体歧义性问题;

(3) 根据已知实体与实体之间的关系实现预测

(上接第 67 页)

开展了快速切割纸箱的实验研究。对视觉系统进行双目标定及手眼标定,获取双目相机内外参数和手眼矩阵关系;在纸箱轮廓识别与定位过程中采用了基于灰度模板匹配基础上金字塔分层搜索算法,提升系统响应速度,能够准确快速地完成目标纸箱识别;通过对比视觉定位计算的结果和实验数据结果,分析了位置精度产生误差的原因。实验结果表明,该系统具有较高的定位精度,能实现良好的工业应用价值。

参考文献

- [1] WANG W, YAMANE S, KOIKE T, et al. Image processing method for automatic tracking of the weld line in plasma robotic welding[J]. The International Journal of Advanced Manufacturing Technology, 2016, 86(5-8):1865-1872.
- [2] 李金义, 杨成, 王京, 等. 基于视觉定位的机器人搬运系统[J]. 制造业自动化, 2011, 33(4):40-42.
- [3] GIOUROUDI I, HOTZENDORFER H, KOSEL J, et al. Development of a microgripping system for handling of microcomponents[J]. Precision Engineering, 2008, 32(2):148-152.

与推理,对知识图进行补充的相关技术。

参考文献

- [1] 林京. 机器信息学:机械产品智能化的学科支撑[J]. 机械工程学报, 2021, 57(2):11-20.
- [2] 何文韬, 邵诚. 工业大数据分析技术的发展及其面临的挑战[J]. 信息与控制, 2018, 47(4):398-410.
- [3] 罗辉, 卢玲. 面向中文命名实体识别的中文字符表示方法[J]. 小型微型计算机系统, 2023, 44(7):1434-1440.
- [4] 曾兰兰, 王以松, 陈攀峰. 基于 BERT 和联合学习的裁判文书命名实体识别[J]. 计算机应用, 2022, 42(10):7.
- [5] 周利琴, 范昊, 潘建鹏. 网络大数据中的知识融合框架研究[J]. 情报杂志, 2018, 37(1):145-150, 197.
- [6] SMIRNOV A, KASHEVNIK A, PETROV M, et al. Context-based coalition creation in human-robot systems: approach and case study [C]//Proceedings of the Interactive Collaborative Robotics: Second International Conference. 2017:12-16.
- [7] 文家富, 郭伟. 基于知识融合的汽车覆盖件模具设计方法研究[J]. 重庆邮电大学学报(自然科学版), 2018, 30(3):423-430.
- [8] RUTA M, SCIOSCIA F, GRAMEGNA F, et al. A knowledge fusion approach for context awareness in vehicular networks[J]. IEEE Internet of Things Journal, 2018, 5(4):2407-2419.
- [9] 沈宙锋, 苏前敏, 郭晶磊. 基于 XLNet-BiLSTM 的中文电子病历命名实体识别方法[J]. 智能计算机与应用, 2021, 11(8):97-102.
- [10] SHAO T, GUO Y, CHEN H, et al. Transformer-based neural network for answer selection in question answering [J]. IEEE Access, 2019, 7:26146-26156.
- [4] MIURA J, IKEUCHI K. Task-oriented generation of visual sensing strategies in assembly tasks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(2):126-138.
- [5] LI Y F, LEE M H. Applying vision guidance in robotic food handling[J]. IEEE Robotics and Automation Magazine, 1996, 3(1):4-12.
- [6] 季旭全, 王君臣, 赵江地, 等. 基于机器人与视觉引导的星载设备智能装配方法[J]. 机械工程学报, 2018, 54(23):63-72.
- [7] 田鹏飞, 杨树明, 吴孜越, 等. 结合精度补偿的机器人优化手眼标定方法[J]. 西安交通大学学报, 2020, 54(8):99-106.
- [8] 毛星云. OpenCV3 编程入门[M]. 北京:电子工业出版社, 2015:464.
- [9] JDZIEROWSKA M, KOPROWSKI R, S WILCZYŃSKI, et al. A new method for detecting the outer corneal contour in images from an ultra-fast Scheimpflug camera [J]. BioMedical Engineering OnLine, 2019, 18(1):115.
- [10] OSTU N. A threshold selection method from gray-histogram [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1):62-66.
- [11] 刘成刚, 刘二林, 姜香菊. 改进 Canny 算子的列车轮对踏面边缘检测算法[J]. 铁道机车车辆, 2021, 41(1):22-25.