



控制与决策

Control and Decision

ISSN 1001-0920, CN 21-1124/TP

## 《控制与决策》网络首发论文

题目：面向制造领域人机物三元数据融合的本体自动化构建方法  
作者：董津，王坚，王兆平  
DOI：10.13195/j.kzyjc.2020.1298  
收稿日期：2020-09-16  
网络首发日期：2021-04-16  
引用格式：董津，王坚，王兆平. 面向制造领域人机物三元数据融合的本体自动化构建方法[J/OL]. 控制与决策. <https://doi.org/10.13195/j.kzyjc.2020.1298>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 面向制造领域人机物三元数据融合的本体自动化构建方法

董津, 王坚<sup>†</sup>, 王兆平

(同济大学 电子与信息工程学院, 上海 201804)

**摘要:** 当前, 智能制造面临的许多问题都具有不确定性和复杂性, 单纯地利用专家经验和机理模型难以有效解决. 本文面向跨层跨域的复杂制造系统网络化协同控制机制, 提出一种基于本体的人机物三元数据融合方法, 研究复杂制造环境下的人机物三元数据融合建模. 在抽取三元组时, 区别于传统的流水线式抽取方式, 提出一种基于实体-关系联合抽取的模型ErBERT. 该模型首先经过预训练模型BERT进行词序列化, 经过最大池化、全连接和Softmax等操作后, 完成实体识别和关系分类任务, 得到抽取完毕的人机物三元组. 将抽取好的三元组按照规则映射至OWL文件, 最终存储在图数据库中, 实现本体模型的构建. 经实验证明, 经过ErBERT抽取出的三元组有较好的准确率, 达到了通过本体来融合人机物三元数据的目标, 并为实现制造企业人机物三元协同决策与优化提供技术支撑.

**关键词:** 人机物; 制造领域; 数据融合; 本体; 三元组抽取

**中图分类号:** TP273

**文献标志码:** A

**DOI:** 10.13195/j.kzyjc.2020.1298

**开放科学(资源服务)标识码(OSID):**



## Automatic Ontology Construction for Human-Cyber-Physical Data Fusion in Manufacturing Domain

DONG Jin, WANG Jian<sup>†</sup>, WANG Zhao-ping

(College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China)

**Abstract:** At present, many problems of intelligent manufacturing are uncertain and complex, which cannot be solved effectively by expert experience and mechanism model. In this paper, the human-cyber-physical data fusion modeling of complex manufacturing environments is studied, and the ontology is proposed as the approach of the human-cyber-physical data fusion. In the extraction of triplets, a model named ErBERT based on entity-relation joint extraction is proposed, which is different from the traditional pipeline extraction. After word serialization by pre-training model BERT, the model completes entity recognition and relationship classification by max pooling, fully connection and Softmax, and obtains the extracted human-cyber-physical triplets. The extracted triplets are mapped to OWL files according to rules, and finally stored in the graph database to realize ontology construction. The experimental result shows that the triplets extracted by ErBERT has a good accuracy and achieves the goal of fusion of human-cyber-physical data through ontology, which provides theoretical method support for realizing the ternary collaborative decision-making and optimization of human-cyber-physical data.

**Keywords:** human-cyber-physical data; manufacturing; data fusion; ontology; triplets extraction

## 0 引言

随着互联网、大数据、人工智能等技术的迅猛发展, 传统制造业正加速向新一代智能制造迈进<sup>[1]</sup>. 快速变化的市场环境及多元化的用户需求使制造业环境日趋复杂, 提升企业应对复杂环境中不确定性因素的控制与决策水平, 是企业向智能工厂转型中亟待解决的重要科学命题.

伴随着制造系统复杂度日益增加、用户个性化

需求不断增长, 以往的制造体系和制造水平已经难以满足个性化、智能化产品和服务增值升级的需求, 制造系统将由以往的机物二元系统发展为人机物三元系统. 在复杂制造系统中, 人(人力资源)具备不完全信息决策能力的优点和获取深度知识能力差的缺点、机(虚拟信息系统)具备处理海量数据的优点和处理不完全信息能力差的缺点、物(生产物理系统)具备执行能力强的优点和缺乏数据强处理能

收稿日期: 2020-09-16; 修回日期: 2021-02-25.

基金项目: 科技创新2030新一代人工智能重大项目课题“数据驱动的人机物三元协同决策与优化”(2018AAA0101800).

<sup>†</sup>通讯作者. E-mail: jwang@tongji.edu.cn.

力的缺点. 新一代人工智能将人的作用引入到系统中, 可极大提高制造系统处理复杂性、不确定性问题的能力, 有效实现产品及其生产和服务过程的最优化, 人机物三元深度融合将会使人的智慧与机器的智能相互启发性地增长<sup>[2]</sup>.

语义网是由Tim Berners-Lee最先提出的一个概念, 可以使异构的数据信息相关联, 组成语义网络, 从而计算机可以理解和处理网络中的语义信息<sup>[3]</sup>. 本体作为语义网的基础, 是一种能在语义及知识层次上描述数据的概念模型, 用于确定领域内被共同认可的概念, 并给出概念间的相互关系, 从而实现海量多元异构数据的集成、共享与重用<sup>[4]</sup>. 传统的本体构建方法主要依靠领域专家手工构建, 一旦构建的领域本体较为庞大, 会耗费大量的时间和精力. 因此, 如何使用自动化的方式从数据源中抽取信息并构建本体, 减少领域专家的参与, 是当前本体研究的热点之一. 异构数据会造成信息交互的问题, 利用本体进行人机物三元数据的集成与融合, 不仅可以解决操作障碍, 减少数据冗余, 还可以加强数据的推理和决策能力. 本文提出一种基于本体的人机物三元数据融合模型, 并在进行本体三元组的抽取时, 区别于传统流水线式抽取方式, 基于预训练模型BERT (Bidirectional Encoder Representation from Transformers)提出一种实体-关系联合抽取模型ErBERT (Entity and Relationship Extraction with BERT), 从而更好的整合实体及其关系之间的信息. 文章末尾以宝钢热轧生产环节为案例, 使用ErBERT自动构建本体. 结果显示, 本文提出的ErBERT模型是行之有效的.

## 1 相关工作

### 1.1 人机物三元数据融合

数据融合的目的是将多元异构数据进行融合, 使得获得的信息能有效地应用到决策中<sup>[5]</sup>. 传统的数据融合技术, 如模糊集理论、概率论理论以及可信度理论, 对于各有其特点的人机物三元数据缺乏有效的融合方法. 目前国内外已有针对人机物三元系统数据融合的研究, Hussein等人提出DSSoT智能服务框架, 将人与物联网的数据通过动态社会物联网的模型进行融合<sup>[6]</sup>; Misra等人提出一种多变量数据融合学习模型, 通过训练朴素贝叶斯、k近邻、决策树和支持向量机四个分类器, 可以改善数据异构, 提高辅助决策预测精度<sup>[7]</sup>; Bu提出了一种基于张量模型的高阶K-means算法用于人机物三元数据的聚类, 以获得更准确的结果<sup>[8]</sup>; Wang等人提出一系列基于张量的数据融合方法, 最后给出一个综合的人机物

数据融合框架<sup>[9]</sup>; Chen等人对基于强化学习算法的三元数据融合进行研究, 但未涉及语义层面的自动化、智能化融合<sup>[10]</sup>. 上述数据融合方法尝试将人机物数据进行融合, 但忽略了人机物数据之间存在的关联与特征. 针对以上问题, 本文采用语义融合的方法, 将人机物三元数据抽象为语义信息, 用本体形式表示语义, 进行人机物三元数据的融合.

### 1.2 本体自动化构建

构建本体的方式可以分为三类: 手动构建本体、半自动化构建本体和自动化构建本体. 自动化构建本体方法由于其有效性, 逐渐成为研究的热点. Hazber等人定义了基于关系型数据库模式自动构造本体的映射规则<sup>[11]</sup>; Zhao等人提出基于多标签学习模型与关联标签传播的原始结构单词提取方法, 以提高本体关系自动识别精度, 优化本体构建<sup>[12]</sup>; Kethavarapu等人采用基于关键字、基于值的抽取方法对日志文件数据进行集成, 再将其转换为OWL (Web Ontology Language)文件, 从而实现了自动本体的生成<sup>[13]</sup>.

实体-关系抽取是信息抽取、知识图谱以及自然语言处理领域的核心任务和重要环节, 同时也是自动化构建本体中最重要的一步<sup>[14]</sup>. 在本体中, 知识以<entity1, relationship, entity2>的格式保存为结构化三元组, 即实体entity1、entity2之间存在relationship关系. 有监督的实体-关系抽取方法可分为流水线式和联合学习式两种. 前者将命名实体识别 (Named Entity Recognition, NER)和关系分类 (Relation Classification, RC)作为两个独立的子任务, 在完成实体识别之后再行关系的抽取<sup>[15-16]</sup>. 目前国内外研究学者在进行三元组抽取从而实现本体自动化构建时, 大多使用流水线式方法, 这种方法忽略了这两个子任务之间的相关性, 且会造成误差累积. 最近的研究表明, 使用联合学习的方法可以更加紧密的交互实体与关系之间的信息, 很好地解决了流水线式方式存在的问题. Zheng等人使用混合BiLSTM-ED和CNN的神经网络模型, 在实体与关系抽取任务上表现优异<sup>[17]</sup>; Luo等人提出了一种基于Att-BiLSTM-CRF的联合学习方法, 用于生物医学实体和关系提取<sup>[18]</sup>. 这些模型基于实体关系联合训练的方法进行三元组的抽取, 但大多使用公开数据集进行训练, 不针对制造业领域. 本文基于BERT预训练模型, 提出一种实体-关系联合抽取模型ErBERT, 针对特定下游任务对模型进行微调, 并针对钢铁制造行业建立数据集HRDT对模型进行训练, 完成制造业领域的人机物三元组抽取.



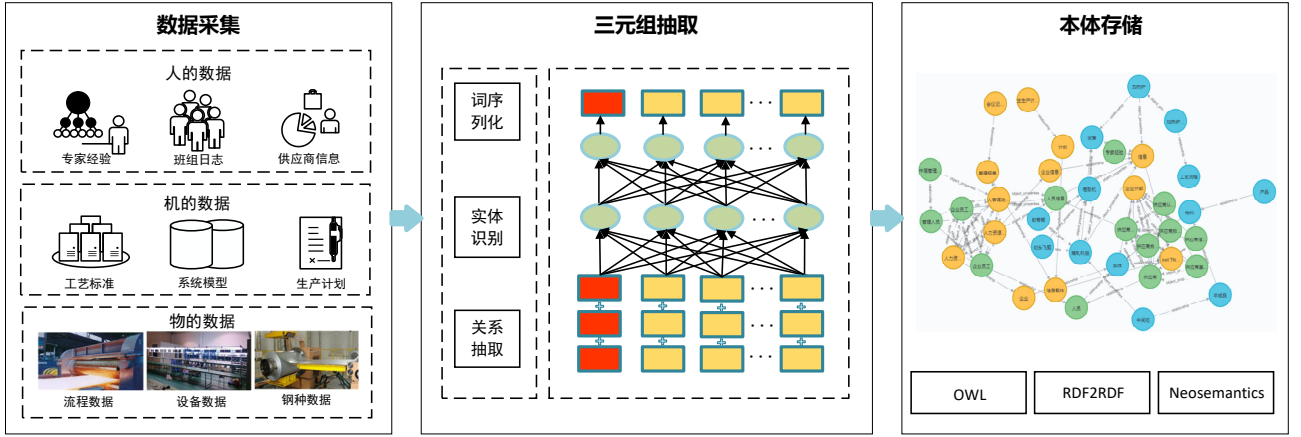


图1 面向人机物三元数据融合的本体自动化构建总体架构

## 2 面向人机物数据融合的本体自动化构建方法

### 2.1 总体架构

本文研究面向制造领域人机物三元数据融合的本体自动构建方法,故立足于制造业,寻找人机物三元数据的各自特征和内在联系是本文研究的必经之路。在制造业领域,人的数据主要指专家经验、供应商及用户信息,同时人具有处理不确定性信息的能力;机可以处理海量数据,但无法处理不确定性知识,数据主要包括数据库里的结构化表单以及企业信息系统内的信息;物的数据来源于客观存在的实体对象,一般指检测设备和生产设备的数据。为了实现人机物三元本体的自动化构建,本文设计了如图1所示的体系架构,包括数据采集、三元组抽取和本体存储三部分。其中,如何从文本数据中自动获取三元组是本文研究的重点内容。

数据采集,是分别以人、机、物作为数据来源,来寻找相关的领域特定概念。人的数据包括专家经验、班组日志以及供应商信息等;机的数据主要来自于企业信息系统,包括工艺制度、生产调度计划以及规则数据等;物的数据由客观存在的数据构成,包括设备数据、钢种数据等。将采集到的文本数据进行切分,以单句的形式输入至三元组抽取模块。

将切分后的单句作为三元组抽取模块的输入,三元组抽取的主要任务是进行命名实体识别和关系抽取。本文提出了基于预训练模型BERT的改进算法ErBERT,实现实体、关系的联合抽取,在下文进行具体介绍。三元组抽取以形如<加热炉,包括,点火器>的三元组形式输出至下一部分。

在完成人机物三元组的抽取之后,将抽取好的人机物三元组按照规则映射至OWL文件,并将其存

储至图数据库中,实现人机物三元本体的自动构建。

### 2.2 基于ErBERT的三元组抽取

三元组抽取是本体的自动化构建任务中极为关键的一步,本文提出的ErBERT模型可以实现实体与关系的联合抽取,其基本流程如图2所示。在词向量化模块中,本文使用了预训练模型BERT进行词序列列化。将可能的实体向量经最大池化处理,在进行全连接和Softmax之后得到实体的类别。关系分类模块在实体抽取的基础上进行,根据上一步实体抽取的结果,将头尾实体向量与头尾实体之间的词向量经过全连接层与Softmax,从而得到关系分类的结果。

对于输入的单句集合,首先使用BERT预训练模型对其进行分词并向量化。BERT是Google AI语言研究人员最近提出的一个预处理模型,在一些自然语言处理任务上表现出超越过往经典模型的优异性能。输出的向量由词向量(Token Embedding)、句向量(Segment Embedding)和位置向量(Position Embedding)组成,相加之后送入双向Transformer结构进行特征提取,最后得到含有丰富语义特征的序列向量。深层双向Transformer的模型结构是BERT模型中的核心,使得模型在进行单词的处理时,能够表征单词在上下文中的具体语义。Transformer编码结构采用多头注意力机制,放弃时间循环结构,可以同时处理整个输入序列,多头注意力机制的输出为:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O. \quad (1)$$

其中,  $W^O$  是一个附加的权重矩阵。且:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V). \quad (2)$$

而:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right). \quad (3)$$

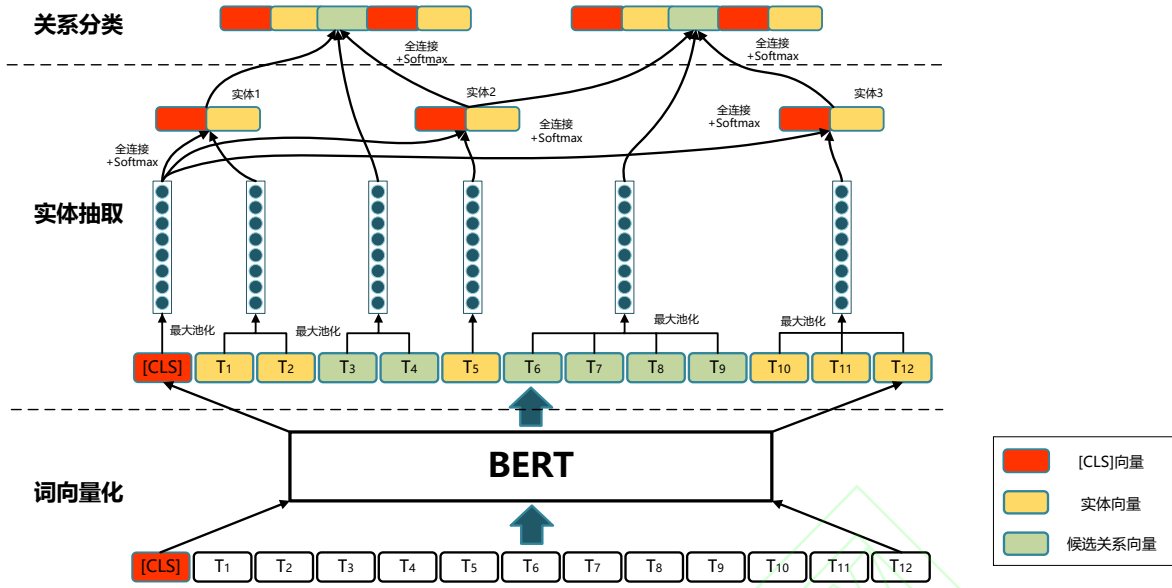


图2 ErBERT模型流程图

公式中,  $Q, K, V$  矩阵为编码器的输入字向量矩阵;  $d_k$  为  $Q, K, V$  矩阵的列数, 即向量维度。

对于一个长度为  $n$  的输入序列,  $t$  位置对应的向量表示以及  $\omega_k$  定义如下:

$$\vec{p}_t^{(i)} = f(t)^{(i)} := \begin{cases} \sin(\omega_k \cdot t), & \text{if } i = 2k \\ \cos(\omega_k \cdot t), & \text{if } i = 2k + 1 \end{cases} \quad (4)$$

$$\omega_k = \frac{1}{10000^{2k/d_k}}. \quad (5)$$

通过BERT预训练模型输出的词序列向量经过最大池化处理, 进行平均和之后, 与 [CLS] 向量进行拼接. [CLS] 是一个能够表示整个文本的语义特征向量, BERT在输入文本前插入 [CLS] 符号, 并将该符号对应的输出向量作为整个文本的语义表示. 与输入文本中已存在的其他字、词相比较而言, [CLS] 作为无明显语义特征的符号, 能够更加公平地融合文本中各个字的含义. 经过拼接得到的实体表示  $x_e$  为:

$$x_e := \frac{\sum_{i=n}^m e_i}{m - n + 1} \circ c. \quad (6)$$

其中,  $e_i$  是经过最大池化处理过后的词向量表示,  $c$  为整个文本的语义表征。

最后, 将得到的实体向量  $x_e$  送入全连接加Softmax层, 全连接层将前面得到的特征做加权和得到每个类别的分数, 再经过Softmax映射为概率, 计算最有可能的实体标记, 公式如下:

$$\hat{y}_e = \text{softmax}(Wx_e + b). \quad (7)$$

完成实体的抽取之后, 进入到关系分类任务. 给

定一组预定义的关系类  $R = \{r_1, r_2, \dots, r_j\}$ , 关系分类模块对处理来自单句中的所有候选实体对, 判断其是否存在来自  $R$  的关系. 关系分类模块输入由两部分组成, 一是在实体识别部分得到的实体向量表示  $x_e$ , 包括词序列向量以及能够表示整个文本的语义特征向量 [CLS]; 二是两个实体之间文本的向量表示, 即候选关系向量, 在进行最大池化处理后与前后的实体向量进行拼接, 送入关系分类器中, 公式如下:

$$x_r = W_s(x_{e1} \circ c_r \circ x_{e2}) + b. \quad (8)$$

$$\hat{y}_r = \text{softmax}(x_r). \quad (9)$$

其中,  $c_r$  是经过最大池化处理过后的词序列向量,  $x_{e1}$ 、 $x_{e2}$  为  $c_r$  前后的实体向量表示。

实体和关系分类问题均使用交叉熵作为损失函数计算loss, 用来衡量同一个随机变量中的两个不同概率分布的差异程度, 即真实概率分布与预测概率分布之间的差异. 交叉熵函数的值越小, 模型预测效果越优异. 关系分类任务的损失函数为:

$$L_r = \sum_{i=0}^m -r_i \log(\hat{y}_r) - (1 - r_i) \log(1 - \hat{y}_r). \quad (10)$$

其中,  $m$  为样本个数,  $r_i$  为类别标签映射而成的one hot向量。

实体-关系联合抽取的损失函数应表示为实体识别损失函数  $L_e$  与关系分类损失函数  $L_r$  之和, 即:

$$L = L_e + L_r. \quad (11)$$

### 2.3 基于图数据库的本体存储

经过 ErBERT 进行三元组抽取得到形如 <加热

炉, 包括, 点火器>、<车间操作工, 相关, 班组工作日志>等实体关系三元组之后, 根据规则建立起人机物三元本体的层级结构, 之后将其映射至OWL文件, 完成本体的构建. 构建完的本体以Web本体语言的形式保存在OWL文件中, 为了实现快速查询, 用于支撑知识推理、知识计算等上层应用, 需要进行有效的本体存储.

图数据库是一种以图论为理论基础的图关系型数据库, 用于存储实体及实体间的关联信息, 其基本组成要素是节点、关系和属性. Neo4j是常用的图数据库之一, 本文采用Neo4j进行本体的存储. 将映射的OWL文件通过开源工具包RDF2RDF转成RDF格式, 再通过Neo4j的功能插件Neosemantics将RDF导入至Neo4j图数据库中, 可完成本体的存储.

### 3 案例验证

#### 3.1 问题描述

随着国家产能政策的优化调整, 供给改革的持续深入, 钢铁行业竞争愈加激烈, 现代生产中多品种、多规格、个性化生产也使得钢铁行业面临日趋复杂的环境. 某钢铁企业连轧产线包括热轧和冷轧两种工艺, 是同时把多个轧机布置在一条生产线上, 从而一次性完成制品的轧制过程, 涉及到的流程复杂, 设备种类与数量繁多, 难以进行有效管理. 热轧生产一般装配步进式加热炉, 连铸板坯先后经加热炉加热、除磷箱去除氧化铁皮后, 进入粗轧机组进行多次往复轧制, 再进行二次除磷后, 进行精轧, 最后由卷取机卷成热轧卷. 冷轧生产以热轧钢卷为原料, 经轧制、退火、酸洗、镀锌及彩图等工序, 最终得到冷轧钢板. 在热轧和冷轧实际生产流程中, 涉及到日生产计划、车间操作人员、库存以及设备信息等人机物数据, 这些数据有的依靠纸张线下传递, 有的通过信息系统存储, 数据之间缺乏关联性, 使得生产车间协同效率低下, 信息同步困难, 决策知识来源单一. 因此, 本课题定位至该企业连轧产线, 收集产线生产过程信息, 研究基于本体的钢铁连轧产线人机物三元数据融合模型, 以增强数据之间的关联性, 旨在促进企业内部信息集成, 从而提高生产车间协同效率, 实现有效的连轧生产管理, 对实现连轧生产协同制造具有重要意义.

#### 3.2 HRDT数据集

目前, 在三元组的抽取训练中, 最常使用的是ACE和CoNLL-04数据集, 但这些数据集没有针对特定工业领域的子集, 且不包含构建本体所需要的层次实体关系. 为了实现人机物三元数据融合的本

体化本体构建, 本文针对钢铁制造行业建立了数据集HRDT. HRDT数据集共有778个单句, 其中有58句来自专家经验, 其他是从网络上爬取或根据企业提供资料整理而成. 此数据集定义了3种类型的实体, 11种类型的关系, 具体如表1、表2所示.

表1 HRDT数据集实体类别信息表

Label	Total	Examples
Person	55	“车间操作工”、“能源供应商”
Cyber	152	“物料库存盘点表”、“物料需求计划”
Physical	126	“传动装置振动信号”、“切头飞剪”

表2 HRDT数据集关系类别信息表

Label	Descriptions	Symmetry
Include	包括...	asymmetric
Belong_2	属于...	asymmetric
Use	使用...	asymmetric
C_about	发生...	asymmetric
C_from	来自...	asymmetric
Is_after	在...之后	asymmetric
Is_before	在...之前	asymmetric
Is_related	与...有关	symmetric
Out	输出...	asymmetric
Enter	进入...	asymmetric
Refer_2	参考...	asymmetric

HRDT数据集按照实体来源将实体类别划分为人、机、物三类; 关系包括层级关系和其他关系. 层级关系有2种, 为Include和Belong\_2, 意为包括、属于, 有明显的层次语义. 在进行本体构建时, 需要定义规则, 将层级关系和其他关系加以区分, 以完成本体的层次结构建立.

#### 3.3 实验结果

本文将HRDT数据集随机分为训练集(80%), 验证集(10%)和测试集(10%)输入至ErBERT模型中进行实验, 采用精确率、召回率和F1值来评价算法的性能, 评价指标定义如下:

$$Precision = \frac{N_{correct}}{N_{all}}. \quad (12)$$

$$Recall = \frac{N_{correct}}{N_{marked}}. \quad (13)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}. \quad (14)$$

其中,  $N_{correct}$ 为预测正确的实体或关系个数,  $N_{all}$ 为预测的实体或关系总个数,  $N_{marked}$ 为标注的实体或关系总个数.

得到实验结果如表3、表4所示. 由表中数据可得, 本文提出的ErBERT模型在实体识别和关系分类任务上有较好的准确率.



表3 实体识别任务实验结果

Label	Precision	Recall
Person	83.06	82.19
Cyber	84.86	84.63
Physical	88.09	87.13
Total	85.34	84.65

表4 关系分类任务实验结果

Label	Precision	Recall
Include	69.75	65.07
Belong_2	66.69	71.52
Use	65.31	69.44
C_about	64.96	63.04
C_from	70.63	67.61
Is_after	64.61	69.01
Is_before	66.63	64.9
Is_related	67.72	65.68
Out	70.38	69.79
Enter	65.9	70.27
Refer_2	67.32	67.98

经ErBERT模型后得到的人机物实体关系三元组通过定义的规则建立起层级结构,映射至OWL文

件后,再将其通过RDF2RDF和Neosemantics存储至Neo4j中。融合后人机物三元数据具体如图3所示,节点的不同颜色代表不同的数据来源,黄色节点为物的数据,包括设备数据、钢种数据等;蓝色节点为机的数据,包括工艺制度、生产调度计划以及规则数据等;绿色节点为人数据,包括专家经验,供应商数据和人员部门信息等。人、机、物的节点相互关联、相互交融,增强了人机物数据之间的关联性,实现了人、机、物三元数据在语义层次上的深度融合。

融合后的本体模型可以解决传统人机物独立运作模式不能充分利用人机物优点、无法发挥协同机制优势的问题。经领域专家的评定,通过ErBERT构建的人机物三元本体起到了融合制造领域人机物三元数据的作用。且研究提出的基于本体的人机物三元数据融合模型能够充分利用本体对多源异构大规模知识的组织和管理优势,有效地进行企业信息集成,提升企业在人机物等更大范畴处理海量数据的综合决策能力,为研究数据驱动的人机物三元协同决策与优化提供了坚实的技术支撑。

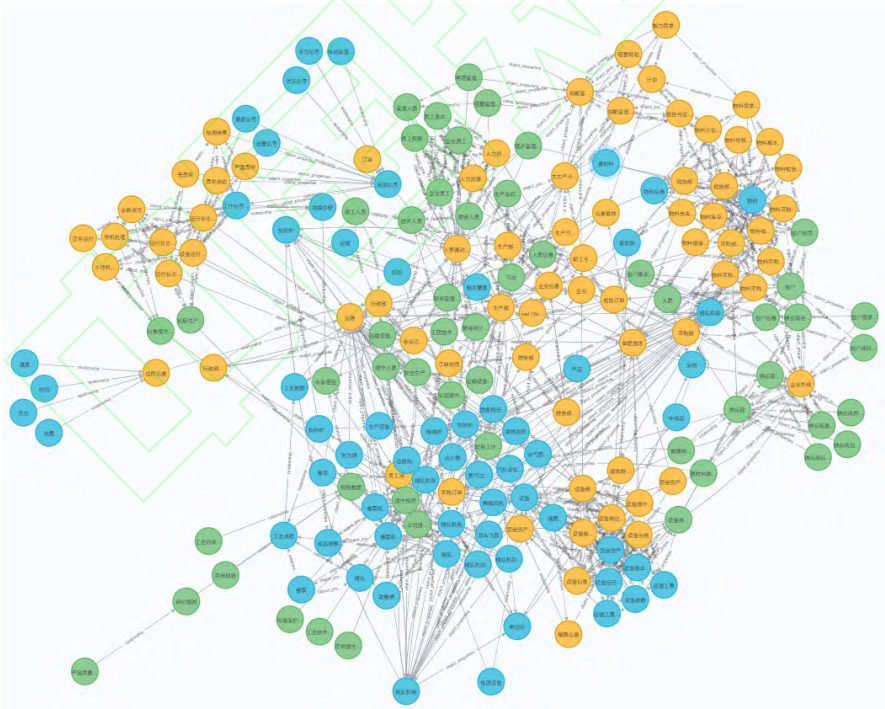


图3 融合后的人机物三元数据

4 结语

本文提出基于本体的人机物三元数据融合模型,并在抽取三元组时,区别于传统的流水线式抽取方式,采用实体-关系联合抽取的模型ErBERT。该模型首先经过预训练模型BERT进行词序序列化,经过最大池化、全连接和Softmax等操作后得到实体与关系的

类别,完成三元组的抽取。随后将抽取好的三元组按照规则映射至OWL文件中,完成本体的构建,最终将三元本体存储在Neo4j中,实现有效的本体存储。

本文提出的实体-关系联合抽取模型ErBERT优势在于使命名实体识别和关系分类任务共享底层神经网络;且在两个任务之间,信息拥有更加紧密的

联系. 实验证明, 经过ErBERT抽取出的三元组有较好的准确率, 且本文最终构建的本体得到了领域专家的一致认可, 达到了通过本体来融合人机物三元数据的目标, 为实现企业人机物三元协同决策与优化提供了技术支撑.

面对复杂多变的制造环境, 基于数据驱动, 发挥人机物协同运行的优势已成为探索自主智能工厂控制与决策科学命题的主导方向. 基于本体的人机物多维工业大数据融合将驱动后续知识发现、智能工厂自学习知识图谱建模及其进化机制、基于自学习知识图谱智能推理的决策与优化理论的建立, 使得企业决策与优化转变为人机物紧密协同合作模式. 在接下来的研究中, 将探索智能工厂自学习知识图谱建模及其进化机制; 同时将进行数据驱动的人机物三元协同决策与优化理论方法研究, 以解决复杂制造环境下企业决策与优化所面临的巨大挑战, 有力地推动企业综合决策与优化科学研究从以专家经验为主向数据驱动的模式转变.

#### 参考文献(References)

- [1] 周济, 周艳红, 王柏村, 臧冀原. 面向新一代智能制造的人-信息-物理系统 (HCPS)[J]. *Engineering*, 2019, 5(4): 71-97.  
(Zhou J, Zhou Y, Wang B C, et al. Human-Cyber-Physical Systems(HCPSs) in the Context of New-Generation Intelligent Manufacturing[J]. *Engineering*, 2019, 5(4): 71-97.)
- [2] 王柏村, 臧冀原, 屈贤明, 董景辰, 周艳红. 基于人-信息-物理系统(HCPS)的新一代智能制造研究[J]. *中国工程科学*, 2018, 20(4): 29-34.  
(Wang B C, Zang J Y, Quan X M, et al. Research on New-Generation Intelligent Manufacturing based on Human-Cyber-Physical Systems[J]. *Strategic Study of CAE*, 2018, 20(4): 29-34.)
- [3] Mohamed A G, Marzouk M M. BIM and Semantic Web-Based Maintenance Information for Existing Buildings[J]. *Automation in Construction*, 2020, 116(2020): 103209.
- [4] Hu L, Wu G, Xing Y, et al. Things2Vec: Semantic Modeling in the Internet of Things with Graph Representation Learning[J]. *IEEE Internet of Things Journal*, 2020, 7(3): 1939-1948.
- [5] Zhang L, Xie Y, Luan X D, et al. Multi-source heterogeneous data fusion[C]. *Proc of the International Conference on Artificial Intelligence & Big Data*: IEEE, 2018: 47-51.
- [6] Hussein D, Park S, Han S, et al. Dynamic Social Structure of Things: A Contextual Approach in CPSS[J]. *IEEE Internet Computing*, 2015, 19(3): 12-20.
- [7] Misra S, Goswami S, Taneja C. Multivariate Data Fusion-Based Learning of Video Content and Service Distribution for Cyber Physical Social Systems[J]. *IEEE Transactions on Computational Social Systems*, 2016, 3(1): 1-12.
- [8] Bu F. A High-Order Clustering Algorithm Based on Dropout Deep Learning for Heterogeneous Data in Cyber-Physical-Social Systems[J]. *IEEE Access*, 2018, 6: 11687-11693.
- [9] Wang P, Yang T, Li J, et al. Data fusion in cyber-physical-social systems: State-of-the-art and perspectives[J]. *Information Fusion*, 2019, 51: 42-57.
- [10] Chen S, Wang J, Li H, et al. Top-Down Human-Cyber-Physical Data Fusion Based on Reinforcement Learning[J]. *IEEE Access*, 2020, 99: 1-1.
- [11] Hazber M A G, Li R, Gu X, et al. Semantic SPARQL Query in a Relational Database Based on Ontology Construction[C]. *Proc of the International Conference on Semantics, Knowledge and Grids (SKG)*, IEEE, 2015: 25-32.
- [12] Zhao L, Ren H, Wan J. Automatic Ontology Construction Based on Clustering Nucleus[J]. *Wuhan University Journal of Natural Sciences*, 2015, 20(2): 129-133.
- [13] Kethavarapu U K, Saraswathi S. A frame work to construct automatic ontology with combine and partitioner logics[C]. *Proc of the 2017 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*: IEEE, 2017: 1-8.
- [14] Hai H, Xiao S, Song M. A Text-Generated Method to Joint Extraction of Entities and Relations[J]. *Applied sciences*, 2019, 9(18): 3795.
- [15] Cai R, Zhang X, Wang H. Bidirectional recurrent convolutional neural network for relation classification[C]. *Proc of the 54th Annual Meeting of the Association for Computational Linguistics*. 2016, 1: 756-765.
- [16] Kazuma H, Makoto M, Yoshimasa T, et al. Simple customization of recursive neural networks for semantic relation classification[C]. *Proc of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013: 1372-1376.
- [17] Zheng S C, Hao Y X, Lu D Y, et al. Joint entity and relation extraction based on a hybrid neural network[J]. *Neurocomputing*, 2017, 257(27): 1-8.
- [18] Luo L, Yang Z H, Cao M Y, et al. A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature[J]. *Journal of Biomedical Informatics*, 2020(103): 103384.

#### 作者简介

董津(1998—), 女, 硕士研究生, 研究方向: 知识图谱、数据融合, E-mail: dongjin@tongji.edu.cn;

王坚(1961—), 男, 教授, 博士, 博士生导师, 研究方向: 工业大数据与工业互联网、工业知识图谱与知识服务等, E-mail: jwang@tongji.edu.cn;

王兆平(1994—), 男, 硕士研究生, 研究方向: 知识图谱、数据融合、智能决策等, E-mail: 1193671148@qq.com.