

文章编号: 1003-0077 (2014) 00-0000-00

## 基于 LM 算法的领域概念实体属性关系抽取\*

刘丽佳<sup>1,2</sup>, 郭剑毅<sup>1,2</sup>, 周兰江<sup>1,2</sup>, 余正涛<sup>1,2</sup>, 邵发<sup>1,2</sup>, 张金鹏<sup>1,2</sup>

(1.昆明理工大学 信息工程与自动化学院, 云南 昆明 650500;

2.昆明理工大学 智能信息处理重点实验室, 云南 昆明 650500)

**摘要:** 针对非结构化自由文本中关系模式比较复杂, 关系抽取性能不高的问题, 该文提出了利用BP神经网络的优化算法-LM算法, 对非结构化自由文本信息中的领域概念实体属性关系进行抽取。首先对语料进行预处理, 然后利用CRFs模型对领域概念的实例、属性和属性值进行实体识别, 然后根据领域中各类关系的特点分别进行特征提取, 构造BP神经网络模型, 利用LM算法抽取相应关系。和适用于二分类问题的SVM相比, 神经网络优化算法自主学习能力强, 识别精度高, 更适用于多分类的问题。通过几组实验表明, 该方法在领域概念实体属性关系抽取方面取得了良好的效果, F值提高了12.8%。

**关键词:** BP 神经网络; LM 算法; 属性关系抽取

中图分类号: TP391

文献标识码: A

## Domain Concepts Entity Attribute Relation Extraction Based on LM Algorithm

LIU Lijia<sup>1,2</sup>, GUO Jianyi<sup>1,2</sup>, ZHOU Lanjiang<sup>1,2</sup>, YU Zhengtao<sup>1,2</sup>, SHAO Fa<sup>1,2</sup>, ZHANG Jinpeng<sup>1,2</sup>

(1. The School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China

2. The Key Laboratory of Intelligent Information Processing, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

**Abstract:** Aim at the problems of complex relation pattern and low relation extraction performance in the unstructured free text, this paper proposed an approach to extract the entity attribute relation from unstructured free text information by applying the LM optimization algorithm of BP neural network. Firstly, preprocessing the corpus, secondly, recognition named entity such as the instance, attributes and attribute values respectively by using CRFs model. And then, extracting the domain features, constructing the BP neural network model, the LM algorithm was used to extract the corresponding relation. Compared to SVM which is suitable for binary classification problems, artificial neural network optimization algorithm is more suitable for multi-classification problems, and its recognition accuracy is higher than SVM. Several groups of tests show that the method in this paper has achieved good effect in the field of entity attribute relation extraction, it improves 12.8% in term of F-score.

**Key words:** BP neural network; LM algorithm; attribute relation extraction

### 1 引言

领域概念实体属性关系(包括实例与属性、实例与属性值、属性与属性值之间的对应关系)是信息检索、社会关系构建、领域本体及知识图谱构建的重要基础。目前国内外的研究主要集中在实例及属性的提取, 或

属性和属性值对的提取。如 David Sánchez<sup>[1]</sup>提出一种以 Web 为检索语料库的自动的、无监督和领域无关的方法对 Web 文本中的领域实例和属性做了提取, 解决了数据稀疏的问题, 在准确率和召回率方面均有所提高; S.Ravi<sup>[2]</sup>利用弱指导的方法从 Web 文档

\*收稿日期: 2014-5-30

定稿日期: 2014-7-28

基金项目: 国家自然科学基金(61175068); 云南省教育厅基金重大专项项目(KKJ201203001); 云南省应用基础研究计划重点项目(2013FA030)

中提取属性，不仅提高了属性提取的精度，还可以自动对提取的属性子集进行标识；李文杰<sup>[3]</sup>利用实例和属性并列结构模式去网页集合中提取同类词语集合，然后再用基于种子的弱指导方法去学习实例和属性共现的上下文模式，最后再通过模式去提取候选实例或候选属性，在进行提取时融入了并列结构这样一种特征，能大大提高系统的召回率；唐伟<sup>[4]</sup>提出了一种基于网页标题和网页结构信息的关系模版自动构建方法，从而实现商品的“属性-值”关系的自动提取，该方法在结构化网页的信息抽取中取得了不错的效果。但是，领域概念的实例、属性和属性值间对应关系不止包括属性-属性值关系，还包括实例-属性关系，实例-属性值关系等等，从完备性角度，只有将这些关系都抽取出来，才能真正满足信息检索、知识图谱构建的实际应用需求。目前关于实例、属性和属性值间的关系抽取主要是在结构化或半结构化文本中进行的，如程显毅<sup>[5]</sup>利用半监督机器学习框架对<实体，属性，值>之间的关系进行抽取，提取过程中利用了Wikipedia的结构化特征和凝聚型层次聚类算法，动态的确定关系类别，完成关系聚类，提高了关系识别系统的性能；杨宇飞<sup>[6]</sup>提出基于弱监督的属性（实例，属性名称和属性值）关系三元组自动抽取方法，利用中文百科条目信息模版中的半结构化属性关系回标条目文本自动获取训练语料，然后利用句子分类器对语料进行训练优化，并在关系抽取过程中加入了触发词词典的特征，提高了关系抽取的召回率，但是只将条目信息作为属性具有一定的局限性。

与Wikipedia、百科条目信息模版中已有的结构化、半结构化文本相比，自由文本的句法表达具有多样性，其中的实例-属性、属性-属性值、实例-属性值等各类关系的模式更复杂，抽取难度更大。目前常用的方法是用支持向量机（SVM）先对实例、属性和属性值等命名实体进行关系判别，把有关系的实体对抽取出来，再制定相应的推理规则，对实体对的具体关系类别进行判别，由于推理规则是人为定义的，不能涵盖所有关系类别的模式，在准确率方面不能有所保

证。

为了解决上述问题，提高关系抽取的性能，本文先对自由文本中的实例、属性和属性值进行命名实体识别，然后尝试利用BP神经网络的方法进行实例-属性、属性-属性值和实例-属性值关系的识别与抽取。BP神经网络是一种前馈反向型网络，其传统算法具有实现复杂的非线性映射的能力、自主学习能力强和可反馈训练等特点，在模式识别、文本分类、图像分类等方面具有广泛应用。但是，传统的BP神经网络算法存在容易陷入局部极小值、收敛速度慢等缺点，为此人们提出了多种改进方案，主要有启发式算法和数值优化算法<sup>[7]</sup>。本文选择优化算法中的LM算法（Levenberg-Marquardt）对各类关系进行识别和抽取。该方法的主要思想是：先对语料进行预处理（分词、词性标注、句子划分、人工标注、命名实体识别等），然后根据预处理的结果枚举找出所有的关系实例（实例-属性关系、属性-属性值关系、实例-属性值关系），并根据领域特点选取实体上下文、类别、位置等体现各类关系的特征构成关系特征集。各类关系的特征作为神经网络的输入样本集，利用LM算法对神经网络的权值和阈值进行调整，然后用实体关系的类别作为分类结果，构造关系分类器。通过多组实验证明，本文提出的方法对自由文本中的实例-属性关系、属性-属性值关系、实例-属性值关系等的识别与抽取具有良好的性能。

## 2 LM 算法介绍

LM算法<sup>[8,9,10]</sup>（Levenberg-Marquardt）是梯度下降法和高斯-牛顿法相结合的一种算法，因此，它既有梯度下降法的全局特性，又有高斯-牛顿法的局部收敛特性，在局部搜索能力上高于传统的BP神经网络算法。并且，LM算法利用了近似的二阶导数信息，比梯度下降法的收敛速度快，比较稳定，其性能优于传统BP神经网络算法，目前在模式识别、信号处理和图像分类等领域的应用均取得较好效果。LM算法简要描述如下。

设 $x_k$ 代表第 $k$ 次迭代的神经网络权值和阈值的输出向量， $x_{k+1}$ 代表新的权值和阈值的输出向量，有：

$$x_{k+1} = x_k + \Delta x \quad (1)$$

对于牛顿法则有:

$$\Delta x = -[\nabla^2 E(x)]^{-1} \cdot \nabla E(x) \quad (2)$$

其中,  $\nabla^2 E(x)$  表示误差指标函数  $E(x)$  的 Hessian 矩阵,  $\nabla E(x)$  表示梯度。

设误差指标函数  $E(x)$  为:

$$E(x) = \frac{1}{2} \sum_{i=1}^N e_i^2(x) = \frac{1}{2} \sum_{i=1}^N (y_i - o_i)^2 \quad (3)$$

其中,  $e_i(x)$  为输出层的误差,

$i = 1, 2, \dots, N$ ,  $y_i$  为输出层的实际输出,  $o_i$  为输出层的期望输出。

则有:

$$\nabla E(x) = J^T(x) e(x) \quad (4)$$

$$\nabla^2 E(x) = J^T(x) e(x) + S(x) \quad (5)$$

其中,  $J(x)$  为 Jacobian 矩阵, 具体表达式为:

$$J = \left[ \frac{\partial e_i(x)}{\partial x_j} \right]_{i \times j} = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \frac{\partial e_2(x)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_N(x)}{\partial x_1} & \frac{\partial e_N(x)}{\partial x_2} & \dots & \frac{\partial e_N(x)}{\partial x_n} \end{bmatrix} \quad (6)$$

$$S(x) = \sum_{i=1}^N e_i(x) \nabla^2 e_i(x) \quad (7)$$

高斯-牛顿法的权值改变公式为:

$$\Delta x = -[J^T(x)J(x)]^{-1} \cdot J^T(x) e(x) \quad (8)$$

LM 算法的权值改变公式为:

$$\Delta x = -[J^T(x)J(x) + \mu I]^{-1} \cdot J^T(x) e(x) \quad (9)$$

其中,  $\mu > 0$  为常数, 是自适应调整参数,  $I$  为单位矩阵。

由上述公式可得, LM 算法的权值调整公式为:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu I]^{-1} J^T(x_k) e(x_k) \quad (10)$$

从公式(10)看出, 当  $\mu=0$  时, LM 算法近似于高斯-牛顿法, 当  $\mu$  比较大时, LM 算法则近似于梯度下降法, 因此  $\mu$  的调整非常重要。对于给定的  $\mu$ , 如果求得的  $\Delta x$  使神经网络的误差  $E(x_k)$  增大, 则  $\mu$  增加, 令

$\mu = \mu_0 \beta$ , 反之, 如果求得的  $\Delta x$  使神经网络的误差  $E(x_k)$  降低, 则  $\mu$  也减小, 令  $\mu = \frac{\mu_0}{\beta}$ ,

$k = k + 1$ , 其中  $\mu_0$ 、 $\beta$  的值是预先定义的。

### 3 关系抽取过程

LM 算法对给定的各类关系进行识别与抽取的流程如图 1 所示。

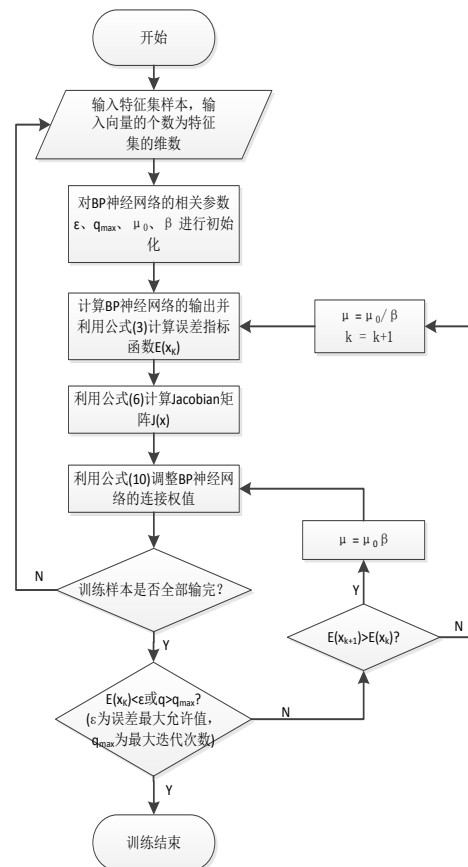


图 1 LM 算法训练关系分类器流程图

下面以旅游领域自由文本中的实例-属性值关系抽取流程为例介绍利用 BP 神经网络的 LM 算法进行关系识别与抽取的原理。

### 3.1 预处理

对语料的预处理包括分词、词性标注、对经过词性标注的语料进行人工标注、命名实体识别、句子切分和提取候选关系实例等。

首先，利用中科院的 ICTCLAS 工具对文本进行分词及词性标注。其次，人工对已经完成分词及词性标注的语料进行命名实体标注，使用最常用的 BMEIO 标注方法，并利用条件随机场（CRFs）对人工标注好的语料进行命名实体识别，并对实体识别结果进行合并。其中，实体的类别是预先定义的，即 in 表示实例，sx 表示属性，pv 表示属性值，根据旅游领域的特点，属性值不仅仅包括地点名词、数字等，很多都是由数字+数量单位组成的，如长、宽、高、海拔等是数字+米/千米，面积是数字+平方千米/平方公里，酒店是数字+星级，门票价格是数字+元/人，人口数是数字+万人等等，因此在命名实体标注的时候，不仅要把数字标注为属性值，数字后面的单位也要标注为属性值，命名实体合并时将它们合并到一起，形成一个完整的属性值。然后，根据文本中的标点符号和上下文等特点，将语料进行切分，分成独立的句子。最后，找出每句话中的所有可能的实体关系对，组成候选关系实例，并对它们的关系类别进行人工标注。其中，预定义的关系包括四类：实例-属性关系、属性-属性值关系、实例-属性值关系和其他关系。至此，语料的预处理完成。如对于句子“萨马贡自然保护区位于维西塔城，面积 243 平方千米，是金沙江上游西岸的重要水源林和野生动物栖息地。”进行预处理之后的结果为：“[萨马贡自然保护区]inB [位于]sxB [维西塔城]pvB, [面积]sxB [243 平方千米]pvB, 是金沙江上游西岸的重要水源林和野生动物栖息地。”，其中，实例-属性值关系的实体有[萨马贡自然保护区]inB、[维西塔城]pvB、[243 平方千米]pvB，它们组成的关系候选实例为：

关系候选实例 1：实体 1--[萨马贡自然保护区]inB、实体 2--[维西塔城]pvB

关系候选实例 2：实体 1--[萨马贡自然保护区]inB、实体 2--[243 平方千米]pvB

关系候选实例 3：实体 1--[维西塔城]pvB、实体 2--[243 平方千米]pvB

### 3.2 构造输入样本集

根据旅游领域自由文本中实例-属性值关系的特点，结合中文领域实体关系抽取中的比较成熟的特征<sup>[11,12]</sup>，本文选取了实体的类别、实体上下文词汇、两个实体的顺序、两个实体的距离、两个实体之间其他实体的个数等特征进行特征集的构造，如上文得到的候选关系实例“萨马贡自然保护区—243平方千米”实体对的特征分别如表1所示。

其他类别的候选关系实例根据各类关系的特点，选取合适的特征进行特征集的构造，过程和上文相同，为了方便构造BP神经网络分类模型，各类关系均选择13类特征。

表1 实例-属性值关系的特征选取实例

特征描述	特征
实体1的类别	实例 (inB)
实体2的类别	属性值 (pvB)
实体1前后两个词的类别	实体1-2: 无(o) 实体1-1: 无(o) 实体1+1: 位于(sxB) 实体1+2: 维西塔城(pvB)
实体2前后两个词的类别	实体2-2: , (o) 实体2-1: 面积(sxB) 实体2+1: , (o) 实体2+2: 是(o)
实体的类别顺序	实例在前, 属性值在后
实体1与实体2的距离	4
两实体间其他实体个数	3

由于特征集样本即为 BP 神经网络的输入样本，因此，特征集样本构造完之后，要转换为向量形式。如对 3.1 中的例子中的关系候选实例：

关系候选实例 1：实体 1--[萨马贡自然保护区]inB、实体 2--[维西塔城]pvB

关系候选实例 2：实体 1--[萨马贡自然保护区]inB、实体 2--[243 平方千米]pvB

关系候选实例 3：实体 1--[维西塔城]pvB、实体 2--[243 平方千米]pvB  
转换为向量形式如表 2 所示。

表 2 输入样本集和期望输出的数值化表达

关系候选实例	输入向量	期望输出
关系候选实例 1	1 3 0 0 2 3 1 2 0 2 1 1 1	3
关系候选实例 2	1 3 0 0 2 3 0 2 0 0 1 4 3	3
关系候选实例 3	3 3 1 2 0 2 0 2 0 0 0 2 1	0

### 3.3 构造 BP 神经网络模型

根据上文构造特征集的过程, 各类关系均选择了 13 类特征作为关系识别的依据, 因此, BP 神经网络模型的输入向量为 13 维。由于关系类别设定有实例-属性关系、属性-属性值关系、实例-属性值关系和其他关系四类, BP 神经网络的输出向量为 4 维, 隐藏层的设定根据经验公式<sup>[13]</sup>

$$h = \sqrt{m+k} + a = \sqrt{13+4} + a \approx 4 + a$$

得到, 其中 a 为 1 到 10 之间的整数, 经过实验验证, 发现当 h=14 时效果最好, 因此, 隐藏层设为 14 个神经元。构造的 BP 神经网络模型如图 2 所示。

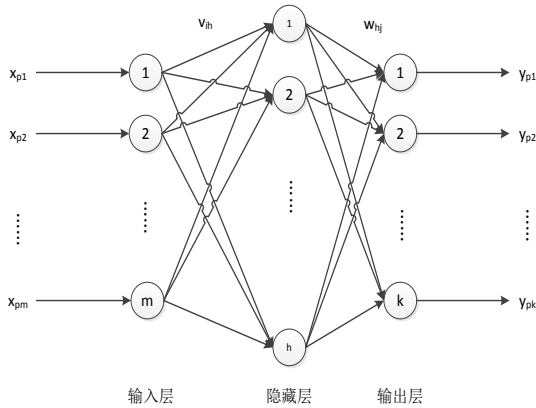


图 2 BP 神经网络拓扑结构图

### 3.4 利用 LM 算法进行关系识别与抽取

LM 算法对自由文本中的实例-属性关系、属性-属性值关系、实例-属性值关系进行识别与抽取的过程为:

(1)对输入样本集数据进行归一化处理, 初始化 BP 神经网络的权值和阈值及相关参数, 迭代次数 q 设为 0, 设定误差最大允许值  $\epsilon$ , 最大迭代次数  $q_{max}$ , 令自适应调整参数  $\mu$  为  $\mu_0$ , k 为 0 等。

(2)根据输入的样本数据计算 BP 神经网络各层的输出结果, 并按照公式(3)计算  $E(x_k)$ ,

(3)利用公式(6)计算 Jacobian 矩阵  $J(x)$ 。

(4)利用公式(9)计算  $\Delta x$ , 并利用公式(10)调整 BP 神经网络的连接权值和阈值。

(5)判断训练样本是否全部输入, 如果已经全部输入, 则执行步骤(6), 反之, 返回(1)。

(6)判断网络误差  $E(x_k)$  是否小于  $\epsilon$ , 或迭代次数 q 是否达到  $q_{max}$ , 如果满足其中一个条件, 则执行步骤(8); 如果两个条件都不满足, 则执行步骤(7)。

(7)判断网络误差  $E(x_k)$  是否增大, 如果增大, 则令自适应调整参数  $\mu = \mu_0 \beta$ , 并且返回步骤(4); 反之, 则令  $\mu = \mu_0 / \beta$ ,  $k = k + 1$ , 并且迭代次数 q 加 1, 返回步骤(2)。

(8)训练结束, 得到最优分类模型。

对于上文中的例子, 利用上述步骤进行关系识别与抽取的结果如表 3 所示。

表 3 实例、属性和属性值的关系抽取结果

实体对	关系
萨马贡自然保护区--面积	实例-属性关系
面积--243平方千米	属性-属性值关系
萨马贡自然保护区--维西塔城	实例-属性值关系
面积--维西塔城	其他关系
.....	.....

## 4 实验设计和结果分析

### 4.1 实验语料

本文的实验语料是旅游领域的自由文本, 是实验室已有语料和在旅游网站上爬取的, 共有2000多篇, 共有88541个实体, 9563个关系实例。采用交叉验证的方法, 随机选取1/10的语料作为测试语料, 9/10的语料为训练语料, 将语料进行预处理, 转化成数值形式的数据, 便于进行BP神经网络分类模型的训练和测试。BP神经网络的相关参数设为:  $\epsilon=0.0001$ ,  $q_{max}=1000$ , 学习率 $\eta=0.01$ , LM 算法的参数设为:  $\mu_0=0.001$ ,  $\beta=10$

### 4.2 评测标准

本文采用最常用的评测标准, 即准确率 P、召回率 R 和 F 值作为最终的实验评价指标。其定义为:

$P = \text{正确抽取的实体关系数} / \text{抽取出的实体关系总数} * 100\%$

$R = \text{正确抽取的实体关系数} / \text{标准结果中的实体关系总数} * 100\%$

$F = 2 * P * R / (P + R)$

### 4.3 实验设计

(1)实验一：为了验证基于LM算法的BP神经网络分类模型在不同特征集和不同大小的训练语料集下对实例-属性关系、属性-属性值关系和实例-属性值关系等对应关系的提取结果，逐渐增加特征集和训练语料集并分别进行对比。实验一结果如表4、表5所

示。

由表4实验结果可以看出，随着特征的不断增多，实验的抽取效果越来越好，说明了扩大特征集会提高关系提取的效果；由表5实验结果可以看出，随着训练语料集的增加，实验的抽取效果也在提高，不过，当训练语料集达到一定的规模时，实验的抽取效果的提高速度会变慢。所以，当语料集达到一定规模后，应该集中到特征的选取和特征集的扩展上。

表4 本文的方法在不同特征集下的实验抽取结果

特征集	关系类别	准确率P(%)	召回率R(%)	F值(%)
实体及上下文词的类别特征 (基准特征)	实例-属性关系	68.2	67.8	68.0
	属性-属性值关系	58.3	58.9	58.6
	实例-属性值关系	59.1	54.8	56.9
	其他关系	66.9	69.5	68.2
	总体评测	<b>63.1</b>	<b>62.3</b>	<b>62.7</b>
基准特征+顺序特征	实例-属性关系	72.8	72.0	72.4
	属性-属性值关系	63.9	67.5	65.7
	实例-属性值关系	76.9	64.9	70.4
	其他关系	79.3	82.0	80.6
	总体评测	<b>73.2</b>	<b>71.6</b>	<b>72.8</b>
基准特征+顺序特征+距离特征	实例-属性关系	83.1	81.4	82.2
	属性-属性值关系	79.8	74.2	76.9
	实例-属性值关系	77.2	76.9	77.0
	其他关系	82.7	83.7	83.2
	总体评测	<b>80.7</b>	<b>79.0</b>	<b>79.8</b>
基准特征+顺序特征+距离特征+结构特征	实例-属性关系	91.3	92.1	91.7
	属性-属性值关系	86.4	83.9	85.1
	实例-属性值关系	87.6	84.6	86.0
	其他关系	87.4	90.3	88.8
	总体评测	<b>88.1</b>	<b>87.7</b>	<b>87.9</b>

表5 本文方法在不同语料集下的实验抽取结果

语料集	准确率P(%)	召回率R(%)	F值(%)
621K	72.9	71.6	72.2
1045K	84.5	82.1	83.3
1809K	88.1	87.7	87.9

(2)实验二：利用BP神经网络的传统算法（梯度下降算法）和LM算法进行实例-属性、属性-属性值、实例-属性值的关系抽取，和SVM与规则相结合的方法的抽取结

果进行对比。SVM方法在实体语义关系抽取方面的应用已经比较成熟，是一种比较好的方法，BP神经网络作为一种新兴的实体关系抽取的方法，应用还不是很广泛，在实体关系抽取方面结果如何，通过和SVM与规则相结合的抽取结果进行对比，确定BP神经网络在实体关系抽取方面是否具有可行性。实验二结果如表6所示。

表6 实例、属性和属性值二元关系实验抽取结果

实验方法	准确率	召回率	F值(%)
	P(%)	R(%)	
SVM+规则	75.4	74.7	75.1
传统的BP算法	55.6	43.8	49.0
本文的方法	88.1	87.7	87.9

由表6实验结果可以看出,在实例、属性和属性值之间的关系抽取方面,传统的BP神经网络算法的分类效果很差,这是因为传统的BP算法即梯度下降法具有易陷入局部最小化和收敛速度慢等缺点,实验过程中,迭代次数达到最大值1000时网络误差仍没达到要求,训练时间为549s。基于LM算法的BP神经网络的分类效果优于SVM分类器与规则相结合的方法的效果,这是由于实例、属性和属性值之间的所有特征组合属于非线性特征组合,而BP神经网络具有复杂的非线性映射能力,LM算法收敛速度比传统的BP神经网络算法快得多,当迭代次数为27时就满足误差指标训练时间只有82s,因此,基于LM算法的BP神经网络更适合解决多分类问题。

## 5 结论

本文提出了利用BP神经网络的优化算法LM算法对非结构化自由文本进行实例、属性和属性值之间的关系进行抽取的方法,该方法利用BP神经网络非线性映射能力强,自主学习和反馈训练的特点,提高了关系提取的准确率,取得了良好的抽取效果,并且通过实验验证了该方法的可行性。下一步工作将在不断收集数据、增加实验语料的基础上,尝试利用半监督的深度学习算法实现对领域实例属性关系的抽取学习,以期进一步提高关系抽取的性能。

## 参考文献

[1] David Sánchez. A methodology to learn

作者简介:



刘丽佳(1988—),女,硕士研究生,主要研究领域为自然语言处理、信息抽取。  
Email: 839856859@qq.com

ontological attributes from the Web[C]. Data & Knowledge Engineering, Volume 69, Issue 6, June 2010, Pages 573-597.

[2] S. Ravi, M. Pasca. Using structured text for large-scale attribute extraction[C]. Proceedings of the 17th CIKM(CIKM 2008), Napa Valley, California, 2008, 1183-1192.

[3] 李文杰, 穗志方. 基于并列结构的实例和属性的同步提取方法[J]. 中文信息学报, 2012, 26(2)

[4] 唐伟, 洪宇等. 网页中商品“属性-值”关系的自动抽取方法研究[J]. 中文信息学报, 2013, 27(1), 21-29.

[5] 程显毅, 朱倩. 未定义类型的关系抽取的半监督学习框架研究[J]. 南京大学学报: 自然科学版, 2012, 48(4): 466-474.

[6] 杨宇飞, 戴齐, 贾真等. 基于弱监督的属性关系抽取方法[J]. 计算机应用, 2014, 34(1): 64-68.

[7] 王建梅, 覃文忠. 基于L-M算法的BP神经网络分类器[J]. 武汉大学学报(信息科学版), 2005, 10: 928-931.

[8] Manolis I. A. Lourakis. A Brief Description of the Levenberg-Marquardt Algorithm Implemented by levmar. Institute of Computer Science Foundation for Research and Technology - Hellas (FORTH). Vassilika Vouton, P. O. Box 1385, GR 711 10 Heraklion, Crete, GREECE. February 11, 2005

[9] 张长胜, 欧阳丹彤, 岳娜, 张永刚. 一种基于遗传算法和LM算法的混合学习算法[J]. 吉林大学学报(理学版), 2008, 04: 675-680.

[10] Janti Shawash, David R. Selviah. Real-Time Nonlinear Parameter Estimation Using the Levenberg-Marquardt Algorithm on Field Programmable Gate Arrays[J]. IEEE Transactions on Industrial Electronics, Volume 60, No.1, January 2013, Pages 170-176.

[11] 董静, 孙乐, 冯元勇. 中文实体关系抽取中的特征选择研究[J]. 中文信息学报, 2007, 21(4): 80-85.

[12] 车万翔, 刘挺, 李生. 实体关系自动抽取[J]. 中文信息学报, 2005, 19(2): 1-6.

[13] Pengyi Gao, Chuanbo Chen, Sheng Qin. An optimization method of hidden nodes for neural network[J]. 2nd International Workshop on Education Technology and Computer Science, ETCS 2010, v2, p53-56.



郭剑毅(1964—),女,硕士生导师,教授,主要研究领域为自然语言处理、信息抽取等。  
通讯作者  
Email: gjade86@hotmail.com