

基于注意力机制的LSTM的语义关系抽取*

王红, 史金钊[†], 张志伟

(中国民航大学 计算机科学与技术学院, 天津 300300)

摘要: 目前关系抽取方法中,传统深度学习方法存在长距离依赖问题,并且未考虑模型输入与输出的相关性。针对以上问题,提出了一种将LSTM(long short-term memory)模型与注意力机制相结合的关系抽取方法。将文本信息向量化,提取文本局部特征,再将文本局部特征导入双向LSTM模型中,通过注意力机制对LSTM模型的输入与输出之间的相关性进行重要度计算,根据重要度获取文本整体特征;最后将局部特征和整体特征进行特征融合,通过分类器输出分类结果。在SemEval-2010 task 8语料库上的实验结果表明,该方法的准确率和稳定性较传统深度学习方法有进一步提高,为自动问答、信息检索以及本体学习等领域提供了方法支持。

关键词: 文本信息; 语义关系; 关系抽取; LSTM; 注意力机制

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2018)05-1417-04

doi:10.3969/j.issn.1001-3695.2018.05.029

Text semantic relation extraction of LSTM based on attention mechanism

Wang Hong, Shi Jinchuan[†], Zhang Zhiwei

(School of Computer Science & Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In the methods of relation extraction, the traditional deep learning method has the problem of long distance dependence and does not consider the correlation between input and output of the model. This paper put forward a new relation extraction model, which combining LSTM and attention mechanism. Firstly, the model embedded the text information and then obtained the local feature. Secondly, it introduced the local feature into the bidirectional LSTM model, and used the attention mechanism to calculate the importance probability between the input and output of the LSTM model to obtain the global feature. Finally, it fused the local feature and the global feature and obtained the result of relation extraction by classifier. Experiments were conducted on the SemEval-2010 task 8 corpus. The results show that the accuracy and stability of the method have been further improved, which provides method support for automatic question answering, information retrieval and ontology learning.

Key words: text information; semantic relation; relation extraction; LSTM; attention mechanism

0 引言

信息抽取^[1,2]是把文本里包含的信息进行结构化处理,变成表格一样的组织形式。信息抽取包含概念抽取和关系抽取两个关键技术,关系抽取^[3,4]是自动识别一对概念和联系这对概念的关系构成的三元组。关系抽取对于海量数据处理、语义网构建以及信息检索等领域具有很大的支持作用,目前主流的关系抽取方法包括传统方法^[5-9]和深度学习方法,深度学习方法相较于传统方法可以获取更多有效的文本特征,其准确率和召回率比较高,国内外学者在深度学习方法上进行了大量的研究。Socher等人^[10]使用递归神经网络(recurrent neural networks, RNN)^[11]的方法来解决关系抽取问题,该方法有效地考虑了句子的句法结构信息,但未考虑两个实体之间的位置信息。Zhang等人^[12]采用基于文本位置信息的RNN方法进行关系抽取,RNN在新的时间状态下不断叠加输入序列会导致前面的输入信息变得越来越模糊即存在长距离依赖问题,并不能有效地考虑上下文语义关系。Zeng等人^[13]提出基于实体位置

信息的卷积神经网络(convolutional neural networks, CNN)^[14]方法进行关系提取,CNN考虑了多个n-gram特征,在一定程度上可以缓解长距离依赖问题,但由于滤波器一般不会选择太大,所以也不能很好地解决长距离依赖问题。另外,RNN和CNN都未考虑模型输入与输出的相关性,使得关系抽取准确率、稳定性差。LSTM是由Hochreiter等人^[15]提出的一种RNN的改进模型,它通过构建专门的记忆单元来存储历史信息,使得每个时间状态都保存了前面的输入信息,可以有效缓解RNN和CNN的长距离依赖问题。该模型已被应用到了自然语言领域,并取得了很好的效果。Sundermeyer等人^[16]提出了一种基于LSTM的语言模型,任智慧等人^[17]采用LSTM进行中文分词,Miwa等人^[18]将LSTM引入SPTree中进行关系抽取,但是LSTM未考虑模型输入对输出的影响。

注意力机制^[19]是由Treisman和Gelade提出的一种模拟人脑注意力机制的模型,它通过计算注意力概率分布,可以突出某个关键性输入对于输出的影响作用,对于传统的模型具有很好的优化作用。因此,本文提出了一种基于注意力机制的LSTM(At_LSTM)的文本语义关系抽取方法,旨在有效解决采

收稿日期: 2017-06-02; 修回日期: 2017-07-17 基金项目: 国家自然科学基金资助项目(U1633110, U1533104, U1233113)

作者简介: 王红(1963-),女,重庆万州人,教授,硕士,主要研究方向为本体技术、数据挖掘与智能信息处理等;史金钊(1992-),女(通信作者),山西运城人,硕士,主要研究方向为人工智能、数据挖掘、自然语言处理等(sjc18234123546@163.com);张志伟(1992-),男,山西运城人,硕士,主要研究方向为人工智能、数据挖掘、自然语言处理等。

用 RNN 和 CNN 进行关系抽取准确率、稳定性差的问题。

1 基于 At_LSTM 的文本语义关系抽取模型

基于 At_LSTM 的文本语义关系抽取模型,采用 LSTM 模型避免传统深度学习方法长距离依赖问题,同时采用注意力机制有效分析了模型输入与输出的相关性,从而获取了更多的上下文语义信息。模型框架如图 1 所示,该模型主要包括两部分:

a) 词向量的生成。对语料库中文本信息进行分词,提取文本实体位置、词性标注、词依赖关系等实体特征,然后将这些实体特征和文本信息转换为词向量,提取出文本局部特征。

b) At_LSTM 构建。将文本局部特征导入双向 LSTM 模型中,采用注意力机制计算注意力概率,对 LSTM 模型输入和输出的相关性进行重要度分析,根据注意力概率获取双向 LSTM 输出特征;对引入注意力机制后 LSTM 的输出特征进行最大池化处理,获取文本整体特征;融合文本局部特征和文本整体特征,再将融合后的特征导入分类器中进行分类,输出分类结果。

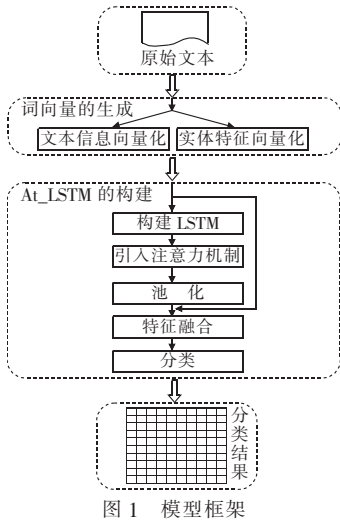


图 1 模型框架

At_LSTM 的构建是研究的核心内容,将 LSTM 与注意力机制结合是研究的难点。

2 模型求解

2.1 词向量的生成

构建文本词向量是将文本信息转换为向量形式,每个句子转换为一个多维矩阵。给定句子 S ,其中包含单词集 $W(w_1, w_2, \dots, w_m)$,提取到的句子 S 的文本特征集 $K(k_1, k_2, \dots, k_n)$, m 表示句子 S 中单词的个数, n 表示每个句子提取到的文本特征个数,第 t 个单词提取到的第 i 个文本特征表示为 $w_i^k (1 \leq i \leq n)$,本文采用的文本特征包括位置标志、词性标注、名实体标志、依赖关系和上下位关系。对文本信息进行词向量化处理:

$$r^w = W^{\text{word}} \times v^w \quad (1)$$

其中: r^w 是词 w 的词向量表示; $W^{\text{word}} \in \mathbb{R}^{l \times |m|}$ 表示文本词向量矩阵,其中 m 表示句子中单词的个数, l 表示词向量的维数,本文采用 word2vec 训练文本词向量矩阵,词向量维数为 300; v^w 是词 w 的 one-hot 表示。

同理,对各个文本特征进行词向量化处理:

$$r^k_i = W^k_i \times v^k_i \quad (2)$$

其中: r^k_i 是文本第 i 个特征的词向量表示; W^k_i 是文本的第 i 个特征的特征分布向量, $W^k_i = (w_1^k_i, w_2^k_i, \dots, w_m^k_i)$ 。每个词对应的

向量化表示为各个向量的连接,第 t 个词对应的向量化表示为

$$x_t = [r_t^w, r_{t1}^k, r_{t2}^k, \dots, r_{tn}^k] \quad (3)$$

最后得到的文本局部特征为 $e = \{x_1, x_2, \dots, x_m\}$ 。

2.2 At_LSTM 的构建

At_LSTM 的构建主要包括构建 LSTM、引入注意力机制、池化、特征融合和分类这五步,其中构建 LSTM 和引入注意力机制的过程如图 2 所示。

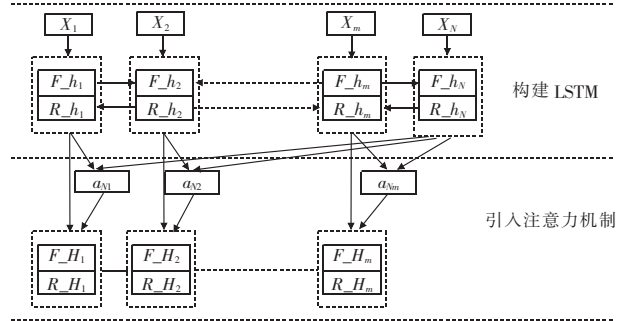


图 2 引入注意力机制的双向 LSTM 网络架构

2.2.1 构建 LSTM

LSTM 模型用来处理序列标注问题,能够充分利用整个文本序列的信息,包括各个词之间的相互关系信息,并将该种信息用于对每个词的处理当中去。一个 LSTM 模型包含很多 LSTM 单元,每个 LSTM 单元包含一个输入门、一个输出门、一个遗忘门和一个记忆单元。设 c 表示 LSTM 模型的记忆单元, x 为 LSTM 模型的输入门, f 为 LSTM 模型的遗忘门, h 为 LSTM 模型的输出门。以文本局部特征 $\{x_1, x_2, \dots, x_m\}$ 作为输入,以第 t 个词为例,激活记忆单元,获取第 t 个词的 LSTM 单元各个状态的特征值如下:

$$\bar{F}_{-c_t} = \text{rec}(W_{xc}x_t + W_{hc}F_{-h_{t-1}} + b_c) \quad (4)$$

$$F_{-i_t} = \sigma(W_{xi}x_t + W_{hi}F_{-h_{t-1}} + W_{ci}F_{-c_{t-1}} + b_i) \quad (5)$$

$$F_{-f_t} = \sigma(W_{xf}x_t + W_{hf}F_{-h_{t-1}} + W_{cf}F_{-c_{t-1}} + b_f) \quad (6)$$

$$F_{-c_t} = F_{-f_t} \odot F_{-c_{t-1}} + F_{-i_t} \odot \bar{F}_{-c_t} \quad (7)$$

$$F_{-o_t} = \sigma(W_{xo}x_t + W_{ho}F_{-h_{t-1}} + W_{co}F_{-c_{t-1}} + b_o) \quad (8)$$

$$F_{-h_t} = F_{-o_t} \odot \tanh(F_{-c_t}) \quad (9)$$

其中: rec 为激活函数; W 表示 LSTM 的权值矩阵; b 表示 LSTM 的偏置向量; σ 为 sigmoid 函数; \odot 代表逐点乘积。

同理,构建反向的 LSTM 模型后,获取的第 t 个词的反向 LSTM 单元各个状态的特征值分别记为 R_{-i_t} 、 R_{-f_t} 、 R_{-c_t} 、 R_{-o_t} 、 R_{-h_t} 。双向 LSTM 每个词的输出特征都包含了整个句子的信息。

2.2.2 引入注意力机制

利用注意力机制计算注意力概率。注意力概率可以突出特定的单词对于整个句子的重要程度,引入注意力机制考虑了更多的上下文语义关联。根据第 2.1 节中生成的文本局部词向量 e 生成新的词向量特征 $e' = \{x_1, x_2, \dots, x_m, x_N\}$,其中 x_N 是前 n 个向量迭加后的平均值。以 LSTM 单元的输出门最后输出为例,将 x_N 导入 LSTM 模型中,获取输出特征值 F_{-h_N} 。第 t 个词对第 N 个词的注意力概率 a_{Nt} 为

$$F_{-h_{Nt}} = U_a \times \tanh(U_b \times F_{-h_N} + U_c \times F_{-h_t} + b_a) \quad (10)$$

$$a_{Nt} = \frac{\exp(F_{-h_{Nt}})}{\sum_{j=1}^m \exp(F_{-h_{Nj}})} \quad (11)$$

其中: U_a 、 U_b 和 U_c 表示注意力机制的权值矩阵; b_a 表示注意力机制的偏置向量。然后获取第 t 个词的新的输出特征值为

$$F_{-H_t} = \sum_{i=1}^m a_{Ni} F_{-h_{ti}} \quad (12)$$

同理,采用上述方法计算第 t 个词的新的 LSTM 单元特征值 F_{-C_t} 以及反向 LSTM 新的特征值 R_{-H_t} 和 R_{-C_t} 。第 t 个词引入注意力机制以后对应的输出特征为 $F_t = \{F_{-H_t}, F_{-C_t}, R_{-H_t}, R_{-C_t}\}$ 。

2.2.3 池化

池化是对模型输出结果进行统计计算,采用最大池化方法对整个句子引入注意力机制以后对应的输出特征 $F = \{F_1, F_2, \dots, F_m\}$ 进行池化。

$$d = \max(F) \quad (13)$$

池化后获得文本整体特征 d , 不论句子长度是多少,池化后的特征维数都是固定的,这样就解决了文本句子长度不一的问题。

2.2.4 特征融合、分类

特征融合是将多种特征融合为一种特征,可以达到多种特征优势互补的效果。将文本局部特征与文本整体特征进行融合,获取新的特征 F' ; 再将融合后的特征 F' 导入分类器进行分类,输出分类结果 $s(x)$ 。

$$F' = e \otimes d \quad (14)$$

$$s(x) = g(V_s(F' \odot M_s) + b_s) \quad (15)$$

其中: \otimes 代表向量拼接; $F' \odot M_s$ 表示分类器的 dropout 处理, M_s 是与 F' 同形状的每轮迭代随机产生的二值向量; V_s 表示分类器的权值矩阵; b_s 表示分类器的偏置向量; g 表示分类器的选择。输出 $s(x)$ 是 l 维的向量,第 i 维表示属于第 i 类的概率。

2.3 模型训练

将关系分类问题看成是一个多分类问题判断。为了获取最优的模型,本文通过最小化负对数似然函数来进行模型训练。假设目标关系得到的关系标签类型 $\mu \in T$, 经分类器处理后为 $s(x)$, T 代表关系标签,经过分类器处理后转换为条件概率。目标关系 y 的标签类型是 μ 的概率为

$$p(y = \frac{\mu}{s(x)}, \theta) = \frac{\exp(s_\theta(x))_\mu}{\sum_{i \in T} \exp(s_\theta(x))_i} \quad (16)$$

采用随机梯度下降算法来最小化负对数似然函数,模型优化的目标函数 $J(\theta)$ 计算如下:

$$J(\theta) = - \sum_{i=1}^D \log(p(y = \frac{y_i}{x_i}, \theta)) \quad (17)$$

其中: $\theta = (W, U, V)$ 表示模型可训练参数; D 表示训练样本个数; (x_i, y_i) 表示训练样本中第 i 个样本 x_i 对应的关系标签是 y_i 。

综上所述,基于 At_LSTM 的关系抽取模型构建的算法步骤如下:

a) 采用 NLPPIR 对原始文本分词,获取文本单词集 W ,提取文本特征获取特征分布矩阵 W^k ,采用 word2vec 训练生成词向量矩阵 W^{word} 。

b) 根据式(3)计算文本词向量,获取文本局部特征 e 。

c) 使用 Xavier 初始化模型参数 θ 。

d) 对 e 进行 dropout 处理。

e) 根据式(7)和(9)计算正向 LSTM 模型的输出特征 F_{-C_t} 和 F_{-H_t} 。

f) 根据式(11)计算注意力概率 a_{N_t} 。

g) 根据式(12)结合注意力概率 a_{N_t} 计算正向 At_LSTM 的输出特征 F_{-C_t} 和 F_{-H_t} 。

h) 重复步骤 c), 提取反向 LSTM 模型的输出特征 R_{-C_t} 和 R_{-H_t} ; 重复步骤 d), 计算反向 At_LSTM 的输出特征 R_{-C_t} 和 R_{-H_t} 。

i) 采用式(13)对 At_LSTM 的输出特征进行池化,计算文本整体特征 d 。

j) 根据式(14)将 e 和 d 进行特征融合,计算结合了文本整体特征与局部特征优势的新特征 F' 。

k) 根据式(15)计算分类结果 $s(x)$ 。

l) 判断是否满足模型优化条件,满足则输出模型,否则返回步骤 e)。

3 实验结果与分析

3.1 实验数据

采用 ACL SigLex 组织提供的 SemEval-2010 task 8^[20] 公共语料库进行实验,该语料库共包括 8 000 个训练样本和 2 717 个测试样本,包含了 10 大类关系。样本关系如表 1 所示,样本示例如表 2 所示。

表1 样本关系

关系类型	样本个数
其他关系(other)	1 864
因果关系(cause-effect)	1 331
整体与部分关系(component-whole)	1 253
实体与目标关系(entity-destination)	1 137
实体与来源关系(entity-origin)	974
生产者与产品关系(product-producer)	948
会员与组织关系(member-collection)	923
实体与主题关系(message-topic)	895
内容与包含关系(content-container)	732
工具与使用者关系(instrument-agency)	660

表2 样本示例

样本编号	样本内容
8001	The most common <e1> audits </e1> were about <e2> waste </e2> and recycling.
8002	The <e1> company </e1> fabricates plastic <e2> chairs </e2>.
8003	The school <e1> master </e1> teaches the lesson with a <e2> stick </e2>.

3.2 实验评价指标

本文采用正确率(precision)、召回率(recall)和 F 值(F_score)三个标准作为模型性能评价指标。指标计算如下:

$$\text{precision} = \frac{\text{out_cor}}{\text{out_all}} \quad (18)$$

$$\text{recall} = \frac{\text{out_cor}}{\text{this_all}} \quad (19)$$

$$F_score = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (20)$$

其中: out_cor 表示输出的判断正确的关系个数; out_all 表示输出的所有关系个数; this_all 表示测试集中的所有该关系的个数。

3.3 参数设置

模型的激活函数选用 relu 函数,隐含层节点数取 300(文献[17]经验值),采用 softmax 作为分类器。为防止模型计算过程中出现的过拟合现象,利用 L2 正则化方法对网络参数进行约束,训练过程还引入了 dropout 策略,其丢码率(dropout rate)取 0.5。另外,采用批量的 Adadelta 优化方法用于模型训练,批量大小取 50(文献[12]经验值),训练轮数取 500。其中,丢码率和训练轮数选取均采用 5 折交叉验证的方法得到,验证集从训练样本中随机获取,实验结果如图 3 所示。

3.4 实验结果

选取 3.1 节语料库的训练样本作为模型训练集进行模型训练,采用其测试样本对训练后的模型进行测试和评估。抽取

10 大类关系的评估结果如表 3 所示。

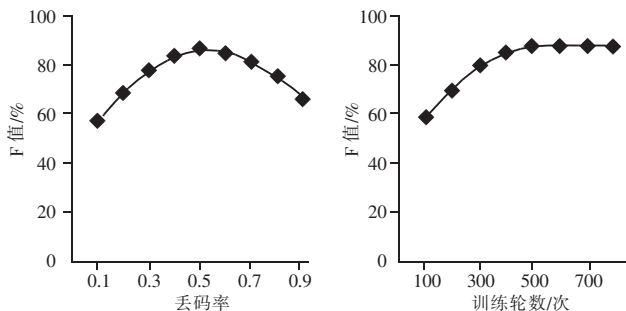


图3 交叉验证

表3 实验结果

关系类型	正确率/%	召回率/%	F 值/%
会员与组织	93.56	96.57	95.04
生产者与产品	91.86	92.64	92.25
实体与目标	90.07	92.47	91.25
整体与部分	89.42	89.74	89.60
内容与包含	89.07	89.58	89.32
实体与主题	86.15	87.01	86.58
实体与来源	83.14	85.44	84.27
因果	79.57	88.72	83.90
工具与使用者	82.69	83.33	83.01
其他	79.30	81.94	80.60
所有	85.90	88.44	87.16

由表 3 可以看出,简单的语义关系(如会员与组织、整体与部分关系等)抽取的效果要比复杂语义关系(如因果、其他关系等)抽取的效果好,这是因为简单语义关系更容易被模型所学习,复杂语义关系存在更多的难以获取的语义关联。

3.5 对比分析

为体现 At_LSTM 在准确度和稳定性上的优势,将 At_LSTM 与 CNN、RNN 方法和 LSTM 方法在同一语料库上进行实验对比。为统一比较标准,所有方法的输入词向量均采用 2.1 节方法生成,所有网络隐含层节点数均取 300,池化、特征融合和分类均采用 2.2 节方法进行。另外,CNN 方法滤波器窗口大小取 2,3,4,每个滤波器的卷积核个数为 100。实验结果如表 4 和图 4 所示。

表4 实验对比

方法	正确率/%	召回率/%	F 值/%
RNN	78.54	78.23	78.38
CNN	81.27	82.35	81.81
LSTM	82.63	84.72	83.66
At_LSTM	85.90	88.44	87.16

由表 4 可知:

a) LSTM 方法比 RNN、CNN 方法体现出更好的性能,因为 LSTM 缓解了 RNN 和 CNN 的长距离依赖问题,提取到的文本特征更有效。

b) At_LSTM 方法比 LSTM 方法效果更佳,该方法结合了注意力机制,突出 LSTM 关键性输入对于模型的影响,获取更多的上下文语义信息,从而证实了本文方法在准确率上的优势。

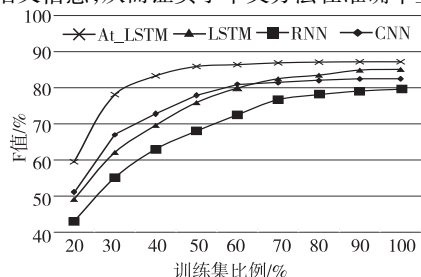


图4 模型稳定性对比图

由图 4 可知,At_LSTM 方法在选取 40% 的训练样本进行训练时就达到了稳定,而其他方法需要在选取 70% 以上的训练样本才能达到稳定。因为 At_LSTM 考虑了更全面的文本上下文语义信息,从而展示了本文方法在稳定性上的优势。

4 结束语

本文针对传统深度学习方法(RNN、CNN)中无法解决文本长距离依赖问题和未考虑模型输入与输出相关性问题的实验结果表明,该方法的准确率和稳定性较传统深度学习方法有了进一步的提升。本文的创新点在于采用 LSTM 模型缓解了长距离依赖问题,并结合注意力机制考虑了模型输入与输出的相关性,进而提取了更有效的文本特征,提高了关系抽取的准确率。但是该方法只能抽取预先设定好的关系集合,如何将深度学习方法引入到开放领域进行关系抽取、自动地发现新的关系还有待进一步研究。

参考文献:

- [1] 赵妍妍,秦兵,车万翔,等. 中文事件抽取技术研究[J]. 中文信息学报. 2008,22(1):3-8.
- [2] 陈宇,郑德权,赵铁军. 基于 Deep Belief Nets 的中文名实体关系抽取[J]. 软件学报,2012,23(10):2572-2585.
- [3] 彭博辉,付琨,黄宇,等. 基于多通道卷积神经网络的实体关系抽取[J]. 计算机应用研究,2017,34(3):689-692.
- [4] Schutz A, Buitelaar P. RelExt: A tool for relation extraction from Textin ontology extension [C]//Proc of the 4th International Semantic Web Conference. Berlin: Springer, 2005:593-606.
- [5] Mintz M, Bills S, Snow R, et al. Distant supervision for relation extraction without labeled data [C]//Proc of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Stroudsburg, PA: Association for Computational Linguistics, 2009:1003-1011.
- [6] Augenstein I, Maynard D, Ciravegna F. Relation extraction from the Web using distant supervision [C]//Proc of International Conference on Knowledge Engineering and Knowledge Management. [S. l.]: Springer International Publishing, 2014:26-41.
- [7] Hoffmann R, Zhang Congle, Ling Xiao, et al. Knowledge-based weak supervision for information extraction of overlapping relations [C]//Proc of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2011:541-550.
- [8] Verga P, Belanger D, Strubell E, et al. Multilingual relation extraction using compositional universal schema [EB/OL]. (2016-06-28). <https://arxiv.org/pdf/1511.06396v1.pdf>.
- [9] Lamurias A, Clarke L A, Couto F M. Extracting microRNA-gene relations from biomedical literature using distant supervision [J]. PLoS One, 2017, 12(3): e0171929.
- [10] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces [C]//Proc of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012:1201-1211.
- [11] Jordan M I. Serial order: a parallel distributed processing approach [C]//Advances in Psychology, 1997:471-495.
- [12] Zhang Dongxu, Wang Dong. Relation classification via recurrent neural network [J/OL]. (2015-12-29). <https://arxiv.org/abs/1508.01006?context=cs>.

(下转第 1440 页)

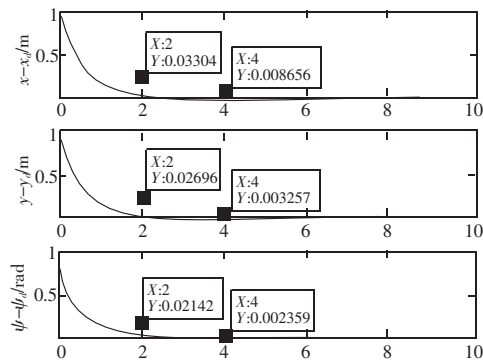


图5 复合自适应扰动观测器控制下无人水面艇的轨迹跟踪误差

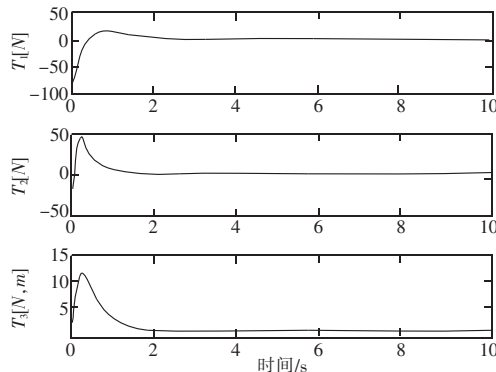


图6 复合自适应扰动观测器控制下无人水面艇的轨迹跟踪控制输入

相较于传统的扰动观测器,采用高增益的复合自适应扰动观测器,其对外界扰动的估计更迅速、更精确,能短时间内对扰动进行精确估计,在轨迹的跟踪控制方面,其具有跟踪速度更快、跟踪精度更高和控制输入快速稳定的优势。

4 结束语

本文研究了带模型参数不确定性和外界扰动的无人水面艇的轨迹跟踪问题,设计了基于复合自适应扰动观测器的轨迹跟踪控制。该复合自适应扰动观测器将估计误差和跟踪误差相结合共同调节自适应参数。通过理论分析和仿真实验研究表明,相较于基于传统扰动观测器的轨迹跟踪控制方法,本文提出的复合自适应控制方法可使无人水面艇具有更高的控制精度和更快的跟踪速度。

参考文献:

- [1] 郭晨,汪洋,孙富春,等. 欠驱动水面船舶运动控制研究综述[J]. 控制与决策,2009,24(3):321-329.
- [2] Do K D, Jiang Z P, Pan J. Global partial-state feedback and output-feedback tracking controllers for underactuated ships[J]. Systems & Control Letters,2005,54(10):1015-1036.
- [3] Švec P, Thakur A, Raboin E, et al. Target following with motion prediction for unmanned surface vehicle operating in cluttered environments[J]. Autonomous Robots,2014,36(4):383-405.
- [4] Dong Wenjie, Yi Guo. Global time-varying stabilization of underactuated surface vessel[J]. IEEE Trans on Automatic Control,2005,50(6):859-864.
- [5] 朱齐丹,马俊达,刘志林. 未知扰动下自主水面船自适应有限时间轨迹跟踪[J]. 系统工程与电子技术,2016,38(2):368-374.
- [6] Lefeber E, Pettersen K Y, Nijmeijer H. Tracking control of an underactuated ship[J]. IEEE Trans on Control Systems Technology,2003,11(1):52-61.
- [7] 程金. 水面船舶的非线性控制研究[D]. 北京:中国科学院研究生院,2007.
- [8] Tee K P, Ge S S. Control of fully actuated ocean surface vessels using a class of feedforward approximators[J]. IEEE Trans on Control Systems Technology,2006,16(4):750-756.
- [9] Zhang Lijun, Jia Heming, Qi Xue. NNFFC: adaptive output feedback trajectory tracking control for a surface ship at high speed[J]. Ocean Engineering,2011,38(13):1430-1438.
- [10] Shim H, Jo N H, Son Y I. Disturbance observer for non-minimum phase linear systems[J]. International Journal of Control, Automation and Systems,2010,8(5):994-1002.
- [11] 贾松涛,朱煜,杨开明,等. 精密工作台扰动观测器的设计[J]. 微细加工技术,2007,25(4):39-42.
- [12] White M T, Tomizuka M, Smith C. Improved track following in magnetic disk drives using a disturbance observer[J]. IEEE/ASME Trans on Mechatronics,2000,5(1):3-11.
- [13] Yang Yang, Du Jialu, Guo Chen, et al. Trajectory tracking control of nonlinear full actuated ship with disturbances[C]//Proc of International Conference of Soft Computing and Pattern Recognition. Piscataway, NJ: IEEE Press,2011:318-323.
- [14] 付明玉,刘佳,吴宝奇. 基于扰动观测器的动力定位船终端滑模航迹跟踪控制[J]. 中国造船,2015,56(4):33-45.
- [15] Pan Yongping, Yu Haoyong. Composite learning from adaptive dynamic surface control[J]. IEEE Trans on Automatic Control,2016,61(9):1.
- [16] 潘永平,黄道平,孙宗海. 欠驱动船舶航迹 Backstepping 自适应模糊控制[J]. 控制理论与应用,2011,28(7):907-914.
- [17] Pan Yongping, Zhou Yu, Sun Tairen, et al. Composite adaptive fuzzy H_∞ tracking control of uncertain nonlinear systems[J]. Neurocomputing,2013,99:15-24.
- [18] Xu Bin, Shi Zhongke, Yang Chenguang, et al. Composite neural dynamic surface control of a class of uncertain nonlinear systems in strict-feedback form[J]. IEEE Trans on Cybernetics,2014,44(12):2626-2634.
- [19] Li Yongming, Tong Shaocheng, Li Tieshan. Composite adaptive fuzzy output-feedback control design for uncertain nonlinear strict-feedback systems with input saturation[J]. IEEE Trans on Cybernetics,2015,45(10):2299-2308.
- [20] Skjetne R, Smogeli, Fossen T I. Modeling, identification, and adaptive maneuvering of cyberShip II: a complete design with experiments[J]. IFAC Conference on Control Application Marine Systems,2004,37(10):203-208.
- [13] Zeng Daojian, Liu Kang, Lai Siwei, et al. Relation classification via convolutional deep neural network[C]//Proc of the 25th International Conference on Computational Linguistics. 2014:2335-2344.
- [14] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the National Academy of Sciences of the United States of America,1982,79(8):2554.
- [15] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.
- [16] Sundermeyer M, Schlüter R, Ney H. LSTM neural networks for language modeling[C]//Proc of the 13th Annual Conference on Inter-Speech. 2012:601-608.
- [17] 任智慧,徐浩煜,封松林,等. 基于 LSTM 网络的序列标注中文分词法[J]. 计算机应用研究,2017,34(5):1321-1324,1341.
- [18] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures[J/OL]. (2016-03-19). <https://arxiv.org/pdf/1601.00770v2.pdf>.
- [19] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J/OL]. (2015-09-22). <https://arxiv.org/abs/1508.04025>.
- [20] ACL SigLex, SemEval-2010 task 8[EB/OL]. https://drive.google.com/file/d/0B_jQiLugGTAKMDQ5ZjZiMTUzMzQ1Yy00YWNmLWJlZDY-tOWY1ZDMwY2U4YjFk/view?layout=list&sort=name&num=50.

(上接第1420页)