

数据到文本生成的神经网络模型研究

曹娟, 龚隽鹏, 张鹏洲

(中国传媒大学, 北京 100024)

摘要:随着人工智能的发展,国内各行各业纷纷涌入人工智能领域,尤其是新闻行业,各大科技公司和科研机构研发自动化新闻生产,掀起了一股新闻机器人的热潮。机器新闻写作,实质上就是一个利用自然语言生成技术进行新闻写作的过程。数据到文本生成属于自然语言生成技术,是指采用结构化的数据例如一张表格作为输入,生成恰当而流畅的文本作为输出来描述数据,是实现机器新闻写作的关键技术之一。近年来随着数据到文本生成研究的不断深入,研究人员将神经网络方法引入该领域,主要基于循环神经网络构建模型,并已取得了不错的研究进展。文中首先介绍了用于数据到文本生成的神经网络模型,并梳理了近年来该领域的研究成果,然后介绍了相关数据集并对比了各模型的实验结果,最后分析了该领域研究存在的问题并提出了未来发展的建议。

关键词:数据到文本生成;机器新闻写作;自然语言生成;神经网络

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2019)09-0007-06

doi:10.3969/j.issn.1673-629X.2019.09.002

Research on Neural Network Model of Data-to-text Generation

CAO Juan, GONG Jun-peng, ZHANG Peng-zhou

(Communication University of China, Beijing 100024, China)

Abstract: With the development of artificial intelligence, various industries in China have flocked to the field of artificial intelligence, especially the news industry. The domestic technology companies and major scientific research institutions have independently developed automated news production, setting off a wave of news robots. Robot journalism is essentially a process of news writing using natural language generation technology. Data-to-text generation belongs to natural language generation technology which refers to using structured data such as a table as input and generating appropriate and fluent text as output to describe data and is one of the key technologies to realize robot journalism. In recent years, with the deepening of data-to-text generation research, researchers have introduced neural network into this field, mainly building models based on recurrent neural network, and have made excellent research progress. In this paper, we first introduce the neural network models used for data-to-text generation and review the research results in this field in recent years, then introduce the relevant datasets and compare the experimental results of each model. Finally we analyze the problems in this field and put forward suggestions for future development.

Key words: data-to-text generation; robot journalism; natural language generation; neural network

0 引言

2016年被公认为“人工智能发展元年”,是人工智能正式提出的60周年,国内各行各业纷纷涌入人工智能领域,尤其是新闻行业,国内科技公司包括行业巨头百度、腾讯、阿里等和各大科研机构各自研发自动化新闻生产,掀起了一股新闻机器人的热潮。机器新闻写作实质上就是一个利用自然语言生成技术进行新闻写作的过程。它是一种自然语言生成引擎,利用算法程

序,通过采集大量的各种题材及高质量的数据,建设各种分类的庞大数据库,借助人工智能实现从数据到认识、见解和建议的提升与跨越,最后由机器自动生产新闻^[1]。

由此可见,机器新闻写作的核心技术是自然语言生成(natural language generation, NLG),并需要大数据和云计算的支撑。自然语言生成是指基于非语言输入自动生成自然语言文本的过程,是人工智能和计算

收稿日期:2018-10-09

修回日期:2019-02-15

网络出版时间:2019-04-24

基金项目:国家科技支撑计划项目(2014BAK10B01)

作者简介:曹娟(1990-),女,博士,CCF会员(44317G),研究方向为机器新闻写作;龚隽鹏,博士,副教授,研究方向为机器新闻写作;张鹏洲,博士后,研究员,CCF会员(07635S),研究方向为内容管理、电视台整体化管理及其应用软件开发。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1047.026.html>

语言学的分支^[2]。自然语言生成中一个典型的问题就是采用结构化的数据例如一张表格作为输入,生成恰当而流畅的文本作为输出来描述数据^[3]。解决这一问题的方法被称为数据到文本的生成(data-to-text generation)。

近年来,随着数据到文本生成研究的深入,研究人员将神经网络方法引入该领域,研究成果主要发表在 INLG(international natural language generation)、ENLG(European natural language generation)、ACL(ACL home association for computational linguistics)、NAACL(The North American chapter of the association for computational linguistics)和 AAI(American association for artificial intelligence)这几个专业学术会议上,目前已经取得了不错的研究进展。文中将介绍用于数据到文本生成的神经网络模型和近年来该领域的研究成果,然后介绍相关数据集并对比各模型的实验结果,最后分析该领域研究存在的问题并提出未来发展的建议。

1 数据到文本生成的神经网络模型

目前用于数据到文本生成的神经网络模型大部分都是基于循环神经网络(recurrent neural network, RNN)的神经语言模型和 Seq2seq 架构,同时使用了注意力模型。下面介绍这几种模型。

1.1 循环神经网络

文本生成本身是一个序列输出问题,RNN 就是一种对序列数据建模的神经网络,可以很好地刻画词汇的前后关联,这也是由 RNN 自身的特点决定的。RNN 可以对前面的信息进行记忆并应用于当前输入的计算中,即隐藏层之间的节点不再是无连接的而是有连接的,并且隐藏层的输入不仅包括当前输入层的输入,还包括上一时刻隐藏层的输出^[4]。图 1 是 RNN 基本模型的示例图。

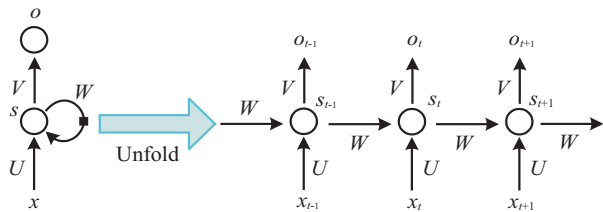


图 1 RNN 基本模型示例

其中 x_t 是 t 时刻的输入,比如单词对应的 one-hot 向量(只有一维是 1,其他维是 0)或者分布式词向量; s_t 是 t 时刻的隐状态,由上一时刻的隐状态和当前输入得到,如式 1 所示, f 一般是非线性的激活函数,在计算 s_0 时,即第一个单词的隐藏层状态,一般置为 0; o_t 表示 t 时刻的输出,比如下个单词的向量表示,如式 2 所示,softmax 是激活函数,用于输出层,输出 0 和 1 之

间的概率值,总和为 1,适合于使用概率刻画输出的情况,例如多分类任务; U 、 V 、 W 是训练时需要学习的参数,在传统的神经网络中,每层网络的参数是不共享的,而在 RNN 中,所有层次均共享同样的参数,大大降低了网络中需要学习的参数量。

$$s_t = f(Ux_t + Vs_{t-1}) \quad (1)$$

$$o_t = \text{softmax}(Vs_t) \quad (2)$$

1.2 基于 RNN 的神经语言模型

语言模型就是指语言产生的规律,一般用来预测所使用语言语序的概率,或者是当前上下文使用某个词语的概率,也就是用来表示语言产生顺序的建模,比如用某个词是否恰当,这样的语序构造句子是否妥当。因此训练一个语言模型需要相当大的样本数据。语言模型可以分为文法型的语言模型、统计模型和神经网络语言模型。统计类的语言模型包括 N-gram、隐马尔可夫链模型和最大熵模型等。

神经语言模型(neural language model, NLM)是一类用来克服维数灾难的语言模型,它使用词的分布式表示对自然语言序列的建模^[5]。基于 RNN 的神经语言模型就是利用循环神经网络(RNN)对语言建模,用来描述语言序列的产生过程。与 RNN 相比,该模型多了词向量层(embedding)和 softmax 层。

1.2.1 词向量层

在神经网络的输入层,每一个单词用一个实数向量表示,这个向量被成为“词向量”(word embedding, 也称为词嵌入)。词向量可以形象地理解为将词汇表嵌入到一个固定维度的实数空间里。将单词编号转化为词向量主要有两大作用:

(1)降低输入的维度。如果不使用词向量层,而直接将单词以 One-hot 向量的形式输入循环神经网络,那么输入的维度大小将与词汇表大小相同,通常在 1 000 以上。而词向量的维度通常在 200 到 1 000 之间,这将大大减少循环神经网络的参数数量和计算量。

(2)增加语义信息。简单的单词编号是不包含任何语义信息的。两个单词之间编号相近,并不意味着它们的含义有任何联系。而词向量层将稀疏的编号转化为稠密的向量表示,这使得词向量有可能包含更丰富的信息。在自然语言应用中学习得到的词向量通常会将含义相似的词赋予取值相近的词向量值,使得上层网络可以更容易抓住相似单词之间的共性。

1.2.2 softmax 层

softmax 层的作用是将循环神经网络的输出转为一个单词表中每个单词的输出概率。为此需以下两个步骤:

(1)使用一个线性映射将循环神经网络的输出映射为一个维度与词汇表大小相同的向量,这一步叫

logits。

(2)调用 softmax 方法将 logits 转化为加和为 1 的概率。

1.3 基于 RNN 的 Seq2seq 架构

机器翻译中最经典的架构就是 Seq2seq,它是

RNN 的变种,也是 Encoder-Decoder 的一种^[6]。基于 RNN 的 Seq2seq 架构包含编码器 (encoder) 和解码器 (decoder),解码部分又分训练 (train) 和推断 (inference) 两个过程,如图 2 所示。

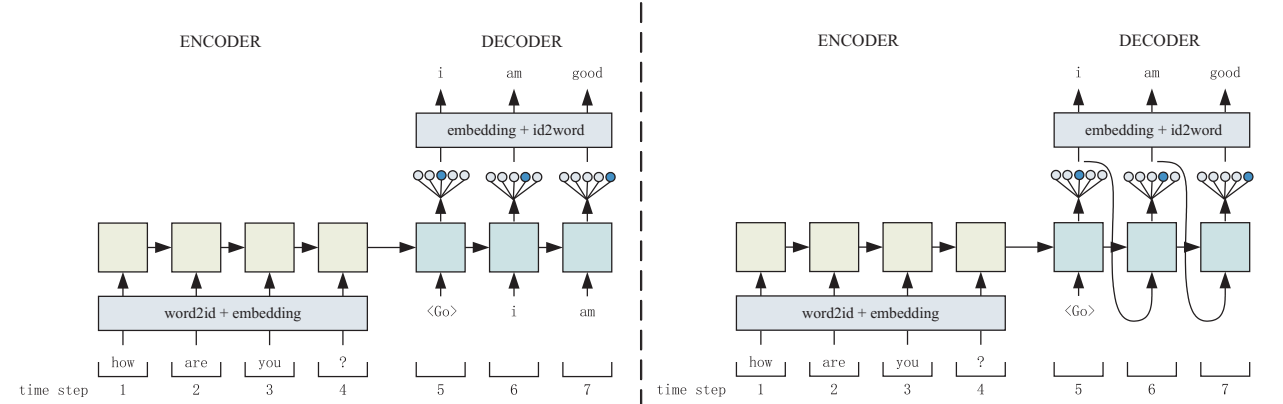


图2 Seq2seq 的训练过程(左)和推断过程(右)

其基本思想就是利用两个 RNN,一个 RNN 作为 encoder,另一个 RNN 作为 decoder。encoder 负责将输入序列压缩成指定长度的向量,这个向量可以看成是这个序列的语义,这个过程称为编码。获取语义向量最简单的方式就是直接将最后一个输入的隐状态作为语义向量。也可以对最后一个隐含状态做变换得到语义向量,还可以将输入序列的所有隐含状态做变换得到语义变量。而 decoder 负责根据语义向量生成指定的序列,这个过程也称为解码。最简单的方式是将 encoder 得到的语义变量作为初始状态输入到 decoder 的 RNN 中,得到输出序列。

1.4 注意力模型

在 Encoder-Decoder 结构中,encoder 把所有的输入序列都编码成一个统一的语义向量再解码,因此,这个语义向量必须包含原始序列中的所有信息,它的长度就成了限制模型性能的瓶颈。

注意力机制 (attention mechanism) 在 Encoder-Decoder 框架内,通过在 encoder 中加入 attention 模型,对源数据序列进行数据加权变换,或者在 decoder 引入 attention 模型,对目标数据进行加权变化,可以有效提高序列对序列的自然方式下的系统表现^[7]。

Attention 模型的作用:减小处理高维输入数据的计算负担,通过结构化的选取输入的子集,降低数据维度;“去伪存真”,让任务处理系统更专注于找到输入数据中显著的与当前输出相关的有用信息,从而提高输出的质量。

2 研究方法

Mei 等在 2016 年提出一种端到端 (end-to-end)

的基于编解码 (encoder-decoder) 框架的神经网络模型 (简称 MBW)^[8],其中用到了基于长短期记忆网络 (long short-term memory, LSTM) 的循环神经网络。Lebret 等在 2016 年介绍了一种建立在文本生成的条件神经语言模型 (conditional neural language models) 基础上的神经模型 (简称 Table NLM),用于根据维基百科人物传记数据集中的事实表格生成人物传记的初始句子^[9]。Sha 等在 2017 年提出一种规划顺序 (order-planning) 的神经网络模型^[10];Chisholm 等在 2017 年针对从维基百科数据生成一句话的人物传记设计了一种自编码器 seq2seq 模型 (简称 S2SAE)^[11];Liu 等在 2018 年提出一种面向结构 (structure-aware) 的 seq2seq 模型^[12];Bao 等在 2018 年提出一种面向表格 (table-aware) 的 seq2seq 模型^[13];Nema 等在 2018 年使用双焦点注意力机制和门控正交化生成结构化数据的描述 (简称 BAMGO)^[14];Wiseman 等在 2018 年设计了一种利用神经网络抽取模板生成文本的方式 (简称 NT-emp)^[15];Freitag 等在 2018 年使用降噪自编码器 (denoising autoencoder) 实现了无监督的自然语言生成过程 (简称 NLGDA)^[16];Kaffee 等在 2018 年提出面向维基百科单句的跨领域多语言的神经网络模型 (简称 UL)^[17]。下面具体介绍这些模型和方法。

2.1 MBW

MBW 模型以 encoder-aligner-decoder 作为框架,使用 LSTM-RNN 和对准器 (coarse-to-fine aligner) 来选择数据并将其转化为文本描述。在编码端,模型将数据记录的全集作为输入,使用了双向 LSTM-RNN 对其编码;然后对准器经过初选和精选决定哪些输入记录出现在描述中;最后由一个 LSTM-RNN 解码器

将挑选出来的数据解码生成自然语言文本输出。在数据到文本生成领域, MBW 是首个使用神经网络方法将内容选择和表层实现视作一体, 实现端到端(end-to-end)的神经网络模型。该模型分别在 WEATHERGOV^[18] 和 ROBOCUP^[19] 两个数据集上进行测试, 并得到了当时最好的结果。但由于这两个数据集中表格结构相对简单, 表格内的数据除了数值类数据, 其他大多都是提前定义好的几种模式, 因此该模型使用了 one-hot 向量作为输入, 适用于这种固定模式表示的数据, 从而不适合更复杂的数据表示的数据集, 比如 WIKIBIO^[9]。

2.2 Table NLM

Table NLM 模型以条件神经语言模型为基础, 针对人物传记数据集 WIKIBIO 进行设计, 关注到了数据记录中的属性(field), 并分别从全局和局部对结构化数据进行解析。全局条件总结了一个人物的所有信息来获取高层主题, 比如这篇传记是关于科学家还是艺术家的; 局部条件根据人物传记数据表格(infobox) 中词之间的关系来描述之前生成的词组。该模型采用了标准的三层架构, 分别是词向量层、线性层和 softmax 层。词向量层又细分为单词向量(word embedding)、表格向量(table embedding) 和聚合向量(aggregating embeddings)。单词向量只包含表格中 value 部分的词组信息, 表格向量体现了局部条件和全局条件, 局部条件主要包含表格的属性信息, 全局条件则将属性和值二者都考虑在内, 聚合向量是将单词向量和表格向量聚合在一起经过线性变换得到。而线性层则是又用了线性变换将聚合向量转换为输出词组的分值, 同时该模型使用了拷贝机制, 通过计算聚合向量与属性和值的关系, 选取表格内最可能替代输出文本中未知词的词组。虽然此模型不适合长距离依赖, 但也获得了当时实验最好的结果。

2.3 Order-planning Model

Order-planning Model 模型使用了 encoder-decoder 框架、RNN 和 attention, 设计了 field linking 机制, 模拟不同 field 之间的关系, 帮助模型更好地规划文本生成的顺序, 同时使用了拷贝机制, 解决少见词的问题(rare words)。该模型由编码器、调度器和解码器组成。编码器将 field 和 content 结合输入到 LSTM 中, 经过基于内容的 attention 和基于 link 的 attention 混合后得到 attention 向量, 再传送给解码器的 LSTM, 最终配合拷贝机制输出文本。该模型在 WIKIBIO 数据集上进行实验, 获得了较之前更好的结果。

2.4 S2SAE

S2SAE 模型是一个基于循环神经网络的 seq2seq 模型, 使用了 attention 机制, 以及自编码器

(autoencoder), 实现了从维基百科人物传记结构化数据到文本单句的生成。该模型选取了 15 个频率最高的 field, 经过三层 GRU 的编码器和解码器, 选取了 10 万个词作为词库, 词向量设定为 256 维。加入自编码器后, 除了向前的编解码, 还有一条回路, 即向后的编解码, 更好地调整误差(loss)。实验证明, 加入自编码器的 seq2seq 比标准的 seq2seq 模型效果提升很多, BLEU 的分值从 33 提高到 41。

2.5 Structure-aware Seq2seq

Structure-aware Seq2seq 模型使用了端到端的 encoder-decoder 架构, 通过本地决策和全局决策来模拟人类选取记录写作的过程, 即首先决定哪些记录应该被选取和如何安排它们的顺序, 其次考虑表格中的哪些单词应该被关注。本地决策主要完成在生成文本的每一步时关注表格中哪些单词, 通过对表格内容进行 LSTM 编码和使用单词级别的 attention 来实现; 全局决策的目的是在生成文本时决定关注哪些记录, 通过对表格的属性进行自定义属性门(field-gating) LSTM 编码, 以及属性级别的 attention 来实现。该模型在 WIKIBIO 上进行实验, 得到了不错的结果。

2.6 Table-aware Seq2seq

Table-aware Seq2seq 模型也使用了 encoder-decoder 架构和 attention 机制, 不同的是, 这里的 RNN 采用了 GRU, 实验对象主要是一个开放领域的数据集 WIKITABLETEXT。WIKITABLETEXT 是第一个针对数据到文本生成的开放领域的数据集, 文本中会出现很多不常见的词汇, 这些词汇通常存在于表格内。为了解决这种少见词的问题, 该模型使用了一个强大的拷贝机制(copying mechanism), 极大地提升了模型在该数据集上的效果。该模型在 WIKIBIO 上也做了实验, 但效果不如 Structure-aware seq2seq 模型, 究其原因, 该模型的设计主要针对 WIKITABLETEXT, 来自不同的领域, 而 WIKIBIO 是专门面向人物传记的, 且其句子平均长度也比 WIKITABLETEXT 中的大很多(26.1>13.9), 因此拷贝机制对其影响较小。

2.7 BAMGO

BAMGO 模型在 seq2seq 基础上做了改进, 加入了双焦点注意力机制和正交化门控机制。双焦点注意力机制结合了宏观和微观层的信息, 即属性和属性内的值, 正交化门控的目的是在生成文本的过程中将已出现的属性在后续生成步骤中被遗忘(never look back)。该模型在英文、法文和德文的 WIKIBIO 上分别做了实验, 在法文和德文上的效果远远不如在英文上的效果好, 究其原因, 法语和德语的数据集本来就比较小, 且其文本长度也比英语的长, 格式也不如英文的固定。由此可见, 模型对于不同语言的文本, 效果也一

定相同。

2.8 NTemp

NTemp 模型是一个基于隐半马尔可夫 (HSMM) 解码器的神经类似模板 (template-like) 生成模型,和 Encoder-Decoder 这种黑盒模型不同的是,它是可解释和可控的神经网络模型。该模型主要通过 HSMM 生成类似模板结构,运用一个转移模型判断两种状态,一种代表不在词汇表中,需要表格中的原始词,通过拷贝机制替换模板生成文本,一种可直接生成文本。生成文本的过程也用了 RNN 来实现。这种方法适合大规模的文本数据集,其效果和 Encoder-Decoder 模型差不多。

2.9 NLGDA

NLGDA 模型基于 seq2seq,向编码器中添加噪音,再使用降噪自编码器 (denoising autoencoder),对生成的语句重新构造,会更好地生成正确的语句。这种无监督的方法适合于包含结构数据全部内容的文本生成,并不适用于所有 NLG 问题和数据集。该模型被用于 E2E 数据集^[20]。E2E 是一个专门为 End-to-End 设计的描述饭店信息的数据集,其中每组数据包含了饭店的最多 8 个不同的信息和描述所有信息的一句话,共 42 061 组数据。

2.10 UL

UL 模型主要针对不同语言的开放领域的维基百科而设计,基于 Encoder-Decoder 架构,使用了无序的三元组作为输入,三元组由主题、属性和对象构成,解码器使用了一层 GRU,最后生成文本利用了拷贝机制。该模型的实验使用了维基百科中两种不同特点的语言,分别是阿拉伯语和世界语。阿拉伯语具有形态丰富和词汇量大的特点,世界语是一种人工语言,其语法严谨,以更容易掌握著称,但世界语的文本数据量少于阿拉伯语。最终测试结果是阿拉伯语的结果比世界语略好,同时二者皆胜过不使用神经网络方式的结果。

3 数据集和实验结果

该研究领域可用的数据集本来就不多,使用神经网络模型做过实验的数据集更是少之又少,下面介绍 WEATHERGOV、ROBOCUP、WIKIBIO 和 E2E 这几个数据集和前面介绍过的方法在这些数据集上的实验结果。

3.1 WEATHERGOV 和 ROBOCUP 的模型实验结果对比

WEATHERGOV 是 Liang 等在 2009 年创建的天气预报数据集。他们从 www.weather.gov 上收集了美国 2009 年 2 月 7 日到 9 日三天的 3 753 个城市的本地天气预报,每个城市每天有白天和晚上两个场景。天气

预报由选定时间段的气温、风速、云量和降雨可能性指标等构成。数据集共包含 29 528 个场景,每个场景平均有 36 条记录和 28.7 个单词。

ROBOCUP 是 Chen 和 Mooney 在 2008 年创建的数据集,包含了 2001 年到 2004 年 Robocup 决赛的 1 919 个场景。每个场景由一句话组成,代表比赛评论的一部分,同时配有评论发出前 5 秒内的一组候选事件记录。每句话最多对应一条记录,Chen 和 Mooney 已经做好了标注。平均每个场景有 2.4 条记录和 5.7 个单词。

WEATHERGOV 和 ROBOCUP 在数据量和结构上有类似之处,相对词汇量小,结构简单,在介绍过的神经网络模型里 MBW 是目前结果最好的,通常与非神经网络模型 KL^[21]和 AKL^[22]对比。KL 是 Konstas 等在 2013 年提出的模型,AKL 是 Angeli 等在 2010 年提出的模型。表 1 展示了各模型在这两个数据集上的实验结果,评价指标为 F_1 值和 BLEU-4。

表 1 WEATHERGOV 和 ROBOCUP 的模型结果

模型名称	WEATHERGOV		ROBOCUP	
	F_1	BLEU	F_1	BLEU
KL	-	36.54	-	24.88
AKL	65.40	38.40	79.90	-
MBW	73.21	61.01	81.58	25.28

3.2 E2E 的模型实验结果对比

E2E 是描述饭店信息的数据集,前面已经介绍过。在上述神经网络模型中,NTemp 和 NLGDA 在其上做过实验。E2E challenge^[23]是 Dusek 提出的监督模型,作为该数据集的基准方法,其结果如表 2 所示。评价指标有 BLEU、ROUGE-L 和 NIST。

表 2 E2E 的模型结果

模型名称	BLEU	ROUGE-L	NIST
E2E challenge	70.20	72.40	8.30
NTemp	89.20	-	-
NLGDA	67.00	72.10	7.80

3.3 WIKIBIO 的模型实验结果对比

WIKIBIO 是 Lebrete 等在 2016 年介绍的数据集。这个数据集包含了 2015 年 9 月从英文维基百科上收集的 728 321 篇人物传记和对应的表格数据,因其数据量大、词汇量大、属性多、结构相对复杂等,也是目前为止在其上实验最多的数据集。使用该数据集获得的实验结果更具有比较的意义。表 3 展示了介绍过的神经网络模型的实验结果,评价指标有 BLEU-4、ROUGE 和 NIST。

表 3 WIKIBIO 的模型结果

模型名称	BLEU	ROUGE	NIST
Table NLM	34. 70	25. 80	7. 98
MBW	35. 10	30. 90	7. 27
Table-aware Seq2seq	40. 26	—	—
S2SAE	41. 00	—	—
BAMGO	42. 03	39. 11	9. 17
NTemp	43. 20	—	—
Order-planning	43. 91	37. 15	8. 85
Structure-aware Seq2seq	44. 89	41. 65	—

4 存在问题和发展方向

目前该领域存在一些问题,需要在未来的研究中解决:

(1)数据集缺乏。目前公开可用的数据集非常缺乏,尤其是能用于训练神经网络模型的数据集,必须保证数据量大。

(2)领域单一。可训练的数据集集中在几个特定领域,一方面需要新开辟其他领域以供研究,另一方面也需要跨领域的研究。

(3)评价方法不统一。研究人员用于各个数据集的评价指标并不完全一致,使用评价方法也不统一,加大了对比的难度。

数据到文本生成在未来的发展中越来越重要,最好能专门针对这一研究领域提供公开数据集和统一的评价方法,以及跨领域、跨语言等不同特点的数据集,供研究人员使用。目前数据到文本生成的神经网络模型虽然比之前监督和半监督的方法提高了不少,但还有进步的空间,需要研究人员大胆尝试不同的方法。比如借鉴相关领域的最新研究如对话系统、机器翻译等,或者加入知识图谱,又或者从模型根本出发,研究结构化数据的特点,等等。

5 结束语

近年来国内新闻行业开始尝试自动撰写新闻,数据到文本生成技术越来越受到重视,加上深度学习的发展,利用神经网络训练数据到文本生成模型的研究也越来越多。当然,一个领域需要其他领域带动着一起发展,只有在神经网络方面的研究不断深入,硬件技术不断提升,存储和处理数据的能力越来越强,数据到文本生成的研究和应用才能有更多的可能性。

参考文献:

[1] 申云. “机器人新闻写作”对新闻采编的机遇和挑战[J]. 今传媒,2016(11):115-116.

[2] 周佳玥. 从 NLG 到机器新闻写作—机器新闻的发展与反

思[J]. 今传媒,2017(10):18-19.

[3] WISEMAN S, SHIEBER S M, RUSH A M. Challenges in data-to-document generation[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. [s. l.]: ACL, 2017: 2243-2253.

[4] 张荣, 李伟平, 莫同. 深度学习研究综述[J]. 信息与控制, 2018, 47(4): 385-397.

[5] 路佳佳, 李晓芳. 神经网络语言模型[J]. 通讯世界, 2017(9): 94-95.

[6] 董陆森. 机器翻译中的常用神经网络模型[J]. 电子技术与软件工程, 2018(10): 147.

[7] 刘洋. 神经机器翻译前沿进展[J]. 计算机研究与发展, 2017, 54(6): 1144-1149.

[8] MEI Hongyuan, UCHICAGO T T I, BANSAL M, et al. What to talk about and how? selective generation using LSTMs with coarse-to-fine alignment[C]//Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics. [s. l.]: ACL, 2016: 720-730.

[9] LEBRET R, GRANGIER D, AULI M. Neural text generation from structured data with application to the biography domain [C]//Proceedings of the 2016 conference on empirical methods in natural language processing. [s. l.]: ACL, 2016: 1203-1213.

[10] SHA Lei, MOU Lili, LIU Tianyu, et al. Order-planning neural text generation from structured data[C]//Proceedings of the thirty-second AAAI conference on artificial intelligence. [s. l.]: AAAI, 2018: 5414-5421.

[11] CHISHOLM A, RADFORD W, HACHEY B. Learning to generate one-sentence biographies from Wikidata[C]//Proceedings of the 15th conference of the European chapter of the association for computational linguistics. [s. l.]: ACL, 2017: 633-642.

[12] LIU Tianyu, WANG Kexiang, SHA Lei, et al. Table-to-text generation by structure-aware Seq2seq learning [C]//Proceedings of the thirty-second AAAI conference on artificial intelligence. [s. l.]: AAAI, 2018: 4881-4888.

[13] BAO Junwei, TANG Duyu, DUAN Nan, et al. Table-to-text: describing table region with natural language [C]//Proceedings of the thirty-second AAAI conference on artificial intelligence. [s. l.]: AAAI, 2018: 5020-5027.

[14] NEMA P, SHETTY S, JAIN P, et al. Generating descriptions from structured data using a bifocal attention mechanism and gated orthogonalization [C]//Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics. [s. l.]: ACL, 2018: 1539-1550.

[15] WISEMAN S, SHIEBER S M, RUSH A M. Learning neural templates for text generation [C]//Proceedings of the 2018 conference on empirical methods in natural language processing. [s. l.]: ACL, 2018: 3174-3187.

[16] FREITAG M, ROY S. Unsupervised natural language genera-