文章编号：1003-0077(2019)07-0056-09

基于领域特征的神经机器翻译领域适应方法

谭 敏，段湘煜，张 民

（苏州大学 计算机科学与技术学院，江苏 苏州 215006）

摘 要：神经机器翻译在资源丰富领域上训练的翻译模型往往在其他资源稀缺领域中表现较差，领域适应是利用资源丰富的领域帮助资源稀少的领域提升翻译质量的一种方法。该文提出基于领域特征的领域适应方法以提升资源稀缺领域的神经机器翻译质量。具体而言，该文尝试构建领域敏感网络以获得领域特有特征，构建领域不敏感网络以获得领域间的共有特征。一个领域判别器被用于区分领域。该文通过训练领域敏感网络使得该领域判别器更易做出准确判断，同时引入对抗机制，使得领域不敏感网络欺骗该领域判别器。最后，提出一种系统集成机制，融合基准神经翻译网络、领域敏感网络、领域不敏感网络以完成神经机器翻译的领域适应。实验结果显示，该方法在中英广播对话领域上和英德口语领域上的翻译效果均有显著提升。

关键词：领域适应；判别器；系统集成

中图分类号：TP391 **文献标识码**：A

Neural Machine Translation Domain Adaptation Based on Domain Features

TAN Min, DUAN Xiangyu, ZHANG Min

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: Translation models trained by neural machine translation system in resource rich areas tend to perform poorly in resource poor areas. This paper proposes domain adaptation based on domain features to improve the quality of neural machine translation with poor resource. Specifically, this paper establishes domain sensitive networks to obtain domain specific features, as well as to build domain insensitive networks to obtain common features between domains. A domain discriminator is used to distinguish the domain. This paper trained domain sensitive network to make it easier for the domain discriminator to make accurate judgements. At the same time, the adversarial mechanism is used so that the domain insensitive network can deceive the domain discriminator. Finally, a system combination mechanism is proposed by combining the base neural translation network, the domain sensitive network, and the domain insensitive network for the domain adaptation task. The experimental results show that this method achieves significant improvement in Chinese-English Broadcast Conversation translation task and English-German Spoken Language translation task.

Keywords: domain adaptation; discriminator; model combination

0 引言

近年来，基于编码器—解码器结构的神经机器翻译系统（neural machine translation，NMT）[1-3]的提出，显著提升了传统统计机器翻译（statistical machine translation，SMT）的性能。NMT 基于平行语料来训练，语料的质量、数量、领域对翻译效果都有很大的影响。NMT 对于训练的语料很敏感，每个领域都有自己的语言风格、句子结构、专业术语等，例如“bank”这个英文单词，在金融领域通常被翻译为“银行”，而在计算机领域，一般被解释为“库”“存储体”等。如果用基于计算机领域语料训练出的 NMT 模型来翻译金融领域的句子，就会导致翻译效果不理想。NMT 系统大多利用大规模新闻领域的平行语料，其他领域如对话、专

收稿日期：2018-09-21 **定稿日期**：2018-10-29

基金项目：国家重点研发计划(2016YFE0132100)；国家自然科学基金(61673289)

利、科技等领域可获得的平行语料规模较小。利用资源丰富的领域语料来帮助语料稀少的领域提升翻译质量，称为机器翻译的领域适应[4]。资源丰富的领域被称为外领域(out-domain)，资源稀缺的领域被称为内领域(in-domain)。

本文提出基于领域特征的神经机器翻译领域适应方法以提升内领域的翻译质量。在机器翻译的语料中，有一些单词只是某个领域特有的，而另外一些单词在内领域和外领域通用，不需要对这些单词区分领域，学习单词的领域特性有助于提升译文质量。为获得和应用这些领域特性，我们首先基于目标端隐藏层信息训练一个领域判别器，使其能够区分当前词属于外领域还是内领域，从而学习到领域特征；继而基于这个领域判别器，提出领域敏感网络(domain sensitive network，DSN)，可以使得领域判别器更加准确；并提出领域不敏感网络(domain insensitive network，DIN)，可以欺骗领域判别器做出错误判断。通过 DSN 可以识别各个领域的特征，通过 DIN 可以识别领域间的共有特征。最后，一个系统集成机制被提出，以融合基准神经翻译网络、DSN、DIN，得到最终的翻译系统。实验结果显示，融合领域特征的网络，在资源稀缺的中英广播对话领域、英德口语领域，均有显著的翻译质量提升。

本文的主要贡献如下：

① 提出基于领域判别器的领域敏感网络和领域不敏感网络，以分别对领域特有特征和领域共有特征进行建模；

② 通过系统集成方法，融合领域特有特征和领域共有特征，共同进行训练，以实现神经机器翻译的领域适应；

③ 本文提出的领域适应方法，在中英数据和英德数据上均显著提升了基准系统的领域适应能力，并优于相关研究的翻译质量。

本文的结构如下：第 1 节介绍相关工作，包括机器翻译的领域适应研究和系统集成研究；第 2 节介绍基准神经翻译系统；第 3 节阐述领域敏感网络和领域不敏感网络以及融合二者的系统集成方法；第 4 节阐述实验结果，并进行实验分析；第 5 节给出总结。

1　相关工作

领域适应方法首先在 SMT 上进行研究，主要方法包括两种：模型适应和数据选择[5]。模型适应主要是将内领域的模型和外领域的模型修改到同一模型级别上；数据选择主要是通过语言模型从外领域的语料里挑选平行语句对来扩充内领域的语料。借鉴 SMT 的领域适应方法，NMT 的领域适应方法也可分为基于数据的方法和基于模型的方法。

基于数据的方法主要是通过训练模型来对外领域的数据进行打分并挑选出得分高的句子来扩充内领域的语料。Wang 等[6]提出计算源端词嵌入向量(word embedding)的中心点，通过词嵌入向量来模拟句子的相似性，对比内领域和外领域句子的词嵌入向量挑选出词嵌入向量相似的句子。Van der Wees 等[7]提出动态数据选择方法，在系统的训练过程中，不同的训练轮数选择不同的训练语料。

基于模型的方法主要是在训练过程中改变训练方法从而得到最优的领域训练目标。Wang 等[8]使用调整实例权重的方法，在计算损失函数时增加内领域的实例损失比重。Wang 同时还提出了一种调整领域权重的方法，在训练过程中将内领域和外领域的数据一起训练，调整每一批(mini-batch)训练里内领域句子和外领域句子的比重。Kobus 等[9]提出在词嵌入向量层加入词级别的领域特征，给每个词加上了领域标签。Luong 等[10]提出“两步训练”的领域适应方法：第一步，用外领域的数据训练出翻译模型；第二步，在第一步训练好的模型的基础上，加上内领域的数据继续训练。

区别于上述相关工作，本文的方法着重于学习领域特征，其中领域共有特征的训练借鉴了生成对抗网络(generative adversarial Networks，GAN)。GAN 最先由 Goodfellow 等[11]提出，随后，Wu[12]、Yang[13]等将其应用于机器翻译中，他们使用卷积神经网络(convolutional neural network，CNN)训练判别器，来区分人类专家的翻译和机器生成的翻译，与此同时，改善生成器让机器生成的译文骗过判别器，在互搏中让生成器和判别器变得更强。本文的判别器用于区分语料的所属领域，DIN 基于领域判别器，通过对抗训练学习到领域共有特征，让生成的译文欺骗判别器，使其判别不出译文的领域。DSN 学习到领域特有特征，增强判别器性能，让神经机器翻译模型在翻译时携带领域信息。

本文学习到的领域特有特征和领域共有特征属于不同的系统，提出一个系统集成机制让翻译系统在翻译时融合两种特征，充分利用领域特性。

神经机器翻译里最常用的系统集成方法是Jean等[14]提出的集成方法(ensemble),在解码时集成多个模型的预测结果并得到最优翻译。Garmash等[15]提出两种融合方法,一种是使用固定的权重向量集成多个目标端的预测概率,另一种是在训练时用门机制动态地控制每个模型对预测概率的贡献,集成的每个模型的源端语言不一样,目标端语言相同。

2 基准系统

本文使用基于循环神经网络的RNMT[1-2]作为基准系统,具体结构如图1所示。

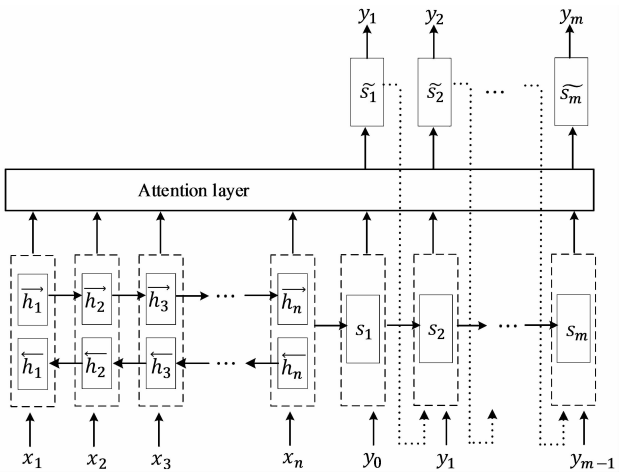

图1 基准神经机器翻译模型

2.1 编码器

编码器用双向循环神经网络对源端输入建模:源端词汇被映射成词嵌入向量序列,得到编码的输入向量序列:$X=x_1,x_2,\cdots,x_n$,n是源端句子的长度。编码器将输入序列编码成隐藏层$h=h_1,h_2,\cdots,h_n$的向量序列表示,每个词的隐藏层向量由双向RNN的结果拼接得到,如式(1~2)所示。

$$h_t=[\overrightarrow{h_t};\overleftarrow{h_t}] \tag{1}$$

$$\overrightarrow{h_t}=f(x_t,\overrightarrow{h_{t-1}}) \tag{2}$$

其中,[;]为向量拼接,$f()$为RNN,本文中使用LSTM,$\overrightarrow{h_t}$是正向编码,$\overleftarrow{h_t}$是反向编码。$\overleftarrow{h_t}$的计算与$\overrightarrow{h_t}$类似,将源端输入序列以相反的顺序送入$f()$。

2.2 解码器

解码器通过注意力向量、目标端隐藏层来预测目标端词汇的生成。图1中的"Attention layer"用来计算上下文向量$\boldsymbol{c}_t$,本文用Luong等[2]提出的全局注意力方法,具体公式如式(3~5)所示。

$$\boldsymbol{c}_t=\sum_{j=1}^{n}\alpha_{tj}h_j \tag{3}$$

$$\alpha_{tj}=\frac{\exp(e_{tj})}{\sum_{i=1}^{n}\exp(e_{ti})} \tag{4}$$

$$e_{tj}=s_t^{\mathrm{T}}\boldsymbol{W}_a h_j \tag{5}$$

其中,$\boldsymbol{W}_a\in\mathbb{R}^{q\times p}$,$p$是源端隐藏层维度,$q$是目标端状态隐藏层维度,$\alpha_{tj}$表示注意力信息。目标端状态隐藏层$\boldsymbol{s}_t$由LSTM得到,上下文向量$\boldsymbol{c}_t$和目标端状态隐藏层$\boldsymbol{s}_t$拼接得到注意力隐藏层$\tilde{s}_t$,最终的目标端词的概率分布$\boldsymbol{y}_t$由$\tilde{s}_t$进行softmax操作得到,如式(6~7)所示。

$$\boldsymbol{s}_t=f(y_{t-1},s_{t-1},c_{t-1}) \tag{6}$$

$$\tilde{\boldsymbol{s}}_t=\tanh(\boldsymbol{W}_c[s_t;c_t]) \tag{7}$$

$$\boldsymbol{y}_t=\mathrm{softmax}(\boldsymbol{W}_y\tilde{s}_t) \tag{8}$$

其中,$f()$为LSTM,$\boldsymbol{W}_c\in\mathbb{R}^{l\times(p+q)}$,$\boldsymbol{W}_y\in\mathbb{R}^{V_y\times l}$,$l$是注意力隐藏层维度,$V_y$是目标端词表大小。源端最终的隐藏层状态初始化目标端隐藏层。

2.3 损失函数

RNMT中,每句话的损失函数定义如式(9)所示。

$$\mathcal{L}_{\mathrm{G}}=\sum_{i=1}^{m}-\log[p(y_i\mid y_{<i};X)] \tag{9}$$

其中,m是目标端译文长度,在RNMT的训练过程中,最终目标是最小化损失函数,使得翻译模型越来越准确。

3 基于领域特征的NMT领域适应

本文提出的系统框架如图2所示。首先,翻译生成器G是基准系统RNMT的翻译模型,生成器G的训练语料是混合了内领域语料和外领域语料的综合语料。其次,在生成器G的基础上固定其网络参数,加入领域判别器,基于注意力隐藏层$\tilde{s}_t$和多层感知机MLP识别当前词所属领域。然后,生成器与领域判别器一起训练,既学习领域特有特征得到领域敏感网络DSN,又通过对抗训练得到具有领域共有特征的领域不敏感网络DIN。最后使用系统集成方法,融合G、DSN、DIN得到最终领域适应系统。

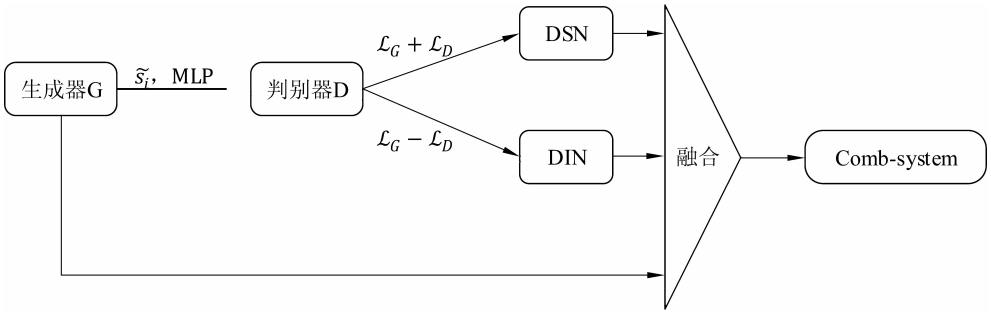

图 2　基于领域特征的神经网络领域适应方法

3.1　领域判别器

本文使用生成器中目标端的注意力隐藏层信息$\widetilde{s_t}$，通过多层感知机训练判别器。注意力隐藏层信息$\widetilde{s_t}$包含了源端的输入信息和目标端 t 时刻之前的译文信息，基于此来预测当前词的所属领域，如式(10～11)所示。

$$p(d_t \mid \widetilde{s_t}) = \text{sigmoid}(V^{\mathrm{T}} \cdot \text{MLP}(\widetilde{s_t})) \quad (10)$$

$$\text{MLP}(\widetilde{s_t}) = \tanh(W_s * \widetilde{s_t} + b_s) \quad (11)$$

其中，d_t 表示 t 时刻单词的领域类别，本文中领域类别只有两种：内领域和外领域。$\boldsymbol{V}^{\mathrm{T}} \in \mathbb{R}^{v \times o}$，$\boldsymbol{W}_s \in \mathbb{R}^{o \times l}$，本文中，$v$ 为 2，o 设置为 250，l 为注意力隐藏层维度。判别器的损失函数设计如式(12)所示。

$$\mathcal{L}_{\mathrm{D}} = \sum_{i=1}^{m} -\log(p(d_i \mid \widetilde{\boldsymbol{s}_i})) \quad (12)$$

其中，m 为当前目标端译文的长度。整个网络在训练领域判别器时固定生成器网络参数，只更新领域判别器参数。

3.2　领域敏感网络 DSN

判别器训练完成后已经具有区分领域的能力，为了利用判别器的领域特征，我们将判别器和生成器一起训练。领域敏感网络借助判别器，训练出携带领域信息的目标端注意力隐藏层向量$\widetilde{s_t}$，系统通过$\widetilde{s_t}$预测目标端单词概率。DSN 的损失函数设计如式(13)所示。

$$\mathcal{L}_{\mathrm{DSN}} = \sum_{i=1}^{m} -\log[p(y_i \mid y_{<i}; \boldsymbol{X})] - \log(p(d_i \mid \widetilde{s_i})) \quad (13)$$

生成器在训练时学习到携带领域特征的注意力隐藏层向量$\widetilde{s_t}$，判别器通过$\widetilde{s_t}$来预测当前训练语句的领域，最小化生成器和判别器的损失，不仅能增强判别器领域判别的能力，还能使生成器通过判别器学习到领域特征，改善翻译效果，让判别器和生成器同时达到最优。

3.3　领域不敏感网络 DIN

与领域敏感网络不同，领域不敏感网络的损失函数是最大化判别器的损失，如式(14)所示。

$$\mathcal{L}_{\mathrm{DIN}} = \sum_{i=1}^{m} -\log(p(y_i \mid y_{<i}; \boldsymbol{X})) + \log(p(d_i \mid \widetilde{s_i})) \quad (14)$$

最大化判别器的损失目的是让判别器区分不出句子的领域类别。在内领域和外领域的语料中，一些词是内领域特有的，很少出现在其他领域；还有一些共有词汇，比如一些功能词，是内领域和外领域的语料中都存在的词汇，具有不因为领域不同，翻译就不同的特性，此时不需要判别器具有准确的判别能力。生成器学习出具有领域共有特征的注意力隐藏层向量$\widetilde{s_i}$，使其能够欺骗到判别器，让判别器分辨不出当前语句的领域，最终训练出领域不敏感网络。

3.4　融合 G、DSN、DIN

为了利用 DSN 的领域特有特征和 DIN 的领域共有特征，本文构建一个集成系统，将 G、DSN、DIN 融合训练。对于领域专有词汇，需要借助领域特有特征，生成其对应领域的正确译文；对于通用词汇，需要借助领域共有特征，避免错误的翻译。G、DSN、DIN 具有不同的特性，预测出的目标端词汇概率分布也不一样，本文的集成系统将不同网络预测出的目标端词汇概率融合在一起，来得到一个综合预测y_t^{comb}，如式(15～16)所示。

$$y_t^{\mathrm{comb}} = w_1\, y_t^{\mathrm{G}} + w_2\, y_t^{\mathrm{DSN}} + w_3\, y_t^{\mathrm{DIN}} \quad (15)$$

$$[w_1; w_2; w_3] = \text{softmax}(\boldsymbol{W}_d) \quad (16)$$

其中，y_t^{G}、y_t^{DSN}、y_t^{DIN}分别是生成器预测的目标端词概率分布、领域敏感网络预测的目标端词概率分布和领域不敏感网络预测的目标端词概率分布。

$\boldsymbol{W}_d \in \mathbb{R}^{1\times 3}$是权重向量，参与网络训练和更新。集成系统中网络的损失计算与基准系统相同。Garmash等[15]使用固定的参数来融合模型，以0.1的步长调试出最优的权重分布，但此方法并不适用于领域适应。领域适应需要在不同的语言、不同的领域进行融合，若采用固定的参数来融合，则需要花费很长的时间去寻找最佳的匹配参数。本文对$\boldsymbol{W}_d$进行随机初始化，通过神经网络训练，来找到最佳的权重组合。使用softmax函数是为了确保综合预测概率y_t^{comb}的概率分布和为1。

4 实验结果和分析

4.1 数据集

本文分别在中英和英德两种语言翻译上验证本文领域适应方法的有效性。语料统计信息如表1所示。

表1 语料统计信息

LDC	Sentences	Src tokens	Tgt tokens
in	20.9k	0.3M	0.3M
out	1.2M	27.9M	34.5M
validation	1.0k	14.8k	17.7k
test1	1.0k	14.4k	17.3k
test2	1.0k	14.7k	17.8k
IWSLT15	Sentences	Src tokens	Tgt tokens
in	207.1k	3.4M	3.2M
out	4.4M	116.1M	108.9M
validation	1.7k	26.4k	24.8k
test1	0.9k	17.8k	16.6k
test2	1.3k	21.1k	20.0k

（注：in表示内领域的语料信息，out表示外领域的语料信息，validation表示验证集信息，test1、test2分别表示两个测试集信息。Sentences表示语料中句子的个数，Src tokens、Tgt tokens分别表示语料的源端单词总数和目标端单词总数。）

中英数据的内领域语料使用LDC（Linguistic Data Consortium）中文广播对话平行语料（LDC2016T09），外领域语料是从LDC语料里抽取的125万句平行语句对，语料包括LDC2002E18、LDC2003E07、LDC2003E14以及LDC2004T07、LDC2004T08、LDC2005T06。本文从内领域分别抽取了1千句作为验证集和测试集（表1中的test1和test2），抽取的验证集、测试集与训练集没有交集。

英德数据上采用IWSLT2015（The International Workshop on Spoken Language Translation）英语到德语的数据集[16]作为内领域的训练语料，该语料所对应的领域是口语领域。外领域语料采用的是WMT2015（Workshop on Machine Translation）英语到德语的数据集。本文把IWSLT2015中的TED tst2012作为验证集，TED tst2013（test1）、TED tst2014（test2）作为测试集。

4.2 实验设置

本文的基准系统是基于Pytorch的神经机器翻译系统Fairseq①。在中英数据上，首先对中文做分词，对英文做tokenization等预处理工作；其次分别对中文、英文实施Byte Pair Encoding（BPE）[17]操作，该方法将训练语料中单词拆分成更为常见的子部分。做BPE处理时，词汇表大小设置为3万。对于英德数据，本文做了tokenization、BPE等处理，做BPE处理时将英文和德文语料混合在一起，生成3万的BPE词表。训练时，中英数据不区分大小写，并且不限制句长，英德数据区分大小写且设置最大句长为50。本文使用multi-bleu.perl评测脚本，评测BLEU[18]值时，中英数据大小写不敏感，英德数据大小写敏感。

实验首先训练出内领域和外领域语料的基准模型，训练这两个模型时源端和目标端使用同一个字典，该字典是混合内领域和外领域语料生成的，也就是公有字典。Fairseq训练参数设置如下：max-token设置为4 000，词嵌入向量维度设为512，源端和目标端隐藏层维度设为512，使用Nag优化方法，初始学习率是0.25，dropout设置为0.2。解码时，使用beam-search方法，beam的大小设置为10，其他参数使用Fairseq的默认参数设置。

4.3 实验结果

本文分别在中英数据和英德数据上实现了基于领域特征的神经机器翻译领域适应，实验结果如表2、表3所示。表2是中英广播对话的领域适应结果，表3是英德的IWSLT15口语领域适应结果。表中Baseline-in为内领域的基准模型，Baseline-out为外领域的基准模型，Baseline-mixed是在混合内领域语料和外领域语料后训练生成的模型，也就是混合领域的基准模型，该模型同时也是生成器G。DSN-mixed和DIN-mixed分别对应领域敏感网络和领域不敏感网络在混合语料上训练的模型，

① https://github.com/pytorch/fairseq/tree/v0.4.0

Comb-G-DSN-DIN-in 为融合 G、DSN、DIN 模型后在内领域语料上继续训练的结果，Comb-G-DSN-DIN-mixed 为融合 G、DSN、DIN 模型后在混合语料上继续训练的结果。

表 2　中英数据实验结果

	Model	test1	test2	avg
Baseline	Baseline-in	5.27	5.80	5.54
	Baseline-out	15.81	14.69	15.25
	Baseline-mixed = G	21.20	21.85	21.53
This work	DSN-mixed	21.75	22.71	22.23
	DIN-mixed	22.17	23.15	22.66
	Comb-G-DSN-DIN-in	22.26	23.12	22.69
	Comb-G-DSN-DIN-mixed	**23.94**	**24.97**	**24.46**
Luong et al. (2015)	Luong-out-mixed	21.22	22.73	21.98
Jean et al. (2015)	Ensemble-G-DSN-DIN-mixed	22.90	23.33	23.12

表 3　英德数据实验结果

	Model	test1	test2	avg
Baseline	Baseline-in	27.94	23.62	25.78
	Baseline-out	23.20	20.14	21.67
	Baseline-mixed (G)	27.63	24.24	25.94
This work	DSN-mixed	28.19	24.61	26.40
	DIN-mixed	28.03	24.63	26.33
	Comb-G-DSN-DIN-in	**31.28**	**26.50**	**28.89**
	Comb-G-DSN-DIN-mixed	29.99	25.85	27.92
Luong et al. (2015)	Luong-out-mixed	28.66	24.12	26.39
Jean et al. (2015)	Ensemble-G-DSN-DIN-mixed	28.08	24.79	26.44

本文比较了 Luong 等[10]和 Jean 等[14]提出的系统集成方法。Luong-out-mixed 复现了 Luong 等人提出的先在外领域语料上训练，再继续在混合语料上训练的方法。Ensemble-G-DSN-DIN-mixed 是在测试时融合 G、DSN、DIN 的预测结果[14]。

本节表中 test1 列为模型在测试集 test1 上的 BLEU 值，test2 列为模型在测试集 test2 上的 BLEU 值，avg 为平均 BLEU 值。

中英领域适应结果

① 从表 2 可以看出，外领域的基准模型(Baseline-out)在测试集上的表现优于内领域的基准模型，混合领域的基准模型(Baseline-mixed)翻译效果在基准模型中最优；

② 学习到领域特有特征的 DSN(DSN-mixed)和领域共有特征的 DIN(DIN-mixed)，各自的翻译性能均有提升，与混合领域的基准模型相比，在测试集上平均提升了 0.70 和 1.13 个 BLEU 值；

③ 本文提出的集成 G、DSN、DIN 的翻译结果，均好于 Luong[10]等提出的“两步训练”法(Luong-out-mixed)和 Jean[14]等提出的“ensemble”方法(Ensemble-G-DSN-DIN-mixed)的翻译结果，其中在混合领域上的集成训练模型(Comb-G-DSN-DIN-mixed)，与混合领域的基准模型相比，平均提升了 2.93 个 BLEU 值。

英德领域适应结果

① 从表 3 可以看出，外领域的基准模型在测试集上的翻译性能没有优于内领域的基准模型(Baseline-in)，混合领域的基准模型在 test1 上的 BLEU 值低于内领域基准模型；

② DSN 和 DIN 的翻译效果均好于内、外领域

基准模型的翻译效果，与混合领域的基准模型相比，平均提升了0.46和0.39个BLEU值；

③ 本文的系统集成方法在英德数据上，内领域上的实验结果（Comb-G-DSN-DIN-in）好于混合领域（Comb-G-DSN-DIN-mixed）的实验结果，与混合领域基准模型相比，提升了2.95个BLEU值。同时，我们的方法也显著优于Luong[10]等的方法和Jean[14]等的方法。

本文提出的领域适应方法，在中英数据和英德数据上的实验结果，与基准系统翻译效果相比均有显著的提升。因此，基于领域特征的神经机器翻译领域适应方法能有效地提升资源稀缺领域的翻译质量。

4.4 实验对比分析

4.4.1 中英与英德实验结果对比

对比表2、表3各实验结果，在中英数据与英德数据上表现有一些不同。中英数据的基准模型中，外领域的翻译模型表现好于内领域的翻译模型，而英德数据中外领域基准模型的翻译效果没有内领域表现得好。本文分析，中英数据内领域的数据量较小，外领域的数据量是内领域的数据量的60倍左右，而机器学习对样本的数据量要求较高，因此内领域的基准模型翻译质量较差，外领域充足的语料有助于内领域翻译效果的提升。这也是系统集成时，中英数据在混合语料上的翻译结果比内领域效果好的原因。英德数据上，内领域的语料量较为充分，所以内领域的基准模型翻译效果较好。同理，结合领域特征，集成系统在内领域翻译效果优于混合领域。

4.4.2 领域判别准确率对比

本文对比了混合语料中，内领域数据占比不同对领域判别准确率的影响。本文进行了4组对比实验，从表4可以看出，无论内领域数据占比多大，判别器和领域敏感网络的领域判别准确率都在0.99左右，可以准确地判别实例所属领域，领域不敏感网络训练时使判别器区分不清句子所属领域，所以准确率很低。当内领域语料所占比例提高时，判别器的acc_in有所提升，acc_out略微下降，acc_mixed总体呈下降趋势，这表明增加内领域数据的比例，会使判别器过拟合外领域的领域特征。将判别器和生成器一起训练时，无论内领域语料的所占比例是多少，领域敏感网络的acc_mixed都维持在0.999左右，具有准确的领域判别能力。test1和test2列对应内领域占比不同的中英混合语料在领域敏感网络和领域不敏感网络的实验结果，avg列为平均BLEU值。从表中可以看出，对于领域敏感网络，当内领域与外领域占比为3∶1时，在test1上取得最好结果；当内领域与外领域占比为1∶1时，在test2上取得最好结果；当内领域与外领域占比增加到6∶1时，平均BLEU值开始下降，且下降明显。对于领域不敏感网络，内领域与外领域占比为1∶1时在test1和test2上均得到最好翻译效果，随着内领域与外领域占比增加，平均BLEU值小幅度波动。由此可见，将一份内领域语料和外领域语料混合在一起，已能充分训练判别器性能，提升翻译质量。

表4 中英数据实验对比

in：out	Model	acc_mixed	acc_in	acc_out	test1	test2	avg
1∶1	D	0.997	0.765	0.999	—	—	—
	DSN	0.999	0.895	1.000	21.75	22.71	22.23
	DIN	0.011	1.000	0.000	22.17	23.15	22.66
3∶1	D	0.993	0.849	0.997	—	—	—
	DSN	0.999	0.886	1.000	22.14	22.40	22.27
	DIN	0.011	1.000	0.000	21.81	22.74	22.28
6∶1	D	0.990	0.900	0.996	—	—	—
	DSN	0.998	0.884	1.000	21.41	20.97	21.19
	DIN	0.011	1.000	0.000	22.11	23.09	22.60
10∶1	D	0.987	0.924	0.994	—	—	—
	DSN	0.999	0.899	1.000	21.21	21.70	21.46
	DIN	0.011	1.000	0.000	21.60	22.69	22.15

（注：单元格中in：out表示内领域语料和外领域语料混合比例，acc_mixed表示判别器对混合语料里句子领域判别的准确率，acc_in表示判别器对混合语料中内领域数据的领域判别准确率，acc_out表示判别器对混合语料中外领域数据的领域判别准确率，-表示未列出结果。）

4.4.3　与“两步训练”法对比

与 Luong 等[10]提出的“两步训练”法(Luong-out-mixed)相比，本文在中英数据上最好结果(Comb-G-DSN-DIN-mixed)比其高 1.34 个 BLEU 值，英德数据上的最好结果(Comb-G-DSN-DIN-in)比其高 2.45 个 BLEU 值。Luong 等人使用预训练的外领域翻译模型在内领域上再训练，导致内、外领域中领域特有单词信息被覆盖，同时缺乏领域特征引导领域共有词汇的翻译，而本文用判别器训练出内外领域的领域特有特征和共有特征，再通过系统集成融合领域信息，指导模型翻译。

4.4.4　系统集成方法对比

Jean 等人的系统集成方法是在测试时融合翻译模型，本文的集成方法则是在训练时融合各翻译系统。表 5 是这两种系统集成方法在多个基准系统和领域网络系统上的融合实验对比，表中的模型均是在中英混合语料上训练，其中 Baseline-1-mixed、Baseline-2-mixed、Baseline-3-mixed 是三个初始化不同的基准模型。从表中可以看出，融合三个基准模型时，本文提出的集成方法 Comb-Baselines-mixed 的实验结果比采取 Jean 方法的 Ensemble-Baselines-mixed 平均提高了 1.39 个 BLEU 值，这表明本文的系统集成方法更能充分融合各模型的预测结果，提高模型泛化能力，这也是 Comb-G-DSN-DIN-mixed 比 Ensemble-G-DSN-DIN-mixed 效果好的原因。此外，用 Jean 的方法融合领域网络的 Ensemble-G-DSN-DIN-mixed 在测试集上的平均得分比融合基准系统的 Ensemble-Baselines-mixed 提高了 1.22 个点，这也从侧面证明了本文的领域网络 DSN、DIN 学习到的领域特征确实能帮助翻译模型提升译文质量。

综上所述，不论是中英数据还是英德数据，通过领域敏感网络和领域不敏感网络提取出领域特有特征和领域共有特征，均能有效地提升翻译质量。并且，在特殊领域的翻译任务上，领域的特有特征能帮助专有词汇的翻译，领域的共有特征能表示通过词汇信息，融合二者可以更有效地提升翻译质量。

表 5　两种系统集成方法在中英混合语料上的对比

	Model	test1	test2	avg
Baseline on mixed	Baseline-1-mixed	21.20	21.85	21.53
	Baseline-2-mixed	19.62	20.83	20.23
	Baseline-3-mixed	19.23	21.11	20.17
Jean et al. (2015)	Ensemble-Baselines-mixed	21.22	22.58	21.90
	Ensemble-G-DSN-DIN-mixed	22.90	23.33	23.12
This work	Comb-Baselines-mixed	22.76	23.82	23.29
	Comb-G-DSN-DIN-mixed	**23.94**	**24.97**	**24.46**

5　总结

本文提出了基于领域特征的神经机器翻译领域适应方法，通过判别器训练领域特征，基于判别器的领域特征分别得到领域敏感网络和领域不敏感网络，使其分别携带领域特有特征和领域共有特征。最后构建集成系统，融合领域特有特征和领域共有特征。实验在中英数据上和英德数据上的翻译效果均显著提升了神经机器翻译的领域适应能力，并优于相关研究的翻译效果。

参考文献

[1] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.

[2] Minh-Thang Luong, Hieu Pham, Christopher D Manning. Effective approaches to attention-based neural machine translation[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 1412-1421.

[3] 李亚超，熊德意，张民. 神经机器翻译综述[J]. 计算机学报，2018,41(12): 100-121.

[4] Rui Wang, Hai Zhao, Bao-Liang Lu, et al. Connecting phrase based statistical machine translation adaptation[C]//Proceedings of the 26th International Conference on Computational Linguistics. Osaka, Japan, 2016: 3135-3145.

[5] Shafiq Joty, Hassan Sajjad, Nadir Durrani, et al. How to avoid unwanted pregnancies: Domain adapta-

tion using neural network models[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 1259-1270.

[6] Rui Wang, Andrew Finch, Masao Utiyama, et al. Sentence embedding for neural machine translation domain adaptation[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 560-566.

[7] Marlies van der Wees, Arianna Bisazza, Christof Monz. Dynamic data selection for neural machine translation[J]. arXiv preprint arXiv: 1708.00712, 2017.

[8] Rui Wang, Masao Utiyama, Lemao Liu, et al. Instance weighting for neural machine translation domain adaptation[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017: 1482-1488.

[9] Catherine Kobus, Josep Crego, Jean Senellart. Domain control for neural machine translation[J]. arXiv preprint arXiv: 1612.06140, 2016.

[10] Minh-Thang Luong, Christopher D Manning. Stanford neural machine translation systems for spoken language domains[C]//Proceedings of the International Workshop on Spoken Language Translation. Da Nang, Vietnam, 2015: 76-79.

[11] Ian J Goodfellow, Jonathon Shlens, Christian Szegedy. Generative adversarial nets[C]//Proceedings of the Advances in Neural Information Processing Systems, 2014: 2672-2680.

[12] Lijun Wu, Yingce Xia, Li Zhao, et al. Adversarial neural machine translation[J]. arXiv preprint arXiv: 1704.06933, 2017.

[13] Zhen Yang, Wei Chen, Feng Wang, et al. Improving neural machine translation with conditional sequence generative adversarial nets[J]. arXiv preprint arXiv: 1703.04887, 2017.

[14] Sebastien Jean, Kyunghyun Cho, Roland Memisevic. On using very large target vocabulary for neural machine translation[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics(ACL). Beijing, China, 2015: 1-10.

[15] Ekaterina Garmash, Christof Monz. Ensemble learning for multi-source neural machine translation[C]//Proceedings of the 26th International Conference on Computational Linguistics. Osaka, Japan, 2016: 1409-1418.

[16] Mauro Cettolo, Jan Niehues, et al. Report on the 11th IWSLT evaluation campaign [C]//Proceedings of the International Workshop on Spoken Language Translation. Hanoi, Vietnam, 2014: 2-17.

[17] Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units[J]. arXiv preprint arXiv: 1508.07909, 2015.

[18] Kishore Papineni, Salim Roukos, Todd Ward. et al. BLEU: A method for automatic evaluation of machine translation[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia, Pennsylvania, 2002: 311-318.

谭敏(1996—),硕士研究生,主要研究领域为自然语言处理、机器翻译。
E-mail: mtan.dzdz@gmail.com

张民(1970—),博士,教授,主要研究领域为自然语言处理、机器学习。
E-mail: minzhang@suda.edu.cn

段湘煜(1976—),通信作者,博士,副教授,主要研究领域为自然语言处理、机器翻译。
E-mail: xiangyuduan@suda.edu.cn