

基于强化学习的古今汉语句对齐研究

喻快 邵艳秋 李炜*

北京语言大学/信息科学学院,
国家语言资源监测与研究平面媒体中心,
北京市海淀区学院路15号, 100083

yukuai_get@163.com yqshao163@163.com liweitj47@blcu.edu.cn

摘要

基于深度学习的有监督机器翻译取得了良好的效果,但训练过程中需要大量质量较高的对齐语料。对于中文古今翻译场景,高质量的平行语料并不多,而粗对齐的篇章、段语料比较容易获得,因此语料对齐很有研究价值和研究必要。在传统双语平行语料的句子对齐研究中,传统方法根据双语文本中的长度、词汇、共现文字等语法信息,建立一个综合评判标准来衡量两个句对之间相似度。此类方法虽然在单句对齐上取得了较好的效果,但是对于句子语义匹配的能力有限,并且在一些多对多的对齐模式上的性能表现不佳。在本文中我们提出尝试利用现在发展迅速且具有强大语义表示能力的预训练语言模型来考虑双语的语义信息,但是单独使用预训练语言模型只能考虑相对局部的信息,因此我们提出采用基于动态规划算法的强化学习训练目标来整合段落全局信息,并且进行无监督训练。实验结果证明我们提出的方法训练得到的模型性能优于此前获得最好表现的基线模型,尤其相较于传统模型难以处理的多对多对齐模式下,性能提升较大。

关键词: 双语对齐; 预训练语言模型; 强化学习; 动态规划

Research on Sentence Alignment of Ancient and Modern Chinese based on Reinforcement Learning

Kuai Yu Yanqiu Shao Wei Li*

Information Science School, Beijing Language and Culture University,
Language Resources Monitoring and Research Center,
15 Xueyuan Road, HaiDian District, Beijing, 100083

yukuai_get@163.com yqshao163@163.com liweitj47@blcu.edu.cn

Abstract

Supervised machine translation based on deep learning has achieved good results, but high-quality aligned corpora are needed in training process. There hasn't been a lot of parallel corpora of high quality for ancient and modern Chinese translation scenes, while coarsely aligned discourse and paragraph corpora are relatively easy to obtain. Therefore corpus alignment is quite valuable and necessary for research. In the study of sentence alignment in traditional bilingual parallel corpus, a comprehensive evaluation criterion is established to measure the similarity between two sentence pairs according to the grammatical information of bilingual text, such as length, vocabulary and context. Although it has achieved good results in single sentence alignment, it has limited ability in sentence matching and poor performance in some many-to-many alignment patterns. We attempt to consider bilingual semantic information by using

* 通讯作者 Corresponding Author

the rapidly developing pre-trained language model with strong semantic representation capabilities. However, the pre-trained language model itself can only cover relatively local information, so we propose reinforcement learning training objectives based on dynamic programming algorithm to integrate global information of paragraphs, and then carry out unsupervised training. Experimental results show that the performance of the model trained by our proposed method is better than that of previous baseline model with best performance, especially compared with many-to-many alignment model, which is difficult to deal with by traditional models.

Keywords: Bilingual Alignment , Pre-trained Language Model , Reinforcement Learning , Dynamic Programming

1 引言

当前基于深度学习的神经机器翻译(Bahdanau et al., 2014; Zoph et al., 2016; Vaswani et al., 2017; Yong et al., 2018)获得了出色的性能表现, 基于端到端神经机器翻译相较于传统的统计机器翻译而言有巨大的优势, 然而这些有监督的模型训练需要大量双语平行语料。如针对英德、英法翻译等其他多语言翻译采用了数百万条平行语料(Ouyang et al., 2020), 模型的翻译能力很大程度建立在超大规模的平行数据上, 而对于一些专门领域的翻译来说, 用人工的方式去获取翻译语料, 覆盖面有限而且代价高昂。因此, 从大规模真实双语语料中自动挖掘双语语料已经成为获取翻译知识的非常重要的途径, 其中一项关键技术就是双语对齐技术。

从20世纪80年代开始有众多学者对双语句子对齐展开了研究, 句子对齐的方法主要有基于长度和基于词汇对齐的方法。Church和Gale(Gale and Church, 1993)提出以文本长度为特征进行文本对齐, 这种方法需要的信息简单, 在法英双语对齐中取得了很好的效果, 但是这个方法有个天然的缺陷, 如果考虑以长度作为特征, 计算机容易把两个1-1模式的正确句子给自动对齐成2-2模式, 并且忽略了文本中的语义信息。Utsuro等(Utsuro et al., 1997)提出使用统计词典信息在英语和日语两种语言之间进行对齐, 规避了仅仅使用长度信息所带来的限制, 提高了句子对齐的准确率, 但是准确率仍然容易受到翻译风格等方面的影响。Lin等(Lin and Wang, 2007)在古文与现代文对齐任务中结合长度信息与共现汉字信息作为特征, 使得对齐性能有了进一步地提高。但是考虑共现汉字这一特征, 往往会产生多对多的模式对齐, 因为考虑的句子个数越多, 共现的汉字个数也多, 所以容易产生多个句子对齐多个句子的情况, 这实际上是与语料中存在绝大多数1-1的对齐模式(刘颖and 王楠, 2013)情况是互相矛盾的。

前人的研究主要是应用语法规则等特征进行对齐, 但是在古代与现代文翻译的语料中, 因为翻译风格的差异, 有些翻译用词会有意译等情况, 单纯的使用语法、句法、共现信息是不够的, 应该尝试考虑语义信息。并且实际语料中会存在一些多对多的对齐语料, 这类匹配模式恰好是语法规则难以覆盖全面的部分。由于不同时代、领域的古文会存在差异并且对应复杂多样的对齐模式, 人工收集各类语料会带来高昂的成本, 因此尝试无监督的方法是非常有必要的。

在深度学习和大数据的支撑下, 自然语言处理技术迅猛发展。EiMo(Peters et al., 2018)、BERT(Devlin et al., 2018)、GPT-2(Radford et al., 2019)等预训练语言模型在下游任务取得了较好的效果, 而基于语言模型预训练和在下游任务精调的框架把自然语言处理带入了一个新的阶段。预训练语言模型的领域适应能力很强, 并且在语义建模方面相较于传统语法规则方法更具有优势, 因此相比语法规则来说, 预训练语言模型更适用于匹配双语句对。

直接使用预训练语言模型进行语料对齐会存在只能建模局部对齐的问题, 模型无法综合的考虑全局文本信息而得出最优的对齐结果; 此外, 如果单纯的使用最长公共子序列(Longest Common Sub-sequence, LCS)算法来构建平行语料进行远程监督训练, 不仅会引入大量噪声, 影响模型的匹配效果, 并且由于多对多对齐模式的出现比例较低, 模型很难通过远程监督信号提升此类对齐的能力。

针对这些问题, 并基于双语语料中绝大多数句子存在对应翻译的假设, 我们提出以覆盖尽量多平行句对为目标的强化学习方法, 来整合段落的全局信息, 预训练语言模型可以以此从段

对齐模式	古文	现代文
1-1对齐	寻此县自不出银，又徭民皆巢居鸟语，不闲货易之宜，每至买银，为损已甚。	考虑到这县不出银矿，加上徭民都住在洞中，说话像鸟语一般，不熟悉交易的方法，每到买银子的时候，损失又很大。
1-2对齐	顷之，世祖命江州众军悉同大举，僧辩乃表皇帝凶问，告于江陵。	不久，世祖命江州各路人马一同大举进攻。//僧辩于是公布皇帝去世的噩耗，把凶讯报告在江陵的世祖。
2-1对齐	城内同时鼓噪，矢石雨下，杀贼既多，贼乃引退。//世祖又命平北将军胡僧率兵下援僧辩。	城内守军同时击鼓呼叫，箭矢码石齐下如雨，杀伤很多叛军，叛军才退兵，世祖又命平北将军曲僧枯率兵沿江而下增援王僧辩。
2-2对齐	又称两受入，易生奸巧，山徭愚怯，不辨自申，官所课甚轻，民以所输为剧。//今若听计丁课米，公私兼利。	另外称两而收，容易导致奸伪狡诈山民。//山民愚蠢而胆怯，无法申明，公家所收的税很轻，徭民却认为很重，如今如果允许按丁收米税，于公私都有好处。

Table 1: 对齐模式示例，//为句子之间的分隔标识

落的全局信息状态来求得最优解，在强化学习过程中，模型对于不同的对齐模式所做出的判断给予相应的奖励信号，模型在决策序列中以最大化双语句对对齐覆盖度这个目标奖励作为探索原则，提升模型在多对多对齐的匹配能力，这样也解决了预训练语言模型在训练中缺乏监督信号的问题，同时模型在大量的无标注的数据中不断地学习，模型对于多对多的对齐模式会有更好的理解。本文的主要贡献有：

- 相较于传统的双语文本对齐方式，本文首次在古文现代文语料对齐任务中引入预训练语言模型来更好的解决古文与现代文的语义匹配问题。
- 本文提出以覆盖尽量多平行句对为目标的全局强化学习训练方法来优化模型，使得模型可以从全局的角度整合段落整体信息，进一步提升了模型在多对多模式的处理能力。
- 我们对于中文古文和现代文对齐的任务构建了大量语料，并且进行了大量的实验和分析，结果表明我们提出的方法在多种对齐模式上的性能均有提升，尤其是在多对多模式的对齐，效果提升尤为明显。

2 相关工作

近些年，在古代和现代文对齐任务中，一种简单结合最长公共子序列与动态规划的算法(Zhang et al., 2018)获得不错的效果。Liu等(Liu et al., 2019)在动态规划的算法基础上结合古文与现代文句子的长度信息、词汇信息、编辑距离的方法，更加全面的考虑到文本中的特征。将依存句法分析融入BiLSTM+CRF(韦希林, 2019)的任务感知模型中，对长句子进行序列标注，使得模型更好的找到长句的分割点，提升句子对齐的准确率。

Gregoire和Langlais(Grégoire and Langlais, 2017)最先提出使用深度学习的方法来解决句子对齐问题，将两个独立的双向循环神经网络对句子进行编码并得到向量化的表示，并将句子向量化表示接入一个全连接层映射成二分类的结果向量，在不依赖手工特征的情况下取得了较好的效果。本文在前人的基础上，引入了更为先进、语义表征能力更好的预训练语言模型BERT对句子进行编码，得到更好的语义匹配能力。

前人的句子对齐工作普遍取得较好的效果是由于数据集中天然存在着大量1-1的对齐模式(刘颖and 王楠, 2013)，该模式对于模型的分类是比较简单的，但是对于数据中存在的部分m-n($0 \leq m, n \leq 2$)匹配模式，如表1所示的2-2、1-2、2-1对齐模式。传统利用语法规则特征的方法主要通过提升1-1的对齐模式来提升模型的性能，却忽略了多对多的对齐模式的准确率。如果仅使用预训练语言模型进行改进，必须根据多种匹配模式进行数据标注，这种从段落对齐的语料库中获取数据的过程是非常繁杂的。并且由于语料中各种匹配模式的比例是不均衡的，要获取数量充足且均衡的样本进行有监督学习是不现实的。本文使用强化学习使得模型尽可能注意到多对多的对齐模式的匹配，以覆盖尽可能多的平行句对为目标，进而避免这一问题。

3 任务与数据

3.1 数据选择和预处理

本文的数据集来源于《二十四史》，我们在互联网⁰上爬取并清洗了300篇古代汉语与相对应的现代文文章，由于段落之间的边界比较容易判断，我们人工的对齐了约9000个段落对齐对，并从中抽取了200对段落对齐语料，进行人工句子对齐，得到1550句对并且标记对应的对齐模式作为本次实验的测试集，测试集的统计数据如表2所示，我们将剩余的人工对齐段落对作为本次实验的训练语料。

在古汉语和现代汉语的数据中，存在多种不同的对齐模式，1-1的对齐模式远远高于其他类别的对齐模式，根据(刘颖and 王楠, 2013)对古文与现代文数据集的统计，1-1的对齐模式占比达到了90.12%，2-2以内的对齐模式的占比达到了98.98%，为了兼顾模型的效率，我们只考虑1-0、0-1、1-1、1-2、2-1、2-2这六种对齐模式，其他出现频率不高的模式没有在考虑的范围之内，并且由于1-0、0-1对齐模式对于模型无法判断，在后续实验中我们将二者归并到2-1、1-2对齐模式里面讨论。

	类别	频率	概率	
主要类别	1-0、0-1	6	0.39%	99.48%
	1-1	1346	86.83%	
	1-2	64	4.12%	
	2-1	106	6.85%	
	2-2	20	1.29%	
其他类别	2-3	2	0.13%	0.52%
	3-1	2	0.13%	
	1-3	4	0.26%	
总计		1550	100%	100%

Table 2: 对齐模式统计

3.2 获取粗对齐平行句对

预训练模型需要构建一定的平行句对语料来对模型微调，使得模型获得初步的粗标能力。将现代汉语翻译成古代汉语存在一个特点，在古代汉语的每一个词都倾向于在现代汉语文本中出现，通常有很强的的顺序性，正确的对齐的语料对通常具有最长的公共子序列的最大长度和，根据这一特点，我们将已有的手工对齐的古文与现代文的段落对作为输入，使用LCS算法来进行段落内部的句子的对齐，构建粗粒度的句对齐的平行语料，并且将其应用到后续的实验。

在微调预训练语言模型之前，模型需要从非平行句对中识别出平行句对，因此必须生成负例，我们使用负采样方法构建训练语料(Conneau et al., 2018)。我们获取包含n个平行句对的平行语料库作为正例，为了构建负例，对于每一个平行句对，在语料库里面进行随机采样，构建新的句对生成m个负例，并用这些数据对BERT模型进行微调，并且记录微调BERT模型在测试集上的结果。

4 我们的方法

本文提出的方法希望通过使用预训练语言模型强大的语义匹配能力，提升模型在古文与现代文句子对齐任务的性能。由于段落中出现的句子的匹配需要结合上下文句子的匹配情况进行考量，本文提出结合动态规划的算法，使得模型能够出于段落的整体角度来考虑全局信息进行匹配，在对句子进行匹配的过程中，模型基于先前的决策做出当前状态的最优解，由于难以直接获得句子级对齐的标签，我们提出采用基于动态规划目标的强化学习方法，其中每一次决策过程就是通过预训练语言模型对当前需要判断的一或多对句子判别是否匹配，而总的目标是使得模型得到的匹配能够覆盖尽量全的句子对。

我们将按照以下两个部分介绍我们的方法：基于预训练语言模型的语义匹配模块、基于动态规划的强化学习序列决策模块。

⁰<http://m.lishishuwu.com/index.php/list/shi24.html>

4.1 基于预训练语言模型的语义匹配

受到BERT(Devlin et al., 2018)在NLI等句对建模任务上成功应用的启发, 本文提出使用类似于BERT中下一个句子判断的形式来建模双语句子级别的匹配预测, 以构建两段文本之间的关系。文本之间关系的判断是一个二分类任务, 在输入层, 我们将古文、现代文的文本分别记为 $x^{(1)}$, $x^{(2)}$, 如式(1)所示:

$$\begin{aligned}x^{(1)} &= x_1^{(1)} + x_2^{(1)} + \dots + x_n^{(1)} \\x^{(2)} &= x_1^{(2)} + x_2^{(2)} + \dots + x_m^{(2)}\end{aligned}\quad (1)$$

我们在输入加入特殊[CLS]来指示文本整体的表示, 使用[SEP]来分隔来自双语的文本片段。需要注意的是, 根据后文中将要介绍的强化学习中具体需要做出的决策的不同, 这里 x_1 和 x_2 可以是一句或者多句单源文本的拼接。经过如下处理得到输入的表示 X 。

$$X = [CLS]x_1^{(1)}x_2^{(1)}\dots x_n^{(1)}[SEP]x_1^{(2)}x_2^{(2)}\dots x_m^{(2)}[SEP]\quad (2)$$

[CLS]表示文本序列开始的特殊标记; [SEP]表示文本序列之间的分隔标记。

对于给定的输入 X , 经过式(3)得到BERT的输入表示 v

$$v = \text{InputRepresentation}(X)\quad (3)$$

在BERT编码层中, 输入表示 v 经过编码, 借助自注意力机制充分学习文本中每个词之间的语义关联, 最终得到输入文本的上下文语义表示 h 。

$$h = \text{BERT}(v)\quad (4)$$

NSP任务只需要判断输入文本 $x^{(2)}$ 是否是 $x^{(1)}$ 的下一个句子, 因此在NSP任务中, BERT使用了[CLS]位的隐含层表示分类预测。具体地, [CLS]位的隐含层表示由上文语义表示 h 的首个分量 h_0 构成, 因为[CLS]是输入序列中的第一个元素。在得到[CLS]位的隐含层表示 h_0 后, 通过一个全连接层预测输入文本的分类概率 P :

$$P = \text{Softmax}(h_0W^p + b^0)\quad (5)$$

W^p 表示全连接层的权重; b^0 表示全连接层的偏置, 最后在得到分类概率 P 后, 与真实分类标签 y 计算交叉熵损失, 学习模型的参数。由于没有直接的监督信号, 单纯将NSP目标移植到本任务上难以进行有效的训练, 所以我们后续将其结合强化学习的方法来构建训练目标, 弥补监督信号缺失的问题。

4.2 基于动态规划的强化学习序列决策方法

为了弥补监督信号缺失的问题, 我们采用动态规划结合强化学习的方法进行优化。强化学习(Francois-Lavet et al., 2018)通过与环境的交互, 在复杂的状态空间序列中不断的决策获得奖励信号, 最大化策略奖励。在我们的任务中, 输入是古文与现代文的段落对齐文本, 环境是动态规划过程每一步根据需要考虑的匹配模式所构建的句对、状态空间是动态规划过程中所有时间步判断的序列、决策是模型对每一步输入的对齐模式句对判断是否匹配。

通过强化学习来学习一个策略网络进行分类, 即4.1节我们提出的基于预训练语言模型的语义匹配, 分类过程可以被看成是一个决策过程, 由于决策过程是离散的, 因此我们使用策略梯度算法(Sutton et al., 2000)来更新参数, 网络模型的参数 θ 根据决策过程的奖励(reward)进行批量更新。

4.2.1 动态规划目标与设定

我们使用动态规划算法, 从段落到段落之间来寻求全局最优解, 动态规划的目标是使得模型预测的匹配能够覆盖尽量全的句子对 (即所有句子都能找到对应的翻译), $D(i, j)$ 表示第1句到第 i 句古文与第1句到第 j 句现代文对齐的分数, 每一步我们需要考虑多种对齐模式, 由

于2-2以内的对齐模式占比达到了极大部分，故我们只考虑2-2以内的对齐模式，具体的状态转移方程如下所示：

$$D(i, j) = \text{Max} \begin{cases} D(i-1, j-1) + \text{BERT}(s_i, t_j) \\ D(i-1, j-2) + \text{BERT}(s_i, t_j \oplus t_{j-1}) \\ D(i-2, j-1) + \text{BERT}(s_i \oplus s_{i-1}, t_j) \\ D(i-2, j-2) + \text{BERT}(s_i \oplus s_{i-1}, t_j \oplus t_{j-1}) \end{cases} \quad (6)$$

其中 $s_i \oplus s_{i-1}$ 表示将句子 s_i 与 s_{i-1} 拼接起来，对于1-0与0-1对齐模式，我们将其归并到2-1、1-2模式，我们仅考虑1-1、1-2、2-1、2-2这四种对齐情况。由此可见，时间复杂度为 $O(MN)$ ，其中 M 是某一古文段落中的句子数， N 是对应的现代文段落中的句子数，随着语料库大小的增长，算法将更为耗时，本文提出的方法更适合段落对齐的语料，而不是篇章对齐语料。

如图1所示，段落对齐的语料中，古文段落被切分为4句，现代文段落被切分为3句，在对齐任务中，每一步的环境取决于动态规划过程中，基于先前的匹配所构建的句对，负责分类的策略网络是我们采用的预训练语言模型-BERT。在动态规划求解的过程中，每一步都会根据判断的四种对齐模式构建对齐句对作为输入，BERT判断句对是否匹配，如图中标记的节点所在时刻，动态规划基于先前语料中已经对齐的1-1状态，根据设定需要考虑的四种对齐模式来构建句对输入BERT进行判断，模型在此刻判断2-1对齐模式是匹配的，记录当前的匹配的对齐模式以及当前状态的总reward，以此类推，直到动态规划算法求解完毕，即获取到动态规划的最后一步所累积的reward，算法详情参考Algorithm 1。

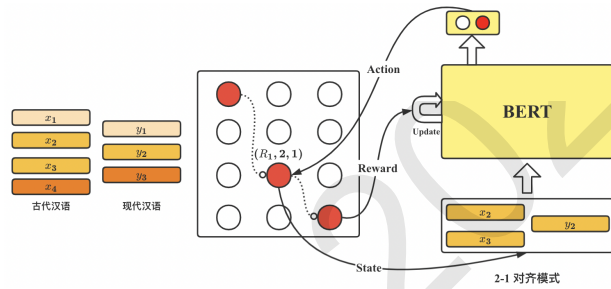


Figure 1: 结合强化学习和预训练语言模型的策略网络模型

4.2.2 Reward目标设定

reward设计总的目标就是使得模型匹配能够涵盖平行句对对齐的覆盖度，BERT模型的决策是一个二分类问题，我们定义为，当 $a_t = 1$ 时，表示当前 s_t 对应的输入句对判断是匹配的；当 $a_t = 0$ 时，表示当前 s_t 对应的输入句对判断是不匹配。

为了使得模型能尽可能的覆盖到多对多对齐模式上，我们对reward函数进行了如下设计，在动态规划的决策过程中的每一个时间步，我们根据不同匹配模式给予不同的reward，具体来说，当 $a_t = 0$ 时，不匹配的情况下reward值为0，当 $a_t = 1$ 时，我们对1-1对齐模式reward值设为2，1-2、2-1、2-2对齐模式reward值分别设为3、3、3，对于后三种对齐模式给予更高的reward是因为我们鼓励预训练语言模型尽可能的在对齐过程中察觉到非1-1的对齐模式，而且模型也会凭此获得更多的多对多对齐模式的监督信号，在训练得到更好的语义匹配能力。

4.2.3 策略梯度训练

策略网络BERT接收到当前对齐模式状态 s_1 ，输出动作 a_1 ，接着环境把 a_1 当成输入，构成新的状态 s_2 作为输入，进而输出决策 a_2 。BERT会不断地接收到动态规划过程中所采集到的状态，并且根据模型内部的参数 θ 而进行决策(action)，进而得到一系列的观测序列 τ ，动态规划完毕后即为完成一次完整的观测序列，每一次完整的观测可以视为一个轨迹 $\tau = (s_1, a_1, s_2, a_2, \dots, s_t, a_t)$ 。

轨迹就是当前的输入的对齐模式的句对、采取的策略，根据 s_1 执行 a_1 的概率 $p_\theta(a_1|s_1)$ ， $p_\theta(a_1|s_1)$ 由BERT模型的参数 θ 决定，该输出是一个分布，BERT会根据这个分布进行采样，决定实际要采取的动作。接下来环境根据动作 a_1 与 s_1 产生 s_2 计算，以此类推，某个完整轨迹 τ 发生的概率为：

Algorithm 1 古文与现代文对齐 → 基于动态规划的强化学习序列决策方法**Require:** 策略网络BERT模型的参数 θ

```

1: 初始化分数  $D[i, j] \leftarrow 0$ 
2: for  $i = 1$  to  $M$  do
3:   for  $j = 1$  to  $N$  do
4:     for  $(d_i, d_j)$  to Align - Pattern do
5:       if  $i - d_i \geq 0$  and  $j - d_j \geq 0$  then
6:         获取当前策略网络BERT根据输入所输出的reward, action, prob
7:         if  $D[i, j] \leq (D[i, j] + \text{reward})$  then
8:            $D[i, j] = D[i, j] + \text{reward}$ 
9:           更新当前的reward, action, prob
10:        end if
11:      end if
12:    end for
13:  end for
14: end for
15: return  $R(\tau) = D[i, j]$ 
16: // 更新策略网络
17: Compute the gradient  $\nabla \bar{R}_\theta$  base on Equation(11)
18: Update  $\theta$  base on Equation(10)

```

$$p_\theta(\tau) = p(s_1) \prod_{t=1}^T p_\theta(a_t | s_t) p(s_{t+1} | s_t, a_t) \quad (7)$$

把环境输出的状态 s 与BERT输出的动作 a 组合起来，并且将动态规划过程中每一步的reward累加起来作为当前回合的总reward，就得到了 $R(\tau)$ ，我们的目标就是调整BERT模型参数 θ ，最大化 $R(\tau)$ 的期望值 \bar{R}_θ 。在决策过程，需要尽可能穷举每一个轨迹 τ ，从分布 $p_\theta(\tau)$ 采样一个轨迹 τ ，计算 $R(\tau)$ 的期望值，并且计算期望 \bar{R}_θ 的梯度，

$$\nabla \bar{R}_\theta = \sum_{\tau} R(\tau) \nabla p_\theta(\tau) \quad (8)$$

计算某一个状态下某一个动作的对数概率 $\log p_\theta(a_t^n | s_t^n)$ ，对这个概率取梯度，在梯度前面乘一个权重，权重就是当前采样轨迹的奖励，计算出梯度后，就可以更新模型，梯度计算如公式(9)所示

$$\nabla \bar{R}_\theta = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^{T_n} R(\tau^n) \nabla \log p_\theta(a_t^n | s_t^n) \quad (9)$$

借鉴(Zinkevich, 2003)的工作，我们使用梯度上升(Gradient Ascent)的方式来更新参数 θ ，来最大化期望奖励，把 θ 加上梯度 $\nabla \bar{R}_\theta$ ，学习率设置为 η ，学习率的调整可以使用Adam(Kingma and Ba, 2014)、AdaGrad(Duchi et al., 2011)等优化器来调整。

$$\theta \leftarrow \theta + \eta \nabla \bar{R}_\theta \quad (10)$$

4.2.4 策略网络初始化

我们从先前手工段对齐语料中抽取300段语料，使用LCS算法构建伪平行语料对BERT-base进行微调后得到的模型，作为本次实验的**基线模型**，并在此基础上进行强化学习训练。

在强化学习的训练过程中，采样轨迹 τ 得到 $R(\tau)$ 一直是正的，由于只采样了部分的动作序列，某时刻的动作可能从来没被采样到，该动作在策略网络的执行概率会降低，但是并不意味着未被采样到的动作不是好的决策，为了让采样得到的 $R(\tau)$ 有正负之分，可以添加基线 b ，使得总奖励为 $R(\tau) - b$ ，如果 $R(\tau) > b$ ，就让当前状态的动作发生的概率升高，反之亦然。通过使用这种带基线的策略梯度算法，提升了训练过程的稳定性，梯度计算如公式(11)所示：

$$\nabla \bar{R}_\theta = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^{T_n} (R(\tau^n) - b) \nabla \log p_\theta(a_t^n | s_t^n) \quad (11)$$

Model	P	R	F
Length (Gale and Church, 1993)	84.7	75.6	79.9
Statistical (Lin and Wang, 2007)	86.1	85.4	85.7
LCS (Zhang et al., 2018)	84.2	87.6	85.9
Proposed Method	88.9	86.1	87.5

Table 3: 本工作的模型和前人的工作在测试集上的性能表现

在强化学习训练过程中， b 设置为基线模型根据输入的段落对齐文本在经过动态规划判断完毕所获得的总reward，reward函数同上4.2.2Reward目标的设定所述。

5 实验部分

5.1 实验设置

对于本实验的基线模型，我们使用LCS与动态规划算法构建的古文与现代文平行句对其进行远程监督训练，本次实验的选用Adam优化器，学习率设置 $2e-4$ ，权重衰减为 $1e-4$ ，并且采用Cosin Warmup(He et al., 2019)策略，学习率按照cosin函数进行衰减，为了提升强化学习的训练效率，我们提出的基于动态规划的强化学习的策略分类模型是在基线模型的基础上进行实验，实验的所有结果均为多次实验求平均值。

5.2 评价指标

本次实验的评价指标采用的是精确率(Precision)、召回率(Recall)、 F_1 -score。假设GB是人工标注的数据集，PB是模型产生的集合，召回率和精确率的计算公式为：

$$Precision = \frac{|GB \cap PB|}{|PB|} \quad Recall = \frac{|GB \cap PB|}{|GB|} \quad (12)$$

召回率和精确率越高，说明句子的对齐效果越好，将召回率和精确率综合起来考虑的 F_1 值，计算公式为：

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (13)$$

5.3 实验结果与对比分析

为了验证本文提出的方法的有效性，我们将本文提出的方法与以下基线模型在测试集上进行比较，包括：

- 基于长度的对齐方法 (Length) (Gale and Church, 1993): 句子对齐与对齐模式和句子长度相关，以句子的长度特征与对齐模式作为特征进行句子对齐。
- 基于综合特征的对齐方法 (Statistical) (Lin and Wang, 2007): 在动态规划的框架下，综合运用长度、对齐模式和汉字信息作为特征，进行句子对齐。
- 基于LCS的对齐方法 (LCS) (Zhang et al., 2018): 将句对之间的最长公共子序列作为句对之间相似度的衡量标准。

实验结果如表3所示，可以看出本文提出的方法相较于其他三种方法在性能上均有较大的提升，其中P值和F值为最优，本文提出的将预训练语言模型与动态规划结合强化学习的方法从两方面提升了古文与现代文对齐的性能。一方面，通过BERT的强大语义表征能力提升了匹配的准确率，更好的兼顾了手工特征难以照顾到的方面。另一方面，使用强化学习构建训练目标的方法使得模型能学习更丰富的匹配模式，提升模型的匹配能力。

5.4 分析

5.4.1 消融实验

为了进一步探究我们提出的方法各个模块对于对齐性能提升的作用，我们进行了消融实验。本次实验包括三个模型，如表4所示，基线模型是使用LCS算法构建的粗对齐的

Model	P	R	F
Baseline (LCS + BERT)	82.0	85.1	83.5
+DP	83.2	86.4	84.8
+DP & RL	88.9	86.1	87.5

Table 4: 消融实验，Baseline采用4.2.4节策略网络初始化的基线模型

	1-1(1346)	1-2(67)	2-1(109)	2-2(20)
Length (Gale and Church, 1993)	78.5(1056)	56.7(38)	55.0(60)	55.0(11)
Stastical (Lin and Wang, 2007)	91.1(1227)	53.7(36)	42.2(46)	40.0(8)
LCS (Zhang et al., 2018)	93.0(1253)	58.2(39)	45.9(50)	40.0(8)
Proposed Method	87.5(1178)	82.0(55)	75.2(82)	60.0(12)

Table 5: 不同模型在多对多对齐模式的准确率

平行句对话料对BERT进行微调，然后再使用微调后的模型对齐得到的结果。+DP是表示在Baseline基础上，在对齐过程以动态规划的全局角度对齐得到的结果。+DP & RL是在使用动态规划的基础上，使用强化学习构建训练目标得到的结果，也就是本文所提出的方法。+DP相较于Baseline而言， P 值、 R 值分别提升了1.2%、1.3%，说明结合动态规划全局信息具有一定的效果。并且可以看到我们提出方法的 P 值、 R 值、 F_1 值相较于Baseline分别提高了6.9%、1.0%、4.0%；相较于+DP而言，我们的方法的 P 值提升了5.7%，召回率 R 下降了0.3%；召回率的下降是因为模型进行强化学习后，模型将小部分1-1的对齐模式错误地进行多对多的对齐，导致召回率偏低，但模型整体的性能得到了较大幅度的提升，也验证了我们提出的应用动态规划与强化学习的方法的有效性。

5.4.2 多对多对齐

为了验证本文提出的方法在多对多对齐模式情况下的效果，我们对不同对齐模式下的句子对齐的数量进行了单独的统计，从表5可以看出，相较于前人的工作来说，我们提出的方法在1-2、2-1、2-2等多对多对齐模式下准确率均为最优，并且有较大幅度的提升，我们的方法在1-1对齐模式上准确率相较于基于综合特征的对齐方法以及基于LCS的对齐方法分别低了3.6%、5.5%，这是因为我们使用强化学习构建目标过程中，使得模型尽可能匹配到非1-1的对齐模式，我们提出的方法在多对多对齐方面的表现更好了，这是符合直觉的，也证明我们在强化学习构建训练目标对于模型的整体性能提升是有价值的。

我们将人工对齐，+DP以及+DP & RL三种方法以图2中样例来进行对比，如图3所示，人工对齐的路径是(0, 0) → (2, 2) → (3, 3) → (4, 5) → (5, 6)，在Baseline的基础上，使用动态规划从全局的角度进行对齐，+DP对齐的路径是(0, 0) → (1, 1) → (3, 3) → (4, 5) → (5, 6)可以看出，因为使用远程监督信号含有大量的噪声，对齐的第一步，+DP就错误地将2-2对齐判断为1-1对齐，直接导致下一步的对齐也错了，尽管后续(3, 3) → (4, 5)的1-2模式匹配正确，但是整体对齐的准确率降低了，而结合我们提出的结合强化学习构建训练方法，+DP & RL方法，在第一步就正确的匹配了2-2模式的句对，后续的对齐路径也是和人工对齐的路径完全重合，这也证明我们提出的方法在多对多对齐上确实是有一定效果的。

5.5 样例分析

尽管本文提出的方法取得了一定的提升效果，但是我们也发现了我们的方法难以解决的语义匹配的例子，如以下样例所示：

• 样例1

- ◇ 古文：斩循及父嘏，并循二子，亲属录事参军阮静，中兵参军罗农夫、李脱等，传首京邑。
- ◇ 现代文：杜军将卢循和他的父亲卢嘏，以及卢的二个儿子，卢的亲属录事参军阮静，中兵参军罗农夫、李脱等人斩首。他们的脑袋被送到京城。

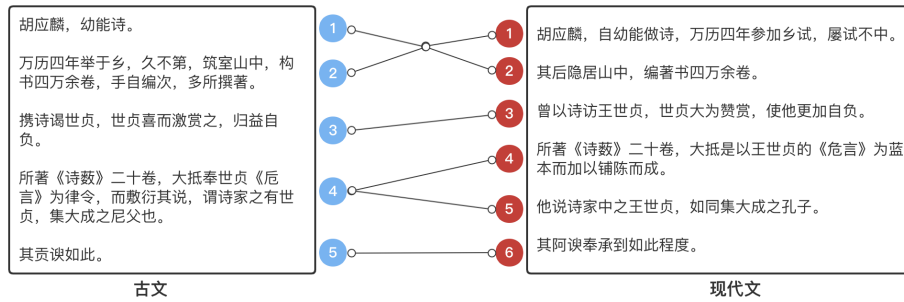


Figure 2: 人工对齐示例；古文第1、2句与现代文第1、2句为2-2对齐模式；古文第4句与现代文第4、5句对1-2对齐模式。

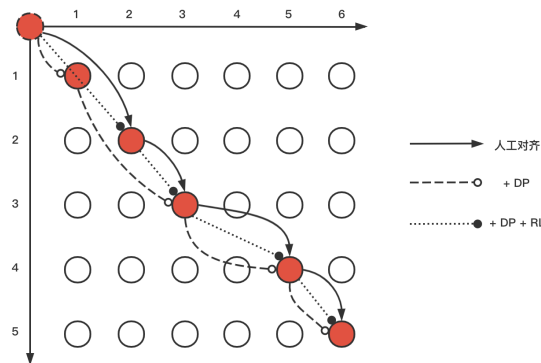


Figure 3: 不同方法的对齐路径对比

• 样例2

- ◇ 古文：魏建国四年，慕容子晃伐之，入自南陔，战于木底，大破钊军。追至丸都。
- ◇ 现代文：魏建国四年，慕容的儿子慕容晃攻打他，从南陔进军，大战于木底，大败钊的军队，追到丸都。

如样例1，该样例的正确的对齐模式是1-2，使用LCS算法可以很好的判断出是1-2对齐模式，但是使用我们提出的方法后，BERT模型却将其判断成1-1对齐模式，因为模型认为现代文中的“**他们的脑袋被送到京城。**”这一句在语义上与对应的古文相关度较低，并且该句的长度相较于前一句较短，构成类似于长句+短句的情况，模型在语义编码的时候，该句占比权重较低，该句的语义编码在整体中容易被忽略，造成了模型的漏匹配，故判断成了1-1对齐模式。

同样的问题在样例2中更为明显，BERT模型同样的将古文中的“**追至丸都。**”这一句给漏匹配，样例2的现代文更为明显的构成了长句+短句的情况，使得模型出现了漏匹配的情况。在先前的阐述中，我们将0-1、1-0的对齐模式归纳到1-2、2-1进行讨论，这样会直接的将一句无关的文本加入到相关的1-1对齐的句对中，给训练样本中注入了小部分的噪声，也间接性的造成了模型对于长句+短句出现漏匹配的情况。

6 结论

针对古代汉语和现代汉语对齐任务，我们引入具有强大语义建模能力的预训练语言模型BERT来更好地进行语义匹配，在此基础上，我们采用动态规划的方法使得模型从全局的角度来匹配文本，将动态规划目标和强化学习方法结合使得模型能够尽可能提高双语对齐的覆盖度。我们构建了古文与现代文对齐测试集数据，并验证了我们提出的方法的有效性，相较于传统方式难以解决的多对多对齐模式，有着明显的提升。

但是，本文模型还存在着不足，虽然预训练语言模型有着强大的性能，在语义匹配能力上还是存在优化的空间，对于多对多的对齐模式，长句+短句的情况下，容易出现漏匹配的情况，在未来的研究中会更加关注如何更好的提升模型的语义匹配能力。除此之外，也会对如何更加有效的获取和利用句对齐的粗粒度的平行语料进行进一步的研究。

致谢

本成果受国家自然科学基金项目(61872402),教育部人文社科规划基金项目(17YJAZH068),北京语言大学校级项目(中央高校基本科研业务费专项资金)(21YBB19, 18ZDJ03),模式识别国家重点实验室开放课题基金资助。

参考文献

- D. Bahdanau, K. Cho, and Y. Bengio. 2014. Neural machine translation by jointly learning to align and translate. *Computer Science*.
- A. Conneau, G. Lample, R. Rinott, A. Williams, Samuel R Bowman, H. Schwenk, and V. Stoyanov. 2018. Xnli: Evaluating cross-lingual sentence representations. *Conference on Empirical Methods in Natural Language Processing*.
- Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.
- Duchi, John, Hazan, Elad, Singer, and Yoram. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*.
- Vincent Francois-Lavet, Peter Henderson, Riashat Islam, MarcG Bellemare, and Joelle Pineau. 2018. *An Introduction to Deep Reinforcement Learning*.
- William A. Gale and Kenneth W. Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75–102.
- F Grégoire and P. Langlais. 2017. A deep neural network approach to parallel sentence extraction.
- T. He, Z. Zhang, H. Zhang, Z. Zhang, and M. Li. 2019. Bag of tricks for image classification with convolutional neural networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- D. Kingma and J. Ba. 2014. Adam: A method for stochastic optimization. *Computer Science*.
- Z. Lin and X. Wang. 2007. Chinese ancient-modern sentence alignment. In *Computational Science - ICCS 2007, 7th International Conference, Beijing, China, May 27 - 30, 2007, Proceedings, Part II*.
- Dayiheng Liu, Kexin Yang, Qian Qu, and Jiancheng Lv. 2019. Ancient-modern chinese translation with a new large training dataset. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*.
- X. Ouyang, S. Wang, C. Pang, Y. Sun, and H. Wang. 2020. Ernie-m: Enhanced multilingual representation by aligning cross-lingual semantics with monolingual corpora.
- Matthew Peters, M. Neumann, M. Iyyer, M. Gardner, and L. Zettlemoyer. 2018. Deep contextualized word representations.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.
- Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation.
- T. Utsuro, M. Yamane, Y. Matsumoto, and M. Nagao. 1997. Bilingual text matching using bilingual dictionary and statistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv*.
- C. Yong, Z. Tu, F. Meng, J. Zhai, and Yang Liu†. 2018. Towards robust neural machine translation.
- Z. Zhang, W. Li, and Qi Su. 2018. Automatic transferring between ancient chinese and contemporary chinese.
- M. Zinkevich. 2003. Online convex programming and generalized infinitesimal gradient ascent. *icml*.

B. Zoph, D. Yuret, J. May, and K. Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*.

刘颖 and 王楠. 2013. 古汉语与现代汉语句子对齐研究. *计算机应用于软件*, 30(11):4.

韦希林. 2019. 基于深度学习的双语长句分割方法研究. Ph.D. thesis, 北京交通大学.

JCL 2022