

基于生成对抗网络的越南语新闻事件共指关系识别方法

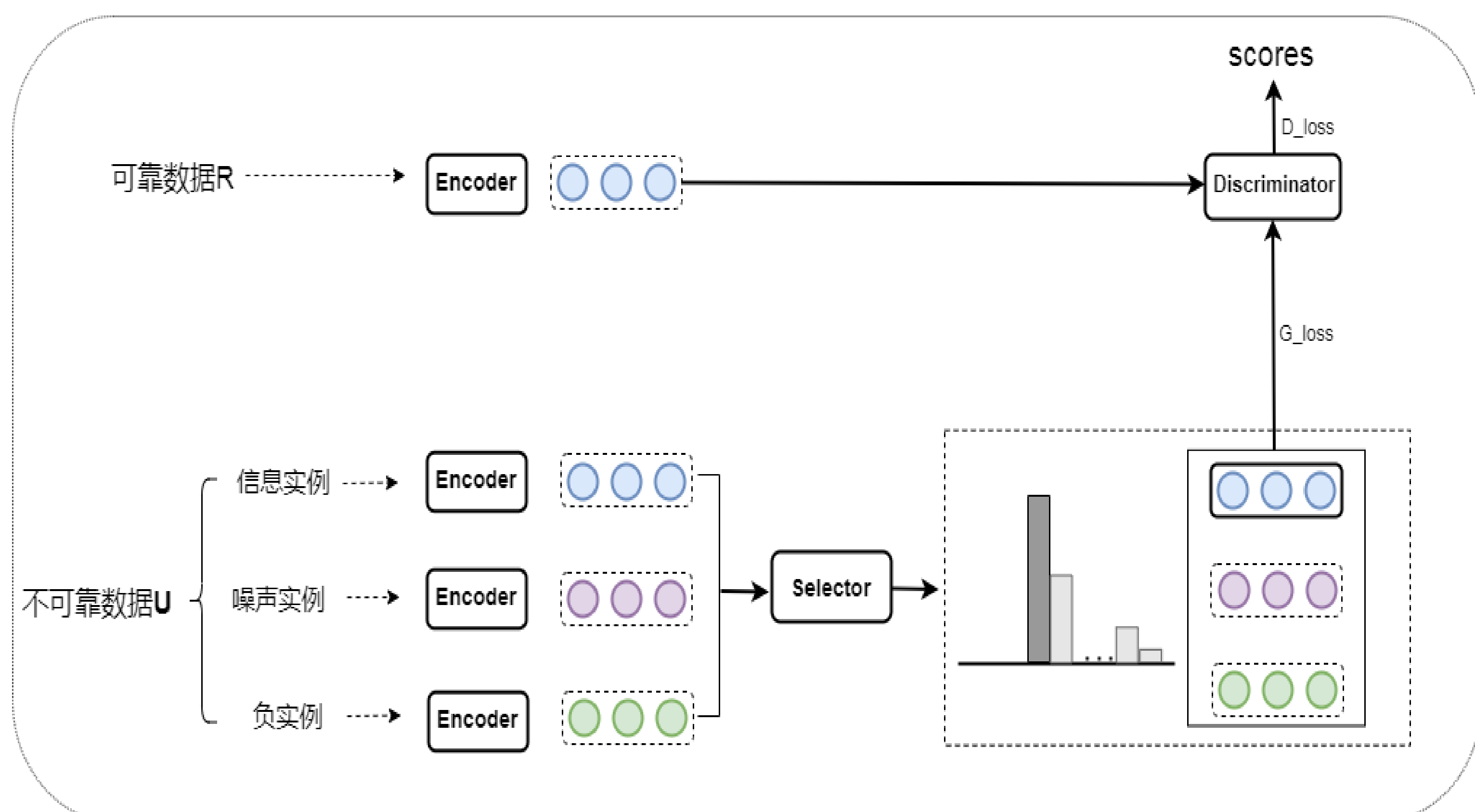
CCKS2022

汪翠、余正涛、梁晨

引言

事件作为信息的重要载体，蕴含着事情发生的时间、地点、参与者等信息。在同一篇文本中通常存在多个事件句之间具有不同的表述方式但却互相关联、指向现实世界中同一件事情的情况，则认为这些事件句之间具有共指关系。然而，在同一篇新闻报道中往往存在不同事件句之间具有相似上下文但不具有共指关系的情况，而容易对模型造成干扰。

方法



(1) 编码器：采用PhoBERT对事件对实例中的目标句子、句子的词性标记、围绕触发词n窗口的上下文、以及事件触发词进行编码与联结来表示事件信息。

(2) 对抗训练：通过选择器和判别器最大最小的博弈训练，使选择器能够关注到不可靠数据U中富含信息的实例，赋予其更高的混淆分数。以及提高判别器的鉴别能力，使判别器在可靠数据R的指导下可以更好地识别来自U中的实例是否具有真实的共指关系。

(3) 通过全局优化算法组建事件共指链：对于每一篇新闻文档而言，以事件对为单位完成共指关系识别后，依据文档内事件之间的传递性属性，以及针对识别误差导致的矛盾关系情况设置整体的全局优化算法以组建单文档内的事件共指链。

实验数据

选取越南《人民报》、越南《通讯社》等新闻网站的文本数据，参考公开数据集标注共7种新闻事件类型。

实验及分析

模型	B3	MUC	CEAF _e	CoNLL
CDGM	58.6	60.1	55.3	58.5
EC_BERT	70.1	65.8	63.9	66.3
EC_NNC	69.4	59.2	62.1	60.7
MLNN	71.2	59.6	65.9	65.5
PGAN	72.1	62.3	66.9	67.1

基于生成对抗网络的越南语新闻事件共指关系识别方法在越南语事件数据集上的性能优于其他基线模型。

结果

- 针对越南语新闻事件共指关系识别任务，本文提出一种基于生成对抗网络的越南语新闻事件共指关系识别方法，采用PhoBERT对事件信息进行表征，并借助生成对抗网络有效提高了模型的降噪能力，采用全局优化算法获取了单文档内的事件共指链。

通讯作者：汪翠
(1987315145@qq.com)

2022全国知识图谱与语义计算大会
China Conference on Knowledge Graph and Semantic Computing
秦皇岛 8.24-8.27