论文修改说明

感谢审稿专家提出的意见（Weaknesses和Comments）。作者根据意见，对论文相关内容进行了修改或说明。

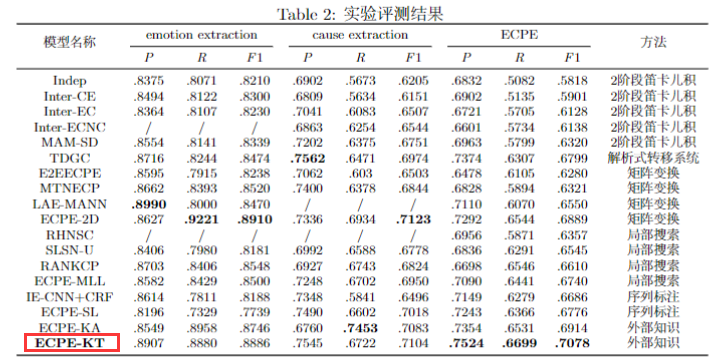
**REVIEWER #1**

**意见1**：

It is better to highlight ECPE models in Table 2 to distinguish them from the selected baselines.

**修改说明：**

（1）已将表2中的ECPE-KT模型名字加粗作区别显示，修改内容如下图。

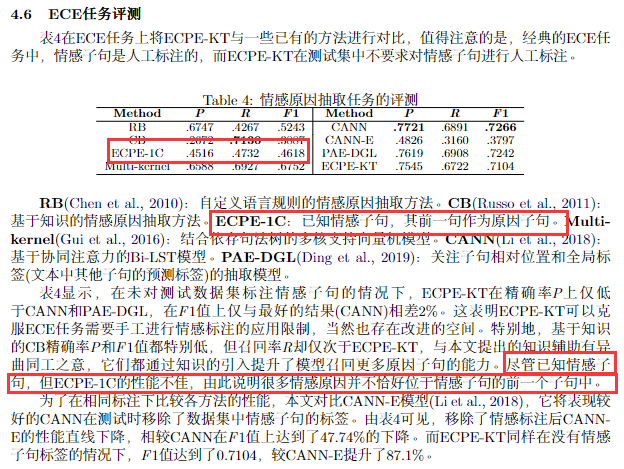


**意见2**：

By looking at the case studies, an interesting observation is that the correct cause sentence is very likely to be the previous sentence. How about adding a simple baseline, and seeing how it performs? This shows the difficulty of this task.

**修改说明：**

（1）根据意见描述，在ECE任务评测中增加一个基准模型（ECPE-1C），其采用的方法是：已知情感子句，其前一句作为原因子句。修改内容如下图。



**意见3**：

Why there are missing values in Table 3 of the ECPE-KT-Trans method?

**修改说明：**

（1）缺失原因：表3中的ECPE-KT-Trans用于消融实验中分析子句间交互的效果，因此只注重表达情感-原因对的抽取，所以未标注其在情感子句和原因子句中的抽取效果。

（2）修改内容：表3中添加ECPE-KT-Trans在情感子句和原因子句中的抽取效果。

**意见4**：

Is it possible to optimize pairing prediction in section 3.5? As mentioned, if there are N sentences, the model still needs O(N^2) effort to make predictions. If this is a bottleneck for model computation? A simple way is to only make predictions in the adjacent k sentences (i.e., k=3).

**修改说明：**

（1）文章3.5节并没有花费O(N^2)去计算，正是采用了窗口机制预测前后k个子句。

**REVIEWER #2**

**意见1**：

本文所引入的LIWC，词性属于常规信息，创新性不足。

**修改说明：**

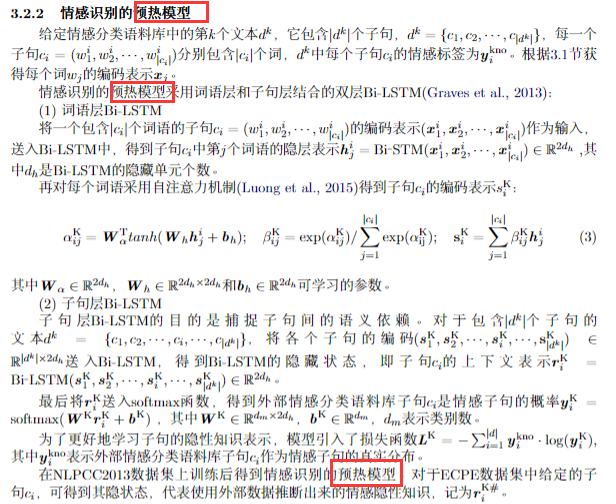
知识有很多种，如人工词典、句法分析、知识图谱、人工规则、甚至一些标注或未标注的相关数据集。本文探索了部分知识在情感原因抽取任务上的有效性，而类似于LIWC这种人工词典是一种重要的外部知识。后期还会探索如何使用新的模型来更充分地利用这些知识。

**意见2**：

3.2.2 中提及的情感识别的预训练模型表达不妥，应该是情感识别分类模型。预训练模型具有专有含义，请参考 pre-trained language model or foundation model。

**修改说明：**

（1）将3.2.2节中的“预训练模型”都修改为“预热模型”，修改内容如下图。



**意见3**：

有错别字，排版也需要改进；模型结构图不够清晰

**修改说明：**

通读全文，对个别字词进行修改，对排版和模型结构进行了优化。

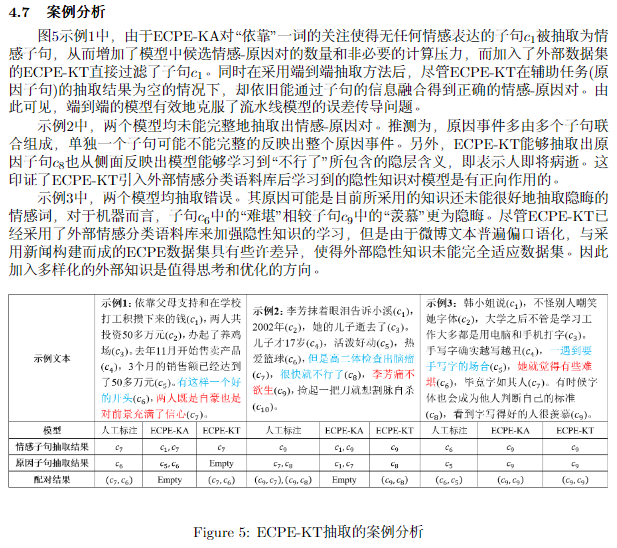
**REVIEWER #3**

**意见1**：

第一，示例2人工标注为2个情感-原因对，ECPE-KT发现了其中一个，且是具有隐含含义的一个。把这种结果归为“错误”，也许要求太高了一点。

**修改说明：**

将4.7.1节和4.7.2节进行合并，将图5和图6进行合并，命名为“ECPE-KT抽取的案例分析”，不再采用“正确、错误”等字眼进行描述。修改内容如下图。



**意见2**：

第二，文中较详细地描述了ECPE-KT模型及其实验结果，但对它为何会比已有模型性能更优越，却较少分析。

**修改说明：**

文章主要贡献概述了本文与已有模型的差异点和产生较优结果的原因，在消融实验中给予以了较充分的分析。

**意见3**：

第三，对199个含有2个或2个以上情感-原因对的测试用例的准确率如何，建议做进一步说明。

**修改说明：**

（1）根据意见描述，在实验评测结果分析段落最后增加含有2个或2个以上情感-原因对的测试用例的准确率，具体内容为：

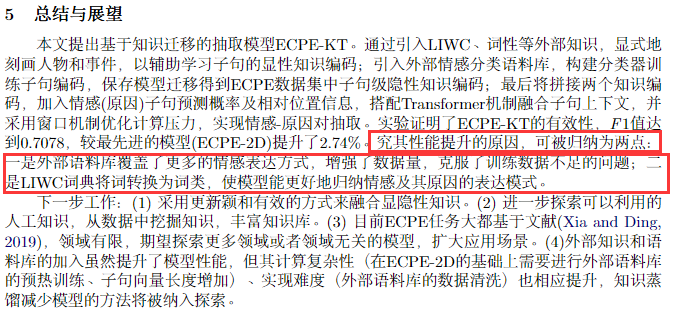
特别地，对于含有2个或2个以上情感-原因对的测试用例，ECPE-KT抽取完整的用例占9.375%，至少抽取出1对的用例占62.5%，而Inter-EC识别出的比例分别为6.25%和为56.52%。可见，两者在多情感-原因对文本上的正确率都有待提高，但是ECPE-KT依然有更好的表现。

**意见4**：

第一，进一步加强ECPE-KT模型为何更具有优势的论证。

**修改说明：**

（1）同意见2，并在“总结与展望”中进一步挖掘ECPE-KT模型加了KT之后带来的改变，以论证其性能提升。修改内容如下图。

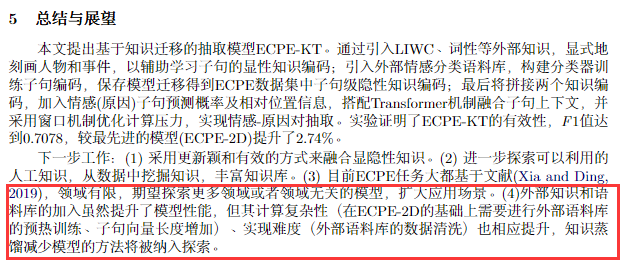


**意见5**：

第二，ECPE-KT通过引入LIWC、词性等外部知识、外部情感分类语料库，最终提高了性能，但其计算复杂性、实现难度可能也会有所提升。除了比较P、R、F1值等指标外，不妨也把新模型与已有模型的计算成本也做一下对比。

**修改说明：**

（1）在“总结与展望”中对模型的计算成本做出定性分析，并提出改进方法的探索方向，修改内容如下图。



**其他修改：**

1.补齐作者信息。

2.添加资助基金号