

Acl2021

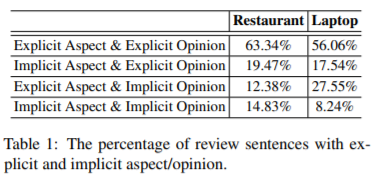
产品评论包含大量隐含的方面和意见。然而，现有的基于方面的情绪分析研究大多忽略了这一问题。本文引入了方面-类别-观点-情感四元提取(aspect-category-opinion-emotion, ACOS)，目的是提取评论句中所有的方面-类别-观点-情感四元，为基于方面的隐含方面和观点的情感分析提供**全面**支持。

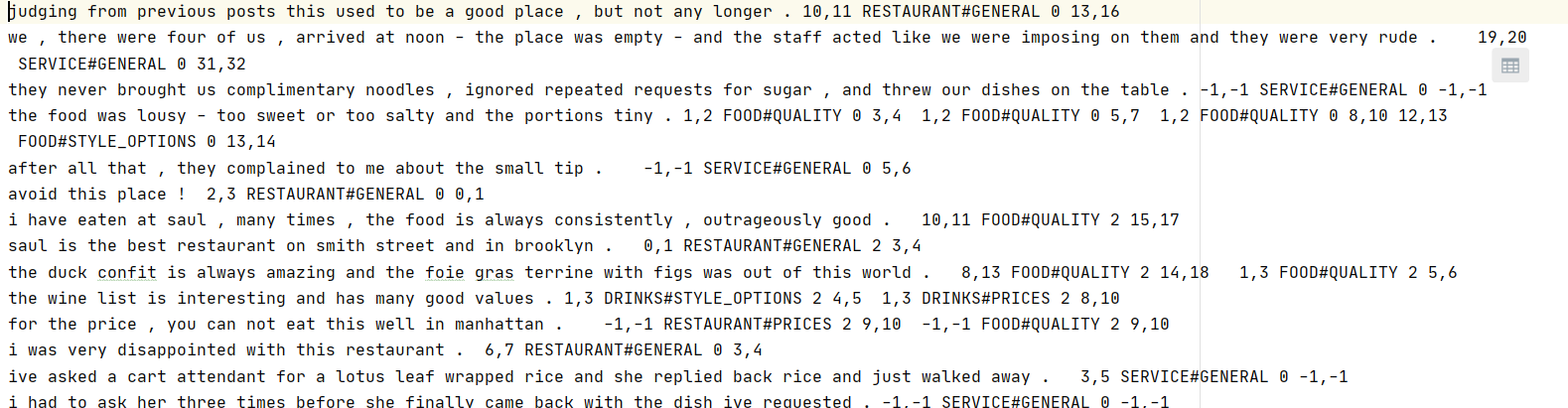
本文创新点：

1. 提出了一个新的任务，Aspect-Category-Opinion-Sentiment(ACOS)四元组抽取任务，以此来支持从评论中挖掘更多的隐含方面和意见。
2. 构建了两个新的数据集来适应这项任务，Restaurant-ACOS和Laptop-ACOS。两个数据集都包含了四元组抽取和隐含的aspects和opinions。
3. 尝试用四个baseline系统去进行测试。

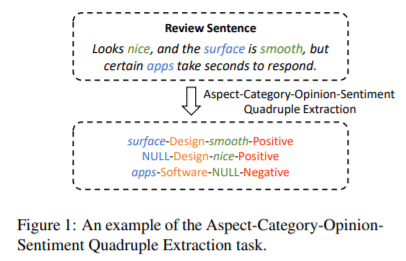
方面级情感分析：

方面级情感分析的核心任务是从产品评论中提取实体和描述方面的opinions，并判断方面的情感极性。方面级情感分析包含许多子任务：aspect-sentiment extraction 和 aspect-opinion extraction和aspect-opinion-sentiment triple extraction（方面-意见-倾向三元组）。但是大多数现有的研究都只能实现显式方面（explicit aspects）和意见（opinion）的抽取，而忽视了评论中隐式方面和意见。从下表中可以清晰地看到评论中这两者的占比。





ACOS 四元组抽取任务：



该任务主要会面临下面两个挑战：

1. 没有标记好的包含所有implicit aspects和opinions的aspect-category-opinion-sentiment 四元组抽取数据集。
2. 如何构建一个好的模型去整合两个抽取任务（方面抽取、意见抽取）、两个分类任务（方面分类、情感分类）。

为了解决这些问题，作者构建了两个新的数据集Restaurant-ACOS（对于SemEval Restaurant数据集增加了四元组抽取标注）和Laptop-ACOS（在亚马逊收集的Laptop数据）。通过将原来的二元组抽取、三元组抽取任务改造成四元组抽取任务，作者提出了四个baseline：Double-Propagation-ACOS, JET-ACOS, TAS-BERT-ACOS and Extract-Classify-ACOS

形式化定义：

方面级情感分析：

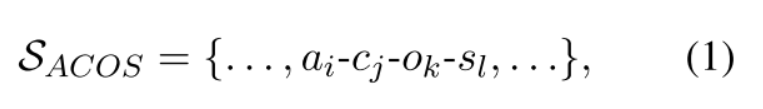
Aspect：意见所指向的实体的属性。

Category：在一个特定的领域中，方面属于的类别。

Opinion：对于方面的形容，通常是形容词或形容词短语。

Sentiment：语义对于方面情感极性的倾向（Positive、Negative、Neutral）

从评论中抽取的四元组可以定义如下：



注意，一个评论通常包含多个方面和观点。ACOS四元组提取任务不仅要识别四个元素，还要将它们组合成一组有效的四元组，同时考虑隐含的方面/意见。由于隐式方面/观点没有明确地表达为一个词或短语，在隐式方面a的情况下，我们设置为NULL并使用分类c描述观点目标，在隐式观点o的情况下，我们设置为NULL并使用情感s来描述语义方向。

数据集：

Restaurant-ACOS数据集是基于SemEval 2016 Restaurant数据集及其扩展数据集构建的。

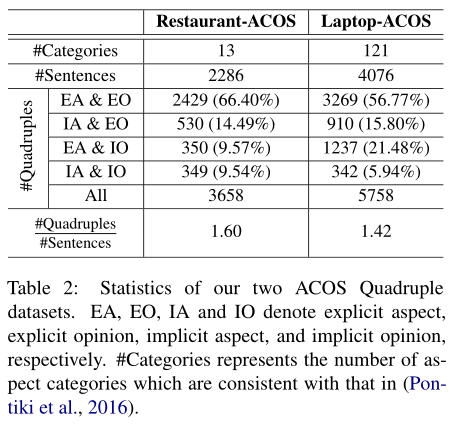
Laptop-ACOS是亚马逊平台收集的2017年和2018年全新的笔记本电脑数据集(涵盖华硕、宏碁、三星、联想、MBP、MSI等6个品牌的10种笔记本电脑)。它包含4,076个评论句子，比SemEval笔记本电脑的数据集大得多。

Q：如何得到标记数据？

SemEval 2016年餐厅数据集用显式和隐式方面、类别和情感进行了注释。其他工作进一步增加了意见注释。我们整合他们的注释，构建方面-类别-意见-倾向四元组，并进一步注释隐含的意见。

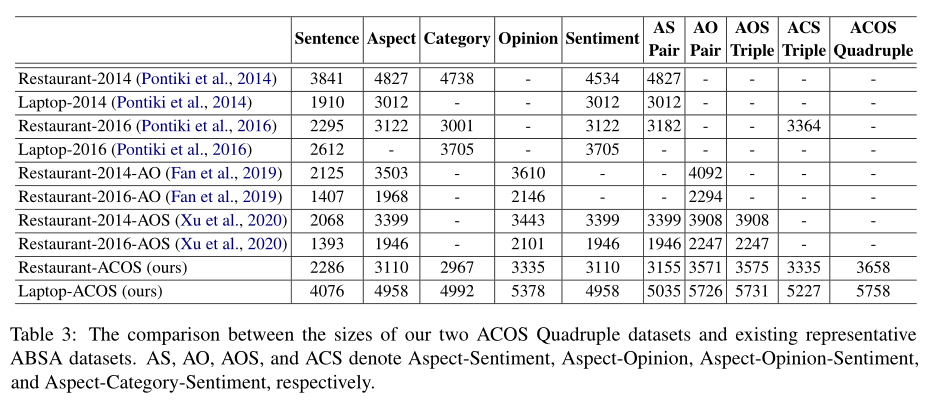
对于Laptop-ACOS，我们完全由自己来注释这四个元素及其对应的四元组。我们使用SemEval 2016 Laptop数据集中定义的方面类别。选择两名熟悉基于方面的情感分析的博士生作为标注者，使用Yang等人介绍的标注工具进行独立标注。两名批注者的四重匹配f1分值为75.86%，表明两名批注者基本一致。如果有不同意见，第三位专家将被要求做出最终决定。

Q：两个数据集区别？



Restaurant-ACOS数据集包含2286个句子，包含3658个四元组，Laptop-ACOS数据集包含4076个句子，包含5758个四元组。正如我们所提到的，四元组中很大一部分包含隐含的方面或隐含的意见。通过比较两个数据集，可以发现Laptop-ACOS隐含意见的百分比高于Restaurant-ACOS。

Q：数据集和其余已知数据集区别？



Restaurant 2014/2016和Laptop 2014/2016分别表示SemEval2014/2016的Restaurant和Laptop数据集。Restaurant2014/2016包含了方面、分类、情感的注解。需要注意的是，两个数据集中的类别定义是不同的。Laptop2014只包含方面和情感的注解，而Laptop2016只包含类别和情感的注解。

Restaurant-2014-AO和Restaurant-2016-AO是基于Restaurant 2014和2016提出的两个方面-意见对数据集。他们删除了隐含方面的句子，增加了意见注释。后来，进一步将原本包含在restaurant 2014/2016的情绪s添加到Restaurant-2014/2016-AO中，并获得了两个方面-意见-情绪三重数据集:Restaurant-2014-AOS和Restaurant-2016-AOS。对于Restaurant-ACOS，我们集成了上述注释来构建ACOS四元组。但需要注意的是，在Restaurant-2016中，我们保留了隐含方面的句子，并进一步注释了隐含的观点。因此，Restaurant-ACOS的大小(包括句子、AO对和AOS三倍)约为Restaurant-2016-AO和Restaurant-2016-AOS的1.6倍。

新的Laptop-ACOS有4076个评论。ACOS注释数量分别为4958、4992、5378、4958。结合这些元素，我们构建了5035对AS、5726对AO、5731对AOS三元组、5227个ACS三元组和5758个ACOS四元组，几乎是restaurant-ACOS的两倍。

方法：

我们采用双传播-ACOS、JET-ACOS、TASBERT-ACOS和提取-分类-ACOS四个基线系统对ACOS四重提取，用于将方面观点对抽取、方面-类别-观点三重抽取或方面-观点-情感三重抽取等代表性方法应用于ACOS四重抽取。

1. 双传播ACOS

Double Propagation(DP)是一种基于规则的方面-意见-情绪三元组抽取方法。该方法首先提取所有的方面-意见-情绪三元组，然后为每个三元组分配方面类别c，以适应ACOS四元组提取任务。我们将这种适应的方法命名为双传播ACOS。

具体来说，我们首先采用DP算法提取方面-意见-情绪三元组，利用方面a和意见o之间的句法关系，在每个评论中迭代提取它们，并依赖情绪词典基于方面和意见来对情绪s进行赋值（即积极情绪、消极情绪、和中立）。其次，为确定每个提取的三元组的方面类别c，我们采用如下策略：如果三元组中的方面a在训练集中，我们将其最常共现的方面类别作为最后的方面类别；否则，我们就采用输入评价中最相近的方面的类别来作为最终的类别。

1. JET-ACOS

JET模型是aspect-opinion-sentiment三元组抽取任务中效果比较好的模型。

JET为该任务引入了端到端框架，将方面、其对应的意见和情绪极性的识别与位置感知标签方案相结合。与Double-Propagation-ACOS类似，我们首先使用JET提取三元组，然后预测每个提取三元组的方面类别，从而使JET适应我们的任务。

具体来说，我们首先基于JET算法得到候选的方面意见-情感三元组，然后设计一个基于bert的模型来得到提取的三元组的方面类别。该模型如下：输入评论句子r，我们首先将其提供给BERT以获得上下文感知的token表示如下：



其中，表示句子r的输出表示。

接下来，给定aos，我们可以获得a和o的信息在句子中的平均值，即：



然后，通过一个全连接层+sigmoid获取c（这个是要训练的网络）：



表示有效四元组，表示无效四元组。

在训练阶段，我们采用标准的二元交叉熵损失进行优化。在推理阶段，我们将从JET中提取的方面-意见-情绪三元组与我们预测的方面类别相结合，得到每个评论句子的所有四元组。

1. TAS-BERT-ACOS

TAS-BERT模型是效果较好的aspect-category-sentiment三元组抽取任务。该方法将基于方面类别的情感分类和方面提取整合到一个统一的框架中，将方面类别a和情感极性s附加到评论句中作为BERT的输入。为了使TASBERT适应我们的ACOS提取任务，我们提出在TASBERT中采用输入转换策略进行基于类别-情绪条件方面-意见共提取，然后过滤无效的方面-意见对，形成最后的四元组。

具体的，通过下面几步完成四元组抽取任务：

首先通过TAS-BERT模型得到候选 category c、sentiment s、aspect。



将r、c、s组成的x传入BERT模型得到隐藏状态H。



其中， 表示句子r的输出，h[CLS]用于category-sentiment验证。

然后，我们通过将其建模为单个序列标记任务来进行方面意见共提取（基于类别-情绪条件方面-意见共提取）。

具体来说，我们采用了一个修改后的Begin-Inside-Outside (BIO)标记方案，它由五个标记组成：{BA,IA,BO,IO,O}，表示方面的开始和内部，意见的开始和内部，以及其他。我们将h\_r输入到CRF层，以提取与输入类别和情绪相关的方面和意见，如下所示：



最后，得到分类结果。

将分类的结果和之前抽取的三元组进行笛卡尔积运算，得到候选四元组。



有个排列组合。A、O表示解析出的方面和建议。C、S表示发现的类别和倾向。

最后，将aspect和opinion的向量平均，传入到全连接层+sigmoid，判断是否有效，得到最终抽取的四元组。

1. Extract-Classify-ACOS

通过两步实现四元组抽取。第一步，在评论中抽取aspect-opinion二元组，然后基于aspect-opinion预测category-sentiment。（刚好和上面相反）

先将评论加上标记放入BERT模型中，得到隐藏状态。



对于explicit aspect-opinion，使用CRF+BIO标记规则来完成。其中我们使用h[CLS]来判断是否implicit aspect or opinion。得到潜在的方面集合S\_A，意见集合S\_O。进行笛卡尔积操作，得到：

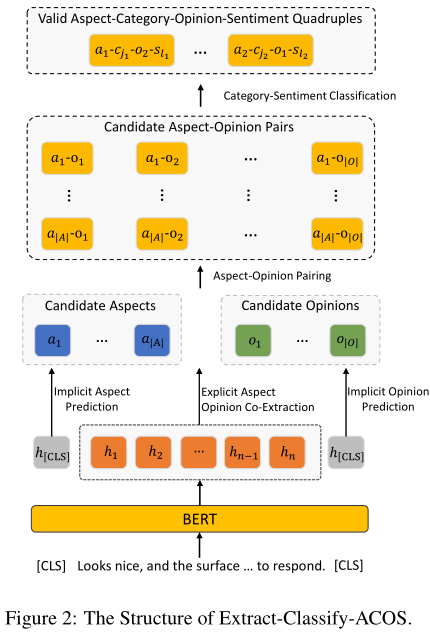


接下来，我们将类别-情感分类建模为一个多类分类问题。具体的，我们对向量取平均输入到全连接层结果。





Extract-Classify-ACOS模型整体架构：



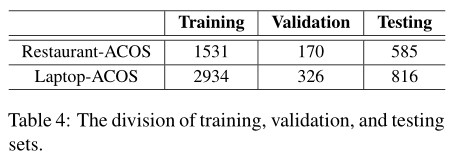
实验设置：

在extract - classifier - acos中，我们采用了bert-base (Devlin et al.，2018)作为基本编码器，它由12个堆叠的Transformer块组成。在训练过程中，我们使用了带有权值衰减修正的BERT的AdamW优化器。评论句子的最大长度设置为128，涵盖两个数据集中的所有句子。

我们将方面意见共抽取和类别情绪分类的批数和学习率分别设为[32,2e-5]和[16,3e5]。dropout设置为0.1。

JET-ACOS类别分类中的批处理大小和学习率以及TAS-BERT-ACOS中的方面-意见对过滤均设置为[8,5e -5]，这两个模块的其他设置与extract - classifier - acos相同。

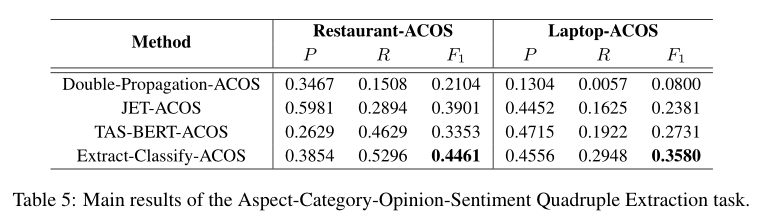
我们根据下表将原始数据集划分为训练集、验证集和测试集。



在评价中，四元组被认为是正确的，当且仅当这四种元素及其组合与金四元组中的元素完全相同。在此基础上，我们计算了Precision和Recall，并使用F1-score作为AOCS四重提取的最终评价指标。

结果：

1. 可以看出，Double-PropagationACOS的性能最低。在评审中，仅使用规则就很难识别多个隐式元素及其复杂的组合，这是合理的。
2. JET-ACOS和TAS-BERT-ACOS在Restaurant-ACOS数据集上性能较好，而TAS-BERT-ACOS在Laptop-ACOS数据集上性能较好。
3. 在四个基线系统中，extract-classifier-acos的性能最好。它在Restaurant-ACOS上的表现比JET-ACOS好5.60%，在Laptop-ACOS上的表现比TAS-BERTACOS好8.49%。extract-classifier-acos的主要优点是可以获得鲁棒性较高的召回分数。相比之下，JET-ACOS的准确率更高或相当，但召回率要低得多。
4. extract-classifier-acos在两个数据集上的f1score都不高(0.4461和0.3580)。这是因为评估指标是基于精确匹配的，而ACOS四重提取比传统的ABSA任务更加复杂。



消融实验：

正如我们所提到的，很大比例的评论包含隐含的方面/观点。因此，有效地建模隐含的方面/意见是非常重要的。为了研究不同系统处理隐式方面/意见问题的能力，在下表中，我们将测试集分成四个子集，并观察在不同子集上的性能:

1)EA和EO表示具有显式方面和显式意见的子集;

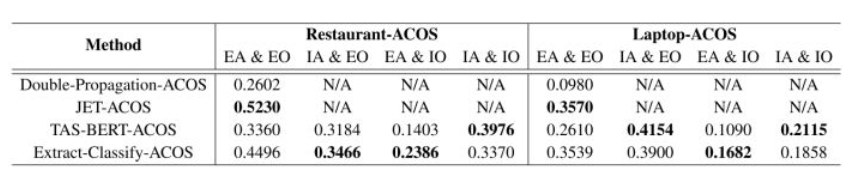
2) IA和EO表示包含隐性方面和显性意见的子集;

3) EA和IO表示显性方面和隐性意见的子集;

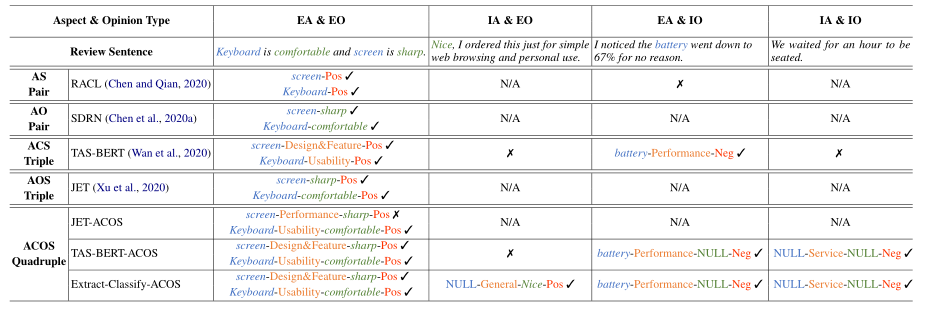
4) IA和IO表示包含隐式方面和隐式意见的子集。

其中，Double-Propagation-ACOS和JET-ACOS只能处理EA和EO，而TAS-BERT-ACOS和Extract-Classify-ACOS可以同时支持隐式方面和隐式观点。他们在建模隐含的方面/观点方面表现出相当的能力。Extract-classification-acos在IA & EO和EA & IO两种情况下表现较好，而TAS-BERT-ACOS在IA & EO和IA & IO两种情况下表现较好。而在EA和EO两种数据集的情况下，extract-classification-acos的性能显著提高。

我们继续比较了不同子集下的表现。总体来看，还是在EA&EO类型的数据中表现最好，这个符合直觉。但是，结果表明，EA和IO的性能最差，相比于IA和IO。一个可能的原因是，在第一个数据集中还超过了IA&EO。这里猜测的可能原因是IA&IO对应的类别比EA&IO更有规律，更容易预测。



个例分析：



表中我们进一步进行了案例研究，通过比较一些有代表性的方法对五个ABSA任务的预测，包括方面-情感(AS)对提取、方面-意见 (AO)对提取、方面-类别-情感(ACS)三重提取、方面-意见-情感(AOS)三重提取和ACOS提取。我们选择了4个不同的句子，根据其方面/观点是显性的还是隐性的，观察通过不同的方法所获得的预测结果。可以观察到:

1. RACL在EA和EO情况下准确提取了AS对，但在EA和IO情况下不支持隐式方面，不能对我们的测试句进行预测
2. SDRN仅能在EA和EO情况下提取方面-意见对;
3. JET在EA和EO的情况下只能提取方面-意见-情绪对
4. 虽然TAS-BERT支持隐式方面或隐式意见的方面-类别-情绪三重提取，但在IA-EO和IA-IO两种情况下，它未能对我们的测试句子给出准确的预测;
5. 对于三个ACOS基线系统，JET-ACOS仅能在EA和EO情况下进行ACOS四重提取，预测存在错误。
6. TAS-BERT-ACOS和extract - classifier -ACOS支持隐式方面和隐式观点下的ACOS四重提取。TAS-BERT-ACOS的性能优于JET-ACOS，但在IA & EO的案例中仍然失败。
7. extract - classifier - acos通常表现最好，在所有情况下产生更准确的预测。

结论：

本文提出了一种新的任务：方面类别-意见-情感(ACOS)四重提取，旨在系统地解决隐式方面/意见问题。我们为此任务构造了两个新的数据集，其中包含隐式方面和隐式意见的ACOS注释。最后，我们使用四个基线系统对任务进行基准测试。实验证明了新任务在基于方面的隐含方面/观点情感分析中的优势。本文的重点是介绍新的任务和数据集。拟议的四个基线系统相对简单，留有许多进一步改进的余地。我们欢迎未来在这方面提出更强模式的工作。我们也欢迎在其他ABSA（Aspect-Based Sentiment Analysis）任务中使用我们的数据集。

双向传播：

1. 基本思想

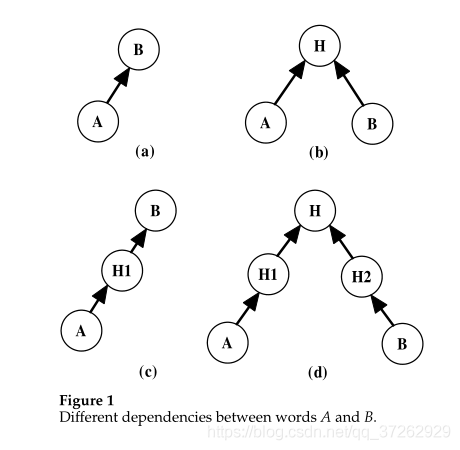
  该方法的基本思想是通过对语法关系的识别，利用已知的和在之前的迭代中提取的意见词和目标，迭代地提取意见词(或目标)。关系的识别是提取的关键。

1. 关系识别

  识别意见词/目标与其他意见词/目标之间的关系是我们意见词汇扩展和目标提取方法的关键。在本节中，我们将详细描述关系标识。在此，为了方便起见，我们将**意见词与目的词之间的关系称为OT-Rel，意见词之间的关系称为OO-Rel，目的词之间的关系称为TT-Rel**。  
  在这项工作中，我们使用依赖语法来描述语法上的关系。在依赖语法中，两个词A和B之间的句法关系可以描述为A(或B)依赖于B(或A)。我们定义了两类来概括两个词在句子中所有可能的依赖关系。

**定义1：直接依赖（DD）**  
  直接依赖关系表明一个单词依赖于另一个单词，而在依赖关系路径中没有任何附加的单词或者它们都直接依赖于第三个单词

**定义2：间接依赖（IDD）**  
  间接依赖指的是一个词通过一些附加词或者它们都通过附加的单词依赖于第三个单词。

在下图中，(a)(b)表示的是直接依赖，©(d)表示的是间接的依赖  


注意，DDs和IDDs仅描述所有可能依赖关系的拓扑。然后对意见词和目标词进行词性标注的约束，并对依赖路径上的句法关系进行约束。  
  使用Stanford进行词性标注，使用Minipar2进行句子解析。在这篇文章中，认为意见词是形容词，目标词是名词/名词短语，现在看来，这种识别方法，可能会造成大量的aspect缺失，近年来很多论文对其进行了改进。

1. 意见词典扩展和目标提取

   利用本文定义的关系，基于传播迭代地执行意见词汇扩展和目标提取任务。为了引导传播，我们只需要一个种子意见词典。  
   我们的提取方法采用基于规则的策略。

3.1基于关系定义的传播规则

在传播的过程中，有四个子任务。

使用意见词提取目标。

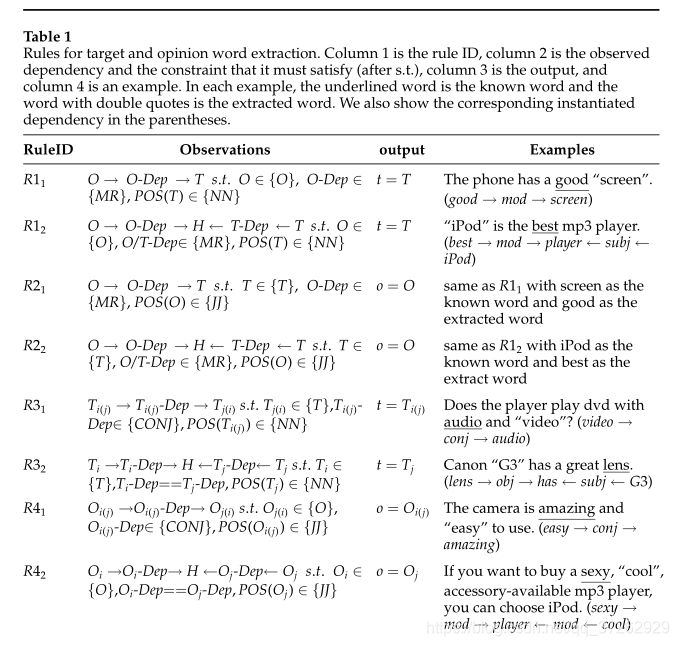
利用提取的目标提取目标

利用提取的目标提取意见词

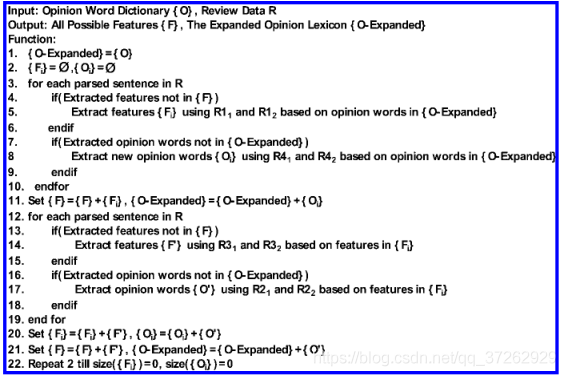
根据给定的意见词和提取的意见词提取意见词

可以看出，任务1和3使用OT-Rels（意见词与目标之间），任务2使用的是TT-Rels（目标与目标之间），任务4是OO-Rels（意见词之间）。

本文在当前方法中仅使用了DD依赖关系（IDD更适合正式的文本），注意这里的规则是独立于域的。

目标词和意见词的提取规则。第1列是规则ID，第2列是观察到的依赖项和它必须满足的约束(s.t之后)，第3列是输出，第4列是一个示例。在每个示例中，带下划线的单词是已知单词，带双引号的单词是提取的单词。我们还在括号中显示了相应的实例化依赖关系。  
  
在表中，o表示提取意见词，t表示提取目标。{O}是意见词的集合，{T}表示的是目标的集合，H代表任何一个词。pos（o or t ）和（O-Dep or T-Dep ）分别表示凡是O or T的pos信息和依赖信息。{JJ}和{NN}分别是潜在意见词和目标的POS标签集。{JJ}包含JJ，JJR，JJS；{NN}包含NN，NNS。{MR}由描述观点词与目标词之间关系的依赖关系组成。{CON}只包含conj（连接词）。箭头表示依赖关系。

3.2传播算法

在该算法中，提供了关于产品的意见词词典O和评论数据R作为输入。这些步骤按照传播顺序设置。当没有新的意见词或目标可以添加时，它就停止了。该算法已在前面的文本中说明。这里我们就不重复了。  


3.3意见词的极性分配

本文提出了一种基于相关证据的极性分配方法。基本的直觉是，人们通常用一致的方法表达他们的观点，除非有明确的想法的词语，如but和however。在实践中，作业实在意见词抽取过程中完成的。在我们描述我们的方法之前，让我们对意见词和目标做一些观察:

评审同一目标的相同极性。评审是一个评审人写的文件。通常情况下，审稿人对同一目标有着相同的情感或极性，尽管这个目标在审稿中可能不止一次出现。

领域语料库中的同一意见词具有相同的极性。通常情况下，同一意见词在一个领域语料库中具有相同的极性。

根据上述观察，我们通过以下规则对提取的意见词和目标进行极性推断：

异构的规则： 对于已知目标提取的意见词和已知意见词提取的目标，我们赋予它们与已知目标相同的极性。（个人感觉，在这里直接赋予情感极性可能会造成误判，现在目标已提取出来了，是不是可以通过其他神经网络的方法进行判断情感极性）。

均匀规则。对于由已知意见词提取的意见词和由已知目标提取的意见词，除非它们之间存在相反的词，否则我们将它们分配为与已知意见词相同的极性。

Intra-review rule 。一些目标从其他评论中提取了一些新的意见词。在目前的审查中，这些目标不应表现出极性，因为它们不符合观察1。因此，意见词不会产生两极分化。如果这些意见只出现在当前的审查中，观察2也不能适用。为了给这样的意见词分配极性，我们利用总体的审查极性来推断。我们假设意见词具有复习的极性，即如果复习是正面的，则评价词具有正极性，反之则具有负极性。评审极性值计算为所含已知意见词的极性值之和(正极性为+1，负极性为- 1)。如果最终和大于0，则评审为正，否则为负。

注意：由于这两种观察结果，一个意见词或目标可能具有多重极性。为了解决冲突，我们将极性值相加。正极性为+1，负极性为- 1。如果和大于0，最后的极性是正的，否则是负的。

3.4 意见目标修剪（删除噪音）

删除噪音（不正确的目标和意见词），文中只给了一个去处目标噪音的方法，并没有对提取的意见进行修剪。