# 蒸馏判别和泛化知识的事件检测通过Δ-表征学习

事件检测系统依靠判别知识来区分模糊触发词，而泛化知识则用来检测未知/稀疏触发词。当前神经事件检测程序专注于以触发器为中心的表示法，它能很好地提取识别知识，但在学习泛化知识方面效果不佳。为了解决这个问题，本文提出了一个∆-学习方法，通过有效的解耦、增量学习和自适应融合事件表征来提取判别和泛化知识。实验表明我们的方法显著优于以前的不可见/稀疏触发字的方法，并且在ACE2005和KBP2017数据集中实现了最新的性能。

曾经的ED系统大多判别能力很强但却缺少泛化能力。（OOV,OOL应对情况会很差）

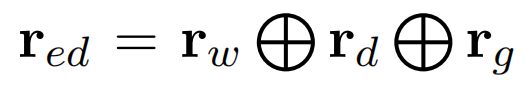
事实上现在通用语言模型（ELMO）相比针对事件检测的模型效果差不多。。

*decouple*, *learn*, and *fuse* alterable **∆**- parts for event representation

解耦，学习，融合可变的事件表示

instead of learning a single comprehensive representation

而非全面的单个事件表示



rw是触发词的预训练词表示，rd是特定于词汇的事件表示，它捕捉识别歧义触发器的识别知识，rg是词法无关事件表示，它捕获用于检测未知/稀疏触发器的泛化知识，而+是融合不同部分的融合函数。

they are independently learned starting from **r***w* and are intended for capturing incremental knowledge for event detection

RD既是与事件相关的部分，又是词汇相关的∆学习部分，目的是学习与事件相关但与词汇无关的泛化知识Rg。

## 贡献：

We propose a new representation learning framework - **∆**-learning

设计了一种新的事件检测方法。通过有效的解耦、独立学习和自适应融合事件表示，

我们将每个令牌ti嵌入为ti=[PW；pp；pe]，其中PW是它的字嵌入（elmo），pp是它的位置嵌入，pe是它的实体标记嵌入。上面则是最初始的嵌入。。

r\_word

是elmo

# Rd（r\_discrimination）

用GRU实现的注意力机制，（q，k，v均是隐藏层输出，且打分函数用多层感知机。多层感知机输入是h0，hj输出用softmax归一）得到中心词对句子的上下文表征。**r***d* = [**h**0; **c**0]

H=gru（t），c=attention（h0，h）

显然这种中心词与上下文强相关的更适合消歧，常规注意力机制是很依赖中心词的。trigger-centric

（至少这篇文章这样认为。我直接的想法是非精确的词向量，或者直接是未知词的词向量会很不准）（实际上是做一个模型融合）（transfomer风格的直接对序列做。）

而这种消歧知识并不适合应对oov，ool（中心词根本没出现过或中心词根本没打过这个标签时这个效果不会很好。）

# Rg

We simply use DMCNN 来作为提取语义无关的模块，但采用了对抗Δ学习方法来消灭语义信息。（希望能语义无关，但事件相关。）（不是，干脆弄一个开始就不输进去的一个网络算了。。他这里还是输进去了。）

一维卷积的做法：向量维度相当于通道数，一个卷积会把不同通道求总和。即不只对当前通道按时间维度相关。并且对各个通道求完后再弄个和作为输出的一个通道。

此处会取中心词。时间长度可能会稍微变长短或不变（取决于你的填充设置。）

这里不知道他的维度细节，并未说明

而他提了一个dynamic-pooling，即把左最大，右边最大的相应维度输出。

# 融合r：

玄学：对抗学习。去耦合，分离与融合。增量学习。

“*[Trigger] to death*”

being a strong trigger pattern for Attack event,

which can be used to detect many different trig

ger words, such as *fifired*, *hacked*, *beat*. In this way,

even an unseen trigger candidate *t* can be easily

identifified by leveraging such knowledge.

解释了为何要语义无关的知识。

# lexical-enhanced

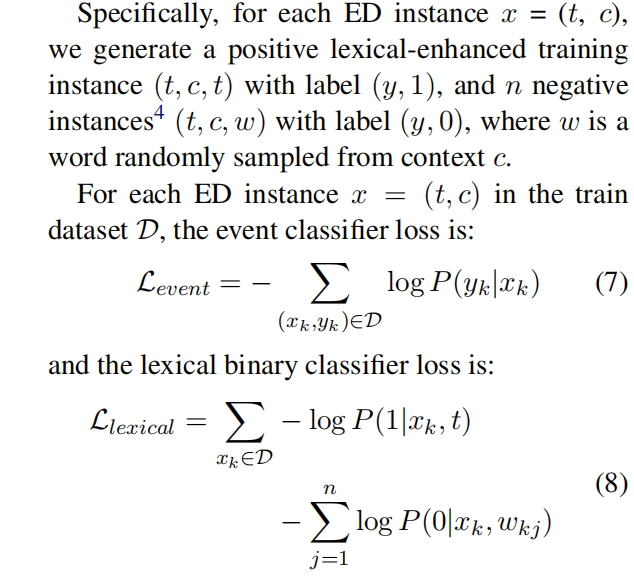
Rd与rw predicate y（event type）与 trigger（因为要trigger—centric嘛）

输出是：event type auxiliary lexical classififier

# Lexical adversarial

灵感：话说极端情况是当这个信息相对语义是噪音时那就

平常我们总是想到不断提取与目标相关的特征，但从来没相关提取与目标完全无关的特征。用这个特征去推断相应结果会非常差。这就意味有一个网络不断想从该特征去推断目标和一个处理网络不断把原始特征处理为目标无关的特征。



行文思路：

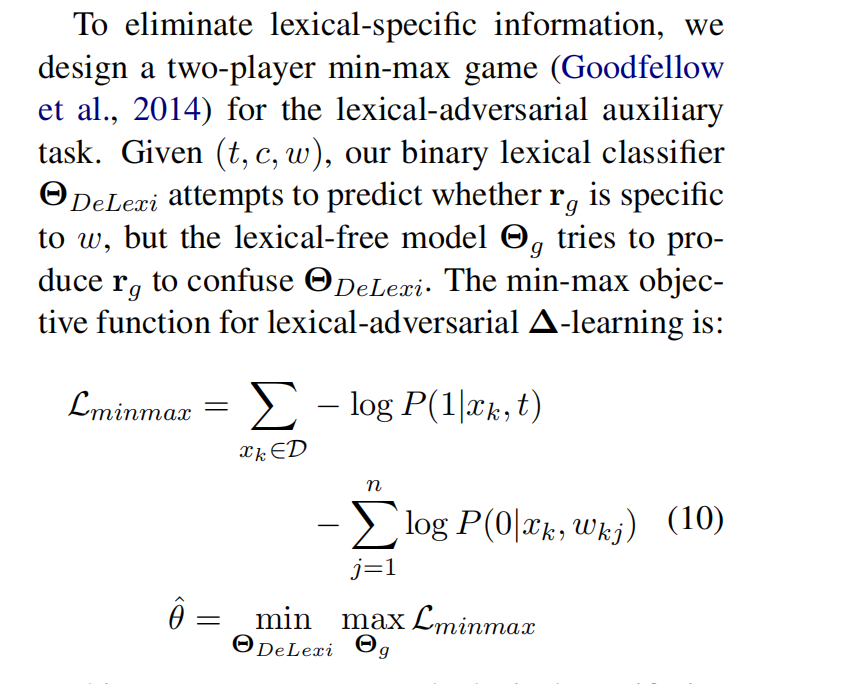
提出自己观点，提出相应问题，现有方法解决不了这个问题，专门方法去解决。

Trigger-Centric与没trigger的融合

感觉该用最大分类熵才能说明该特征不好用，不然二分类问题会在颠倒级性后到最优。特征任务无关的最大值应该在最大熵处取得（或类似。），而非最大交叉熵。

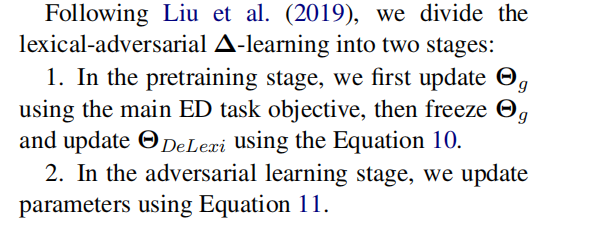
语义对抗学习。

lexical-adversarial auxiliary task



判别网络尽可能判断rg是否是w相关的，但selta g尽可能产生的rg用来困惑分类误差。

最佳的分类在百分之五十取得。



GCN-ED