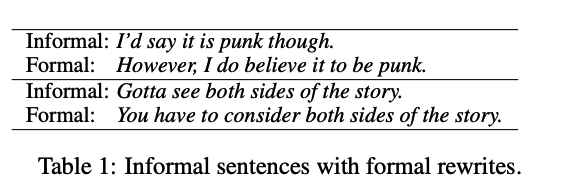
阅读论文《GYAFC Dataset》

论文来源： 2018 NAACL

论文主要内容：

本文主要发布了一个formality的数据集—GYAFC Datatset,下面是一个例子：

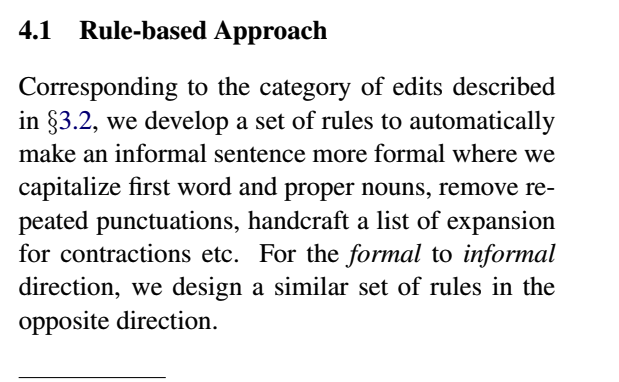


文中还提供了数据集下载地址：https://github.com/raosudha89/ GYAFC-corpus,但是这只是一个说明链接，要想下载还需要给作者发邮件=w=，

这篇论文的relate work可以作文formality的综述论文，可以反追溯前人的工作。

作者采用了几个beanchmark，分别是：

* Rule-based model:作者只是简要介绍了下，没具体将



* Phrased-based machine translation:作者采用的是前人的工作,Paraphrase for style. 2012 Coling
* NMT baseline:是个双向LSTM再加上attention。
* NMT copy：Shakespearizing mdern language using copy-enriched sequence-to-sequence models. 2017 EMNLP
* NMT Combined：这个作者只是讲了个大概，说具体细节在supplementary material中，但是这个附件论文后也没给=w=。

评估方法：文中进行评估分为两种方式，一种是采用人工的方法，一种是自动评估。但是都是评估formality，fluency，meaning preservation，overall Ranking这四个指标：

主要记录下，自动评估这四个指标采用的方法：

* Formality：采用的是论文An empirical analysis of formality in online communication 2016 TACL。
* Fluency：采用的论文是Predicting grammaticality on ordinal scale. 2014 ACL
* Meaning Preservation：作者采用的也是别人的方法，提供了一个github地址：https://github.com/castorini/MP-CNN-Torch
* Overall Ranking：BLEU，PINC指标