

# Comment Abuse Classification with Deep Learning

(2017) Theodora Chu, Kylie Jue, Max Wang

<https://web.stanford.edu/class/cs224n/reports/2762092.pdf>

どんなもの？

Abuse コメントを識別するためにニューラルネットワークを使って自動判別する手法を提案。Stanfordの授業のレポート

議論はある？

文字ベースのCNNはワードベースのCNNに比べてトレーニングに時間がかかる。文字ベースのCNNとLSTMを組み合わせることで、ハイパーパラメータの数を減らしてチューニングしやすくすることがFuture Workとして挙げられている

どうやって有効だと検証した？

先行研究で挙げられているベースラインモデルと比較して、良いF値が出るモデルを構築することで有効性を検証している。  
使用しているデータセットはWikipediaが提供する、編集の議論の際のコメントセット

先行研究と比べて何がすごい？

RNNが自然言語処理では高性能だと思われていたが、文字ベースのCNNを使って既存の結果をoutperform

技術の手法や肝は？

文字ベースのCNNにより既存手法をoutperform。コメントのような文章は通常短く、スラングや省略などが頻繁に使われるため、wordベースのembeddingを取るよりも成果が良かったのではないかと筆者は推察している。

次に読むべき論文は？

Wulczyn, Ellery, Nithum Thain, and Lucas Dixon.  
"Ex Machina: Personal Attacks  
Seen at Scale." arXiv preprint arXiv:1610.08914 (2016).