

# Yahoo!知恵袋からの 対話コーパス抽出による傾聴型対話システム

橋口友哉 山本岳洋 大島裕明 (兵庫県立大学)

目的 ユーザが共感できる文を見つけたい



## 悩みの類似文検索: TF-IDF, BM25, BERT[1]



### BERT訓練データの例

[CLS]で文の関連性を見る

[CLS] 愚痴を聞き [MASK] もらえない。[SEP]/なので、[MASK] が溜まります。[SEP]

[CLS] [MASK] を聞いてもらえない。[SEP]/今日は、[MASK] で憂鬱だ。[SEP]

### BERTの入力例



[1] J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding., NAACL2018.

悩み: 「愚痴を聞いてもらえないので、ストレスが溜まります…」に対する類似文検索の結果

	TF-IDF	BM25	BERT 日本語 pretrained (黒橋研)
1位	でもストレス溜まります。	でも私の愚痴は聞いてもらえないのがちょっと…と思いました。	どこへも出掛けられずストレスがたまっています。
2位	ストレスも溜まります。	でもストレス溜まります	どうしてもストレスがたまってしまうます。
3位	ストレスが溜まります…。	その全面的に私を信じてもらえない言動にストレスが溜まります	ちょっとした音なのにイライラして気分が悪くなってしまい、自分に対しても嫌気がさします。