

Yahoo!知恵袋からの 状況が類似する悩みの検索

兵庫県立大学応用情報科学研究科
大島研究室M2 橋口友哉

YAHOO! 知恵袋
JAPAN

「子育ての悩み」では



毎日子供に振り回され、夕方には
ヘトヘト疲れ果てて気力もないため、
ろくに夕飯作れてません。

質問者

(悩み相談者)



旦那が...



姑が...

ピア（同じ悩みの人） 同士の支えあい

自分だけじゃないんだ…



毎日子供に振り回され、夕方にはヘトヘト疲れ果てて気力もないため、ろくに夕飯作れてません。

私もたまに夕飯も作る気力もなく子供達をお風呂に入れてやれない日もあります。



ピアは重要なサポート [小野2007]

共感できる悩みの発見

入力

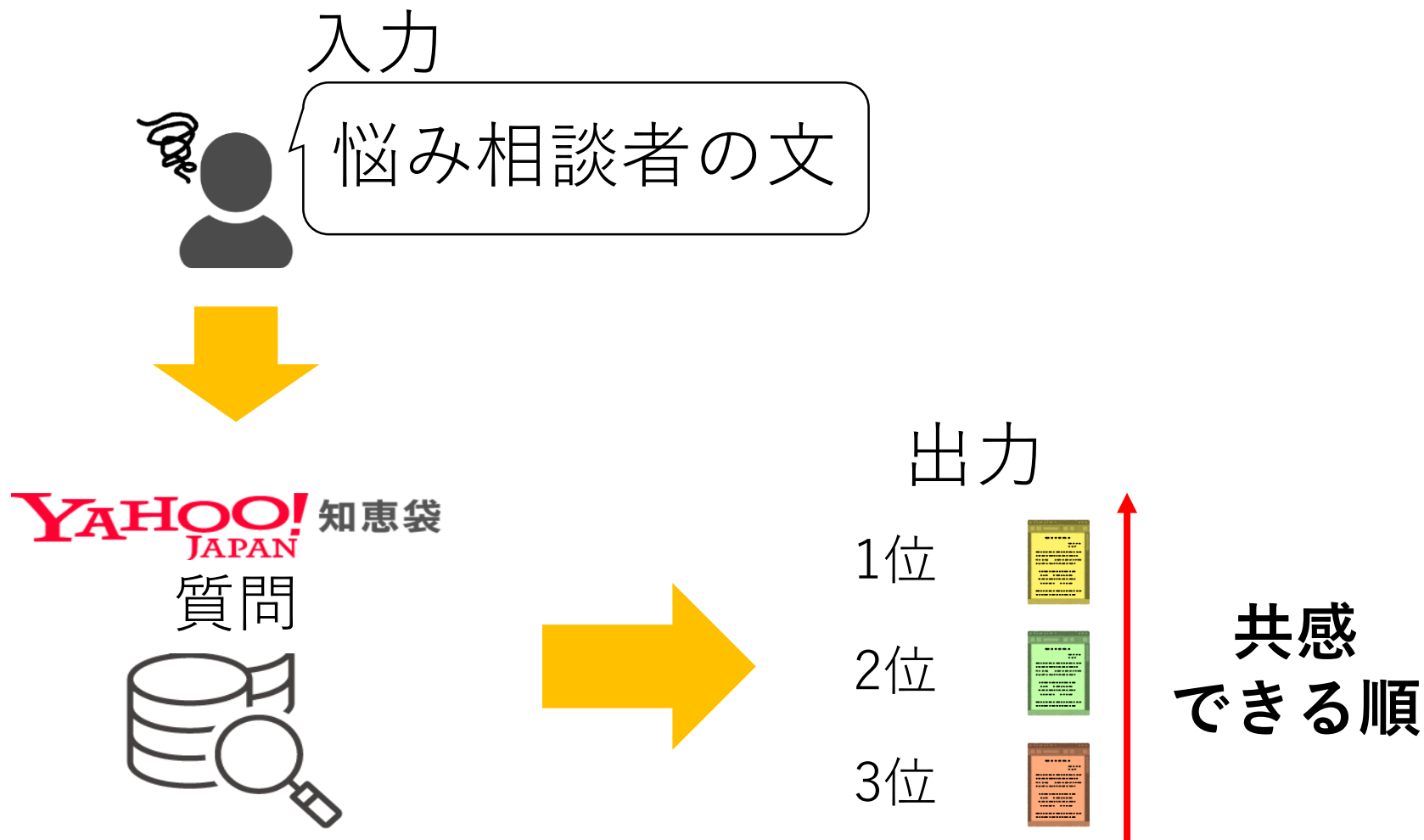
毎日子供に振り回され、夕方にはヘトヘト
疲れ果てて気力もないため、ろくに夕飯
作れてません。

たまに夕飯も作る気力もなく子供達を
お風呂に入れてやれない日もあります。

出力

BERTを使った検索

悩み相談者が共感できる質問をランキング



BERTの入出力



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います。
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 なってしまう、自分に対しても嫌気がさします。

BERTをファインチューニング

- 文章に注目
- 単語に注目

検索の仕方を変更

- 文間の類似度
- 重要な単語でベクトル化

検索評価

- nDCG

悩みの**状況が類似**＝共感できる



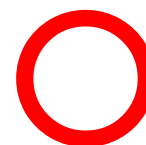
毎日**子供に振り回され**、夕方にはへとへと疲れ果てて気力もないため、ろくに夕飯作れてません。



姑に振り回されたので、夕飯作れてない



子供の相手をして疲れたので、夕飯できてない

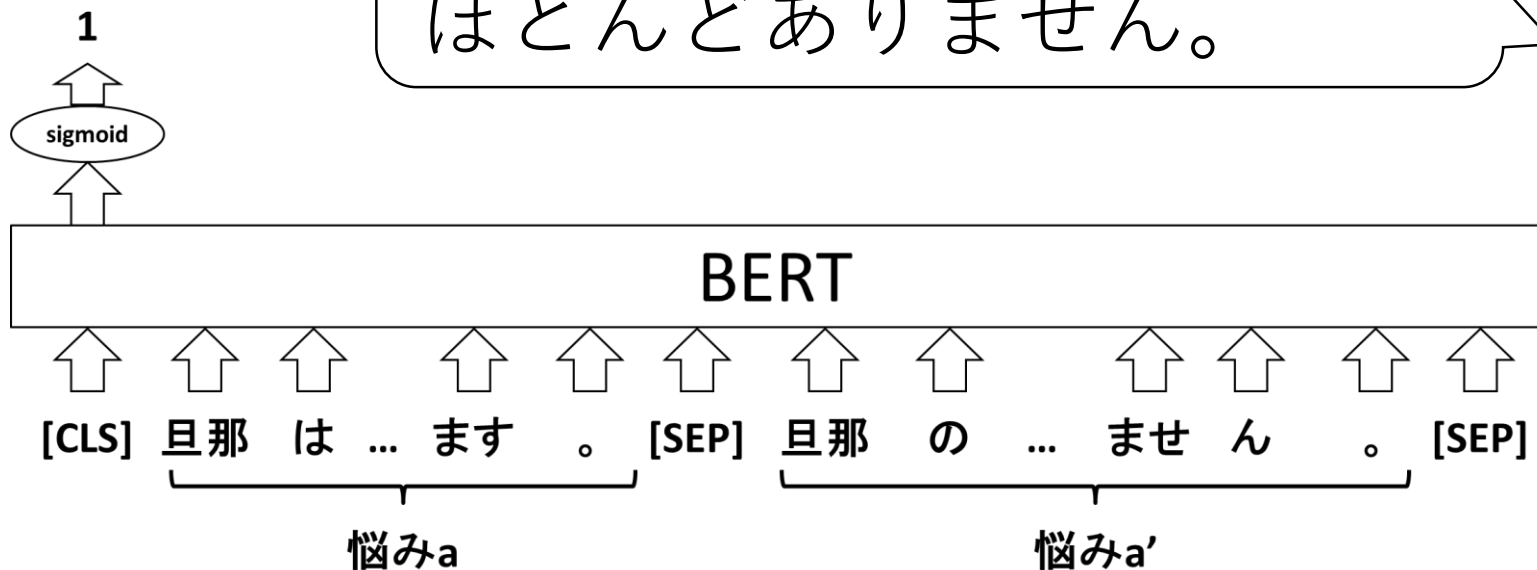


状況類似/非類似のペア



旦那は疲れて帰ってくるので、最近はだれかと会話することなく一日を終えます。

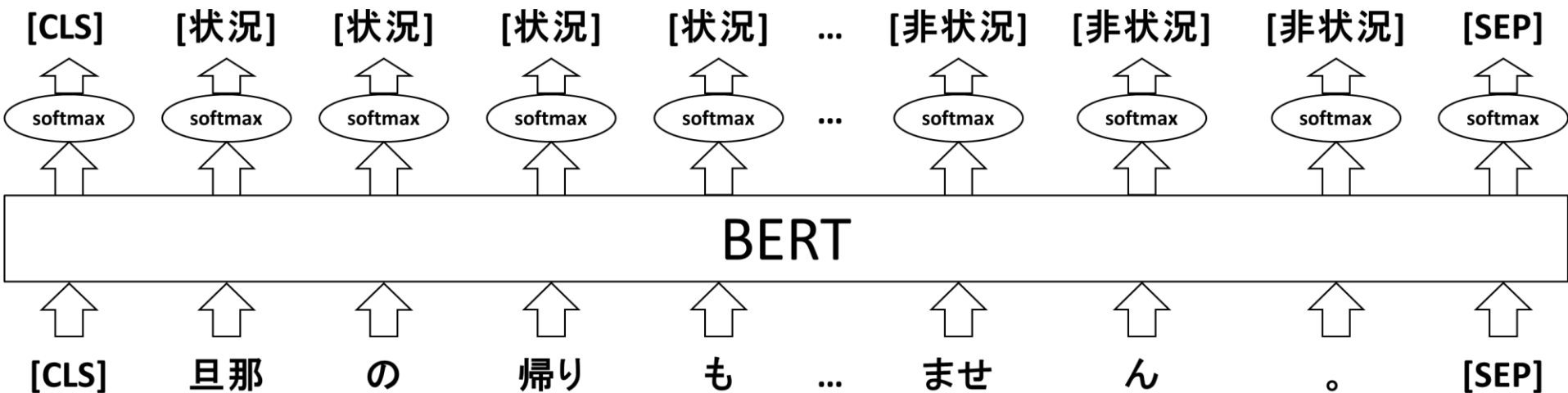
旦那の帰りも遅く、会話はほとんどありません。



旦那の帰りも遅く、会話はほとんどありません。

悩みの状況

悩みの非状況

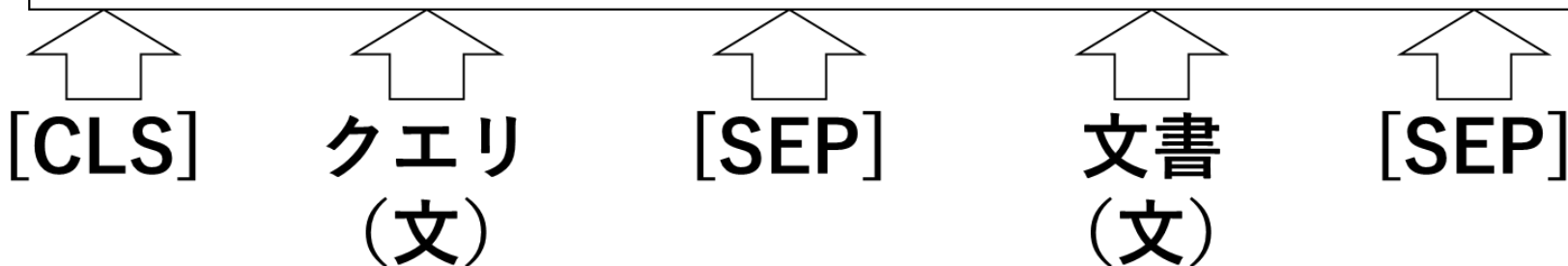


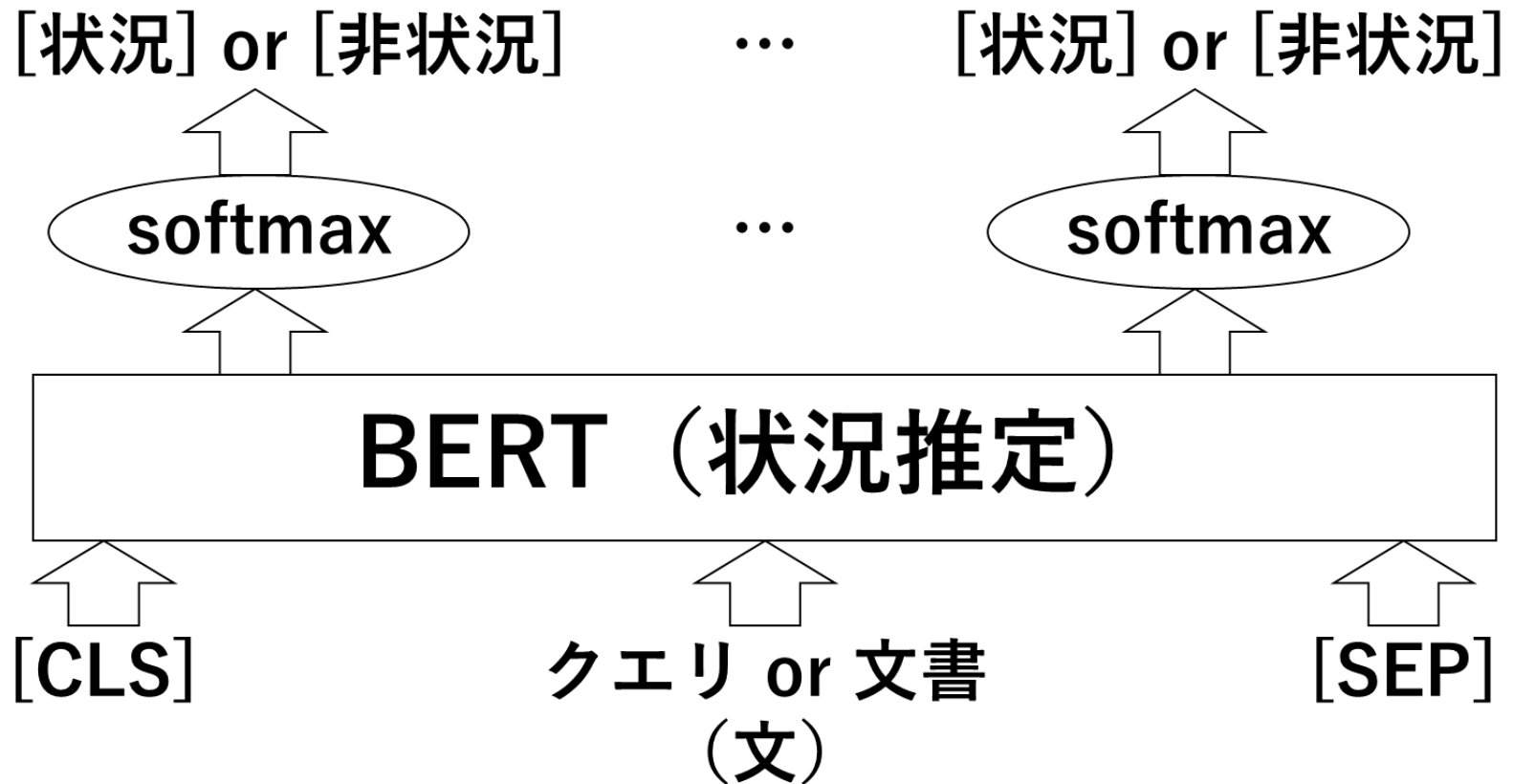
検索対象の文書：**文分割**

類似度 (0~1)



BERT (状況類似文判定)





文ベクトル：**[状況]単語の平均**

評価：クラウドソーシングの選択人数 13

共感できる質問 選択人数 2値適合性 多値適合性



24人



1

2



3人



1

1



2人



1

1



1人



0

0



0

0



0

0



	2値適合性			多値適合性		
	nDCG@1	nDCG@3	nDCG@5	nDCG@1	nDCG@3	nDCG@5
TF-IDF	0.250	0.238	0.253	0.125	0.172	0.210
Okapi BM25	0.150	0.210	0.229	0.100	0.185	0.209
BERT (事前学習のみ)	0.150	0.222	0.273	0.075	0.178	0.236
BERT (状況類似文判定)	0.250	0.262	0.307	0.125	0.204	0.265
BERT (状況推定)	0.300	0.227	0.216	0.200	0.196	0.201

ファインチューニングで精度向上

クエリ



義母はたいへん親切な人なのですが、私に赤ちゃんが産まれてから、おせっかいに拍車がかかって困っています。

普段だったら、用意してもらえてありがたく思う所なのですが・・・私の実家も初孫でベビー用品とかも買いに行くのを楽しみにしていたのに・・・とどめになったのが、両方の両親に赤ちゃんが産まれて落ち着いてから病院に来てほしいとお願いしたのですが主人のお母さんは病院で赤ちゃんが産まれるのを待ちたいようです。

[状況]
[非状況]

1位



まとめ

- 共感できる質問の検索問題
- BERTのファインチューニング、検索の仕方を変更
- 共感できる文が得られた
 - 単語に注目：nDCG@1で0.300
 - 文章に注目：nDCG@5で0.265

今後の課題

- 子育て以外の悩みへの適用
- ファインチューニング用データの半自動化作成

- 特集号：「Webインテリジェンスとインタラクション」採録
- タイトル「CQAコンテンツからの状況が類似する悩みの検索」