
Yahoo!知恵袋 ニューラルネットワークで 検索モデルの精度改善に挑戦

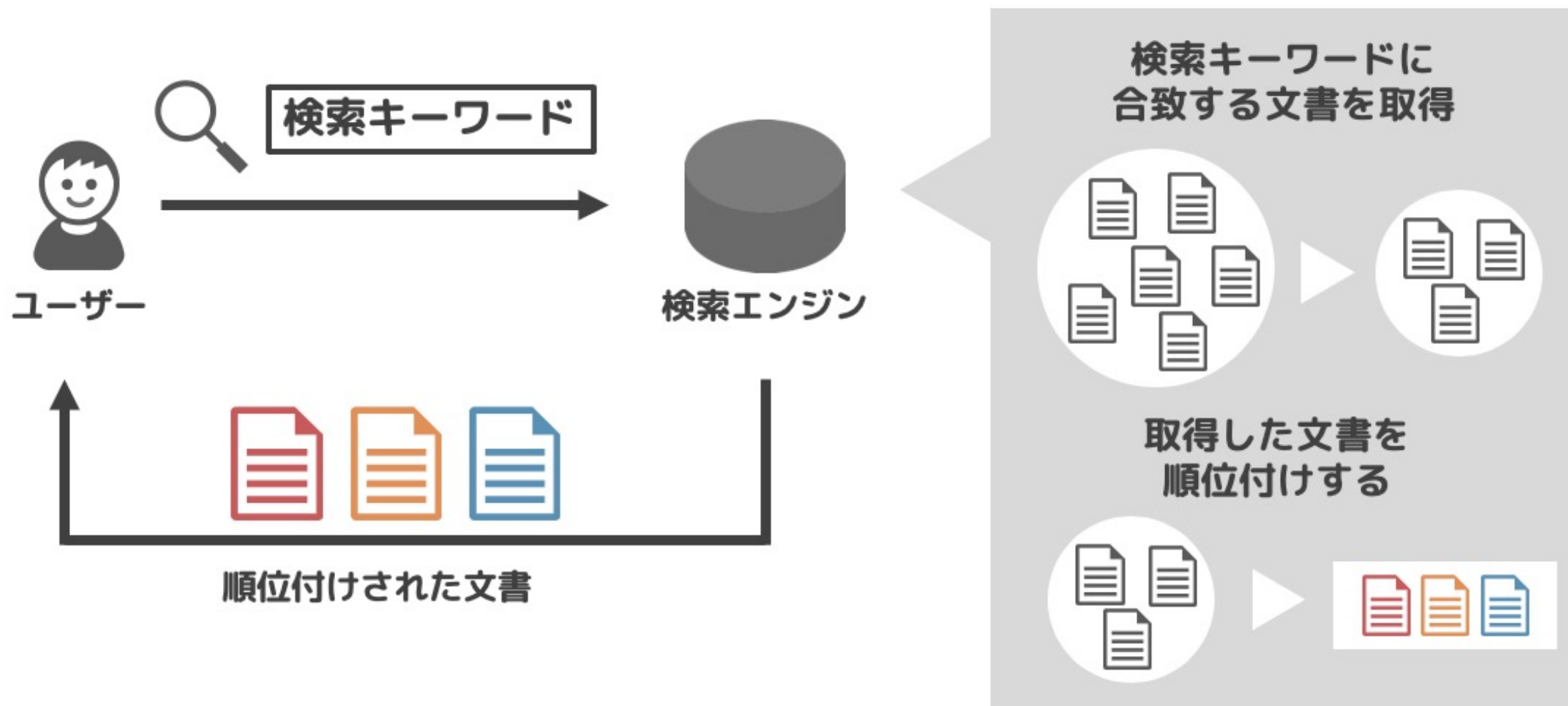
大阪大学大学院情報科学研究科

門谷 宙

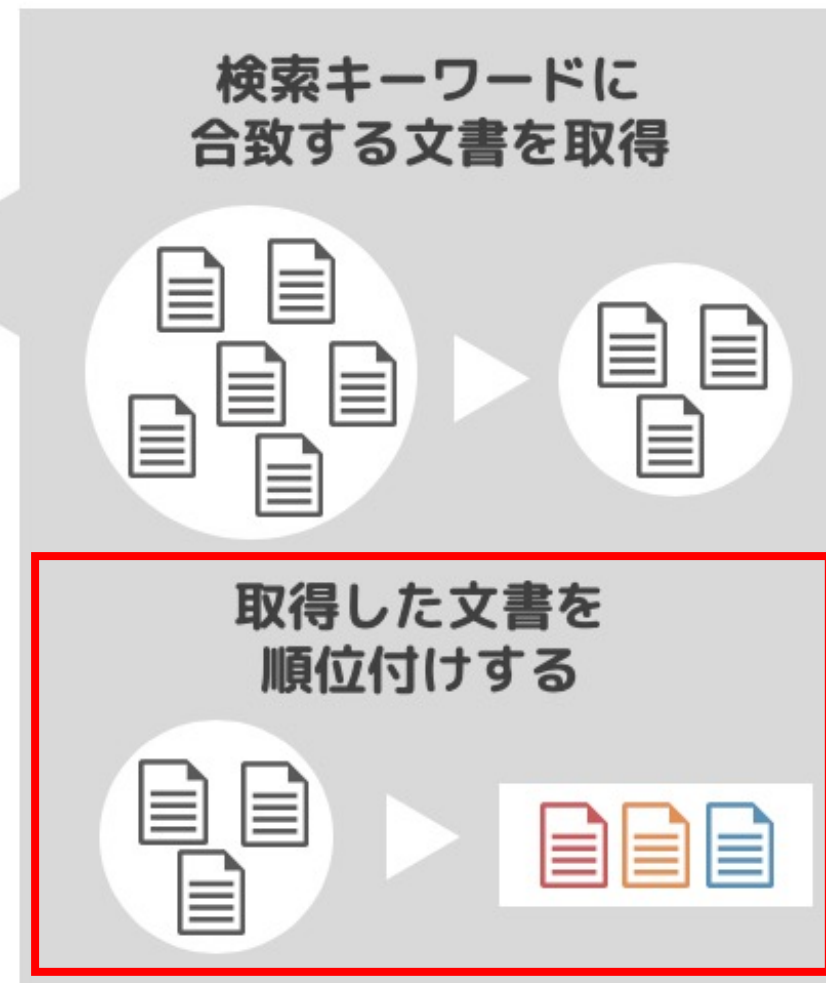
もくじ

1. 背景
2. 手法1: BERTをfine-tuningする手法
3. 手法2: BERTScoreを適用する手法
4. 最終結果
5. 結論

検索モデルの概要



検索モデルの概要



リランキングモデル

ベースラインモデル: LightGBM [1]

- 決定木に基づく機械学習モデル
 - 入力: クエリと文書に関する様々な素性
(e.g. テキストマッチスコア, 文書のクリック実績)
 - 出力: 最終的な関連度スコア
 - テキストマッチスコアの課題
 - 表記揺れや意味が近い単語を考慮できない
 - 単語の順番を考慮できない
- テキスト情報からニューラルネットワークで素性を抽出

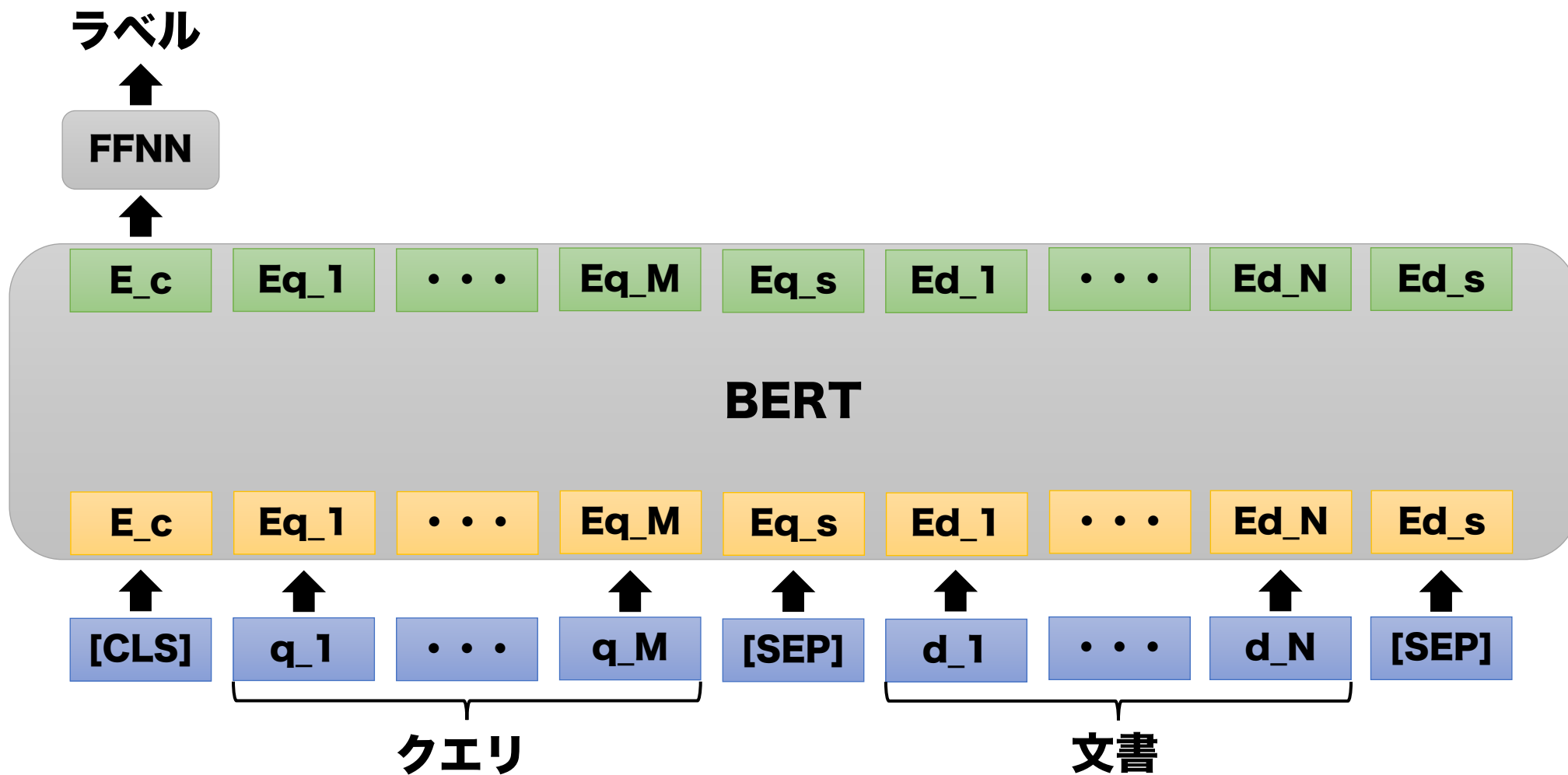
タスク詳細

- 題材: Yahoo!知恵袋
- 利用データ: ユーザーの検索ログ
(クエリ: 検索ワード, 文書: 質問の本文, ラベル: クリックの有無)
- 比較手法: テキストマッチスコア
(TF-IDF, BM25, アドホックテキストマッチスコア)
- 目標:
 - 比較手法を上回る精度の素性を抽出
 - LightGBMの素性に追加することで精度改善
- 評価指標: MRR, nDCG@10

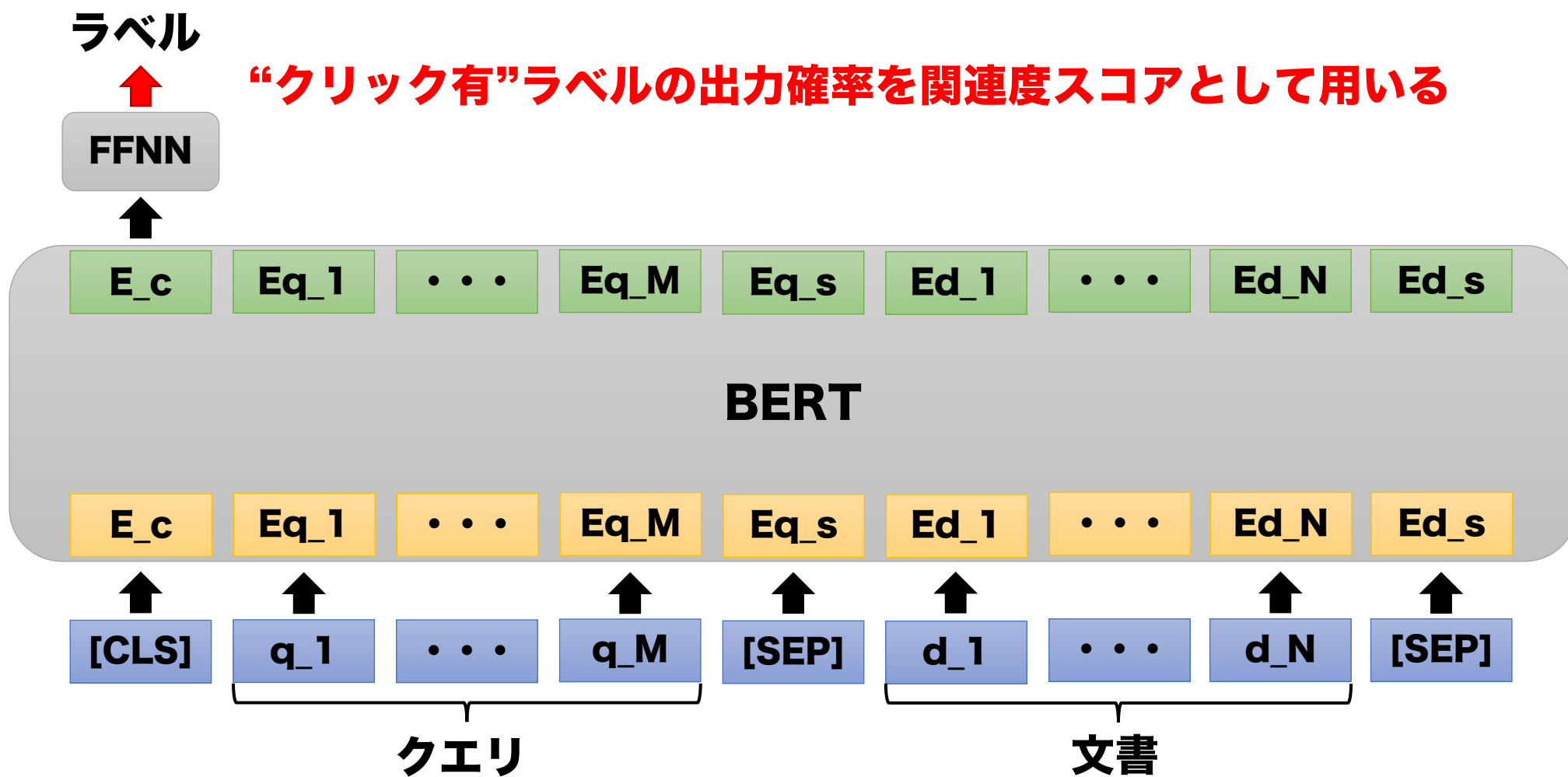
BERT [2]

- 自然言語処理タスクのための汎用的な事前学習済みモデル
- 事前学習後の重みを初期値として, 目的タスクのデータを学習 (fine-tuning)
- 特徴
 - 様々な自然言語処理タスクに応用可能
 - 少しfine-tuningするだけでも高性能なモデルを作成可能

検索モデルのためのfine-tuning



検索モデルのためのfine-tuning



検証実験

実験設定

- 検索ログ1日分を1エポックだけ学習
- 検証手法
 - BERT:
負例の文書を同リクエストの
非クリックにしてfine-tuning
 - BERT (random):
負例の文書を別リクエストから
ランダムサンプリング

実験結果

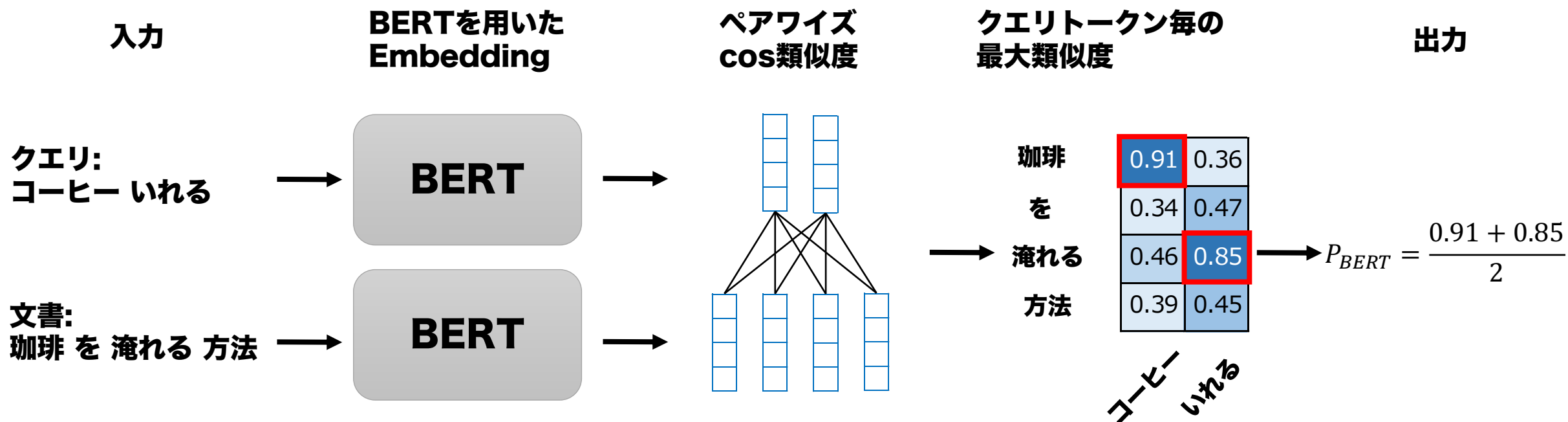
- BERT (random) が既存
- 素性を大きく上回る精度を達成

baselineに対する相対値→	MRR	nDCG@10
TF-IDF	0.917 (-8.31%)	0.951 (-4.89%)
BM25	0.994 (-0.58%)	0.998 (-0.24%)
アドホックテキストマッチスコア	1.000 (±0.00%)	1.000 (±0.00%)
BERT	0.942 (-5.82%)	0.965 (-3.46%)
BERT (random)	1.057 (+5.67%)	1.032 (+3.18%)

BERTScore [3]

- BERTを利用した生成タスクの自動評価指標 (precision, recall, F1がある)
- クエリと文書の関連度スコアとしても効果が見込めそうなアルゴリズム

→ 検索タスクに初めてBERTScoreを使用



検証実験

実験設定

- BERTはfine-tuningしない
- 検証手法
 - BERTScore (precision)
 - BERTScore (recall)
 - BERTScore (F1)

実験結果

- BERTScore (precision) が既存素性を上回る精度を達成
- BERTScore (recall) は, 文書にノイズとなるトークンが含まれるため精度が低い

baselineに対する相対値→	MRR	nDCG@10
TF-IDF	0.917 (-8.31%)	0.951 (-4.89%)
BM25	0.994 (-0.58%)	0.998 (-0.24%)
アドホックテキストマッチスコア	1.000 (±0.00%)	1.000 (±0.00%)
BERTScore (precision)	1.003 (+0.30%)	1.003 (+0.26%)
BERTScore (recall)	0.953 (-4.69%)	0.976 (-2.42%)
BERTScore (F1)	0.988 (-1.16%)	0.996 (-0.40%)

BERTScoreのメリット

- fine-tuningなしでもある程度の精度が出る
 - fine-tuningでさらなる精度改善が期待できる
- 検索エンジンに現実的な計算量で載せられる
- クエリと文書を連結してBERTに入力する手法 (手法1)
 - クエリと文書のペアをBERTに通す
 - ヒットした文書の数だけBERTに通す必要あり
- BERTScoreを適用する手法 (手法2)
 - クエリと文書はそれぞれ別にBERTに通す
 - クエリは1回だけBERTに通せば良く、文書は事前に計算できる

手法と結果

- BERTをfine-tuningする手法 と BERTScoreを適用する手法 を検証
- 両手法とも, 現行素性を上回る精度を達成
- LightGBMの素性に追加することで, 現行モデルの精度を改善

baselineに対する相対値→	MRR	nDCG@10
Baseline	1.0000 ($\pm 0.000\%$)	1.0000 ($\pm 0.000\%$)
+ BERT (random)	1.0050 (+0.505%)	1.0033 (+0.330%)
+ BERTScore (precision)	1.0005 (+0.052%)	1.0004 (+0.035%)
+ BERT (random) + BERTScore (precision)	1.0056 (+0.561%)	1.0036 (+0.358%)

まとめ と 今後の課題

- まとめ
 - BERTをfine-tuningする手法 と BERTScoreを適用する手法 を検証
 - 両手法とも, 既存素性を上回る精度を達成
 - LightGBMの素性に追加することで, ベースラインモデルの精度を改善
- 今後の課題
 - 手法1: BERTをfine-tuningする手法
データ・学習時間の追加やハイパーパラメータ探索でさらなる精度向上
 - 手法2: BERTScoreを適用する手法:
使用するBERTをfine-tuningすることでさらなる精度向上