

AI王 ～クイズAI日本一決定戦～

ICS Lab. (株式会社ベルシステム24ホールディングス)

2021/03/19

ICS Lab. (株式会社ベルシステム24ホールディングス) について

BELSYSTEM24[®]

主な事業：

コンタクトセンターアウトソーシング

事業規模：

- 国内37拠点, 18,000席超の席数
- 30,000人以上のオペレーター
- 年間200,000,000コールを受けるクラウド基盤



日々, “自然言語”で業務を遂行し,
“自然言語”の実践的な課題が生まれる会社

株式会社ベルシステム24ホールディングス



INNOVATION
COMMUNICATION
SCIENCE LAB.

沿革：

- 2018～ Sony CSLとの共同研究開始
- 2020/4 ICS Lab.設立
- 2020/7 Mopas[®], Knowledge Creator[®]提供開始

設立趣意：

現場での運用ノウハウ
x 機械学習・自然言語処理の実務適用
→ 「次世代コンタクトセンターの構築」

ICS Lab.

(イノベーション&コミュニケーションサイエンス研究所)

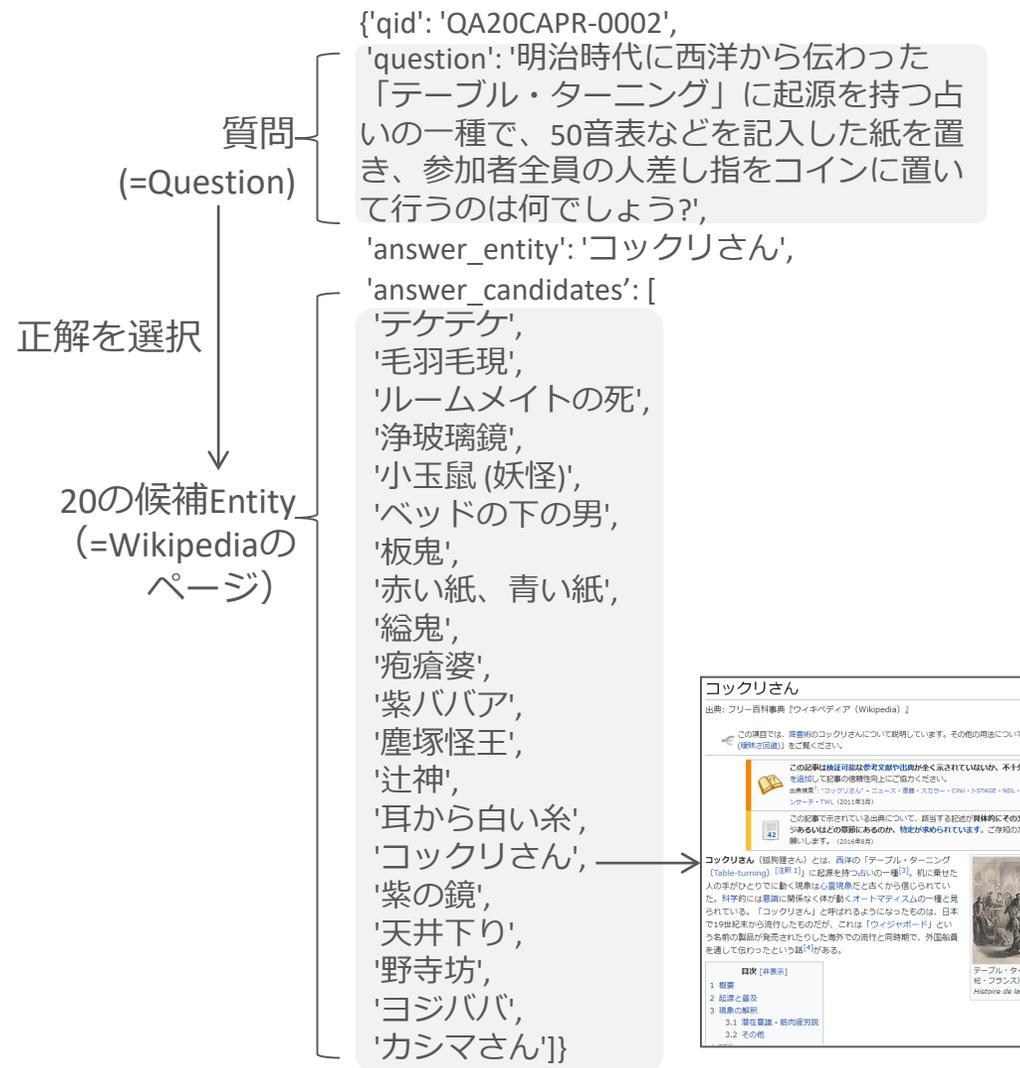
問題設定の概要

• 問題設定

- 質問 (=Question) の正解を,
20の候補Entity (=Wikipediaのページ)
から選ぶ

• その他のルール

- 評価値 = 1,000問の正答率
- 候補はWikipediaの92万Entityから与えられる
- Wikipediaを含め, 一般公開されている,
もしくは公開できるデータのみ利用可能
- 外部リソース (インターネット検索など)
は利用禁止



1つの問題例

処理の流れ, および課題と対策

課題

- Entityの本文だけでは情報が不十分

対策

- Entityに関する, Questionにまつわる情報を集める

Information Gathering

課題

- BERTで扱えるTokenの数が限られる

対策

- Questionに関係しそうな情報だけ選択する

Passage Selection

課題

- BERTを学習する度に結果・精度がぶれる

対策

- 多様な入力を与えるとともに複数のモデルを利用する

Train, Inference

Information Gathering / 課題

• 例-1: Entityのつけ間違え

Q

スマート、エアタッチ、ノンダスト、ライトなどの種類がある、トンボ鉛筆から発売されているロングセラーの消しゴムは何でしょう？

A

MONO

→消しゴムではなく、音楽バンドを指している

• 例-2: Entityに情報がない

Q

1957年の映画『十二人の怒れる男』で、陪審員7番は何という野球チームの試合に行きたいため、議論を早く終わらせようと画策する？

A

ニューヨーク・ヤンキース

→ニューヨーク・ヤンキースのページでは『十二人の怒れる男』には言及されていない

課題

- Entityの本文だけでは情報が不十分

対策

- Entityに関する、Questionにまつわる情報を集める

MONO
出典: フリー百科事典『ウィキペディア (Wikipedia)』

← その他の"MONO"については「モノ」をご覧ください。

この記事は検証可能な参考文献や出典が全く示されていないか、不十分です。出典を追加して記事の信頼性向上にご協力ください。
出典検索: "MONO" - ニュース・書籍・スカラー・CINII・J-STAGE・NDL・dlb.jp・ジャパンサーチ・TWL (2015年7月)

MONOは、1999年結成した日本・東京の4人組ポストロック/インストゥルメンタル・ロック・バンド。
オーケストラとシューゲイザー、ギターノイズを合わせたMONOの楽曲スタイル。

目次 [非表示]

- 1 略歴
- 2 メンバー
 - 2.1 旧メンバー
- 3 受賞
- 4 ディスコグラフィ
 - 4.1 アルバム
 - 4.2 EP・シングル・コンピレーション
 - 4.3 スプリットアルバム
 - 4.4 DVD
 - 4.5 その他のリリース
 - 4.6 サウンドトラック
- 5 オフィシャルビデオ
- 6 ライブ歴
 - 6.1 2018年
 - 6.2 2017年
 - 6.3 2016年

MONO (2016年)

基本情報

出身地 ● 日本

ジャンル インストゥルメンタル、アンビエント、オルタナティブ・ロック、ロック、ポストロック、現代音楽

活動期間 1999年 -

レーベル テンポラリー・レジデンス・リミテッド
Pelagic Records
Human Highway Records
ツァディク

公式サイト www.monooofjapan.com

メンバー Takaakira 'Taka' Goto

Information Gathering / 対策

課題

- Entityの本文だけでは情報が不十分

対策

- Entityに関する、Questionにまつわる情報を集める

Q	1957年の映画『十二人の怒れる男』で、陪審員7番は何という野球チームの試合に行きたいため、議論を早く終わらせようと画策する?
A	ニューヨーク・ヤンキース

MONO

出典: フリー百科事典『ウィキペディア (Wikipedia)』

← その他の"MONO"については「モノ」をご覧ください。

この記事は検証可能な参考文献や出典が全く示されていないか、不正確な内容を追加して記事の信頼性向上にご協力ください。
出典検索?: "MONO" - ニュース・書籍・スカラー・CINiiサーチ・TWL (2015年7月)

結成した日本・東京の4人組ポストロック/インディ・ロック・バンド

対策-1
曖昧さ回避
なども見る

対策-2
Questionで検索して、回答候補
EntityにリンクがあるEntityも見る

十二人の怒れる男

出典: フリー百科事典『ウィキペディア (Wikipedia)』

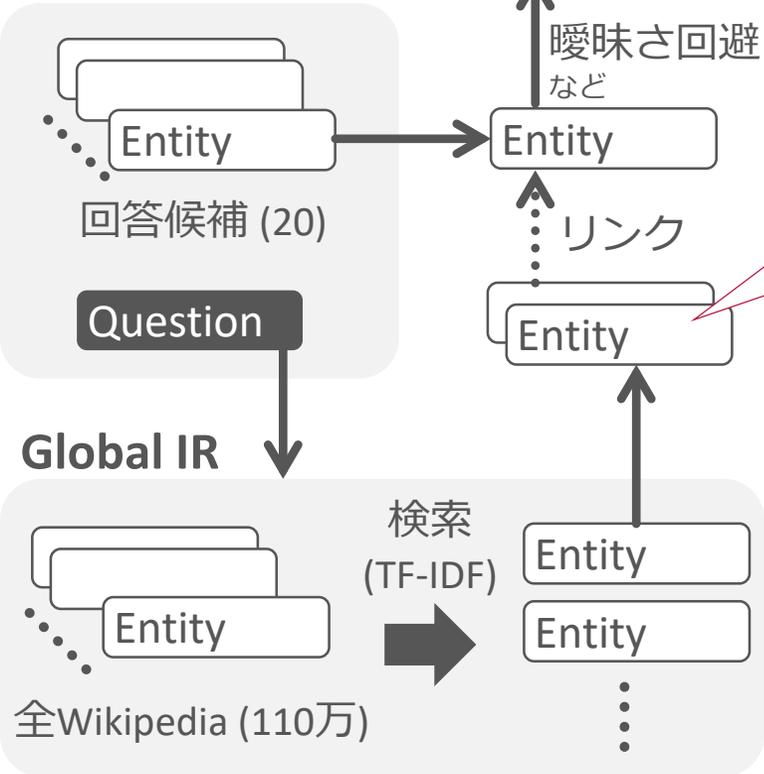
『十二人の怒れる男』（じゅうににんのいかれるおとこ、12 Angry Men）は、1954年製作のアメリカのテレビドラマ。またそのリメイクである1957年製作のアメリカ映画。これ

陪審員6番
塗装工の労働者。義理、人情に篤い。

陪審員7番
食品会社のセールスマン。裁判にまったく興味がない。ヤンキースの試合を観戦予定で時間ばかり気にしているが、夕立で試合が流れたため面倒くさくなる。

陪審員9番

入力



Passage Selection / 課題と対策

課題

Wikipediaの本文
(この例では3,000文字程度)

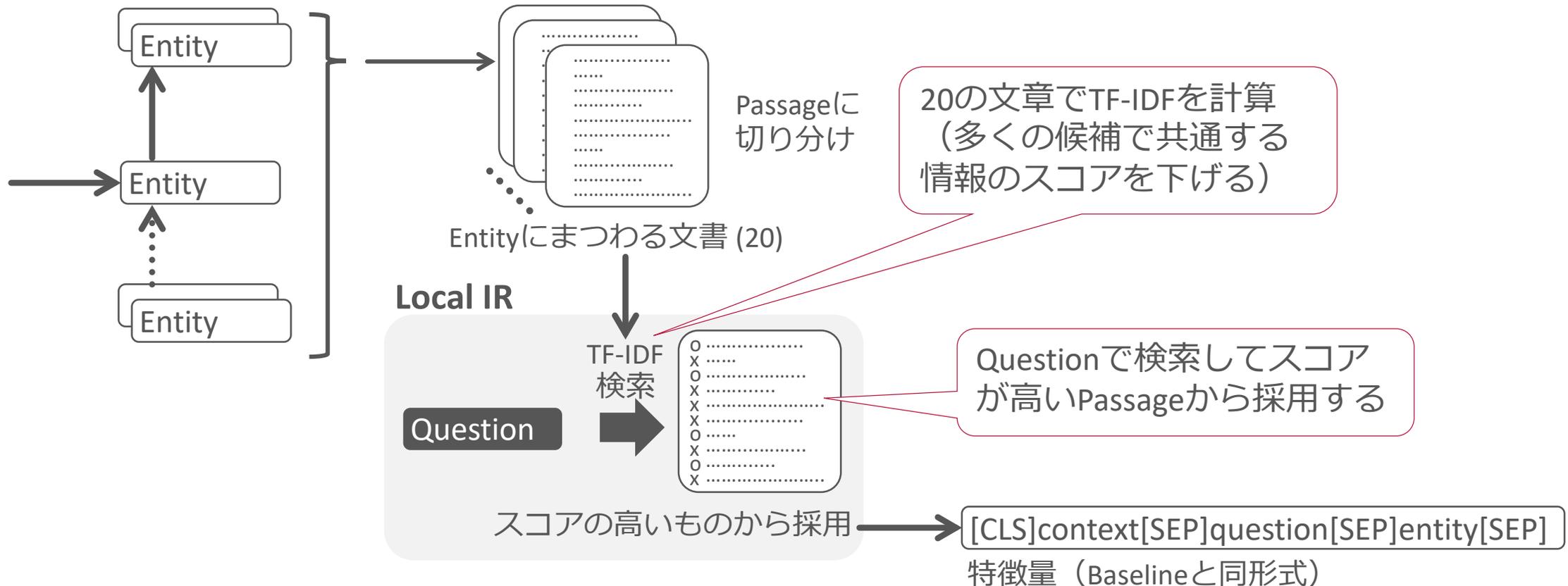
700文字程度

課題

- BERTで扱えるTokenの数が限られる

対策

- Questionに関係しそうな情報だけ選択する



Train, Inference / 課題と対策

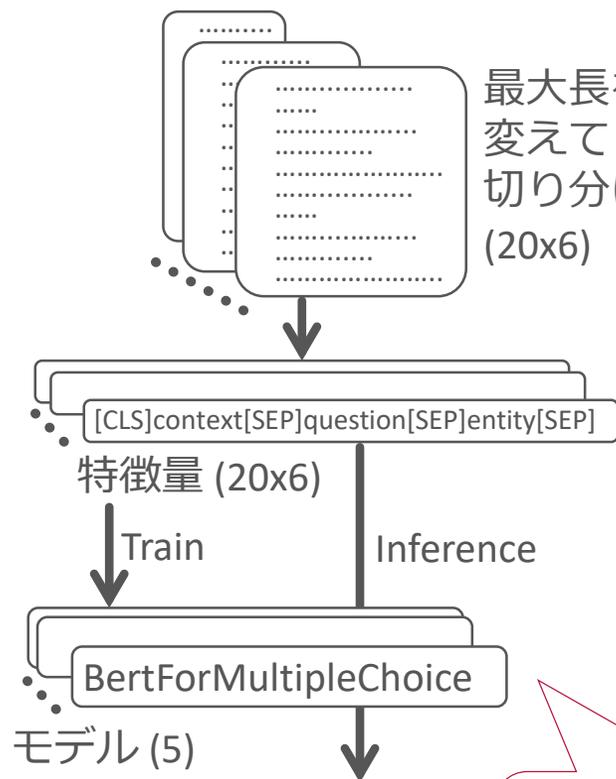
課題

- 同じ特徴量を学習させても，評価値が0.01近くぶれる

対策-1

Passage selection する前の切り方を多様にする

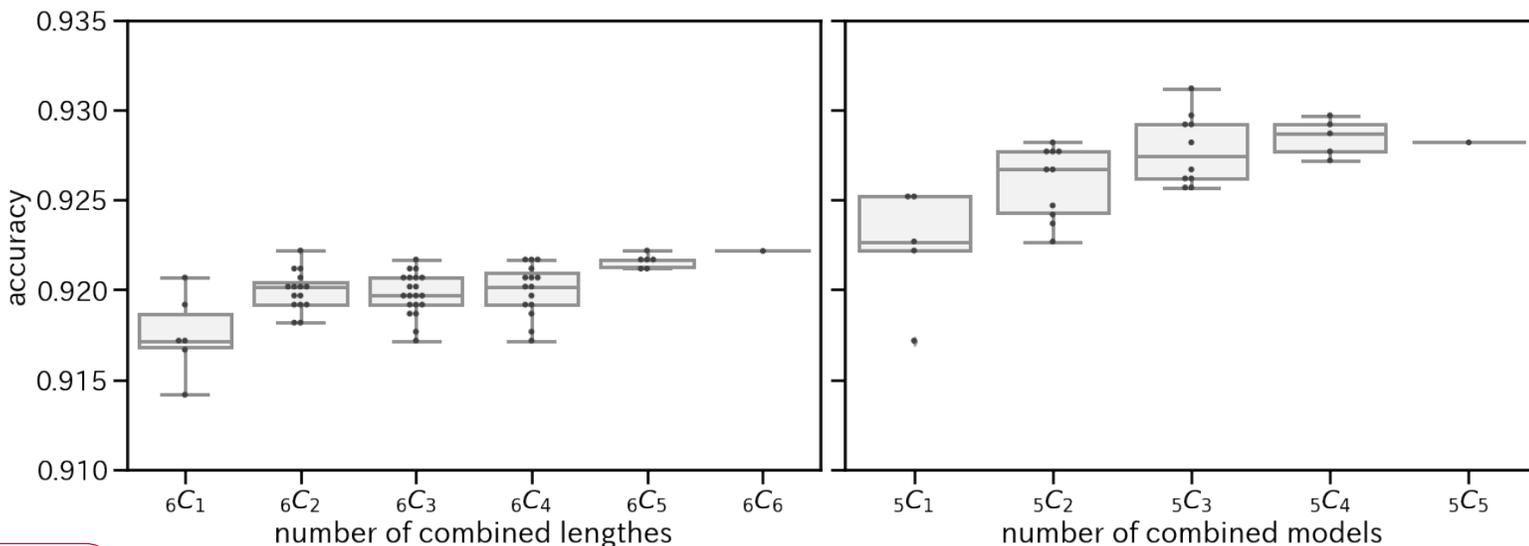
最大長を変えて切り分け (20x6)



6x5のスコアの平均
→20候補でソート
→ルール処理

対策-2

同じ条件 (特徴量) で複数モデルを作る



長さを複数利用する効果

6 lengthsで1 modelを学習. 推論時に利用するlengthの組みを変えながら, dev1+2で評価.

モデルを複数利用する効果

6 lengthsで5 modelsを学習と推論. 推論時に利用するmodelの組みを変えながら, dev1+2で評価.

課題

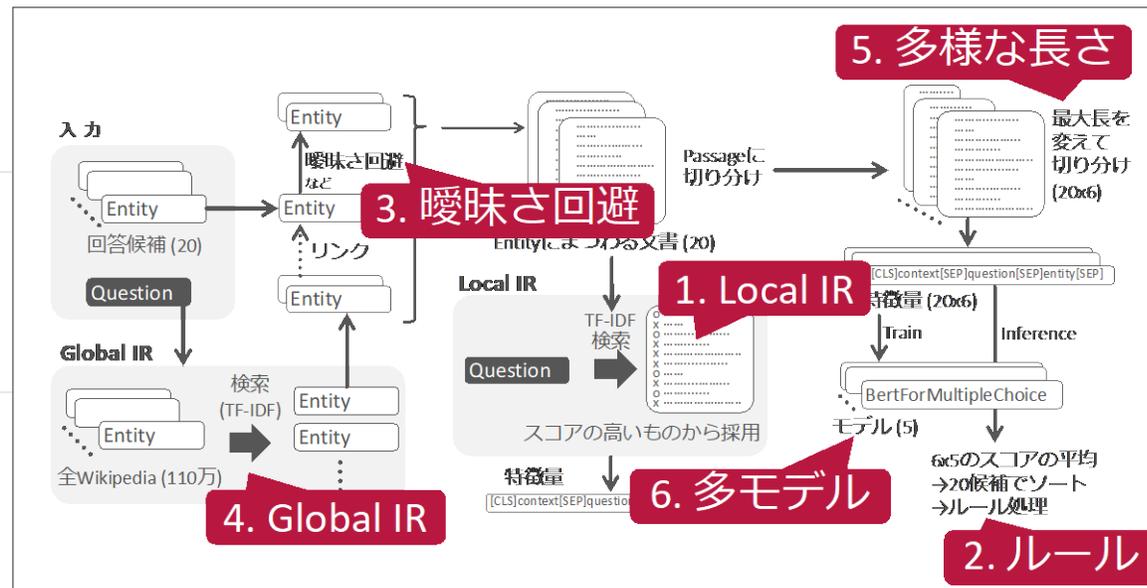
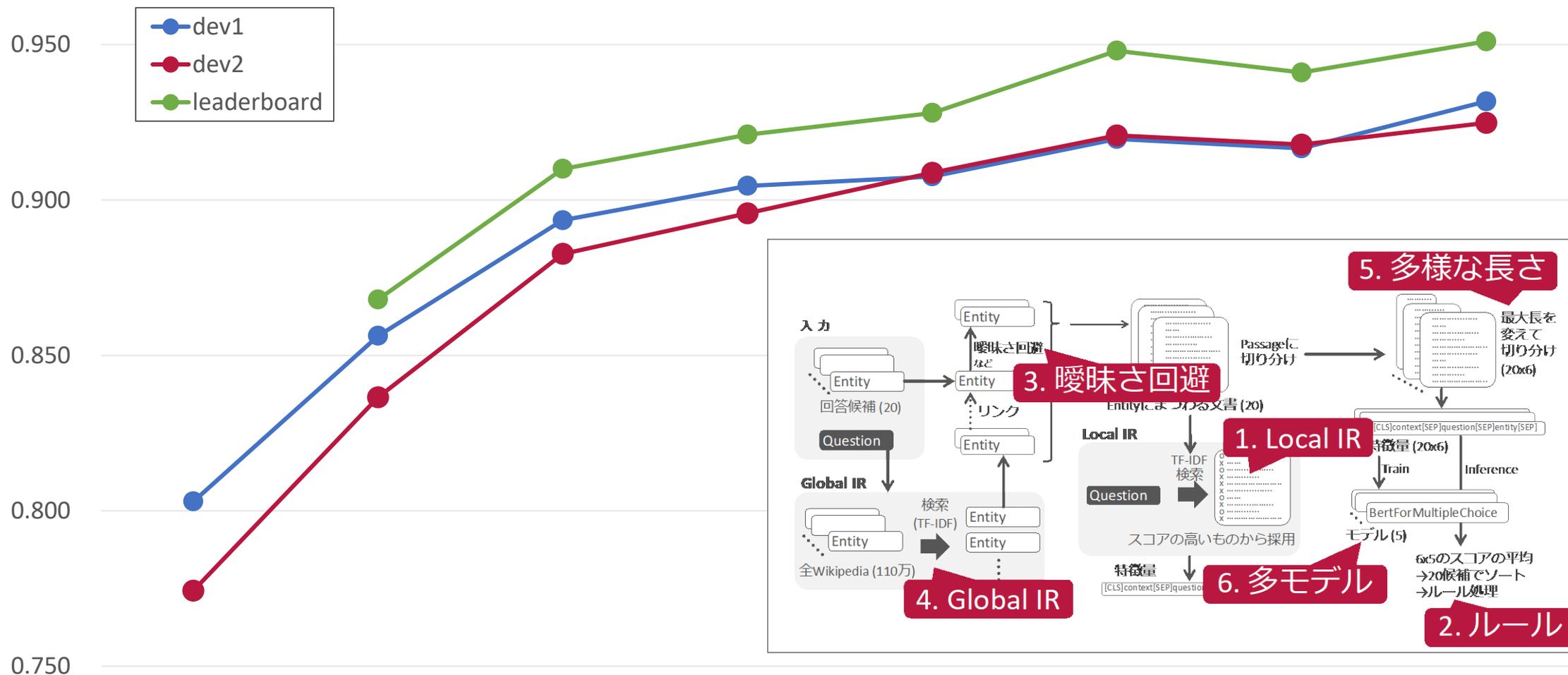
- BERTを学習する度に結果・精度がぶれる

対策

- 多様な入力を与えるとともに複数のモデルを利用する

(ご参考) 各施策とスコアの推移

[*] 各評価では、ここで言及しなかった施策・調整も含む



ベースライン

20候補で学習

1. Local IR

2. ルール

3. 曖昧さ回避

4. Global IR

5. 多様な長さ

6. 多モデル

