SecHack365

プロジェクトの更新があればここに追記予定

ERUSTA

Extension that Recommends Useful and Secure Technical Articles

学習駆動コース坂井ゼミ

竹内 悠人

～有用な技術記事を推薦する拡張機能～

ERUSTAとは？

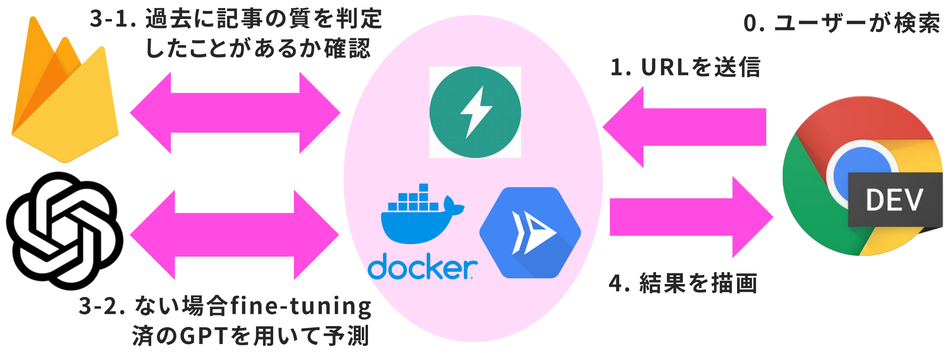
有用性の低い記事たち

- アフィリエイト目的
- タイトル詐欺
- 内容の薄い記事
- 危険なドメイン

削除

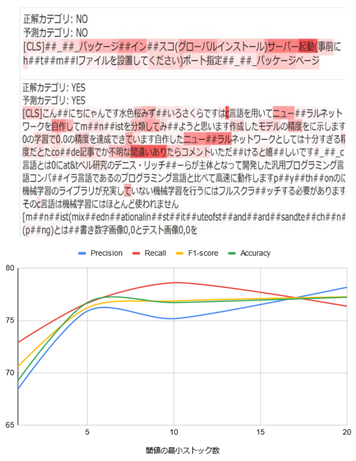
- 質の高い記事のみを表示
- (将来的には) 個々や団体の技術力に合わせた推薦

～システム構成～



～SecHack365での一年～

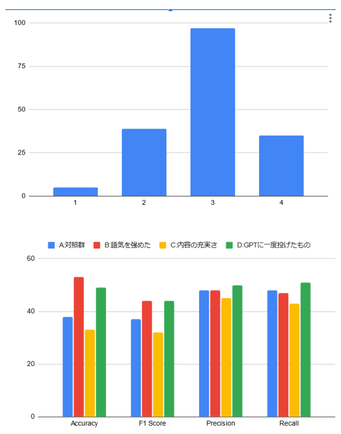
実験1: 記事の良し悪しって機械にわかるの？



- 初期のデータセットとして、Qiitaのストック(お気に入り)数を利用
- 2万件の記事を2種のラベルに分類して、データセットを作成
- このデータセットで日本語BERTモデルを学習しAttentionを可視化
- 最終スコアは、最小ストック数の閾値を20にした時が一番高かった(77.3%)
- Attentionの可視化により、記事の有用性を特徴づける文章自体の抽出は難しいことがわかった

⇒精度に限界！ 有用さはストック数だけでは判断できないが、記事の良し悪しはいい感じに学習できそう

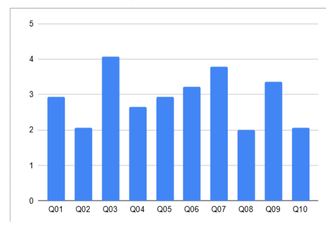
実験2: 良し悪しの判断を人間に近づけられるか



- 実験にエンジニア15人が参加
- ランダムに生成した20個の検索ワードを用意
- 検索ワードでの結果上位10件の記事を収集
- 各記事ごとに有用度を3人のエンジニアが2段階で評価
- 左の4つの条件でGPT-3.5 Turboを学習
- 結果どれも性能が伸び悩んだ(50%)

⇒二値分類に帰着したのに精度がおしまい 厳しい

実験3: 拡張機能のプロトタイプのユーザビリティ調査



Q01 このシステムをもっと頻繁に使いたい
Q02 このシステムは必要以上に複雑だ
Q03 システムがシンプルで使いやすい
Q04 このシステムを使うには、技術的なサポートが必要だ
Q05 システムがスムーズに機能し、うまく統合されていると感じる
Q06 システムに矛盾がとても多いと感じた
Q07 このシステムは、ほとんどの人がすぐに習得できると思う
Q08 このシステムは手間がかかると思う
Q09 このシステムを自信をもって操作できた
Q10 このシステムを使いこなすには事前にたくさんの知識が必要だと思う

- 拡張機能のプロトタイプを作成した
- 実験に15人が参加
- ユーザビリティをSUSを用いて評価した
- システムのシプルさや操作性で高評価を得た
- Q1, Q5, Q6, Q9が低かった一番の原因は、品質の推論結果が表示されるまでに時間がかかってしまったからと考えられる

⇒推論に時間がかかってしまうことが問題

反省① 推論に時間がかかってしまっている

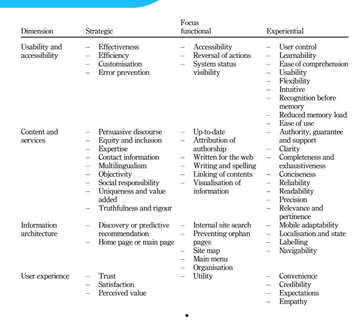
⇒推論を並列化 + 過去に一度推論した記事はDBに保存

反省②作成したデータセットのサイズが小さい

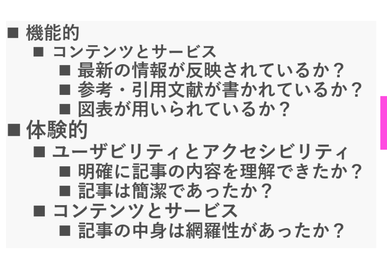
⇒検索キーワードを大量に用意(客観性保持) + 評価軸を増やす

“有用性”を再考する

- 有用性を特徴づける要素は多種多様であり、どれを評価軸として用いることが適切なのかわからない
  ⇒先行研究をあたりたい
- 先行研究:記事の有用性の特性をおおまかに、「戦略的」「機能的」「体験的」に分けた時、記事の有用性を特徴づけるのに良く使われるパラメーターは120個存在する(右図)[1]
  ⇒この中から技術記事での有用性に使えそうなものを選定
- 加えて技術記事に特化させる目的でパラメーターを追加した

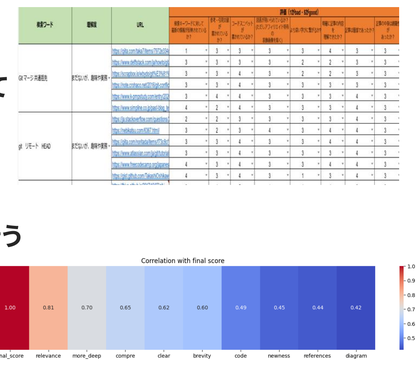


Dimension	Strategic	Focus functional	Experiential
Usability and accessibility	– Effectiveness – Efficiency – Customisation – Error prevention	– Accessibility – Reversal of actions – System status visibility	– User control – Learnability – Ease of comprehension – Usability – Flexibility – Intuitive – Recognition before memory – Reduced memory load – Ease of use
Content and services	– Persuasive discourse – Equity and inclusion – Expertise – Contact information – Multilingualism – Objectivity – Social responsibility – Uniqueness and value added – Truthfulness and rigour	– Up-to-date – Attribution of authorship – Written for the web – Writing and spelling – Linking of contents – Visualisation of information	– Authority, guarantee and support – Clarity – Completeness and exhaustiveness – Conciseness – Reliability – Readability – Precision – Relevance and pertinence
Information architecture	– Discovery or predictive recommendation – Home page or main page	– Internal site search – Preventing orphan pages – Site map – Main menu – Organisation	– Mobile adaptability – Localisation and state – Labelling – Navigability
User experience	– Trust – Satisfaction – Perceived value	– Utility	– Convenience – Credibility – Expectations – Empathy



検証

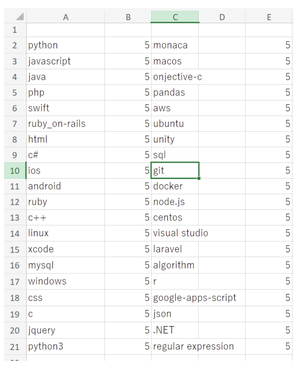
- 情報系の大学生4人+エンジニア17人がこの1200件に記事を先ほど紹介した9個の有用性を特徴づける要素について5段階での点数付けを行った
- 検索キーワード、記事の本文、検索結果の順位のみを用いてGPT3.5のFine-tuningを行うと3段階、5段階評価でそれぞれF1scoreはそれぞれ38%, 19%
  ⇒そのままでは有用な記事の推薦は困難
- 有用性を特徴づける9個の要素を追加して同様の操作を行うと、F1scoreはそれぞれ74%, 52%
  ⇒高確率で良質な記事を推薦することが可能！
- また最終的なスコアとの相関が高い要素を調べたところ関連性、より深い理解につながるか、網羅性があったかなどが高い相関をもっていることが確かめられた



クエリ作り直し

- タイトルからクエリを作りやすい
  &質問にいいねを付ける文化があるため質問の質を評価できることから日本語Stack Overflowを採用
- 右に示す人気タグ40個から
  上位5件の質問のタイトルを検索クエリ化
- 先ほど作成した検索クエリをGoogle検索に投げ、上位3件+18,19,20番目の記事を取得(質の悪い記事も集めるため)

⇒40テーマ×5クエリ×6記事
＝1200記事を取得



	A	B	C	D	E
1					
2	python	5	monaca		5
3	javascript	5	macos		5
4	java	5	onjective-c		5
5	php	5	pandas		5
6	swift	5	aws		5
7	ruby_on-rails	5	ubuntu		5
8	html	5	unity		5
9	c#	5	sql		5
10	ios	5	git		5
11	android	5	docker		5
12	ruby	5	node.js		5
13	c++	5	centos		5
14	linux	5	visual studio		5
15	xcode	5	laravel		5
16	mysql	5	algorithm		5
17	windows	5	r		5
18	css	5	google-apps-script		5
19	c	5	json		5
20	jquery	5	.NET		5
21	python3	5	regular expression		5

まとめ・今後の展望

- 有用な技術記事を推薦するブラウザ拡張機能を開発し、その上で様々な実験や検証・考察を行った
- 本拡張機能により、検索時間の短縮化及び危険なサイトへのアクセスを未然に防ぐことができる
- またSEOによって埋もれていた良質な記事を発掘できる
- あるグループやコミュニティが見るような記事の推薦
  ⇒その分野の初級者への知の高速道路となる
- 個々へのFine-tuning機能
  ⇒その人の理解度に応じた記事の推薦を

外部での発表(IPSJ2024)

RAGを用いた有用性の高い技術記事を推薦するモデルとデータセットの開発
竹内悠人・菅原朔

引用・謝辞

[1]Alejandro Morales-Vargas-2023
Website quality evaluation:a model for developing comprehensive assessment instruments based on key quality factors
Table 6.Most frequently mentioned website quality parameters,organized by focus

実験1は国立研究開発法人科学技術振興機構グローバルサイエンスキャンパス(GSC)「情報科学の達人」育成官民協働プログラムの支援のもと実施しました.また実験費を支援してくださった株式会社ラックサイバー・グリッド・ジャパン　ITスーパーエンジニア・サポートプログラム「すごラボ」並びに様々な助言をくださったSecHack365のトレーナーやアシスタントの方々に多大なる感謝を申し上げます

NICT National Institute of Information and Communications Technology