

自然言語処理を利用した類似障害情報の抽出と活用方法の提案

2021年2月26日（金）

研究コース5 人工知能とソフトウェア品質 SKUチーム

研究員：上田 良太（日本ユニシス株式会社）
栗原 崇至（エプソンアヴァシス株式会社）
杉本 智（NTTコミュニケーションズ株式会社）
主 査：石川 冬樹（国立情報学研究所）
副主査：栗田 太郎（ソニー株式会社）
徳本 晋（株式会社富士通研究所）

目次

- 1. 研究概要**
- 2. 課題設定**
- 3. 課題解決に向けたアプローチ**
- 4. 評価**
- 5. 考察**
- 6. まとめ**

1. 研究概要

1. 研究概要

- AI技術の発展に伴い、日常業務におけるAI技術活用を検討
- AI知識、開発スキルが乏しくても、容易にAIを活用する方法の検証
- 自然言語で書かれた文書の活用
- AI技術、中でも自然言語処理を手軽に利用し活用する方法の提案



「既存AIツールを活用した
品質向上への可能性」



2. 課題設定

2. 課題設定

- 蓄積する過去の障害事例（障害管理票）の有効活用
- 一般に公開・普及している自然言語処理技術およびサービス（特に日本語処理）を利用することで、扱いの難しい非定型文章を活用



2. 課題設定

- 日常業務で作成された文書は、プロジェクトまたはプロダクトに依存し書き方も多種多様な状況 . . .
- AI利用や導入は専門的な知識がなければ難しい . . .



3. 課題解決に向けたアプローチ

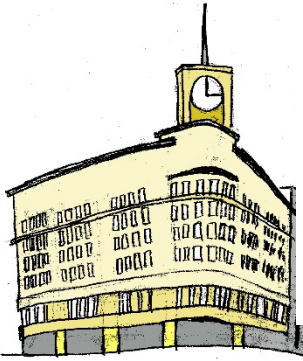
3. 課題解決に向けたアプローチ

~~銀座~~

やさしい

~~バイパス~~

~~解決方法を
ググる！~~



違います！

oogle

3. 課題解決に向けたアプローチ

NLP (自然言語処理)



+

やさしい・簡単

Python
プログラム



```
# 【解説】
# GiNZAによるトークナイズ（形態素解析）と併せて
# 同時に文章の長さに依存しないベクトルを生成する。
# GiNZAの次元数はデフォルトで300次元 (V4.0.0)。
# <引数>
# text: 文章テキスト
# <戻り値>
# GiNZAによる300次元 (V4.0.0) の文章ベクトル
# =====
def compute_vector(text):
    doc1 = nlp(text)
    tokens = []
    print("変換中==> ", doc1)
    for sent in doc1.sents:
        for token in sent:
            tokens.append(token.orth_)
    return doc1.vector # 文章の長さ(単語数(Token数)分)に依存しない 1 x 300次元のベクトルを生成
```

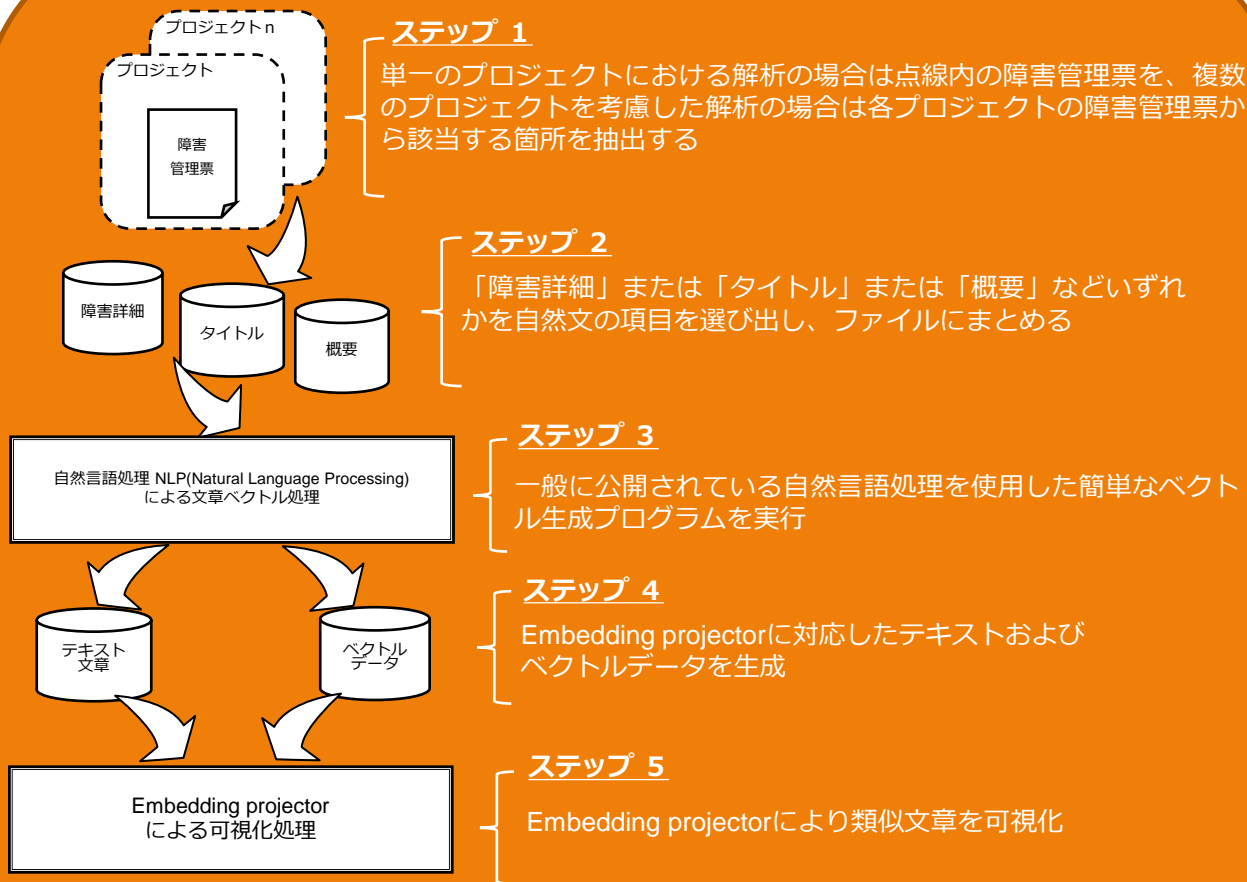
+

Google

**Embedding
Projector**



3. 課題解決に向けたアプローチ



4. 評価

4. 評価

■ 実験の目的

- 被験者が目視で似ていると判定した事例とAIツールで抽出した上位10事例との一致率を比較
- 被験者に対するアンケートにより、目視で抽出に要した時間、および、AIツールが算出した結果に対する納得度を回答してもらい、有効性を検証

■ 実験の概要



4. 評価

■ 評価のポイント

- AIツールで抽出した上位10事例と被験者が目視で似ていると判定した事例の一致率のほかに、以下の観点でも評価する
 - ① AIツールで算出した類似度に閾値等で機械的な判断が可能か
 - ② 記述フォーマット（記述内容、文字数等）による一致率への影響
 - ③ 専門用語の有無による一致率への影響

4. 評価

■ 実験に使用したデータ

- 4つのプロジェクトで使用した障害管理票から、自然言語で記載されているタイトルや障害内容等の部分を結合し、障害データを作成
- 文章内で障害に直接関係ない以下のような記述は削除している
 - 証跡や障害管理システムへの参照情報
 - デバグトレースなどのシステム固有の情報

プロジェクト	障害管理票の特性	障害データ構成	特徴	比較元データの抽出条件
A	システム開発PJの結合試験の障害管理票	「タイトル」 「障害内容」	・ 平均130文字	被験者(該当PJ未参加)にて無作為に5件ピックアップ
B	アプリケーション製品検査の障害レポート	「タイトル」のみ	・ 6～94文字 ・ 専門用語有	被験者(該当PJ未参加)にて実運用で課題があると意見があった5件をピックアップ
C	システム開発PJのユーザー試験の障害管理票	「タイトル」 「障害内容」 「対処方法」	・ 平均213文字 ・ 記述フォーマット統一	被験者(該当PJ参加)にて実際にPJ参加期間中に調査した障害3件をピックアップ
D	システム開発PJの運用開始後にユーザーから申告された障害管理票	「タイトル」 「障害内容」 「対処方法」	・ 平均195文字 ・ 専門用語有	被験者(該当PJ参加)にて実際に対応を行った障害3件をピックアップ

4. 評価

■ 実験結果

- 被験者の抽出した結果とAIツールで抽出した結果の一致率は約45% (49/108)
実験ケースごとに見ると、一致率が70%を超えたのは5/16
- アンケート結果として、「納得できる」という回答があった実験ケースが9/16と半数以上を占める
 - 「納得できる」 : 9
 - 「納得できない」 : 4
 - 「どちらとも言えない」 : 3
- プロジェクトCでは記述フォーマット（記述内容）が統一された障害管理票を利用しており、一致率が高い（※1）
- プロジェクトBでは文字数の極端に少ないデータ（6文字）を使用しており、一致率が極端に低い（※2）
- プロジェクトBとプロジェクトDでは記述内容に専門用語を含むデータを使用しており、一致率が極端に低い（※3）

4. 評価

■ 実験結果

A-1			A-2			A-3			A-4			A-5		
被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出	
類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)
3781	3781	0.257	3730	3603	0.386	3883	3669	0.21	3617	3617	0.13	3850	3916	0.282
3860	3864	0.288	3603	3904	0.387	3669	3883	0.231	3941	3941	0.164	3933	3933	0.295
	3768	0.307	3779	3600	0.459	3840	3823	0.407	3715	3715	0.477	3948	3850	0.299
	3915	0.42	3907	3839	0.517		3840	0.409		3782	0.477	3893	3931	0.346
	3808	0.423		3853	0.518		3972	0.421		3701	0.495	3951	3951	0.371
	3586	0.472		3730	0.55		3862	0.471		3837	0.506	3681	3789	0.383
	3860	0.473		3896	0.605		3824	0.487		3764	0.523	3702	3681	0.454
	3691	0.49		3793	0.618		3721	0.497		3806	0.53		3893	0.485
	3654	0.493		3671	0.623		3582	0.508		3756	0.551		3601	0.495
	3758	0.497		3645	0.654		3586	0.531		3661	0.554		3807	0.527
一致率	100% (2/2)		一致率	50% (2/4)		一致率	100% (3/3)		一致率	100% (3/3)		一致率	71% (5/7)	

B-1			B-2			B-3			B-4			B-5		
被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出	
類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)
115	115	0.502	344	267	0.400	116	353	0.253	175	146	0.472	145	45	0.408
249	96	0.525	267	344	0.565	70	395	0.440	146	220	0.543	147	36	0.478
284	66	0.623	61	449	0.574	408	390	0.494	11	88	0.562	44	176	0.519
455	189	0.624	87	274	0.588	212	354	0.552	88	144	0.646	38	174	0.552
251	120	0.636	107	427	0.616	391	356	0.553	44	171	0.718	311	53	0.567
93	59	0.638	228	313	0.649	✖2	453	0.610	220	11	0.724	✖3	296	0.571
131	120	0.639	254	292	0.668	265	260	0.610	171	251	0.743	289	288	0.579
96	113	0.644	111	63	0.693	239	394	0.612	394	44	0.747	36	139	0.590
3	203	0.646	274	463	0.705	160	361	0.623	239	243	0.751	37	272	0.635
113	79	0.651	292	418	0.706	27	404	0.629	393	449	0.752	385	399	0.639
一致率	30% (3/10)		一致率	40% (4/10)		一致率	0% (0/10)		一致率	60% (6/10)		一致率	10% (1/10)	

4. 評価

■ 実験結果

C-1			C-2			C-3		
被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出	
類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)
102	120	0.169	111	111	0.154	98	98	0.157
120	102	0.224	23	50	0.29	35	6	0.334
12	45	0.327	38	44	0.328	74	30	0.341
15	69	0.34	50	77	0.355		61	0.372
59	22	0.422	68	93	0.371		11	0.404
※1	12	0.489	91	23	0.433	※1	90	0.429
	15	0.511	101	66	0.491		28	0.475
	88	0.567	105	2	0.522		85	0.502
	37	0.602	124	91	0.565		52	0.548
	71	0.632	132	101	0.589		35	0.612
一致率	80% (4/5)		一致率	50% (5/10)		一致率	66% (2/3)	

D-1			D-2			D-3		
被験者選定	Ginza算出		被験者選定	Ginza算出		被験者選定	Ginza算出	
類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)	類似案件番号	類似案件番号	Cosine類似度 (cosine距離)
169	160	0.141	279	281	0.128	382	180	0.428
177	161	0.143	281	285	0.128		406	0.469
191	205	0.261	285	297	0.142		289	0.521
205	401	0.285	291	291	0.144		171	0.533
260	169	0.346	297	279	0.156		441	0.567
298	409	0.366	344	299	0.391	※3	399	0.614
351	311	0.369	356	154	0.401		423	0.632
392	266	0.388	405	381	0.489		201	0.688
401	450	0.423	431	166	0.522		259	0.698
433	390	0.424	447	405	0.538		300	0.716
一致率	30% (3/10)		一致率	60% (6/10)		一致率	0% (0/10)	

5. 考察

5. 考察

■ 評価ポイント①

「AIツールで算出した類似度に閾値等で機械的な判断が可能か」

類似度（cosine距離）が約0.3 以下の場合に、類似障害として一致判定できる確率が高い

類似障害として判定できる件数 / cosine距離が0.3以下の件数
⇒約73% (19/24)



AIツールの算出した類似度（cosine距離）を、類似障害の機械的な判定のための基準（目安）として利用できる可能性あり

cosine距離が小さいにもかかわらず、類似しているとはいえない障害情報については、その理由を考慮し、改善する必要あり

例) 文書整形が過剰に行われ、文章の意味的な特徴が除かれていた

⇒障害情報の特徴が除かれないような、文章整形のルール化が必要

5. 考察

■ 評価ポイント②

「記述フォーマット（記述内容、文字数等）による 一致率への影響」

障害内容の文章構成が揃っているプロジェクトでは、類似障害を検出しやすい（cosine距離が0.3以下で検出され類似判定できた）

障害情報の文字数の少ないプロジェクトB(94文字以下)では一致率が低く、それ以外のプロジェクト(100文字を超える)では比較的一致率が高い



記述フォーマットを定め文章構成を統率することがAIツールによる判定に大きく影響する

類似文章を抽出するためにはある程度の文章の長さが必要

AI導入の前段階で、あらかじめ記述フォーマット、文章構成、文章の長さ等をルール化して合わせていく取り組みが有効

5. 考察

- 評価ポイント③
「専門用語の有無による一致率への影響」

専門用語が多く含まれるデータ群での実験では、今回用意したAIツールでは効果的に類似判定できていなかった



専門用語が含まれる文章にも対応するには、前処理として辞書作成等の実施が必要

今回は、より簡単にAIを活用することを重視し、特別な処理として追加学習や辞書の使用は行っていなかったため、専門用語が使われている文章で類似判定ができないことは想定通り

5. 考察

■ アンケート結果から得られた知見 効率面、業務知識・経験の補完の効果

被験者による類似障害検索で、1障害あたり平均12分を要したデータ総量が増えればさらに手間がかかると予想される

抽出の際に「業務経験に基づく解釈」や「ソフトウェアの振る舞いからの推測」といった業務やソフトウェアに対する知識・経験が必要



AIツールにより機械的に類似障害を抽出できれば、類似障害検索の手間を大幅に削減する効果が期待できる

AIツールでは、業務やソフトウェアに対する知識・経験に関係なく類似障害抽出でき、経験の浅い担当者でも効果的に類似障害の検索を行えるようになる

6. まとめ

6. まとめ

- **GiNZA、Embedding Projectorの2つのオープンソースのAIツールを活用し、類似障害を検出することができた**
 - AIを活用した品質向上の第一歩となった
- **AIツールが算出する類似度（cosine距離）を類似障害絞り込みの指標・基準として利用できる可能性を確認できた**
 - 膨大な情報の中から、AIツールを使い、指標によるフィルターをかけて情報の絞り込みを行うような使い方が有効
- **障害情報の記述フォーマットを統一し、文章構成を合わせることでAIツールによる類似判定の精度が向上することを確認できた**
 - 既に文章構成が統率されているようなプロジェクトにターゲットを絞ると早期に効果が得られる
- **今回作成したプログラムは簡易なものであったが、十分有用な結果得られた**
 - 高度なAI知識、開発スキルを保有していない場合でも、AIを活用できるアプローチを提案することができた

6. まとめ

蓄積する過去の障害事例から、類似障害を検出することで原因や対応策など参考となる情報を導き出せる！

高度な知識がなくとも、身近なAI技術をソフトウェア品質の課題解決に活用できる！



6. まとめ

■ 今後の課題

- 専門用語が含まれる文章については辞書などによるチューニングが必要であるが、AIに関する専門的な知識が必要、手間がかかる
 - ➡ 専門的な知識がなくともチューニング作業が容易にできる仕組みがあれば、さらに活用の幅が広がる
- AIツールの指標による機械的な類似障害判定を提案しているが、最適な閾値について本稿では検証できていない
 - ➡ 最適な閾値については今後検証が必要
- 記述フォーマット・文章構成の統一の取り組み行なっているプロジェクトはまだ少ない
 - ➡ 例えばBotなどによりヒアリング形式で記述していくことで自然と文章構成が統一されていくような仕組みの検討が有効と考える

ご清聴ありがとうございました