

シンポジウム

「ビジネス日本語の研究と教育 ―文書改善に必要な条件を探る―」

(2021.9.25)

FUJITSU

shaping tomorrow with you

AI による効果的な文書表現の発見

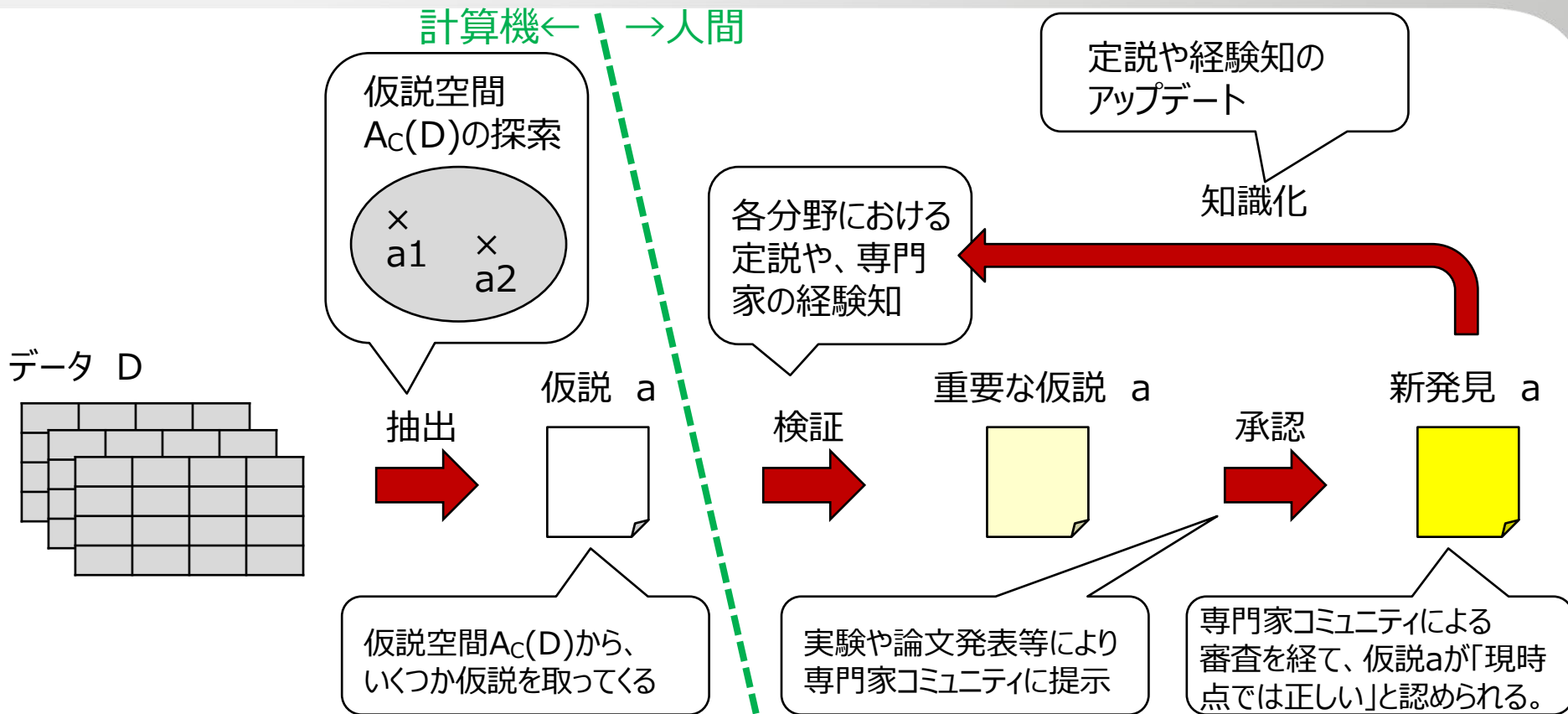


富士通株式会社
研究本部人工知能研究所
シニアリサーチャー

浅井 達哉

- AIによる科学的発見のプロセス
- 共同研究の進め方と役割分担
- 富士通の説明可能なAI「Wide Learning™」
- AIによる効果的な文書表現の発見
 - 仮説空間の設計
 - データクリーニング
 - AIの適用と発見事例

AIによる科学的発見のプロセス



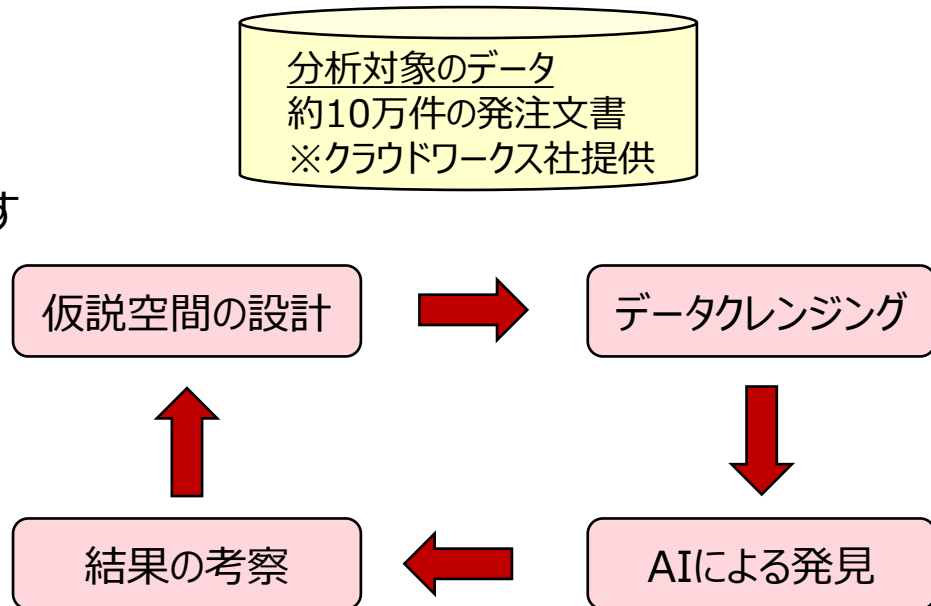
共同研究の進め方と役割分担

■ 進め方

- 言語学の専門家（言語班）と AIの専門家（AI班）が右図のサイクルを試行錯誤しながら繰り返す

■ 役割

- 仮説空間の設計：言語班・AI班
- データクレンジング：AI班
- AIによる発見：AI班
- 結果の考察：言語班



富士通の Wide Learning技術

Wide Learningとは

ある商品を買う人と買わない人はどのような人?

データから重要な仮説を網羅的に発見!

購買データ



	富士通 太郎 Taro Fujitsu
性別	男性
婚姻	既婚
住まい	持ち家
年齢	35歳
年収	¥650万



全組合せの中から 自動で学習・発見できる



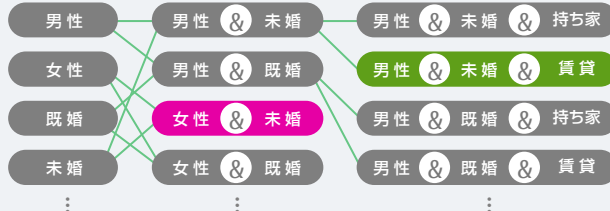
買う人の特徴

男性 & 未婚 & 賃貸



買わない人の特徴

女性 & 未婚



得られた仮説を使って予測!



新たな顧客に当てはめて 根拠と一緒に予測できる



この人は買う? or 買わない?

予測



根拠

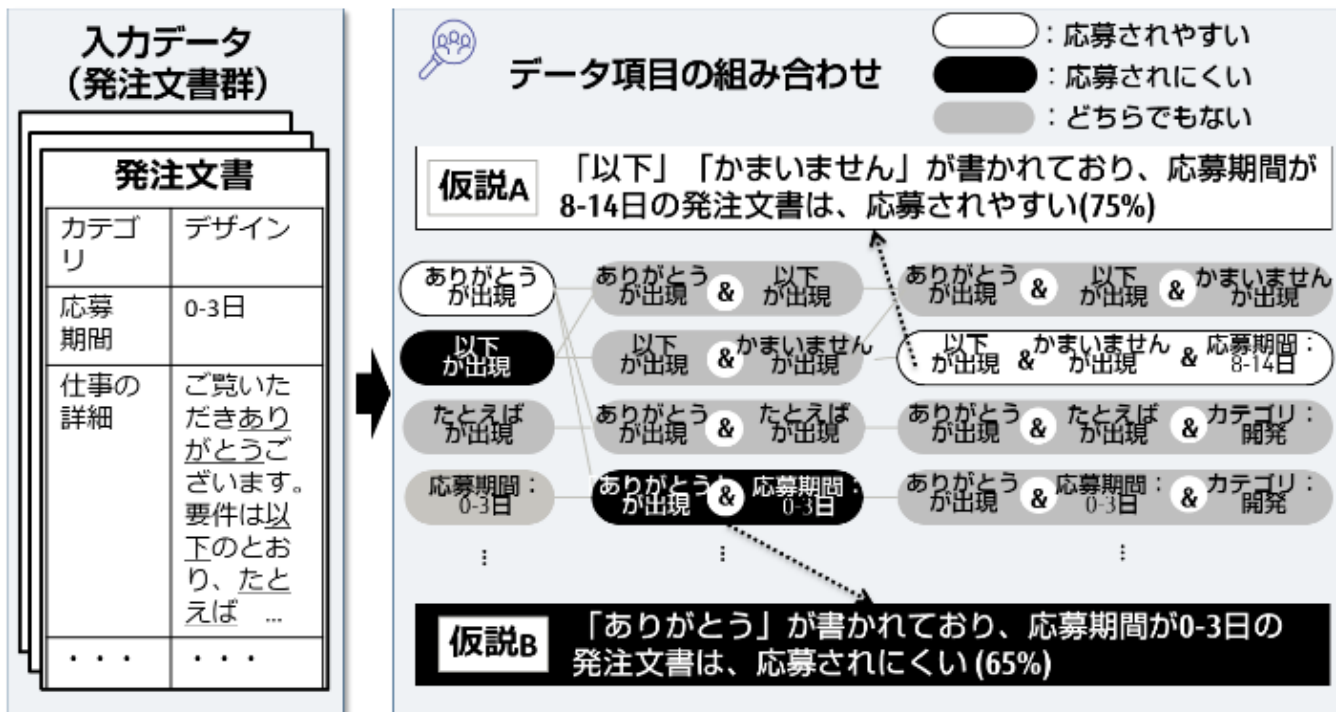
男性 & 未婚 & 賃貸

資料請求有 & 年収500万

貯金なし & 事務職

30代 & 車所有

- 発注文書から作成したデータ項目を組み合わせることで、
よい発注文書と悪い発注文書を判別するための重要な仮説を発見する



出典: 浅井・横野・柳瀬・岩崎・井上・田中・石黒「クラウドソーシングの言語表現 —ビジネス日本語研究におけるAI技術の活用—」
 専門日本語教育研究(22), pp. 9-16, 2020年12月.

発見の準備

仮説空間の設計とデータクレンジング

仮説空間の設計（言語属性）

- 発注文書を精査し、文書の印象に影響を与えると考えられる283属性をピックアップ
- ピックアップした言語表現（抜粋）

大分類	小分類	事例
発注者側	一人称	わたし、私、当方
	発注者の行為	いたします
受注者側	二人称	あなた、そちら
	受注者の行為	ご応募ください
社交表現	感謝	ありがとう、感謝です
	謝罪	失礼しました
接続表現・参照表現	接続表現	そのため、しかし
	参照表現	以下、上記
程度量副詞	量副詞	おおよそ、すこし
	頻度副詞	たまに、よく

仮説空間の設計（非言語項目）

- 数値またはカテゴリで表現される属性に関して、下記の19項目を作成

種類	項目
募集人数	1人、2-5人、6-10人、11人以上
応募期間	0-3日、4-7日、8-14日、15日以上
報酬（下限）	5,000円未満、5,000円、10,000円、50,000円、100,000円、300,000円、500,000円
カテゴリ	ライティング、開発、デザイン、ビジネス事務

302属性（言語属性283種類、非言語項目19種類）の
任意の組み合わせを仮説空間とする

■ 発注文書のコピペ

- 同じような案件で、似たような文章が大量に存在
- **頻度に基づく分析においては、悪影響を及ぼすおそれがある**

【1記事100円】 **かに通販「北釧通販」** の口コミを書いてください。（300字）

良かった点や味、おすすめの商品、対応などを書いてください。

※文章は改行を入れてください。

【注意事項】

- ・文字数を稼ぐ行為は禁止とします。
- ・他記事のコピペが発覚した場合は、報酬をお支払いできません。

またお仕事のご依頼はその時点で中止させていただきます。

【1記事100円】 **カニの浜海道** の口コミを書いてください。（300字）

良かった点や味、おすすめの商品、対応などを書いてください。

※文章は改行を入れてください。

【注意事項】

- ・文字数を稼ぐ行為は禁止とします。
- ・他記事のコピペが発覚した場合は、報酬をお支払いできません。

またお仕事のご依頼はその時点で中止させていただきます。

発注文書のコピペの例

■ データクレンジングの手順

■ 応募率の上位25%と下位の25%にラベル（正例または負例）を付与

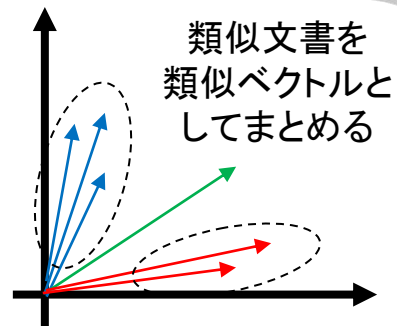
• 応募率 = 応募者数 / 閲覧者数

■ レコード間の類似度を計算し、類似レコードをクラスタ化

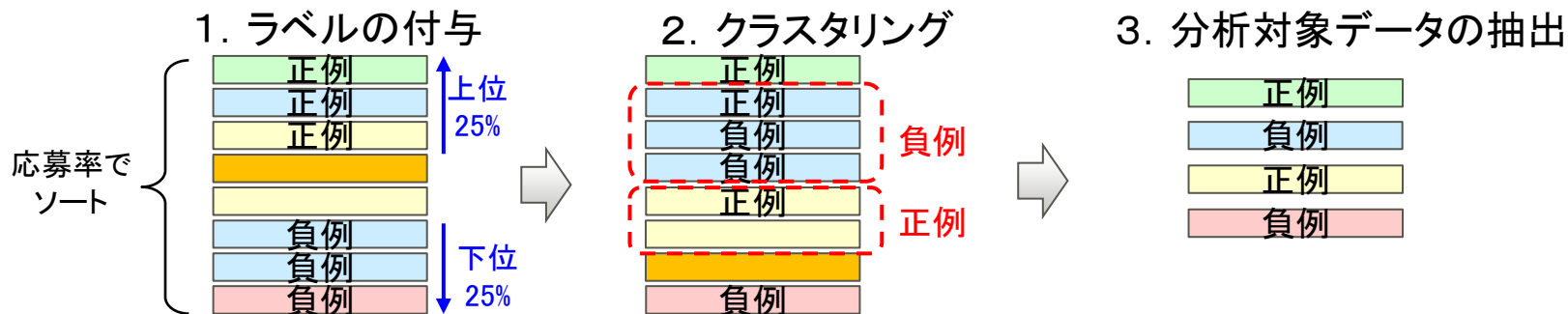
• 発注文書をbigramでベクトル化し、コサイン類似度でクラスタリング

• 各クラスタ内のレコードの応募率平均をみて、クラスタそのものの正例/負例を決める

■ 各クラスタから代表レコードを1件ずつ抽出し、それらを分析対象データとする。



■ データクレンジングの結果 正例：2802件 負例：3250件



(参考) クラスティング結果の抜粋

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	1511990	かに通販のおすすめのお店を紹介してください 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 】 ----- 																
2	1528083	【1記事50円】かに本舗の口コミを書いてください。 (300字?) 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 】 ---																
3	1528086	【1記事100円】かにまみれの口コミを書いてください。 (300字?) 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 】 																
4	1511926	かに通販のおすすめを紹介してください。 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 】 ----- ・文																
5	1528095	【1記事100円】かに通販「北廻り」の口コミを書いてください。 (300字?) 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 																
6	1528101	【1記事100円】カニの北海道の口コミを書いてください。 (300字?) 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 】 																
7	1528107	【1記事100円】北国からの贈り物の口コミを書いてください。 (300字?) 良かった点や味、おすすめの商品、対応などを書いてください。 ※文章は改行を入れてください。 【 注意事項 																
8	-----																	
9	1512814	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
10	1524503	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
11	1521977	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
12	1542916	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
13	1555487	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
14	1534747	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
15	1526946	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
16	1555486	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
17	1576947	▽お仕事内容 会社名のリストをお渡し、その会社の事業内容などの情報が記載されたサイトがあるかどうかを調べていただきます。 公式サイトや求人サイト、それに準ずるサイトがあればURLを入力し、角																
18	-----																	
19	1569235	表題の通り、現在楽天市場GOLDのデザイン・コーディングをそのままポンパレモールとWowmaで同じように表示させたいです。 全く同じで構いません。 なるべく早く仕上げて頂ける方、また安価にお受け																
20	1552578	表題の通り、現在楽天市場GOLDのデザイン・コーディングをそのままポンパレモールとWowmaで同じように表示させたいです。 全く同じで構いません。 なるべく早く仕上げて頂ける方、また安価にお受け																
21	1521822	表題の通り、現在楽天市場GOLDのデザイン・コーディングをそのままポンパレモールとWowmaで同じように表示させたいです。 全く同じで構いません。 なるべく早く仕上げて頂ける方、また安価にお受けいた																
22	-----																	
23	1526649	【コンテンツの具体的な内容】 「 冊子印刷 」について記事・コンテンツを作成する為、タスク形式で依頼致します。 冊子印刷の方法、業者に依頼するメリット、 冊子印刷のシステムについて、精																
24	1550736	【コンテンツの具体的な内容】 「 お名前シール 」について記事・コンテンツを作成する為、タスク形式で依頼致します。 お名前シールって何だろう、お名前シールの活用方法、使うメリット、 料金の精																
25	1538170	【コンテンツの具体的な内容】 「 封筒印刷 」について記事・コンテンツを作成する為、タスク形式で依頼致します。 封筒印刷の方法、業者に依頼するメリット、 封筒印刷のシステムについて、精																
26	1539698	【コンテンツの具体的な内容】 「 名刺印刷 」について記事・コンテンツを作成する為、タスク形式で依頼致します。 名刺印刷の方法、業者に依頼するメリット、 名刺印刷のシステムについて、精																
27	1526651	【コンテンツの具体的な内容】 「 お名前シール 」について記事・コンテンツを作成する為、タスク形式で依頼致します。 お名前シールって何だろう、お名前シールの活用方法、使うメリット、 料金の精																
28	1541082	【コンテンツの具体的な内容】 「 冊子印刷 」について記事・コンテンツを作成する為、タスク形式で依頼致します。 冊子印刷の方法、業者に依頼するメリット、 冊子印刷のシステムについて、精																

発見例と考察

よい文書と悪い文書に特徴的な言語表現

「よい文書」に多く見られる特徴的な言語表現

順位	言語表現	正例数	負例数
1	お待ちしております	1081	313
2	気軽に	828	322
3	たくさん	667	231
4	こちら	855	406
5	コツコツ	122	19
6	弊社	510	315
7	以降	182	63
8	致します	670	487
9	いただければ	384	226
10	いたします	1037	909
11	しっかり	165	71
12	尚	105	30

「悪い文書」に多く見られる特徴的な言語表現

順位	言語表現	正例数	負例数
186	あなた	126	419
185	してください	669	1193
184	必ず	180	481
183	禁止です	63	220
182	して下さい	121	306
181	以下	479	789
180	自由に	48	150
179	以上	633	964
178	なぜ	18	68
177	お願いします	861	1185
176	どう	104	202
175	できません	110	206

■ ポジティブな言語表現を組み合わせたら、ネガティブな印象を与える例

- 「ぜひ（是非）」+「いたします（致します）」
 - 是非、よろしくお願いいたします。
 - 是非協力よろしくお願いいたします。
- 「なにとぞ（何卒）」+「いたします（致します）」
 - 何卒、宜しくお願い致します。
- 「どうぞ」+「いたします（致します）」
 - どうぞ、宜しくお願い致します。

■ 考察

- 強調する副詞をお願いする表現とともに使用すると、業務量の多い仕事や難度の高い仕事を押しつけるような印象を与え、閲覧者に好まれなくなる

■ ネガティブな言語表現を組み合わせたら、ポジティブな印象を与える例

- 「あらかじめ（予め）」+「できません（出来ません）」

- ……できませんので予めご了承ください。

■ 考察

- できないことを単に伝えるだけでなく、前もって知っておいてほしい、理解してほしいという気持ちが感じられ、閲覧者に好印象を与える

発見例（3）

■ 状況によって印象が異なる言語表現

	正例数	負例数
[言語表現]ありがとう	492	374
[非言語項目]応募期間0-3日	399	417
[言語表現]ありがとう & [非言語項目]応募期間0-3日	43	77


急ぎの案件においては、閲覧者は内容を素早く把握したいので、「ありがとう」と書くことは逆効果。

	正例数	負例数
[言語表現]以下	479	789
[非言語項目]ビジネス事務	677	929
[言語表現]以下 & [非言語項目]ビジネス事務	124	139

ビジネス事務案件においては、「以下」のような無機質な表現がそれほど嫌われない。

- 科学的発見の方法論にしたがい、AIを用いて効果的な言語表現を発見した
 - 言語の専門家と、AIの専門家とのコラボレーション
 - データサイエンスのサイクルを試行錯誤しながら繰り返した
 - 仮説空間の設計
 - データクレンジング
 - AIによる発見
 - 結果の考察
 - いくつかの発見例の提示と考察

次の発表（柳瀬）で、分析結果から得られた発注文書の改善方法を紹介します



FUJITSU

shaping tomorrow with you