

人工知能技術を活用した動画や画像などのメディアを対象としたファイル検索システム Search Mission の事例紹介

Case study of a file search system “Search Mission” of a target of media of a movie and image, etc., using AI Technology

兼安 勉^{1*}

Tsutomu Kaneyasu¹

¹ アルファテクノロジー株式会社

¹ Alpha Technology Co., Ltd.

Abstract: As the usual provision of products and services with AI technology of explosive leap, in order to solve the problem of time and cost, and the problem of engineer shortage, many small and medium enterprises are using API services of other companies' AI technology.

In addition, based on the shift to a media medium due to normalization of smartphones, etc., and based on "work style reforms" required for today's companies, this paper will introduce system examples to be implemented to efficiently search files using AI technology.

1 はじめに

爆発的に飛躍中の人工知能技術搭載の製品やサービス提供が当たり前になる中で、時間とコスト、更に技術者不足の課題を解決するべく、多くの中小企業は他社の人工知能技術の API サービス活用を選択する。

また、スマートフォンなどの常態化によるメディア媒体へのシフト、昨今の企業に求められる「働き方改革」を踏まえ、本書では、人工知能技術を活用し、従来対象外であったファイル検索を効率的に実施するシステム事例を紹介する。

2 課題

Web 上から欲しい情報を検索する仕組みは、Google などを筆頭に浸透している。

ただし、企業内の検索には多くの課題が存在し、システム単位での検索利用となっているのが現状である。

*連絡先：アルファテクノロジー株式会社
事業統括部 製品開発課

〒150-0043 渋谷区道玄坂 2-10-7 新大宗ビル 1 号 9F
E-mail: kaneyasu.tsutomu@alphatc.co.jp

課題としては、以下 4 点となる。

- ①資料が見つからない
- ②色々なシステム利用の不便さ
- ③色々なメディアの活用拡大
- ④社内、社外でも検索、閲覧したい

2.1 資料が見つからない

図 1 に示すように、人は時間経過するごとに記憶したことを忘れてしまう。^{*1} 図 2 の赤枠には、企業の業務において、必要な文書を容易に取り出せない問題から生じる対処に費やす時間の割合を示す。^{*2}

1 年に換算すると、約 3 ヶ月分となる。



図 1：エビングハウスの忘却曲線

	1週間に費やす時間	費やす時間の割合	1週間に無駄にする時間	無駄にする時間の割合	企業の生産性が失われる割合
複数のファイルやフォーマットに存在する情報を1つの文書に集約する	4.1	8.2%	1.0	2.0%	1.7%
紙の文書で発生する問題や手間のかかるタスクに対処する	3.9	7.8%	1.0	2.0%	1.6%
文書を探すが見つからない	2.7	5.5%	2.7	5.5%	4.4%
文書の最新バージョンまたは正しいバージョンが見つからないか失われてしまったため、再作成する	2.3	4.6%	2.3	4.6%	3.7%
合計	12.9	26.1%	7.0	14.1%	11.4%

n = 140 (国内のインフォメーションワーカー)
 Note:
 ・調査方法は補遺/補遺資料を参照。割合は、Figure 1の作業に費やす1週間当たりの52.4時間という回答と、従業員の81%がインフォメーションワーカーであるというデータに基づく。
 ・列の合計値が合わない場合は、四捨五入による誤差である。

図2：問題への対処に費やす時間

2.2 色々なシステム利用の不便さ

Amazon EC2 サービス開始, Microsoft Azure に登場により、クラウド市場が2006～2008年にかけて大幅に拡大してきた。^{*3}

以前は企業内でサーバ構築、運用するオンプレミスの形態が主流であった。

上記の流れから、当初はオンプレミスから全てクラウドへ移行し、統合システムを期待されていた。実際には、図3に示すように、システム統合が困難で、業務に適したシステム活用を目指し、オンプレミスとクラウドを混在したハイブリッドクラウドが2015年以降から徐々に割合を増やしている。^{*4}

そのため、企業内では様々なシステムの混在化は継続し、それに伴う情報連携への課題が残ったままである。

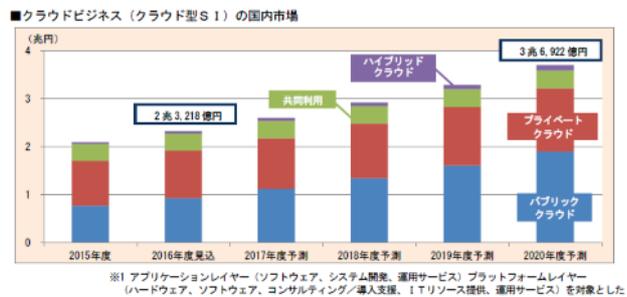


図3：クラウドビジネスの国内市場

ここで文書について示す。文書とはe文書法^{*5,*6}で定めている。

代表的な文書としては、

- ・ 会計帳簿
- ・ 契約書、見積書等
- ・ 振替伝票
- ・ 営業報告書
- ・ 財産目録
- ・ 規約等
- ・ 議事録等
- ・ 設計書

などが挙げられる。

また、e文書法における基本要件としては、

- ①見読性
- ②完全性
- ③機密性
- ④検索性

が挙げられる。

項目2.1, 2.2で重要な要件としては、上記の③機密性、④検索性が該当する。

2.3 様々なメディアの活用拡大

企業内の業務は、紙媒体からPDFやOffice文書などのドキュメントファイルの電子媒体が中心になった。昨今では、SNSの台頭、主流により、図4に示すように、動画のメディア活用が拡大している。^{*7}

今後は更に様々なメディア活用は拡大していき、ドキュメントファイルの利用率と同等になってくるとも考えられる。

ここで企業における動画活用の主なシーンは、

- ・ 販促、営業
- ・ マニュアル
- ・ カスタマーサポート
- ・ 社員教育
- ・ 採用活動

となる。

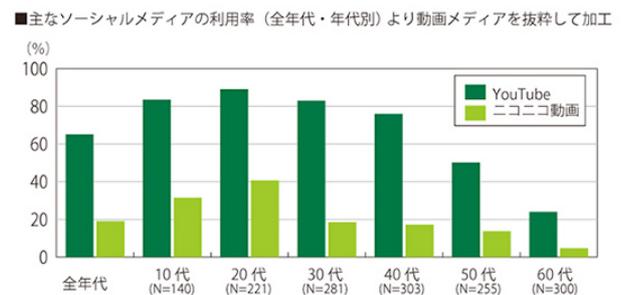


図4：SNSによる動画メディア加工

2.4 社内、社外でも検索、閲覧したい

スマートフォンやタブレットの利用率は全世代でも7割を超え、企業内でも必需品となった。

主な利用は、図5に示すように、

- ・ 営業活動の効率化
- ・ 業務時間の短縮
- ・ ファイル・データの共有

などである。^{*8}

どの項目も社内、社外の場所を問わず、どこでも簡単に欲しい情報を検索し、閲覧できるといった要素を含んでいる。

また、社外からの閲覧はセキュアな環境が求められているため、項目 2.2 にも記載している基本要件の重要な点の機密性、検索性にも密接に関連する。

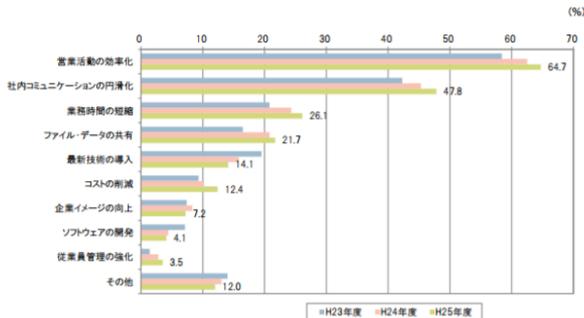


図 5：スマホやタブレットの業務利用シーン

3 提案方法

3.1 従来の検索方法

テキストを抽出できる PDF や Office 文書などのドキュメントファイルから全文テキストを抽出し、インデックス検索するために再構築したデータベースを利用する。

インデックス検索とは、以下 3 点のポイントがある。

- ①事前に文書から単語を抽出し、各単語の頻度を数値化
- ②検索キーワードとなるクエリ頻度と、上記①で抽出した文書頻度をコサイン類似度で測定
- ③コサイン類似度から定めたルールに応じて(例：高い値)、ヒット文書を抽出

インデックス検索の例としては、下図 6 が存在するとき、各文書のコサイン類似度は以下となる。

ここでは文書 3 を抽出する。

	私	あなた	年齢	性別	扉	空	雲	雨	晴
文書1	0	0	0	0	0	0	0	0.71	0
文書2	0	0	0.71	0.48	0.18	0	0	0.48	0
文書3	0.64	0.44	0	0	0.16	0.44	0	0	0
クエリ	0	0	0	0	0.35	0.94	0	0	0

図 6：インデックス検索の数値

文書 1 とクエリのコサイン類似度 =

$$0*0 + \dots + 0*0.35 + 0*0.94 + \dots = 0$$

文書 2 とクエリのコサイン類似度 =

$$0*0 + \dots + 0.18*0.35 + 0*0.94 + \dots = 0.06$$

文書 3 とクエリのコサイン類似度 =

$$0*0.64 + \dots + 0.18*0.35 + 0.44*0.94 + \dots = 0.47$$

3.2 畳み込みニューラルネットワーク手

法によるタグ抽出

画像を複数のカテゴリに分類するように学習し、タグ付けを実現するには、一般的な手法である畳み込みニューラルネットワーク (Convolutional Neural Network) を利用する。

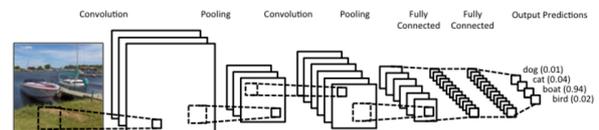


図 7：畳み込みニューラルネットワーク

図 7 に示すように、Clarifai^{*9} が提供している畳み込みニューラルネットワークは、①畳み込み、②プーリング、③畳み込み、④プーリング、⑤全結合、⑥全結合で構成されている。

畳み込みは図 8 に示すように、画像内のピース単位のピクセル値に、フィルタとなるピクセル値を掛け、特徴のピクセルの総数で割ることで求める。^{*10}

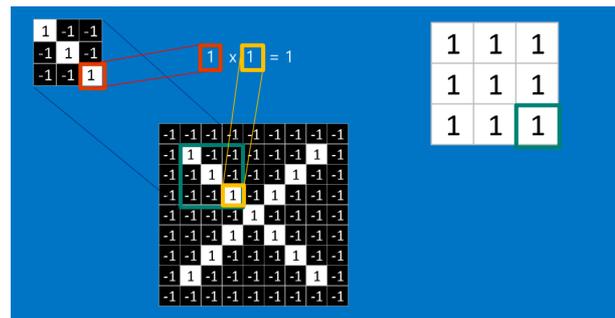


図 8：畳み込み

プーリングは図 9 に示すように、大きな画像を重要な情報を残しつつ縮小する方法である。画像内を小さなウィンドウに区切り、区切った各々のウィンドウ内の値から最大値を求めるものである。

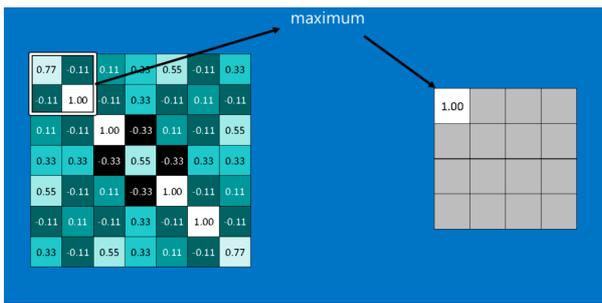


図 9：プーリング

本稿では 1024 次元の全結合層となる。

前記畳み込みニューラルネットワーク手法で学習したモデルを活用することで、図 10 に示すように、画像を入力すると、「幸せ」や「純粋」タグといった表情などを認識すること、「猫」タグといった人間以外や職業なども認識することができる。



図 10：タグ抽出

3.3 本提案の検索方法

本提案におけるインデックス検索用のデータベースへ抽出したタグを登録するフローを図 11 に示す。

アップロードしたファイルに対し、処理を行いやすいようにファイルを変換する。ここでは、ドキュメント類は PDF へ、動画類は mp4 へ、画像類は jpg とする。

次に、ファイルのフォーマットタイプ毎にサムネイル画像の抽出方法が分岐する。

PDF の場合は、先頭ページの全体をサムネイル画像で抽出する。

mp4 の場合は、動画内の各画像から抽出した解像感の高いものをサムネイル画像で抽出する。解像感での抽出方法は別途記す。

jpg、その他の場合は、画像自体、または、所定画像をサムネイル画像で抽出する。

本提案では抽出したサムネイル画像に対し、前項 3.2 で記した畳み込みニューラルネットワークモデルを用いて、認識値の高い上位 10 個のタグを抽出する。

ここでの畳み込みニューラルネットワークの事前学習で活用している情報は、人物、モノ、感情などを含めてタグ数 11,000 個となる。

PDF 以外は、タグのみでインデックス検索用のデータベースへ登録する。PDF は全文テキストを抽出し、タグと連結した情報でインデックス検索用のデータベースへ登録する。

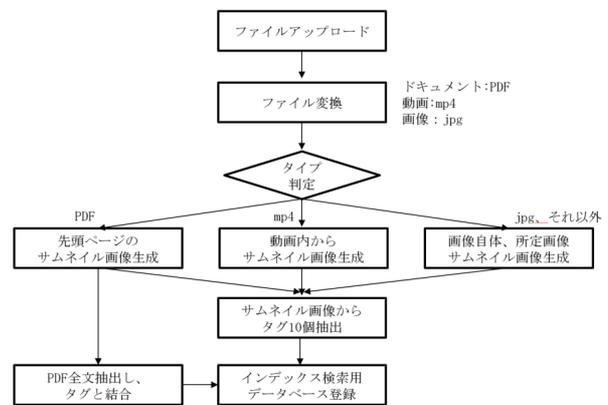


図 11：インデックス検索 DB へのタグ登録手法

解像感での抽出は下記①、②を処理する。

①画像から輝度値の 2 次元配列を生成する。

$$\text{輝度値} = 0.298912 * r + 0.586611 * g + 0.114478 * b$$

P11	P12	P13	P14	P15	P16	...
P21	P22	P23	P24	P25	P26	
P31	P32					
P41	P42					
P51						
P61						
.						
.						
.						

②横同士、縦同士の全組合せの輝度値の差の比を足し合わせて平均を取得する。

$$R[11-12] = |P11 - P12| \div (P11 + P12)$$

$$\text{SUM[横方向]} = R[11-12] + R[12-13] + \dots + R[21-22] + R[22-23] + \dots$$

SUM[縦方向] = R[11-21] + R[21-31] + … + R[12-22]
 + R[22-32] + …

評価値 = (SUM[横方向] + SUM[縦方向]) ÷ 個数

4 事例

4.1 アニメ業界

アニメ業界では背景やモノなどの素材画像を再利用して活用するケースが多い。ただし、素材画像の管理が外付けハードディスクを利用し、乱雑な状態である。また、素材画像に対しても正しいタグ付けが行われていないという課題がある。

このような課題を解決するため、本ターゲットに絞り込んだ畳み込みニューラルネットワークでカスタマイズ学習したモデルも利用することにする。

まず本提案では、図 12 に示すように、AI 技術を活用した検索システム『Search Mission (サーチミッション)』を基盤とする。Search Mission 自体は Windows Server の Web システムである。

システム構成としては、

- Web サーバ： Nginx 1.11.4
- PHP： 7.0.10
- PHP フレームワーク： CakePHP3.0
- DB サーバ： PostgreSQL 9.5.7-1

となる。

エンドユーザーは、ブラウザ (Internet Explorer 11 以上, Firefox 24.0 以上, Google Chrome 29.0 以上, Safari iOS のみ 5.0 以上) さえあれば、デバイスを問わずに操作することができる。



図 12： Search Mission 構成図

図 13 に示すように、一般モデルを活用して抽出した 10 タグ (コンセプト) だけでなく、カスタマイズしたモデルからピンポイントとなる 1 タグ (コンセプト) を合わせて抽出し、インデックス検索用のデータベースへ登録する構成する。

今回は図 14 のように 8 タグ (コンセプト) に対し、Positive, Negative 画像を各々学習した。1 タグ当たり画像 30 枚程度が推奨という中で、Positive に関して

は、十分な枚数を確保することが出来た。

結果として、本事例での認識値は低い。認識結果の中での上位タグを抽出すると、望むタグを識別することは出来ている。

ただし、今後の改善点は、

- タグ (コンセプト) の数を増やしたときの識別
- 複数キャラクタなどが含まれているときの識別
- カスタマイズモデルのリアルタイム更新が挙げられる。

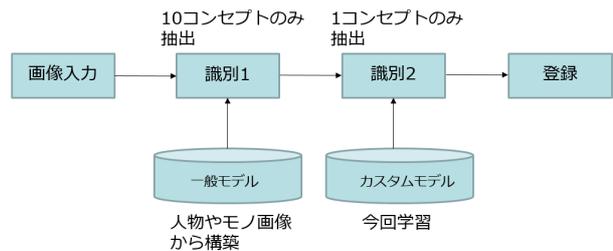


図 13： 識別 (タグ抽出) 構成

コンセプト	Positive (ファイル数)	Negative (ファイル数)
キャラクタA	79	21
キャラクタAの口	14	11
刀	52	0
女性A	53	50
女性B	49	37
男性C	13	10
男性D	12	22
黒い車	28	0

図 14： 学習データ数

4.2 納品 & FAQ 提供

企業には、ファイル管理システムや転送媒体の暗号化システムを利用するケースが多い。ただし、ファイル管理を社外との共有で利用することができないことや、媒体転送はあくまでも社外へ一方的に送信するだけで社外からの負担や転送方法などが存在しないといった課題がある。

また、製品紹介やFAQは別システムや別部署の担当者で管理されるため、構築のハードルが高く、大企業以外で提供されるケースは少ない。

本事例では図 15 に示すように、Explorer 風のフォルダ管理を実装している。各フォルダを社外ユーザー単位で管理し、アクセス権限を付与することで、各フォルダ内に納品物、プロジェクトに必要なデータを互いにアップすることが出来るため、社内外で

の共有を実現することが可能となる。

また、製品、会社紹介の動画を自動タグ付けでアップロードすることができるため、ユーザーが自然と他の情報を目にするような環境を構築できる。

最後に、製品に関する FAQ データは、企業内で蓄積している。Search Mission へ質問、回答の 1 グループ単位をテキストデータでアップすることで、容易に FAQ システムを構築することができる。

多くの FAQ データを人手でアップするには工数を費やしてしまうが、NTT データ製 WinActor^{*11} といった RPA (Robotic Process Automation) ツールを活用すると、数分でデータアップを完了することができる。

RPA とは、認知技術 (ルールエンジン、機械学習、人工知能等) を活用した、主にホワイトカラー業務の効率化、自動化の取組みである。人間の補完として業務を遂行できることから、仮想的労働者 (Digital Labor) とも言われている。^{*12}



図 15 : Search Mission UI

むすび

本稿では、企業内の検索に存在する 4 個の課題、①資料が見つからない、②色々なシステム利用の不便さ、③様々なメディアの活用拡大、④社内、社外でも検索、閲覧したい、の各々について整理した。

前記課題に対して、人工知能技術で一般的に利用される畳み込みニューラルネットワーク手法、インデックス検索手法の基礎について説明した。その後、ドキュメントだけでなく、画像や動画といったメディアも含めた検索方法におけるタグ登録フローを中心に紹介した。

最後に、アニメ業界、納品&FAQ 提供の 2 事例を活用して、人工知能技術搭載のファイル検索システムを紹介した。

今後としては、

- ・タグ (コンセプト) の数を増やしたときの識別
 - ・複数キャラクタなどが含まれているときの識別
 - ・カスタマイズモデルのリアルタイム更新
- の改善点に加え、動画全体に渡るタグ抽出や、音声に対するタグ抽出なども実現していきたい。

謝辞

本稿作成にあたりましては、某社の情報の御提供、ならびに種々の御助言を賜りました。ここに深甚なる謝意を表します。

参考文献

- [1] フリージュニアアカデミー, <http://free-academy.jp/junior/index.php?エビングハウスの忘却曲線>
- [2] 日本国内のインフォメーションワーカーが抱える生産性ギャップを埋める: IT 部門の新たな課題と機会, IDC, 6月, 2012
- [3] 丸山不二夫, クラウド時代の始まり, <https://thinkit.co.jp/article/1024/1>, 2009
- [4] 2017 クラウドコンピューティングの現状と将来展望, 株式会社富士通キメラ総研, 2017
- [5] e-文書法によって電磁的記録が可能となった規程, 経済産業省,
- [6] 文書の電磁的保存等に関する検討会, 文書の電子化の促進に向けて, 経済産業省, pp.48-64
- [7] 平成 26 年情報通信メディアの利用時間と情報行動に関する調査 報告書, 総務省, pp.84, 2014
- [8] 平成 26 年度我が国情報経済社会における基盤整備報告書, 経済産業省, 2014
- [9] Clarifai, <https://www.clarifai.com/>
- [10] 「畳み込み」と「プーリング」, <http://postd.cc/how-do-convolutional-neural-networks-work/>
- [11] WinActor, <https://winactor.jp/>
- [12] Wikipedia, <https://ja.wikipedia.org/wiki/ロボティック・プロセス・オートメーション>

ホワイトカラーの非コア業務効率改善に向けた 個人支援 RPA の検討

Study on Personal-Assistance Robotic Process Automation for Improving Efficiency of White-Collar Non-Core Operations

濱本真生^{1,2} 櫻井隆雄¹ 恵木正史¹

Masaki Hamamoto^{1,2}, Takao Sakurai¹, and Masashi Egi¹

¹ 日立製作所 デジタルテクノロジーイノベーションセンタ

¹ Center for Technology Innovation - Digital Technology, Hitachi, Ltd.

² 日立アジアマレーシア 研究開発センタ

² Research & Development Center, Hitachi Asia (Malaysia) Sdn. Bhd.

Abstract: Recently, workload of white-collar job is getting larger and larger and it became one of the major social problems in Japan. To solve the problem, improvement of business-operation efficiency in each worker is necessary. As a tool for improving the efficiency, Robotic Process Automation (RPA) technology is emerging. However, current RPA tools are aimed at back office operations but not at white-collar's various business operations in each worker. In this report, a personal-assistance RPA (PA-RPA) which assists input operation in business applications is proposed. Our rapid prototyping demonstrates that the proposed PA-RPA can reduce operating time by 57% when it's applied to settlement of travel expenses through in-house business application in our company.

1. はじめに

近年、ホワイトカラーの過重労働が大きな問題になっており、2016年9月に日本政府は「働き方改革実現会議」を立ち上げた[1]。また2016年10月には厚生労働省から「過労死等防止対策白書」が発刊された[2]。このようにワークライフバランス、残業時間縮減の実現に向けて労働者一人ひとりの業務効率改善がますます大きな社会課題となっている。

従来、システムエンジニア(SE)による定型業務のITシステム化によってホワイトカラーの業務効率は改善されてきた。しかし、部署ごとの運用ポリシーの違いや、長期運用途中での制度や業務オペレーションの変化、短期間の試験的業務オペレーションなどに対しては、その都度ITシステムをリプレースすることが金銭的な理由により実施できず、現場の労働者が手作業で対応する実態があり、その業務効率改善が十分されない問題があった。

このような問題を解消する方法として、昨今 Robotic Process Automation (RPA) と呼ばれる取り組みが注目されている。RPAは、現場労働者が自動化したい業務プロセスをスクリプトとして書き出し、専

用の業務ロボット・ソフトウェア（以下、業務ロボットと略す）に自動実行させることで業務を効率化する現場主導型の取り組みである。RPAはOCR（光学的文字認識）やパタン認識などAI技術の進展を背景に急速に普及しつつあり、その市場規模は年率20%で成長し、2024年には50億ドルに上ると予想されている[3]。しかし、現在のRPAは決済業務や監査業務などバックオフィス業務をターゲットとしており、多様性を内包する労働者一人ひとりの業務効率改善には対応できない。

一方で、近年ユーザー一人ひとりの情報を蓄積し活用するインタフェースとしてチャットボットと呼ばれる対話ロボットが大きく進展している。対話ロボットは個人情報と複数のサービスアプリケーション（以下、サービスと略す）を一元管理し、さらにITリテラシが不要で誰でも操作可能な自然言語インタフェースを提供できることが強みである。対話ロボットは複数のサービスの窓口として機能し、そこを通過する全ての個人情報を記憶できる。そのため、昨今急速に進展しているAI技術と連携することで労働者一人ひとりの業務支援と非常に相性が良いと考えた。

そこで本報告では対話ロボットと業務ロボットを用いた個人支援 RPA を提案し、そのシステムアーキテクチャを検討する。個人支援 RPA のターゲット業務としてホワイトカラーの典型的な非コア業務である申請/登録業務を設定した。試作した提案 RPA を社内の旅費精算システムに適用し、業務効率改善効果をキータイプ数、マウスクリック数、作業時間の観点で定量的に評価する。

2. 個人支援 RPA の提案

本章では RPA の概要と提案する個人支援 RPA の位置づけを示す。また、提案 RPA 実現のための課題を明らかにする。

2. 1 RPA の概要

図 1 に RPA がターゲットとする業務プロセスを示す。従来、ホワイトカラーの業務効率改善は SE/SI が開発する Enterprise Resource Planning (ERP) や Customer Relationship Management (CRM) などの業務アプリケーションを導入することで達成されてきた。一方、既存業務アプリケーション間の情報転写、顧客からの問合せ対応、監査などの業務プロセスは人手によって達成されてきた。これは、これらの業務の実施頻度が相対的に低頻度である、あるいは業務プロセス自体が複雑（変更が多い、専門性が必要など）であり、経営者にとって業務アプリケーション導入の費用対効果が小さいため SE/SI による大きな IT 投資を正当化できないことに起因する[4][5]。RPA のターゲットはこのような従来 IT 投資をすることができず人手で実施してきた業務プロセスであり、特に API がない Web インタフェースの操作や OCR による文書情報取り込みの自動化を強みにしている。

図 2 に RPA の有無に対する業務プロセスの比較を示す。図 2(a) は RPA が無い場合の業務プロセスを示しており、業務アプリケーション 1 を参照して業務アプリケーション 2 に入力するなど、各業務アプリケーションをつなぐ作業を労働者が担っている。一方、図 2(b) は RPA を導入した場合の業務プロセスを示しており、各業務アプリケーションをつなぐ作業を業務ロボットが担う特徴がある。ここで、業務ロボットに実行させるスクリプトは、GUI ベースのツール[6][7]を用いて現場の労働者自身によって作成されることが大きな特徴であり、SE/SI への委託を必要としないため低コストに実現可能である。

このように、RPA は(1)従来 IT 投資をすることができず人手で実施してきた業務プロセスをターゲットとし、(2)現場主導で業務ロボットを開発/運用して業務効率を改善する特徴を有する。

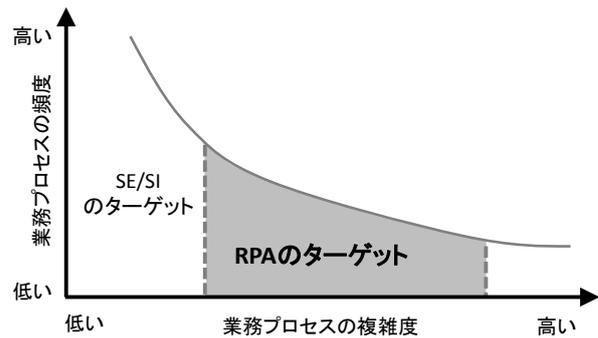


図 1: RPA のターゲット業務[4]

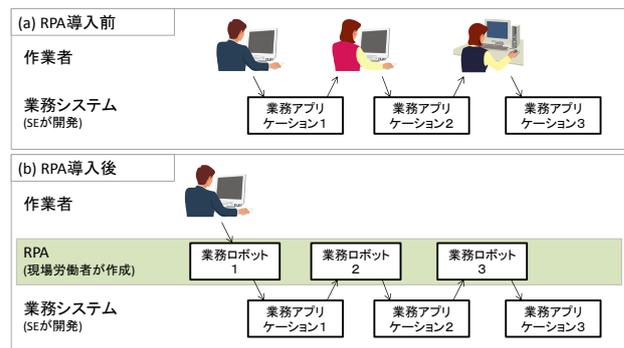


図 2: RPA 導入前後の業務プロセスの比較

2. 2 個人支援 RPA の位置づけ

図 3 に提案する個人支援 RPA の位置づけを示す。また、表 1 に現行 RPA と個人支援 RPA の比較を纏める。現行 RPA の主なターゲットは注文伝票処理、決済処理、顧客情報更新などバックオフィス業務の効率化である[8][9]。バックオフィス業務では、作業員がその業務プロセスをコア業務として毎日実施する専門家であり、その業務は始業時間から就業時間までの定期実行型のバッチ処理として実行され、さらにフロントオフィスからの申請手続きを集約して処理するため処理内容の偏りが小さくなる特徴がある。一方、個人支援 RPA の主なターゲットは旅費申請、資産棚卸、購買発注などホワイトカラーの特に申請/登録業務に関する非コア業務の効率化である。ホワイトカラーの非コア業務では、作業員がその業務プロセスを非コア業務として実施する非専門家であるため、本人にとって初めての業務や頻度が低い業務が発生した場合、コールセンターへの質問や知見がある同僚を探すなど時間損失が発生し、一つの業務手続きに大きな時間を費やすことがある。これを解決するためには非専門家を支援する質疑対応機能が効果的である。また、ホワイトカラーの非コア業務は作業員にとって必要なタイミングでインタラクティブ処理として実行され、その処理内容は個人の

属性や業務パターンによって大きく偏る特徴がある。そのため、このような業務には個人の属性や業務パターンを考慮した個人特化型の業務支援が必要になる。すなわち、バックオフィス業務をターゲットとする現行 RPA に対し、個人支援 RPA は次の 3 つの特性が必要になると分析した。

- (1) 非専門家を支援する質疑応答機能
- (2) 複数候補から入力値を一意に絞り込む対話型処理機能
- (3) 個人の業務パターンを反映する個人特化機能

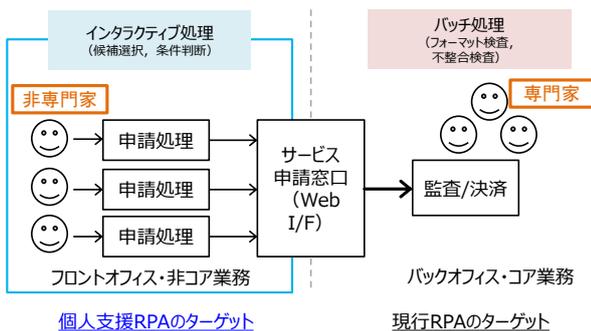


図 3: 個人支援 RPA の位置づけ

表 1: 現行 RPA と個人支援 RPA の比較

No	項目	現行 RPA	個人支援 RPA	個人支援 RPA への要求
1	作業者	専門家	非専門家	質疑応答機能
2	作業タイプ	バッチ	インタラクティブ	対話型処理機能
3	作業の偏り	小さい	大きい	個人特化機能

2. 3 課題抽出

個人支援 RPA を実現するためには 2.2 節で述べた 3 つの特性が必要になる。ここで、(1)質疑応答機能は 2011 年に「IBM Watson¹」が発表されて以来、その技術開発は大きな注目を浴びている[10]。しかし、実用に足る汎用的な質疑応答機能を実現するには極めて高度な AI 技術が必要であり、現在の技術ではこれを実現する方法は見つかっていない。また、作業者の入力作業量を削減する機能とは独立の機能要素である。そのため、本研究では(2)対話型処理機能と(3)個人特化機能を解くべき課題として以下のように捉え、これらを提供するシステムアーキテクチャを検討する。

¹ IBM Watson は International Business Machines Corp. の商標です。

課題 1 : 対話型処理機能

経路検索、商品検索などを含む業務アプリケーション実行のために、複数の候補から作業者が選択する仕組みを提供する。

課題 2 : 個人特化機能

個人の業務パターンを学習して業務アプリケーションへの入力値を推定するなど、現在の入力作業に比べて作業負荷を低減する仕組みを提供する。

3. アーキテクチャ検討

本章では対話型処理機能と個人特化機能を提供するアーキテクチャの構成とその適用例を示す。

3. 1 システムアーキテクチャ

図 4 に提案する個人支援 RPA のシステムアーキテクチャを示す。本アーキテクチャは対話ロボットと業務ロボットから構成される。近年急激に技術が進展している対話ロボットは個人支援 RPA の課題である対話型処理機能と相性が良い。さらに、様々なアプリケーションを統合した窓口として機能するため、個人の嗜好や行動パターンに関する情報を蓄積しやすく個人特化機能とも相性が良い。そのため、本システムでは対話ロボットをユーザインタフェースとして採用した。また、対話ロボットは一つ以上の業務ロボットと繋がり、ユーザとの会話から適切な業務ロボットを選択して処理を実行する。

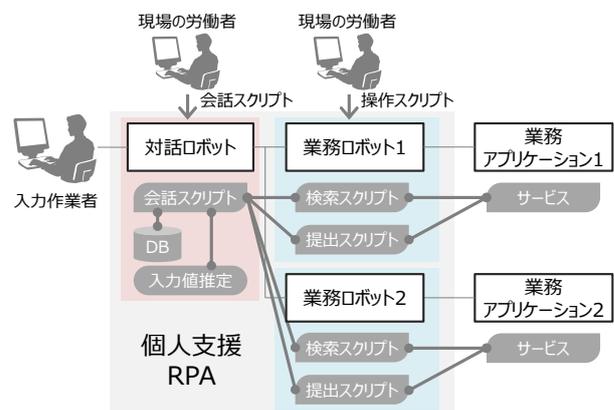


図 4: 個人支援 RPA のシステムアーキテクチャ

以下、対話型処理機能と個人特化機能を提供する方法および現場労働者の開発項目について詳述する。

(A) 対話型処理機能（入力候補絞り込み機能）

経路検索、商品検索などユーザとの対話的な意思

決定を含む業務アプリケーションに対応可能な対話型処理の方式を定義するために、社内業務アプリケーション（旅費精算、勤休管理、購買管理、資産管理など）の業務パターンと操作手順を分析した。その結果、図5に示すように対話ロボットは入力処理フェーズ、提出処理フェーズを持ち、これに対応して業務ロボットは業務アプリケーション上で入力候補を検索する検索スクリプト、最終提出する提出スクリプトを有することで個人支援 RPA に必要な対話型処理機能を実現できることが分かった。そこで、本アーキテクチャでは以下の方法で複数の候補から対話的に作業員が選択する仕組みを提供する。

入力処理フェーズ：

対話ロボットは複数候補が発生し得る業務アプリケーションへの入力要素について業務ロボットの検索スクリプトを実行し、得られた候補を作業員に提示して入力値を決定する。作業員が提示候補以外の値を入力した場合は再検索となる。

提出処理フェーズ：

対話ロボットは業務アプリケーションへの入力値に対する最終確認をユーザに求め、業務ロボットの提出スクリプトを実行する。

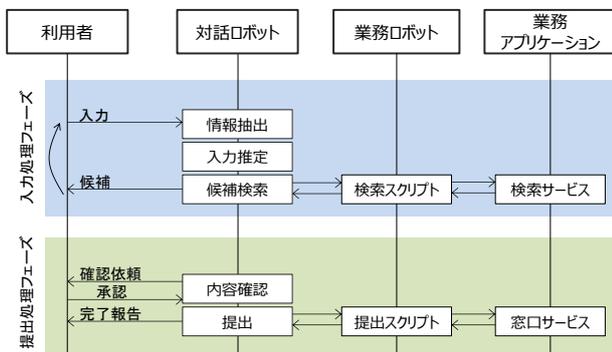


図5: 個人支援 RPA のシーケンス図

(B) 個人特化機能（入力値推定機能）

申請/登録業務に関する業務アプリケーションへの入力値推定に必要な情報を定義するために、社内業務アプリケーションへの入力値を分析した。その結果、個人情報、組織情報、環境情報、行動履歴情報に関する4つのデータベースを有することによって大部分の入力値を推定できることが分かった。そのため、対話ロボットは前記4つの情報と最終的に業務アプリケーションに提出した申請情報との関連に基づいて、次回以降の申請処理時の入力値を推定する。

個人情報：居住地、家族構成、所属、予定など作業員個人に関する情報

組織情報：自社、パートナー、顧客に関する拠点、組織体制、連絡先など組織に関する情報

環境情報：天気、気温、事故、イベントなど外部環境に関する情報

行動履歴情報：移動経路、申請内容、アプリ起動など作業員個人が起こしたアクションに関する情報

(C) 現場労働者の開発項目

上記機能を有する個人支援 RPA を導入するために、現場の労働者が開発すべき項目は次の2つに大別できる。

操作スクリプト：業務アプリケーションへの一連の入力操作手順を記述したスクリプト。

会話スクリプト：業務アプリケーションへ入力する情報を得るための会話手順を記述したスクリプト。必要に応じて各業務アプリケーションへの入力値推定の方法や業務アプリケーション間の連携を定義する。

現状、各ロボットのスクリプトを作成するツールとしてコード記述が不要な GUI ベースの作成環境が商用化されているが、その作成作業には技術コンサルタントの指導が必要である。しかし、近年、対話ロボットと業務ロボット(RPA)の技術は急激に発展しているため、現場の労働者のみでも容易に各ロボットのスクリプトを作成できる設計環境が今後整備され、この問題は徐々に解決されていくと予想する。

3. 2 適用例

提案した個人支援 RPA を社内の旅費精算へ適用した例を図6に示す。ここでは、旅費精算サーバにアクセスして操作する業務ロボット1と、社内のグループウェアサーバにアクセスして操作する業務ロボット2を用いた。旅費精算の経路検索では乗車駅と降車駅に対する複数の路線候補が出力されるため、業務ロボット1の検索スクリプトとして電車の経路検索を設定した。また、作業員の予定情報を用いて旅費申請への入力値を推定するために、業務ロボット2の検索スクリプトとして作業員の予定確認を設定した。

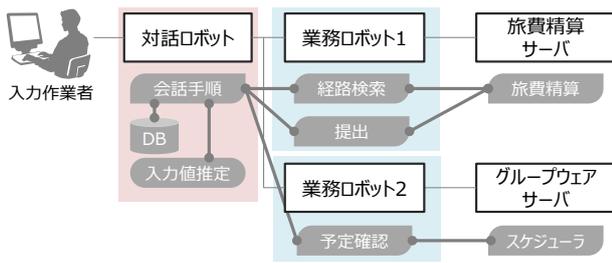


図 6: 個人支援 RPA の旅費精算への適用例

旅費精算の入力要素の一部とその入力例を表 2 に示す。本試作ではこれらの入力要素を対話ロボットを介して入力する。なお、電車移動経路の入力要素については、乗車駅と降車駅を入力すると詳細経路候補が作業者に提示され、作業者がその候補から経路を選択する仕様とした。

表 2: 旅費精算の入力要素

入力要素	入力例
出張日	2017/09/12
出張目的	案件 A の打合せ出席
用務先名称	中央研究所
都道府県名称	東京都
都市名称	国分寺市
用務内容	打合せ
電車移動経路	戸塚, 国分寺
出張経費	昼食
コメント	なし

本適用例で、入力値推定に用いる情報源と推定対象となる入力項目の対応を表 3 に示す。個人情報にはユーザの居住地、勤務地と予定の情報を含み、居住地と勤務地は経路検索の乗車駅と降車駅の推定に使用される。予定情報は「用務先名称」と「用務内容」の推定に使用される。組織情報は自社の事業所とパートナーの事業所の所在地情報を含み、経路検索の乗車駅と降車駅の推定および用務先の「都道府県名称」と「都市名称」の推定に使用される。入力値の過去履歴情報は過去に申請した移動経路情報を含み、乗車駅と降車駅を与えた時の詳細移動経路を推定する。詳細移動経路が正しく推定された時には対話ロボットは業務ロボットに経路検索を依頼する必要がなくなるため、処理時間を短縮できる。なお、本適用例では環境情報は利用していない。

表 3: 情報源と推定対象の対応

入力値推定のための情報源	推定対象
個人情報 (居住地, 勤務地, 予定)	電車移動経路, 用務先名称, 用務内容
組織情報 (事業所の所在地)	都道府県, 都市名称, 電車移動経路
行動履歴情報 (過去の電車移動経路)	電車移動経路

4. 有効性の検証

本章では、個人支援 RPA の試作システムを社内の旅費精算へ適用し、その業務効率改善効果を定量的に評価する。

4. 1 評価方法

本評価では、社内の旅費精算システムへの入力作業を作業者による直接入力と試作システムを介した入力についてそれぞれ実施し、その作業量を比較する。作業量の評価指標はキータ입数、マウスクリック数と作業時間であり、試作システムは 3.2 節に示した方式に準拠する。また、いずれの入力作業も Web ブラウザを介して行われる。そこで、Web ブラウザの起動、入力作業、終了までを 1 回の申請作業と定義した。また 10 秒間以上キー入力がない時間を別の業務をしている時間として作業時間から除外した。評価に使用したテストデータは一人の従業員が 1 年間に申請した旅費精算の入力データ 74 件分である。上記評価条件を表 4 に纏める。また、試作システム構築に使用したソフトウェアツール群を表 5 に纏める。本試作では対話ロボットに Hubot²[11], 業務ロボットには Windows Script Host³ (WSH)を使用した。ユーザと対話ロボットを繋ぐ対話プラットフォームには Rocket.Chat⁴[12]を使用した。

表 4: 評価条件

項目	条件
評価指標	キータ입数, クリック数, 作業時間
試作システムの方式	3.2 節に記載
作業時間の定義	Web ブラウザの起動から終了まで (但し, 10 秒間以上キー入力がない時間は除外)
テストデータ	過去申請データ 74 件

² Hubot は GitHub, Inc の商標です。

³ Windows は Microsoft Corp.の登録商標です。

⁴ ROCKET.CHAT は Engel, Gabriel の商標です。

表 5: 使用したソフトウェアツール群

構成要素	使用ツール
対話ロボット	Hubot[11]
業務ロボット	WSH
対話プラットフォーム	Rocket.Chat[12]

4. 2 評価結果

試作システムの評価結果を表 6 に示す。カッコ内の数値は Web インタフェースへの直接入力(従来法)に対する割合を示している。試作システムは直接入力に対し、キータイプ数を 54%削減、クリック数を 90%削減、作業時間を 57%削減する結果を得た。ここで、キータイプ数の削減は入力値推定機能(個人特化機能)による効果である。また、提案法では対話型処理機能を担う対話ロボットがチャット形式のインタフェースを提供するため、画面遷移などが不要となりクリック数が大幅に削減された。

なお、ホワイトカラー労働者一人当たりの非コア業務時間を平均 15 分/日とし、営業日数 240 日/年とすると、今回得られた 57%の作業時間削減は年間 34.2 時間の業務時間削減に相当する。

表 6: 評価結果

評価項目	Web インタフェース (従来法)	個人支援 RPA (提案法) [従来法比]
キータイプ数	8284	3798 [45.8%]
クリック数	2344	235 [10.0%]
作業時間 (sec)	9670	4068 [42.1%]

5. 結論

本報告ではホワイトカラー一人ひとりの業務効率を改善するため、非コア業務アプリケーションへの入力支援を行う個人支援 RPA を提案し、そのシステムアーキテクチャを示した。また、簡易試作を通じて提案 RPA が従来の直接入力に対し、作業時間を 57%削減可能であることを示し、その有効性を確認した。

本報告では、非常に簡単な入力支援機能を実装し評価した。今後は、より幅広い情報源と機械学習や確率統計モデルを導入し、より高度な入力支援機能を実装してその有効性を検証する。また、現場の労働者にとって業務ロボットの操作スクリプト作成はまだまだ障壁が大きい。そこで、業務アプリケーションの操作ログを活用し、労働者が業務を実施しているだけで自動的に操作スクリプトが作成できるなど、スクリプト自動作成技術についても検討する。

参考文献 :

- [1] 首相官邸, 働き方改革の実現,
<http://www.kantei.go.jp/jp/headline/ichiokusoukatsuyaku/hatarakikata.html>.
- [2] 厚生労働省,平成 28 年版 過労死等防止対策白書, 2016.
- [3] Global Market Insights, Inc., “Robotic Process Automation Market Size By Tools, By Operation, By Service, By Solution, By Organization Size, By Application, Industry Analysis Report, Regional Outlook, Application Potential, Price Trends, Competitive Market Share & Forecast, 2017 – 2024,” 2017.
- [4] Capgemini Consulting, “Robotic Process Automation - Robots conquer business processes in back offices,” 2016.
- [5] Leslie Willcocks, et.al, “Paper 15/05 The IT Function and Robotic Process Automation,” The Outsourcing Unit Working Research Paper Series, 2015.
- [6] RPA Technologies Inc., BizRobo!,
<https://bizrobo.com/products/series.html>.
- [7] David Chappell, “Introducing Blue Prism,” 2017.
- [8] Mary Lacity, et.al, “Paper 15/02 Robotic Process Automation at Telefonica O2,” The Outsourcing Unit Working Research Paper Series, 2015.
- [9] Thoughtonomy, Ltd., “Robotic Process Automation 6 real world use cases,” 2016.
- [1 0] David Ferrucci, et.al, “Watson: Beyond Jeopardy!,” Artificial Intelligence, volume 199-200, pages 93-105, 2013.
- [1 1] GitHub, Inc., Hubot, <https://hubot.github.com/>.
- [1 2] Rocket.Chat, <https://rocket.chat/>.

軌跡の直接操作に基づく時系列データの 視覚的分析インタフェースの提案

Proposal of Visual Analytics Interface for Time Series Data based on Direct Manipulation of Trajectory

高見 玲¹ 高間 康史¹
Rei Takami¹ Yasufumi Takama¹

¹ 首都大学東京 システムデザイン学部

¹ Faculty of System Design, Tokyo Metropolitan University

Abstract: Recently, various temporal data have been collected in many fields, and visual analytics interface is expected to be useful for utilizing such data. However, when temporal data is visualized using animation, collision would occur between movement of time series data itself and movement caused by interaction with users. This paper focuses on trajectory, which can handle temporal and spatial changes uniformly, and proposes a visual analytics interface based on it. This paper also shows a use case of applying prototype interface to time series data.

1 はじめに

情報技術の普及に伴い、近年、センサや SNS のストリームデータを代表とする時系列的特徴を持つデータが蓄積されるようになってきている。これらのデータを活用するために、クラスタリングやトレンド・外れ値の発見などのデータマイニング技術の活用が期待される。しかし、それらの技術を適用する前段階として、分析者自身によるデータの傾向や因果関係の解釈や、知識形成が重要である。この作業を支援するために、対象データの可視化に基づく視覚的分析システムの研究が行われている [1]。

視覚的分析において、可視化オブジェクトの直接操作とフィードバックにより、データに対する理解を促進し、知識形成を支援する手法が研究されてきた [2]。その中でも、背後に存在するモデルやパラメータを意識せず、可視化されたオブジェクトの直接操作によりモデルパラメータを修正し、知識形成を促進する Semantic Interaction が提案されている [3]。しかし、時系列データに対してこのような直接的アプローチを適用する場合、データ自身の変化とユーザ操作による変化の衝突や、異なる可視化手法間のトレードオフの存在などが問題点として想定される。衝突の例として、ズーム操作における、空間もしくは時間のいずれを対象としているかの解釈の困難性が挙げられる。トレードオフの例として、時系列データの可視化手法としてアニメーションや Small Multiple[4] が一般的だが、前者はデータの時間的変化を発見しやすい反面、探索時間の増大や見落とし

が生じる。一方で、後者では探索の時間と正答率に関しては有効であるが、スケーラビリティの問題がある。

上述の問題に対処するために、本稿ではデータの時間的変化を空間上に表現することで、操作との衝突を回避できる軌跡を用いた可視化に着目する。そして、軌跡の直接操作によりデータの時空間的変化の特徴把握を促進し、知識形成を支援する視覚的分析インタフェースを提案する。また、提案インタフェースをプロトタイプとして実装し、時系列データに適用した事例を示す。

2 関連研究

2.1 視覚的分析

一般に、データの分析者であるドメイン専門家は分析対象については豊富な知識を持つが、必ずしもデータ分析手法の専門家ではない。そのため、クラスタリングや次元削減アルゴリズム等の背後に存在するパラメータ調整の困難性が指摘されている [5]。この問題に対処するため、データを可視化して提示することで、ドメイン知識による解釈を行い、データに関する洞察を獲得するプロセスを支援する視覚的分析が研究されている [1]。視覚的分析のモデル化に関して、Sacha らはデータ探索、仮説検証、知識形成の 3 つのループからなるモデルを提案している [6]。分析者はデータ探索ループで可視化されたデータの特徴的な視覚的傾向に着目し、仮説検証ループで自身の有する知識に基づき視覚的傾向から仮説を検証、形成する。そして、知識形成ループで仮

説の検証と積み上げによりデータに関する知識を形成する。

また、視覚的分析におけるインタラクションのモデル化の例として、Endertらは可視化オブジェクト自体への直接操作、フィードバックに基づく非明示的な仮説形成ループ (Human is the loop) を提案している [14]。このモデルでは、直接操作の原則に従いインタラクションは逐次的 (例えば、検索結果などが即時フィードバックされる)、高速、かつやり直し可能である必要がある [7]。データオブジェクト自体への直接操作を行う手法として、Semantic Interaction のコンセプトが提案されている [3]。このコンセプトでは、システム側が可視化オブジェクトに対するユーザの操作内容を解釈し、その結果に基づき可視化の背後に存在する次元削減やクラスターリングアルゴリズムのモデル構成やパラメータを調整する。調整したモデル・パラメータを用いてデータを処理した操作結果はユーザに即時フィードバックされる。これによって、分析者がモデルに関する知識を有しない場合でも、データの操作、可視化から知識形成へのシームレスな移行が可能になるとしている。

2.2 時系列データの分析と可視化

時系列データの主な可視化手法として、時間軸を含む静的可視化のほかに、変化を移動として表現するアニメーション、時系列に従って複数のグラフを表示する Small Multiples、時間的変化を1つのグラフに重ねて表示する軌跡 (Trace) などが存在する [4]。アニメーションを用いて時系列データの分析を行う際に、ユーザはアニメーションで全体の傾向を掴んだ後に、重要だと判断した時点に対して UI を用いたインタラクティブな探索を行うことが示されている [8]。また、視覚的分析におけるユーザの認知特性に関して、インタラクティブ性は探索速度や正確さを失うことなくユーザエクスペリエンスを向上させることが示されている [9]。

軌跡による時系列データの可視化例として、Kondo, Collins はデータの変化軌跡をなぞる操作に基づく軌跡上の各時点への直感的な時間的ナビゲーションを提案している [10]。この手法によって、スライダなどの間接的なユーザインタフェースを使用せず、オブジェクト中心の直接操作による時系列データの探索が可能になるとしている。また、散布図に適用した場合には探索速度と正確性について既存手法に対する有効性が示されている [10]。また、Bachらは、時系列データの変化パターンの把握や、視覚的概要の提示を行う手段として軌跡に着目し、データの各時点における自己類似性を基準とした軌跡の描画手法を提案している [11]。これらの既存研究では、時系列データの変化特性の概要把握を支援するために軌跡を用いているが、より詳細なデータ分析

やデータマイニングへの応用に関しては考慮されていない。

3 時系列データの可視化における問題点

時系列データに対して、直接操作による視覚的分析を導入する際に、以下のような問題点が想定される。

3.1 時間軸上でのインタラクションの衝突

データの可視化手法にアニメーションを採用した場合には、視覚的分析におけるフィードバックに伴うオブジェクト配置の変化と、アニメーションの再生に伴うオブジェクト配置の変化は共に時間軸上で発生する。この様に、ユーザが行うインタラクションに関する時間軸と、時系列データが内包する時間軸の衝突が生じる。これに伴い、操作結果とデータ本来の時系列変化の混同が生じ、ユーザの混乱の原因となりうる。

3.2 時空間軸間でのインタラクションの衝突

時系列データを分析する際にユーザが行う操作は、オブジェクトの配置変更など空間に対し行われる他、再生速度の変更など、時間軸に対しても行われる。その結果、操作適用対象に関して時間および空間軸間の衝突が生じる。例えば、時系列データに対して詳細を確認するために拡大 (ズームアップ) 操作を行った場合、空間的に拡大表示したいのか、再生速度を遅くしたいのか、二通りの解釈がありうる。採用された解釈により、望まない変更が行われた場合には分析の妨げとなりうる。

また、時系列データの変動をアニメーションの様に時間軸の変動として描画するモード、折れ線グラフの様に空間的配置により描画するモードを用意し、相互に切り替え可能なシステムも考えられる。このとき、モード間の遷移をユーザの操作を解釈により暗黙的に行う場合には、ユーザは自身が現在どちらのモードに存在するかを判断しづらくなる。一方で、明示的に移行する場合、必要なインタラクション数が増加し、ユーザビリティ低下の原因となりうる。

この他にも、Yiらのユーザ要求に基づくインタラクション分類 [12] の内の以下においても、時空間軸間でのインタラクションの衝突が発生すると考える。

- Select: 特定対象のマーク
- Abstract/Elaborate: 可視化の詳細度変更
- Filter: 条件に合致するオブジェクトを表示
- Connect: 関連アイテムを表示

3.3 可視化手法間のトレードオフ

時系列データのアニメーションを用いた可視化は、聴衆にストーリーを理解させるプレゼンテーションのように変化を発見、理解する場面では有効といえる。しかし、データの時空間的な変化傾向の把握に用いた場合、アニメーションの再生/静止を繰り返すによる認知的負荷の増大や、探索の正確性や速度の低下が問題点として報告されている [4]。一方で、軌跡 (Trace) や Small Multiple は探索の正確性に関してはアニメーションと比べて有効である反面、前者は多数のオブジェクトが存在する場合に重なりにより探索が困難となる。また、後者の場合には占有する画面領域が広くなり、探索時間や認知的負荷が増大する問題点が指摘されている [4]。

このように、時系列データに対して一意に最適な可視化手法は存在しないため、分析目的や対象に応じて複数の手法を組み合わせることが有効だといえる。

4 提案手法

前述の背景と問題点を踏まえ、本稿では、時系列データを二次元散布図および軌跡として可視化し、直接的操作による視覚的分析を通じた知識形成を支援する可視化インタフェースを提案する。提案インタフェースが備えるべき要件として、以下に着目する。

1. 従来の視覚的分析手法との互換性 (時空間的な変化特性の把握)
2. 時間軸上でのインタラクションの衝突 (3.1 節) の回避
3. 時空間軸間でのインタラクションの衝突 (3.2 節) の回避
4. 可視化手法間のトレードオフ (3.3 節) への対応

1 点目については、散布図、軌跡による可視化や、詳細ビューといった、可視化手法として一般によく用いられるものを採用する。2 点目は、軌跡を用いて、時空間的な変化を空間上に表現することで対応する。3 点目については、軌跡表現により両軸への操作を統一することで対応する。加えて、一部の明示的に分離すべき操作に関しては探索モードの導入により対応する。4 点目に関しては、探索モードの切り替えにより各可視化手法の利点を生かすことで対応する。

4.1 対象データ

分析対象とする時系列データの種類として、以下を想定する。

- 数値的データ: ヒストグラムや円グラフなど。汎用的な可視化手法を適用可能であり、解釈が容易。
- 集合/関係的データ: テキストストリームやネットワーク。そのまま数値的に表現できないため、類似度関数などの距離メトリクスを用いて可視化される [2]。
- 複合的データ: 複数の数値的データや集合/関係的データから構成されるデータ。

集合/関係的データや複合的データを対象とした場合でも、テキスト解析や次元削減などのデータマイニング技術を適用することで平面などの低次元空間上の座標として描画可能である。そのため、提案インタフェースでは、可視化手法として散布図を採用する。また、座標空間上の近接性が類似性を示すことや、モデルパラメータの調整や値の推定に関するタスクについて、長さや位置、距離エンコーディングが面積や濃度と比較してタスクの正確性や達成速度に関して有効であることが報告されている [13]。そのため、散布図はデータの時空間的特性の把握や、視覚的傾向に基づく知識形成の点でも適していると考えられる。

4.2 軌跡を用いた可視化とインタラクション

提案インタフェースでは、時系列データの散布図による可視化に加えて、データの時空間的な変化を軌跡として表現する。これにより、時間的な変化を空間上のみで表現可能となり、ユーザとのインタラクションに伴う変化との衝突を防げると考える。また、時系列データを二次元平面上で扱えるため、時間および空間軸に対する時間的的操作を統一できる。そのため、Select、Filter 操作などについてインタラクションの衝突を回避する効果も期待できる。時系列データの軌跡を用いた可視化の例を図 1 に示す。軌跡は対象データの各時点における属性値の二次元空間座標間を結ぶ曲線により描画する。このとき、各点間の補完に cubic Catmull-Rom Spline 曲線 [15] を用いる。

軌跡は、散布図上のオブジェクトへのマウスオーバー操作に基づき描画される。また、軌跡を構成する各点のマウスオーバーにより、その時点での散布図が図 2(2) のように薄く重畳表示される。時間的な文脈を失わずに、他時点データ分布を確認できる。その状態でクリック操作を行うことで、図 2(3) のようにその時点への時間的遷移が実行される。

また、軌跡のパスをクリックすることで、軌跡の選択 (画面上への固定) が行える。選択された軌跡は赤色でハイライトされる。これによって、データの時系列的変化による見落としを軽減できると考える。また、図 3 に示すように、データが線形的に増加した場合には軌跡は



図 1: 軌跡による可視化

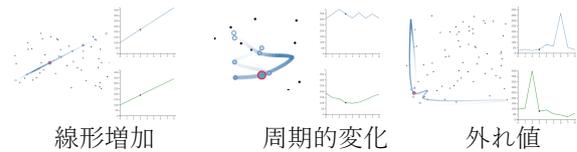


図 3: 軌跡の幾何学的形状の例

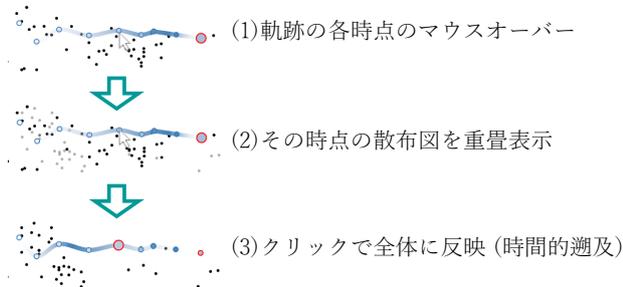


図 2: 軌跡操作に基づく時間的ナビゲーション

直線として現れる。水平あるいは垂直軸のどちらかで周期的に変化した場合には、その軸において往復するような軌跡が、外れ値を含む場合には広域に渡り急峻な変化をする軌跡が描かれる。このように、軌跡の形状はデータの時間的変化に対応し、データの視覚的概要の提示手段として有用であるといえる。

4.3 探索モード

アニメーションは時系列データの変化の発見や理解について有効だが、操作数増加などの原因により探索に関しては不向きである。一方で、軌跡や Small Multiple による可視化は探索効率と正確性に関しては有効だが、スケーラビリティの問題が存在する。

本稿では両者の利点を活かすために、アニメーションと静止の二つの探索モードの導入によるインタラクションの分岐を行う。以下に、各モードにおけるユーザの分析行動(探索目的や着目軸)の例を示す。

- アニメーション: 時系列データの概要としてアニメーションを再生し、特徴的な変化を検出したり、データの時間軸上の変化特性に基づき、興味のあるオブジェクトを選択、追跡する。
- 静止: アニメーションで見出した概要に基づき、空間軸上にて軌跡の直接操作による詳細な探索や視覚的発見の検証を行い、知識を形成する。

ここで、探索はアニメーションモード(視覚的傾向の発見)と静止モード(仮説検証)間の遷移の繰り返しで行われると仮定する。表 1 に探索モードによる操作とインタラクションの対応関係を示す。モードの変更により、暗黙のうちに探索目的や着目軸に対応したインタ

表 1: モードによる操作とインタラクションの対応関係

操作	モード	適用インタラクション
ズーム	アニメ	時間的(再生速度の)ズーム
	静止	空間的ズーム
データ選択	アニメ	データ点のトラッキング
	静止	軌跡の集合、離散の表示

ラクションを選択・適用することで、操作を複雑にすることなく両可視化を活用可能である。また、モードの導入に伴う着目軸の分離によって Abstract/Elaborate 操作などにおける時空間的なインタラクションの衝突を回避できると考える。

4.4 オブジェクト/軌跡のグルーピング

提案インタフェースでは、図 4(a)(b) のようなスケッチベースの入力に基づくオブジェクトの選択とグルーピング、追跡が可能である。

このとき、複数オブジェクトを選択すると、それらを含む凸包を画面上に描画する。図 5 に示すように、凸包の面積はデータ間の空間的な近接性(類似性)に対応する。つまり、データ点同士の凸包は各時点でのデータ群間の類似性を表現し、軌跡同士の凸包はデータ間の時空間的類似性を表現する。複数オブジェクトに対する選択処理に対して、アニメーションの再生時には、その時点での選択済みデータを含む凸包を計算し、描画する(図 4(a))。これにより、着目データ群の時間的な集合/離散を確認できる。また、選択済みデータは色と大きさでハイライトされる。そのため、アニメーションによる位置の変化や、動くデータ点への追従に伴う認知的負荷を軽減できると考える。

対して、静止モード時には、選択されたオブジェクト群の全時点におけるデータ座標に対して凸包を計算、表示する(図 4(b))。また、複数軌跡に対して選択操作を行った場合にも、同様の凸包を描画する(図 4(c))。前者では、空間座標の範囲に基づく対象データの暗黙的な指定により、時間軸全体にわたるデータ群の変化傾向の把握が促進されると考える。一方で、後者では明示的に指定した軌跡のグループに対して凸包を重畳することで、軌跡間の空間的座標の近接性(類似性)を直感的に把握可能になる。これによって、分析者が興味を持ったデータの詳細な探索を支援できると考える。

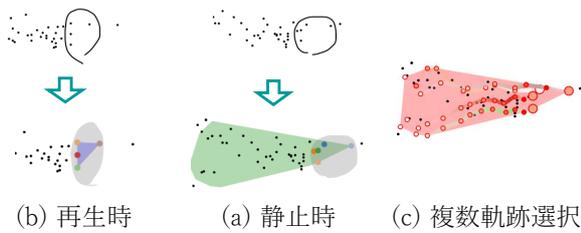


図 4: オブジェクト, 軌跡のグルーピングと追跡
 (灰色はスケッチされた領域)

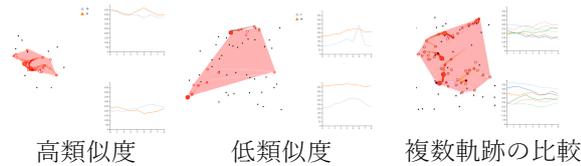


図 5: 凸包による類似性の表現

4.5 相補的なビュー

前述のように, 軌跡による可視化は時系列データの概要の提示手段として有効な一方で, 描画時の曲線補完などによりデータの正確性は喪失する. また 散布図による可視化は, 複数データ間の類似性や関係の判断に関しては有効だが, 視覚的傾向に基づく仮説検証においてはツールチップ等を用いて生データを提示した場合の方が有効であることが示されている [9]. そのため, 提案インタフェースでは散布図ビューに関する詳細情報を提示する詳細ビューを提供する. すなわち, システムは散布図と詳細ビューの協調的マルチビューによって構成される. 詳細ビューでは, 散布図ビューで選択されたオブジェクトに対する操作に基づき, 対応する各属性値の時間的変化を線グラフにより表示する. このとき, 一方のビューへの操作は他方へも適用され, インタフェースの一貫性を保持する. 例えば, 複数軌跡を選択した場合には, それぞれの軌跡における各属性値の比較グラフが詳細ビュー上に表示される. また, 散布図ビュー上で特定の軌跡に対してマウスオーバー処理を行ったときには, 対応するグラフがハイライトされる.

5 ユースケース

提案インタフェースの有効性検証のために, プロトタイプインタフェースを用いたユースケースを示す. インタフェースは Web アプリケーションとして実装し, フレームワークには Ruby on Rails 5, マークアップ言語には HTML と CSS, 可視化とインタラクションの実装に jQuery および D3.js を用いた.

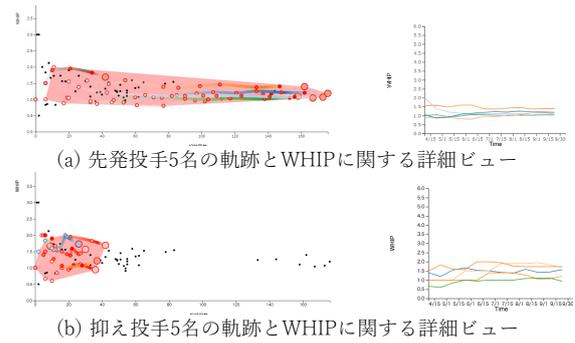


図 6: 先発投手と抑え投手における視覚的傾向の相違

分析対象とする時系列データにはプロ野球マルチデータ¹で公開されている 2017 年日本プロ野球パリーグの登板数が 10 試合以上の投手 103 名の投手成績を 15 日間隔でサンプリングしたものを使用した. ここで, 散布図の各軸には WHIP(投球回あたりの与四球と被安打数の合計) および投球回数を使用した.

5.1 投手の評価指標検討のための使用例

提案インタフェースの使用例を, プロ野球における投手成績の評価をタスクとした場合の例を示す.

最初に, 分析者はデータの概要を把握するため散布図ビューを静止モードで各時点について閲覧したり, アニメーションを用いて全体的な傾向の把握を試みた. この時, 投球回数, WHIP をそれぞれ横軸, 縦軸にとった散布図において, 投球回数の多い先発投手と, 少ない中継ぎ投手では空間的位置に明確な違いが見られた. 次に, 分析者は得られた概要や自身のドメイン知識に基づき, 静止モードにて軌跡の探索を用いてシーズンで優良な成績を残した先発/中継ぎ選手の傾向を探索した. その結果, 図 6(a) に示すように, 良い成績を残した先発投手は, 早期に WHIP が特定の値に収束したのに対して, 抑え投手の場合は図 6(b) のように軌跡の変動に個人差が観測された.

その後, 軌跡の選択によるグルーピングなどを行い分析を進めた結果, 同様の最終成績を持つ投手同士でも, 成績の収束時期やパターンに差があることを発見した. 最終的に, 探索で得られた視覚的傾向や洞察に基づき, 分析者は WHIP が早期に収束し, 投球回数が多い先発投手を安定感があるとして評価した. 一方で, 抑え投手に関して, WHIP は投手ごとの傾向の違いは確認できるが, 時系列的に共通する傾向は確認できないため, その他の評価指標を検討すべきであると結論付けた.

このように, 提案インタフェースを利用することで, 時系列データの視覚的分析を支援できると考える.

¹<http://lcom.sakura.ne.jp/NulData/>

6 おわりに

本稿では、時系列データの直接操作による視覚的分析を対象とし、時間軸上のインタラクションの衝突などの問題点解決を目的として、軌跡の直接操作に基づく視覚的分析インタフェースを提案した。また、提案インタフェースをプロトタイプとして実装し、利用例を示した。

今後は、ユーザビリティテストによる提案インタフェースの定量的・定性的な有効性の検証を行う必要がある。また、今後の展望として、知識形成促進のための Semantic Interaction への応用や、データマイニング技術を活用した可視化の品質保証やインタラクション解釈の不確かさへの対処、操作のガイダンス [5] による探索の支援などが挙げられる。具体的には、大規模データセットへの拡張を踏まえた DBSCAN などのクラスタリング手法や、軌跡の形状に基づくハイライトやフィルタリング、類似性判定などの導入を検討している。

参考文献

- [1] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, H. Ziegler, Visual analytics: Scope and challenges, S. J. Simoff, M. H. Böhlen, A. Mazeika (eds.), Visual Data Mining, Springer, pp. 76-90, 2008.
- [2] A. Endert, L. Bradel, C. North, Beyond control panels: Direct manipulation for visual analytics, IEEE computer graphics and applications, Vol. 33, No. 4, pp. 6-13, 2013.
- [3] A. Endert, P. Fiaux, C. North, Semantic interaction for visual text analytics, Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 473-482, 2012.
- [4] G. Robertson, R. Fernandez, D. Fisher, B. Lee, J. Stasko, Effectiveness of animation in trend visualization, IEEE Transactions on Visualization and Computer Graphics, Vol. 14, No. 6, pp. 1325-1332, 2008.
- [5] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, D. A. Keim, What you see is what you can change: Human-centered machine learning by interactive visualization, Neurocomputing, 2017.
- [6] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, D. A. Keim, Knowledge generation model for visual analytics, IEEE Transactions on Visualization and Computer Graphics, Vol. 20, No. 12, pp. 1604-1613, 2014.
- [7] B. Shneiderman, Direct manipulation: A step beyond programming languages, IEEE Computer, Vol. 16, No. 8, pp. 57-69, 1983.
- [8] S. Kriglstein, M. Pohl, C. Stachl, Animation for time-oriented data: An overview of empirical research, 16th International Conference on Information Visualisation (IV '12), pp. 30-35, 2012.
- [9] M. Adnan, M. Just, L. Baillie, Investigating time series visualisations to improve the user experience, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5444-5455, 2016.
- [10] B. Kondo, C. Collins, DimpVis: Exploring time-varying information visualizations by direct manipulation, IEEE Transactions on Visualization and Computer Graphics, Vol. 20, No. 12, pp. 2003-2012, 2014.
- [11] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, P. Dragicevic, Time curves: Folding time to visualize patterns of temporal evolution in data, IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No. 1, pp. 559-568, 2016.
- [12] J. S. Yi, Y. A. Kang, J. Stasko, J. Jacko, Toward a deeper understanding of the role of interaction in information visualization, IEEE Transactions on Visualization and Computer Graphics, Vol. 13, No. 6, pp. 1224-1231, 2007.
- [13] B. Saket, A. Srinivasan, E. Ragan, A. Endert, Evaluating interactive graphical encodings for data visualization, IEEE Transactions on Visualization and Computer Graphics, 2017.
- [14] A. Endert, M. S.Hossain, N. Ramakrishnan, C. North, P. Fiaux, C. Andrews, The human is the loop: new directions for visual analytics, Journal of intelligent information systems, Vol. 43, No. 3, pp. 411-435, 2014.
- [15] C. Yuksel, S. Schaefer, J. Keyser, Parameterization and applications of Catmull-Rom curves, Computer Aided Design, Vol. 43, No. 7, pp. 747-755, 2011.

深層学習による日本語キャプション生成システムの開発

Development of Japanese caption generation system by deep learning

小林豊¹ 鈴木諒¹ 谷津元樹² 原田実²

Yutaka Kobayashi¹, Ryo Suzuki¹, Motoki Yastu², Minoru Harada²

¹ 青山学院大学大学院 理工学研究科

¹ Graduate school of Science and Engineering, Aoyama Gakuin University

² 青山学院大学理工学部情報テクノロジー学科

² Department of Integrated Information Technology, College of Science and Engineering,
Aoyama Gakuin University

Abstract: For the purpose of developing a dialogue system to dialogue after visually understanding the surrounding situation. We developed Japanese Caption generation system *Deep Watcher* and image datasets with captions. We used the Show and Tell model using CNN and LSTM to generate captions. We also evaluated the coincidence rate of caption content and five feature items manually. As a result the coincidence rate of the contents of the generated caption was 41.6%, the highest characteristic item was gender and was 86.9%. The coincidence rate of the caption contents were not high by over learning, but we could show the possibility of application to the dialog system for the feature item of gender.

1 はじめに

近年、深層学習を用いた画像キャプション生成が発展しており[1,2], コンピュータに入力された画像に映る状況を説明文(キャプション)として出力することが可能となっている。例として、MicrosoftのWeb上で画像のキャプションを生成するボットCaption Bot¹が挙げられる。

深層学習を利用した画像キャプション生成技術を、周辺状況を視覚的に理解した上で対話する対話システムに応用することを考えた。これは、対話システムが対話相手の性別や人数、服装、持ち物、動作などを中心に対話に応用するというものである。また、過去の深層学習によるキャプション生成の研究[2,3]では、英語のみを使用しているので、日本語キャプションの生成が可能かどうかを試みる為に、日本語キャプション生成システムDeep Watcherを作成した。対話内容に応用することができるキャプション生成について調査を行った。

本研究では日本語でのキャプション生成を考える。しかし、既存のデータセット[4,5]では、付与されているキャプションは英語のみである。また、本目的に沿っていない画像も含まれてしまうため新たに人物を中心とした日本語キャプションが付与され

た学習データセットを作成した。

2 Deep Watcher

この節では作成したキャプション生成システムDeep Watcherについて説明する。Deep Watcherの実行イメージを図1に示す。Deep WatcherはGUIを用いたシステムであり同時に複数の画像についてキャプション生成を行うことができる。

2.1 キャプション生成モデル

Deep Watcherには、Show and Tell Model[3]をキャプション生成モデルとして採用する。このモデルは畳み込みニューラルネットワーク(CNN)[6]と長短期記憶ニューラルネットワーク(LSTM)[7]の組み合わせで構成されている。モデルの構成を図2に示す。入力画像からCNNによって画像の特徴量を抽出したものと、画像の説明文をWord Embeddingにより単語ベクトルに変換したものをLSTMに入力して、次の単語の出現確率を計算し、最大確率の語をつなげて画像のキャプションを生成する。

モデルの構築には、Deep LearningのフレームワークであるChainerを用いる。CNNではILSVRC-2014

¹ <https://www.captionbot.ai/>

model with 19 weight layers(VGG19) の学習済みモデルを用いる。LSTM では、本研究で作成した学習データセットの訓練データから言語モデルを生成する。



図 1 Deep Watcher の実行イメージ

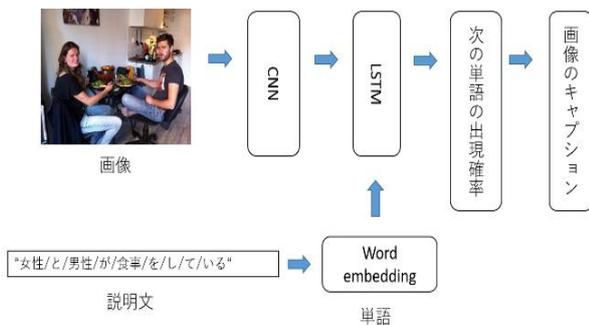


図 2 キャプション生成モデルの構成

2.2 モデルの学習

モデルの学習には、学習データに画像 1 枚に対し 5 文の日本語説明文をペアとして使用する。

順伝播では、まず画像を CNN に入力し画像特徴量を抽出し、それを LSTM に入力する。次に、日本語説明文の単語を Word Embedding に入力し単語ベクトルを生成し、それを LSTM に入力して次の単語の単語出現確率を出力する。

誤差逆伝播法を用いてパラメータの調整を行うため Softmax-cross entropy 関数を用いて、次に LSTM に入力する単語と単語の出現確率から誤差を算出する。誤差を LSTM と Word Embedding の各層に伝播させて、勾配降下法に基づいてパラメータを調整する。順・逆伝播を文末の単語まで繰り返し、学習させていき、テキスト及び画像を入力とする言語モデルを生成する。

入力には、CNN を用いて画像から抽出した画像特徴量ベクトル(4096×2000)と、日本語説明文の各単

語(1×単語種類数次元の one-hot ベクトル)を Word Embedding により変換させた分散表現ベクトルを用いる。出力では、各単語の出現確率を持つ単語出現確率ベクトル(1×単語種類数次元)を得る。

2.3 キャプション生成

まずキャプション生成させたい画像を CNN に入力し画像の特徴量を抽出して LSTM に入力する。次に Word Embedding を用いて文開始記号<S>を単語ベクトルにしたものを LSTM に入力し、LSTM の出力を Softmax 関数に入力し、単語出現確率ベクトルを求め、文開始記号<S>における次の単語の出現確率を求める。BeamSearch アルゴリズムより、単語の出現確率が高い上位 M 位と確率値の対数を保持する。次に単語出現確率が上位 M 位の単語 M 個を、LSTM に入力し出力を Softmax 関数に入力し、得られた確率を対数に変換後、和を計算して M×M 個の単語列を求める。そして、再び BeamSearch アルゴリズムより、単語列の出現確率の和が高い上位 M 位と確率値の対数を保持する。次に単語列の出現確率が高い上位 M 位の単語列の最後の単語を LSTM に入力し、上記と同じ処理を繰り返す。最後に、BeamSearch アルゴリズムより求めた、出現確率の和の平均が高い上位 M 位の単語列全てに文終端記号</S>が出現したら処理は終了する。作成された M 個の単語列の中から確率の和が高い上位 5 件をキャプション生成として出力する。



図 3 学習データセットの画像例

女性/と/男性/が/食事/を/し/ている
 女性/は/左側/にいる
 男性/は/椅子/の/上/に/胡座/を/かいて/いる
 女性/は/右手/に/フォーク/を/持っ/ている
 顎髭/を/生や/した/男性/が/いる

図 4 学習データセットの日本語説明文の例

3 学習データセット

本研究で作成した学習データセットについて説明する。作成したデータセットは画像とその画像の日本語説明文から構成されている。

3.1 画像

画像は、人物が中心に映っているものを MPII Human Pose Dataset より 1,400 枚, 室内で人物が中心に映っているものを MS-COCO より 600 枚の計 2,000 枚を人手により選択した。画像のサイズは最大 1,920×1,080 画素から最小 300×168 画素と制限はない。学習データセットの画像の例を図 3 に示す。

3.2 日本語説明文

日本語説明文は、計 20 名で作成した。1 人 1 画像に対して 5 文を担当し、何をしているのか、持ち物は何か、何を着ているのか、人物の数、性別、場所に着目して 1 人あたり 100 枚分作成し合計 10,000 文を作成した。

また、考案した日本語説明文は、日本語形態素解析²:無料 WEB 便利ツール²により、形態素解析を行い、形容詞、形容動詞、感動詞、副詞、連体詞、接続詞、接頭辞、名詞、動詞、助詞、助動詞、特殊の 12 の形態素ごとに分かち書きを行った。分かち書きされた語を 1 単語単位として使用する。日本語説明文の例を図 4 に示す

4 実験

4.1 ハイパーパラメータ

実験を行う前に、より精度の良い文章を生成するためのハイパーパラメータとデータセットの調整を行った。

ハイパーパラメータの調整には、3 節で作成した学習データセット(画像:2,000 枚, 日本語説明文:10,000 文)を使用した。ハイパーパラメータは、学習に使用する単語の最低出現回数の `Word_mini_count`, 勾配降下法の最適化, 過学習抑制の `DropOut`, LSTM の入出力と `Word Embedding` の出力との共通した次元数の `com_dim`, ミニバッチの学習サンプルのサイズ数の `batch_size` の 5 つに関して調整を行い、データセットでは、データセット内の訓練データとテストデータの比率を調整した。

調整の結果、本実験では、各ハイパーパラメータの値は、`Word_mini_count=1`, 最適化は `Adam`, `DropOut=0.5`, `com_dim=512`, `batch_size=256` とし、データセットに対しては、訓練データを 1,200 枚, テストデータを 600 枚(2:1)に振り分けたものを使用した。

4.2 実験設定

今回の実験では、3.1 節と同条件で新たに 100 枚の画像を集め、100 枚の画像に対して 1 画像につき 5 文, 合計 500 文のキャプションを生成し、以下の 2 つの評価実験を行った。評価実験に使用する画像は、学習データセットに使用した画像と同様に集めた。

● 内容一致

キャプション生成システム `DeepWatcher` で生成した文章と画像の内容がどのくらい合っているかを「完全に合っている」「一部分合っている」「間違っている」の 3 項目に第 1 著者が人手で分類し評価を行った。

● 特徴一致

キャプション生成システムで生成した文章が画像に写っている人間の、性別、人数、持ち物・容姿、動作、場所の特徴を出力することができるのか、正しく出力しているかを第 1 著者が人手で次のように分類し評価を行った。

① キャプション生成システムの生成した各文において、人物に関する性別、人数、持ち物・容姿、動作、場所についての特徴の出現頻度を数える

② ①の特徴のうち画像と特徴が一致した正しいものの数を数え、その割合を正解率とする。例えば「男性と女性が食事をしている」という文章では性別について出力された特徴数は「男性」と「女性」という単語から 2 つ。動作については「食事をしている」から 1 つ。持ち物・容姿、場所については特徴の出力なしとし、画像とキャプションの特徴が一致したものを数える。

² <http://tools.metro-bb.com/api/keitaiso>



女性はキッチンにいる
 彼女はキッチンで土台の焦げを落としている
 彼女は右手にスプーンを持っている
 彼女はカーディガンを羽織っている
 彼女は誕生日ケーキの蝋燭に息を吹き掛けようとしている

図5 テストデータ内の画像およびキャプションの具体例

4.3 実験結果

図5に生成されたキャプションの具体例を示す。

- 内容一致
 表1に結果を示す。

Deep Watcherにより生成されたキャプション自体は自然な日本語文という印象を与えるものであった。しかし文章の内容としては500文中、完全に合っていたものが137文、一部合っていたものが71文、間違っているものが292文となり41.6%が内容に合っている又は一部内容に合っており、58.4%が内容と間違っているキャプションを生成した

- 特徴一致

表2に結果を示す。

性別の特徴は90%以上、動作の特徴は60%以上の文で出力されたが、人数、持ち物・容姿、場所については、50%を下回り、出現頻度は低かった。正解率としては性別、場所、人数、行動の順に高く、持ち物・容姿が一番低く45%となった。生成されたキャプションの具体例のように性別について5文出力している画像が63枚あれば、性別、人数、持ち物・容姿、行動、場所すべての特徴を1つ以上出力した画像は3枚あった。

表1 内容一致の評価実験

評価項目	一致率
完全に合っている	137(27.4%)

一部合っている	71(14.2%)
間違っている	292(58.4%)

表2 特徴一致の評価実験

特徴	出力された特徴数	正解数	正解率
性別	459	399	86.9%
人数	107	66	61.6%
持ち物・容姿	201	77	38.3%
動作	329	153	46.5%
場所	72	49	68.0%

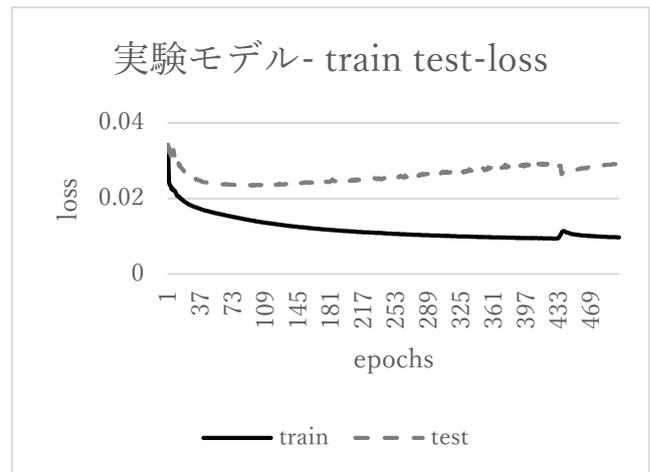


図6 実験モデルのepoch数によるtrainとtestのloss値の変化

5 考察

Deep Watcherが生成したキャプションは、文章と画像の内容が一致する確率が低く、Deep Watcherには改善の余地がある。しかし、「カヌー」、「釣り」、「キッチン」、「体育館」など正確に画像の特徴をとらえているものもあった。

キャプションの内容の正確性を阻んでいる要因の一つに、過学習が考えられる。図6に実験モデルのtrainとtestのloss値を示す。図6より、trainとtestのloss値が乖離しているため過学習を起こしていたと考えられる。評価画像のうち水辺の画像は10枚(50文)であり、内容に合っている・一部内容に合っている割合は、水辺の画像が52%、水辺以外の画像は、40.4%と11.6%の差があったため、水辺の画像に対しての特徴を強く学習してしまい過学習が起きたといえる。

特徴一致では多くの文に性別について出力された。これは学習データセットのキャプションのほとんどに性別についての記述があったためであると考えら

れる。学習データセットのキャプション 10,000 文中性別の記述があったものは 8,180 文であった。

6 おわりに

本研究では、周辺状況を視覚的に理解した上で対話するロボット搭載型対話システムの開発を目的として、日本語キャプション生成システム Deep Watcher を開発した。また対話に応用できるキャプションについて調査を行った。

今後の課題は、キャプション生成した文章と画像の内容がどのような画像を選んだ場合にも合うように、また正確に特徴をとらえるように、さらなる文章出力精度の向上が挙げられる。そのためには、学習データセットの増加による過学習の改善と CNN に対して、より特徴量の認識精度が高い Deep Residual Learning(ResNet)[8]による画像特徴量の抽出を検討する。また人物が中心的な周辺状況を視覚的に理解する対話システムの開発目的に沿うよう、既存の学習済み CNN を使用するのではなく、人物認識に特化・限定した独自の学習済み CNN の作成も検討する。

謝辞

データセット構築に協力していただいた青山学院大学の学生有志に感謝致します。

参考文献

- [1] A.Farhadi,M.Hejrati,M.A.Sadeghi,P.Young,C. Rashtchian, J. Hockenmaier, and D. Forsyth. Every picture tells a story: Generating sentences from images. In ECCV. 2010.
- [2] A.Karpathy,L.Fei-Fei:Deep Visual-Semantic Alignments for Generating Image Descriptions. In CVPR,pages 3128-3137, 2015.
- [3] O.Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell:A neural image caption generator. In CVPR,pages 3156-3164, 2015.
- [4] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollar,and C.L.Zitnick.Microsoft COCO captions: Data collection and evaluation server. arXiv preprint,1504.00325,2015..
- [5] P. Young, A. Lai, M. Hodosh, and J. Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. Transactions,of,the Association for Computational Linguistics,2:677-78,2014.

- [6] K.Simonya and A.Zisserman:Very Deep Convolutional Networks,for, Large Scale,Image,Recognition.arXiv:1409.1556.2014
- [7] F. A. Gers, J. Schmidhuber, and F. Cummins.Learning to forget: Continual prediction with LSTM. Technical Report IDSIA-01-99, IDSIA,Lugano, CH, 1999.
- [8] H. Kaiming, Z.Xiangyu, R.Shaoqing, and S.Jian . "Deep Residual, Learning, for, Image, Recognition." In CVPR, pages 770-778, 2016.