Twitter 発言の分析による Web サービス障害の影響調査

プロジェクトマネジメントコース　矢吹研究室　1442012　岩瀬翔

1. 序論

複数のメンバが同時に開発を行うソフトウェア開発プロジェクトにおいて，「GitHub」のような Web サービスが使われることがある．

Web サービスの停止は，それを利用しているプロジェクトに大きな影響を与えると思われる．実際，2016 年 1 月 28 日の GitHub の停止時には，そのせいで仕事が進められなくなったというようなツイートが Twitter 上で複数観測された [1].

2. 目的

Twitter の発言を収集するためのツールを開発し，それを用いてソフトウェア開発で利用される Web サービスの停止が開発に与える影響を調査する．

3. 手法

2016 年に発生した GitHub の障害発生に関するツイートをデータとして収集する．Twitter の API には，1 週間以上前のツイートは取得できない制限がある [2]．そこで，制限無く検索できるブラウザの Twitter を利用する．この検索結果は画面を最下部にスクロールすることで古いものが読み込まれていく．検索する日付は GitHub を継続的に監視している「GitHub Status」を参照する．

データを収集するツールを開発するために 2 つのプログラムを作成する．1 つ目では，ブラウザの自動操作ができるライブラリである「Selenium WebDriver」を使用し，画面スクロール後，HTML ファイルを保存する．2 つ目では，Python のライブラリである「BeautifulSoup4」でデータを抽出する．

4. 結果

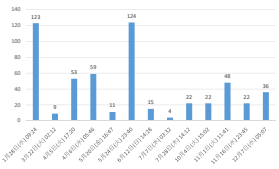

図 1　サービス停止から復旧までの間隔

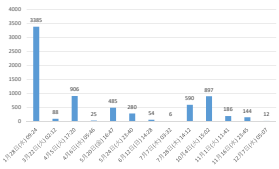

図 2　サービス停止中に投稿されたツイートの数

2016 年に GitHub で発生した 13 回のサービス停止を調査対象とする．各障害のサービス停止から復旧までの間隔を分単位でグラフにしたものが図 1 である．そして，各障害のサービス停止中に投稿されたツイート数をグラフにしたものが図 2 である．

5. 考察

サービス停止時の曜日や時間帯によってツイート数に差はあるが，調査した全ての障害で反応が観測された．特に多かったのは平日の日中で，土日祝日や深夜でもツイートが観測されていることから，その影響が出ていることがわかる．Web サービスが停止してしまうと 1 日のタスク確認やチーム内のコミュニケーションが取れなくなってしまうため，そのリスクを考慮する必要があると考える．

また，GitHub Status に記されているアナウンスよりも，平均約 7 分早くサービス停止に関するツイートが観測されていた．これは，サービスの状態について運営元が発表している情報が必ずしも正しくはないことを示唆している．

6. 結論

Twitter のブラウザでの検索結果を保存するツールを開発し，それを用いて GitHub や Slack などの Web サービスの停止に対する開発者の反応を調査した．その結果，日中はもちろん深夜でもサービス停止の影響は大きいこと，サービス運営元による停止時間についての発表は実際のそれとはずれていることがわかった．このように Web サービスの障害とその影響を調査することが，それを利用するソフトウェア開発のマネジメントにおいて有用な知見となることが期待される．

参考文献

[1] 岩城俊介, @ IT.　Github がダウン、「ぜんぶ○○のせいだ」のような惨状　今後の課題も.　http://www.atmarkit.co.jp/ait/articles/1601/28/news126.html（2016.06.29 閲覧）.

[2] 鳥海不二夫. Twitter 上のビッグデータ収集と分析. 組織科学, Vol. 48, No. 4, pp. 47–59, 2015.

SNS においてフェイクニュースを拡散するユーザの特徴抽出

プロジェクトマネジメントコース　矢吹研究室　1442014　岩橋瑠伊

1. 序論

スマートフォンなどの普及と共に，Twitter を始めとしたマイクロブログが普及している．Twitter は 2011 年 3 月 11 日に発生した東日本大震災時に，携帯電話がつながらない状況下での有用な連絡手段として活躍した．しかし，その有用性はデマや誤情報も大量に拡散させる手助けとなりえる．例えば東日本大震災時には数十種類のデマや誤情報が情報として拡散されてしまい，日本中を混乱させた．震災時のように連絡手段が限られた状況はこれからも発生する可能性は十分にあり，対策が必要である [1].

2. 目的

デマが拡散されることを防ぐために，デマツイートをリツイートしているユーザの特徴抽出を行う．

3. 手法

デマツイートをリツイートするユーザーとそれ以外のユーザの違いを見つけ，その違いが偶然生じたものではないことを示す．

1. 調査対象とするデマツイートを決める．
2. ユーザ ID を乱数で指定し，日本人ユーザ 50 人をランダムサンプリングする．
3. TwitterAPI を用いてデマツイートをリツイートしたユーザ 50 人を取得する．
4. TwitterAPI を用いて集めた各ユーザの最新 100 ツイートに含まれるリツイートの数を調べる．
5. 日本人ユーザ 50 人とデマツイートをリツイートしたユーザ 50 人の直近 100 ツイートに含まれるリツイートの数の平均の差が，偶然的な誤差の範囲にあるものかどうかを判断する為に 2 標本 t 検定を行う．

4. 結果

ランダムサンプリングした日本人ユーザ 50 人の直近 100 ツイート中の平均リツイート数は 20.04 人，デマツイート（4 件）をリツイートしたユーザ 50 人の直近 100 ツイート中の平均リツイート数は，56.68 人，62.64 人，58.46 人，57.92 人となった．

F 検定を行い分散が等しいか等しくないかを確かめる．等分散の場合の 2 標本 t 検定と不等分散の場合の t 検定を F 検定の結果に基づいて行った結果，全ての組み合わせで有意差が確認できた（有意水準は 5 %）．

5. 考察

デマを拡散するようなユーザに共通する特徴として，リツイート数に着目しランダムサンプリングしたユーザと，デマツイートをリツイートしたユーザの直近 100 リツイート内のリツイート数を比較した結果，それらの平均には違いがあることがわかった．この結果からデマを拡散するようなユーザはリツイート機能を多用する傾向にあり，ツイート内容の真偽を確かめる前にリツイートをし，デマ拡散者の一員となっていると考えられる．

自分がデマ拡散者にならない為の手段として，デマ拡散ユーザリストにあるユーザと，リツイートの多いユーザを排除することが有効だと考えられる．

6. 結論

本研究では，デマツイートをリツイートしているユーザの特徴抽出としてリツイート数の調査を行った．その結果，デマツイートを拡散するユーザの特徴として，ツイートに占めるリツイートの割合が高いことが確認できた．このような知識を活用することで，Twitter を閲覧する際に，デマツイートを真に受けて拡散してしまうリスクを下げられることが期待できる．

参考文献

[1] 榎本光, 内田理, 鳥海不二夫. O-054 東日本大震災時のツイート分析によるデマ判別に有用な特徴抽出 (O 分野:情報システム, 一般論文). 情報科学技術フォーラム講演論文集, Vol. 12, No. 4, pp. 649–650, 2013.

ブロックチェーンによるゲーム内乱数の信憑性確認法の提案

プロジェクトマネジメントコース　矢吹研究室　1442020　大木崇雅

1. 序論

2017 年 11 月 15 日に株式会社 Akatsuki が提供しているソーシャルゲームで有料アイテム抽選装置の確率の不正が疑われ，会社の時価総額が暴落した事件があった．このような事件をデータの改ざんが困難であるブロックチェーン技術を用いて解決できるのではないかと考えた．ブロックチェーンとは分散型のコンピュータネットワークであり，データベースを中央に置かずに分散して取引記録を管理している [1]．ブロックチェーンは利用者がそれぞれ同じデータを保有することで，単一のシステムや管理組織に依存しない新たなシステム基盤技術である．

データの改ざんが困難な理由は 2 つある．1 つはあるコンピュータ上に存在するブロックを不正に書き換えても，他のコンピュータ上の記録と異なるブロックを多数決で判断して排除する為だ．全体の 50 %以上のコンピュータ上の記録を書き換えないと改ざんできない仕組みである [2]．もう 1 つの理由は常に新しいブロックが増え続けるからだ．新たなブロックが生成される速度を上回る速度でブロックを書き換える計算能力を持ったコンピュータがなければブロックを改ざんする事は不可能である．

ブロックチェーンの応用分野は仮想通貨などの金融サービス業に限らず，「改ざんできないデータを共有する」メリットがある業務は対象になり得る．本研究ではブロックチェーンの，データの改ざんが困難であるという特徴に重点を置いて研究を進める．

2. 目的

ソーシャルゲームの有料アイテム抽選装置での抽選結果をブロックに書き込んでユーザー間で共有・閲覧できるようにすることが本研究の目的である．今回は有料アイテム抽選装置をサイコロで代替し，サイコロの出目を複数のノード間で共有・閲覧可能な環境を再現する．サイコロの出目が記録されたそれぞれのブロックからデータを取り出し，集計して乱数に偏りがないか調査する．

3. 手法

疑似乱数列生成器の 1 つであるセルメンヌツイスタを用いてサイコロの疑似乱数を発生させ，乱数データを CSV ファイルに保存するプログラムを作成する．乱数データの入ったブロックを P2P ネットワークで共有できる naivechain のブロックチェーン上に追加する．

4. 結果

naivechain のプログラムを入れていないノードからでも，同一ネットワーク上で繋がっているノードであればデータを共有する事が可能になった．ホストノードの URL を指定することで，ブロックチェーン上の全ブロックの閲覧と，ホストノード上にあるブロックチェーンにブロックを作成して追加する事ができた．

5. 考察

データの改ざんが困難なブロックチェーン上で誰もが記録を閲覧できる本研究は，Akatsuki のようなソーシャルゲームサービスを提供している会社の意図的な不正と，ユーザー側の一方的な誤解を防ぐ証明として役立つのではないかと考えられる．

6. 結論

以上の結果から，ブロックチェーンに記録されたゲーム内乱数の信憑性は高いと言える．今後は抽選結果をコンピュータが自動的に判断してブロックチェーンに追加するシステムを実装することが今後の課題である．

参考文献

[1] 広田望. ブロックチェーン (Blockchain). 日本経済新聞社, 2016.

[2] 丸山和子, 愛敬真生. 文系でもわかるブロックチェーン. 日経 BP 社, 2017.

文書自動添削システムによる学生の文書改善履歴の調査

プロジェクトマネジメントコース　矢吹研究室　　1442031 小山隆太郎

1.　序論

学生が行う研究では，研究だけではなく文書を作成する時間が長い．卒業論文は文量が多く，執筆形式も指摘される．大量の文書を人の目で添削を行うことには限界があり，かかる労力は大きい．

また，文書を自分以外が読んでもわかりやすく書く必要があり，文が長いほど理解が難しくなってしまう場合や，口語が混じり，文書の質が落ちてしまうことがある．

そこで，継続的インテグレーション [1] を用いることで，文書添削を自動化できないか考えた．継続的インテグレーションとは，プログラム全体を常に統合し，動作する状態を指している．

文書自動添削ツールで活用されている RedPen を執筆環境に導入することで，文書の質が向上すると考えた．継続的インテグレーションと RedPen を組み合わせ，文書添削を自動化するツールを構築する．

2.　目的

RedPen が提供する添削機能は，利用する組織のルールに対応できるように設定が柔軟に行える仕様になっている．RedPen の文書添削機能を確立し，学生が書く文書の質の向上と，作成時間の短縮を図ることを目的とする．

3.　手法

本研究の手法について以下に記述する．

1. 文書自動添削ツールの添削機能を作成する．
2. GitHub にアップロードした文書の添削を自動化する．
3. 作成した添削機能を用いて，文中のミス数の推移を記録する．

4.　結果

矢吹研究室に所属する 3 年生が書いた課題研究の概要文の添削を行った際の，エラー数の推移は図 1 のとおりである．各折れ線が文章 1 つのミス数の推移を表している．ミス数が減った文書の修正は以下のように行われた．

1.「の」,「が」等の接続詞の多用や，同一単語の複数回利用を抑えたことで，文長を短くした．「丁度」,「ちょうど」といった同じ言葉や，数値，アルファベットの表記を統一し，文書を修正した．
2.「これ」,「あれら」等の指示語の利用を抑えた．「感じる」,「思う」といった感嘆符を使用している文書は，断定系に修正された．

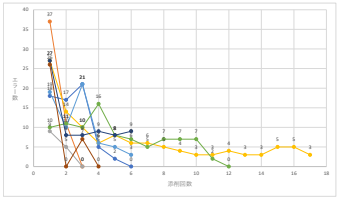

図 1　添削ツールを使用した文書の添削項数の推移

5.　考察

文書自動添削ツールの添削機能を作成し，執筆に使用したところ，専門用語を用いて解説する文書を多く見ることができた．「感じる」,「考えられる」等の感嘆符の利用を避け，「考える」と「である調」を使用したことで，研究内容を詳細に解説することに役立った．

6.　結論

文書添削ツールを使用し，文書添削をしたことで，文中のミスを削減できた．文書添削ツールを利用することで，文書作成の効率化が実現できるか調査することが今後の課題である．

参考文献

[1] 技術評論社.　継続的インテグレーションと文書執筆. http://gihyo.jp/lifestyle/serial/01/redpen/0002(2018.1.18 閲覧).

分散型 SNS におけるユーザの潜在要求分析

プロジェクトマネジメントコース　矢吹研究室　1442037　加藤 健弥

1. 序論

スマートフォンなどの普及により，手軽にインターネットへの接続が可能になった．そのため，Twitter や Facebook などの様々な SNS（ソーシャルネットワークサービス）が注目されるようになった．近年では Mastodon という新たな SNS の利用者が増えてきている．

Mastodon とは 2016 年に公開されたオープンソースソフトウェアであり，誰でも自由にサーバを立てて運用できる．そのため，Twitter や Facebook のような利用者が一つのサーバにログインする中央集権型のサービスに対して Mastodon の利用者は管理者も設置場所も異なるサーバにあるインスタンスにログインする分散型のサービスである．

インスタンスとは，Mastodon を運用しているそれぞれのサーバのことである．そのため，利用者は別のインスタンスの利用者とはつながっていない．しかしインスタンス同士が連合という形で結びつくことができるため，別のインスタンスであっても連合であれば利用者同士でつながることができる [1].

2. 目的

Twitter と Mastodon で，投稿される話題に違いがあるかを，つぶやきを定量的に分析することによって調査する．

3. 手法

Twitter API，Mastodon API を使用し，Twitter と 30 の Mastodon のインスタンスから 1 つのインスタンスごとに無作為に 100 のつぶやきを集める．その集めたつぶやきを Word2vec によってベクトル化する．その結果を Twitter と Mastodon の各インスタンス同士で主成分分析する．

4. 結果

Twitter と 30 の Mastodon のインスタンスを対象に調査した．図 1 は Twitter と話題が自由なインスタンスである mstdn.jp の 100 のつぶやきをベクトル化し，主成分分析をした結果である．図 2 は Twitter とスプラトゥーンの話題が中心のインスタンスである ika.queloud.net の 100 のつぶやきをベクトル化し，主成分分析した結果である．

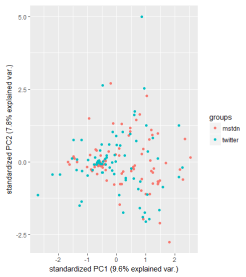

図 1　話題が自由なインスタンス

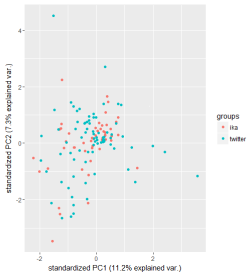

図 2　スプラトゥーンが話題の中心のインスタンス

5. 考察

主成分分析の結果を可視化したバイプロットでは，話題が幅広い Twitter のつぶやきは拡散し，話題が限定されている Mastodon のつぶやきは局所化することが予想されたのだが，分析結果は図のように，両者に明確な違いは見られなかった．このことは，Word2vec と主成分分析という方法では，人間が簡単に理解しているような，話題の違いを検出できないことを示唆している．

6. 結論

本研究で用いた Word2vec と主成分分析という手法で話題の広さの違いを識別することは困難だということが分かった．つぶやき単体ではなく，大量のつぶやきをまとめてベクトル化する手法を試みることが今後の課題であろう．

参考文献

[1] 小林啓倫, コグレマサト, いしたにまさき, まつもとあつし, 堀正岳．マストドン 次世代ソーシャルメディアのすべて．株式会社マイナビ出版, 2017.

Twitter におけるデマ拡散のシミュレーション

プロジェクトマネジメントコース　矢吹研究室　1442043　川崎貴雅

1. 序論

Twitter はリアルタイムな情報を手軽に多くのユーザへと伝播できるため社会に影響を与えている．#MeToo というハッシュタグの投稿により性的被害やセクハラについて考えるきっかけが，世界中に広がった事が挙げられる．しかし悪い影響を与えてしまう場合もある．例えば東日本大震災時のライオンの脱走や北朝鮮のミサイルの目撃デマが挙げられる．このようなツイートの拡散をシミュレーションで再現することを試みる．本研究では Twitter のデマ拡散をシミュレーションで再現することができるかの調査を行う．

2. 目的

本研究では現実のデマ拡散に近い状況を再現できるシミュレーションの開発である．

3. 手法

デマの拡散をシミュレートするためには，ユーザ同士のネットワーク作成，つぶやきの頻度，RT の頻度を求めることが必要なため，以下の手順で行う．

1. ツイートの拡散する様子をシュミレートする手法を確立するために，ランダムグラフでの RT シミュレーションを試みる [1].
2. TwitterAPI を用いて 50 万人のユーザから 1 日のツイート数取得を行い，それをもとに 1 日あたりのツイート数の分布を出す．
3. ユーザから 1 日の RT 数の取得を行い，分布を出す．
4. ネットワークの作成のためユーザーのォロー数の平均を出す．

4. 結果

Twitter ユーザ 50 万人分のデータを使って，1 日あたりのツイート数の確率分布を描くと図 1 のようになる．これによくフィットする関数を探索すると $C/(1+\exp(t-1))$（C は定数）であった．全確率が 1 になるように $C = 1/\log(1+e)$ とし，ツイート数の期待値を求めると約 1.38 となった．またグループ構築にランダムグラフを使ったツイート拡散のシュミレート手法も確立できた．しかし 1 日あたりの RT 数の分布と 1 ユーザのフォロー人数の平均が出せなかったため現実的なシミュレーションを行うことはできなかった．

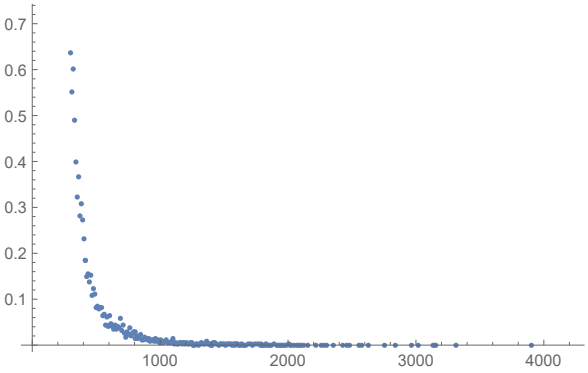

図 1　1 日のツイート数に対する割合

5. 考察

10 人でのシミュレーションのメンバ全てがツイートを確認できるようになるのは互いに繋がってる確率が 0.5 で RT する確率が 0.6 のときである．この場合フォローしている人間は 5 人で，その中で 3 人の人間が RT をすると考えられる．

6. 結論

本研究では，1 日のツイート数の分布とツイート拡散のシュミレートをする手法の確立を行った．その結果 1 日のツイート数の分布確認，ツイートの拡散シミュレートの手法の確立が行えた．この結果に 1 日あたりの RT 数の分布と 1 ユーザのフォロー人数の平均が取得できれば現実に近いシミュレーションを行うことが期待できる．

参考文献

[1] アルバート＝ラズロ・バラバシ. 新ネットワーク思考. NHK 出版, 2002.

顔画像を用いた性別予測と SNS 上での行動予測

プロジェクトマネジメントコース　矢吹研究室　1442045　川辺明俊

1. 序論

ネットリテラシー とは，情報ネットワークを正しく利用することができる能力のことである．リテラシーとは，もともとは識字能力のことで，文字や言語に対する能力の意味である．それに「ネット」と付け加えることで，インターネットを使いこなす基本的な能力を指す言葉として「ネットリテラシー」が定着した [1]．このネットリテラシーが不足していると，インターネットを使用する際に，不正サイトでクレジットカード番号を盗まれたり，コンピューターウイルスの感染により，個人や会社などの情報を流出されたり，ネット上にある嘘の情報に騙されてしまう．実際に，熊本地震でライオンが動物園から脱走したと，デマ情報を流し Twitter 上で拡散した男性は，偽計業務妨害容疑で逮捕された．

Twitter とは，インターネット上で「ツイート」と呼ばれる 140 文字以内のメッセージや，画像，動画，URL を投稿できる情報サービスである．日本では，現在（2017 年 10 月）利用者が 4,500 万人にもなるソーシャル・ネットワーキング・サービス（SNS）と見られている．

Twitter では日々，大量の情報がツイートされる．もちろんデマ情報などの，ネットリテラシーを問われるような，情報も錯綜している．そこで私は，Twitter のデマ情報に騙され，ネットリテラシーが不足している人の情報を解析し，分析できるのではないかと考えた．

2. 目的

Twitter のデマ情報を信じ，情報を拡散してしまうのに男女間で差は生まれるのかを調べる．

3. 手法

研究方法は以下のとおりである．

1. 機械学習で男女の性別を，判定できるように wikipedia のプロフィール画像をもとに，Neural Network Console を使用し，男女の顔画像を判別するための学習済みモデルを作成する．
2. デマ情報のツイートをリツイートした人のプロフィール画像を集める．
3. 集めたプロフィール画像を，始めに作成した学習済みモデルを使用し，機械学習で性別を判定させる．
4. 男女の数を集計し，どのぐらい差が生じるか調べ，考察する．

4. 結果

本研究の結果として，Neural Network Console を使用し，男性と女性の判別をした場合，80 %程度の精度しか出なかった．デマ情報を拡散する Twitter のユーザーと，正確な情報を拡散するユーザーには関わらず，男性と女性では差あるのかは，分がからなかった．

5. 考察

Twitter ユーザーのプロフィール画像は，自分の顔画像を使用していることはとても少なく，データが不足してしまうことが分かった．そして，写真から男性と女性を学習させた学習済みモデルで，画像認識を使用した場合，機械学習の性能が低いため，性別を判別するのは，難しいと考えた．

6. 結論

本研究ではネットリテラシーの不足している人には，どのような特徴があるのかを調べた．だが，Twitter では，プロフィール画像で自分の写真を使用しているのはわずかであり，機械学習を使用した画像認識では，性別を判別するのは難しいことだと分かった．

参考文献

[1] ネットリテラシー -インターネット用語辞典- | OCN. http://www.ocn.ne.jp/support/words/naline/83l83b83g838A83e838983V815B.html.

ブロックチェーン技術を用いたマネジメント法の提案

プロジェクトマネジメントコース　矢吹研究室　　1442068　　鈴木 博文

1.　序論

当研究では，急速に発展が拡大するブロックチェーン技術を，プロジェクトマネジメント（以下，PM）学科内の研究室において利用した際の利点・難点を調査する．

ファイナンスとテクノロジーを掛け合わせた造語である「フィンテック」の分野における企業買収や設立が昨年から緩和された [1]．中でも世界的に流通が拡大しているのがビットコインを始めとした「仮想通貨」である．これが通貨として機能し，サービスが成り立つために必要な技術がブロックチェーンだ．

ブロックチェーンは「仮想通貨」だけでなく投資や投票など，他の分野でも活用が試みられている．

Jack du Rose 氏が運営する Colony 社は，ブロックチェーンを会社経営・マネジメントに応用し，インターネット上での組織の自律的な運営を試みている [2].

2.　目的

ブロックチェーン技術を，マネジメントに応用した例を参考に，PM 学科内の研究室において同技術を利用した際の利点・難点を研究する．出欠情報・プロジェクト内での作業時間・成績情報・経歴情報など，存在証明が必要とされるドキュメントを管理するプロトタイプを実装し記録改ざんの複雑化と存在証明の効率化を図る．

3.　手法

ブロック作成に関する契約行動を自動実行出来る，スマート・コントラクトが構築可能な Ethereum を利用し，以下の手順でプロトタイプ開発を行う．

1. 管理者と記録者・閲覧者のアカウントを作成する．
2. 記録者が，キーに対する必要な情報を登録する．
3. 情報へのアクセスが可能な閲覧者を設定する．
4. アクセス許可を持つ閲覧者が情報を閲覧する．

4.　結果

閲覧者のアカウントで，記録者が登録した本人の経歴情報を図 1 の通り確認出来た．

記録者の氏名・生年月日・所属している組織を取得した．最終行は，閲覧者があらかじめ設定された閲覧期限を過ぎた場合の実行結果である．

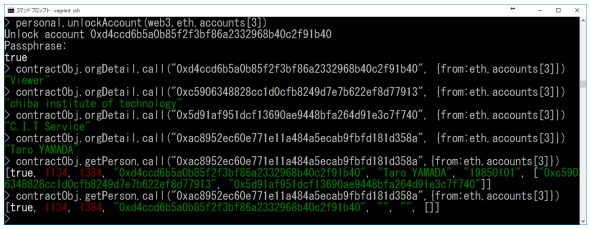

図 1　登録された経歴情報の取得結果

5.　考察

ブロックチェーンを用いた存在証明をスマート・コントラクトにて構築でき，PM 学科内においても幅広く利用する価値があるのではないかと考えた．

電子記録の存在証明はプロジェクト内の成果物において利用することも重要であるため，構築したプロトタイプの強化も有用と考える．

6.　結論

証明が必要となるドキュメントをブロックチェーンで管理することで，改ざんを複雑化しデータの信頼性を向上させることが出来た．

マネジメントに応用する点で独自性が低いため，具体的に利用する内容の検討が必要である．

参考文献

[1] 東洋経済新聞社. フィンテックで何が起こるか知っていますか. http://toyokeizai.net/articles/-/166765（2018.01.19 閲覧）.

[2] 北田淳. ブロックチェーンは世界を変える. In *WIRED VOL.25*, pp. 54–55. コンデナスト・ジャパン, 2016.

Word2vec を用いた文章構造の解析手法

プロジェクトマネジメントコース　矢吹研究室　1442069　氏名　須山 武弘

1. 序論

レポートや論文を書く際には，読みやすく，論理的な文章を書くことが大切である．論理的文章を書くための書き方として，世界で標準的なパラグラフ・ライティング（Paragraph writing）がある [1]．パラグラフ・ライティングは，英語文章の一般的スタイルであり，序論，本論，結論の 3 部構成となっている．序論でトピックとなる文が示され，本論は序論に続く支持文となり，最後に結論で文章をまとめる．冒頭にトピックとなる文章を示すと伝えたいことが明確になり，速読が可能となったり，内容の理解が深まるなど多数のメリットがある．

言語を定量的に表すツールとして，Word2vec がある．Word2vec は，単語をベクトルへ変換することができるため，文章の話題の方向性を解析し，文章作成の補助ができるのではないかと仮説を立て，本研究に取り組んだ [2].

2. 目的

Word2vec を用いて文字列である文章をベクトルへ変換し，定量的に文章構造を解析することでパラグラフ・ライティングができているかを調査する．

3. 手法

文章が論理的でパラグラフ・ライティングの原則に沿って書かれているか確かめ，実際に Word2vec による文章解析ができるかを検証する．

矢吹研究室で過去に書かれた文章データや，新聞記事などの文章データを解析対象とし，以下の手順で研究を進めた．

1. MeCab を使い，文章の形態素解析をした．
2. 日本語 Wikipedia エンティティベクトルのコーパスを使用し，Word2vec によって文章をベクトルへ変換した．
3. データ解析ツールを使用し，多数の文章で主成分分析を行い，比較，考察をした．

4. 結果

私が 3 年次に課題研究の概要として書いた文書の二段落（6 文章）を分析した結果が図 1 である．

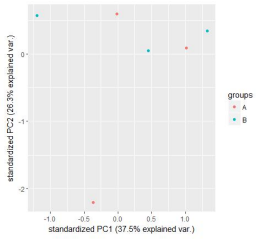

図 1　課題研究概要の分析結果

5. 考察

同段落内の文章は同じ話題でなければならないため，文章のベクトルも同じ方向性である必要がある．

図 1 では，一文章ごとの数値がグラフにプロットされている．このことから，文章の方向性が同じならば，タグ A とタグ B に対応する点がそれぞれ別々に集まることが期待される．

本研究での分析結果は，同段落内の文章にもかかわらず，それぞれのタグに対応する点が散らばって分布している．従って，解析対象文章の話題の方向性はバラバラであったと考えられる．

6. 結論

今回の研究から，Word2vec を用いてベクトルへ変換した文章を定量的に検証することで，個人の主観による添削だけでなく，定量的な文章の添削を行うことが期待される．

参考文献

[1] 倉島保美. 論理が伝わる 世界標準の「書く技術」. 講談社, 2012.

[2] 西尾泰和. word2vec による自然言語処理. 株式会社オライリー・ジャパン, 第 2 版, 2017.

プロジェクトで発生するリスクの MBTI を用いた事前予測

プロジェクトマネジメントコース　矢吹研究室　1442085　中村 真悟

1. 序論

MBTI（Myers-Briggs Type Indicator）という自己理解メソッドがある．MBTI とはカール・グスタフ・ユングの心理学的類型論の指標（内向：I-外向：E，感覚：S-直感：N，思考：T-感情：F）に判断的態度：J-知覚的態度：P の指標を加えて，4 指標 16 タイプとして性格を分類する．主に相談場面や教育現場，企業の組織編制，人事政策などに利用されている [1].

2. 目的

本研究の目的は，メンバの MBTI のタイプの相互作用がプロジェクトのリスクにどう影響を及ぼしているのかを調べ，メンバ間で発生しやすいリスクを予測することである．

3. 手法

以下の手法で研究する．

1. グループワークで課題に取り組んでもらう．
2. グループワーク後に，性格検査と発生したリスクについてのアンケートを行う．
3. 集めた回答結果をトレーニング用とテスト用にデータを分ける．
4. トレーニング用データをアソシエーション分析し，確信度が一定の値（閾値）を超えたルールを採用する．
5. テストデータを使い，ルールの精度と再現率（後述）を求める．
6. 精度と再現率の調和平均（F 値）を求め，値が最も高くなるルール抽出の閾値を求める．

採用されたルールをテストデータで検証する．テストデータ中に存在する MBTI タイプの組み合わせにルールが適合したら，そのルールに対応するリスクが発生すると予測する．発生が予測されたリスクのうち，実際に発生したものの割合を精度，実際に発生したリスクのうち，予測できたものの割合を再現率とする．

4. 結果

講義のグループワークで性格検査とアンケートを実施した．集めた 39 グループのデータを，トレーニングデータとテストデータに分け，トレーニングデータからルールを抽出した．抽出されたルールはたとえば，「MBTI のタイプ ESFJ と ESFP のメンバがいるとリスク 20 が発生する（発生率 0.2，確信度 1)」というものである。

ルールを採用する基準とする確信度の閾値を 0.8 にしたときに，F 値が最高値 0.388 となった（精度は 0.25，再現率は 0.864）．

5. 考察

今回の結果から，特定の MBTI のタイプが揃うとリスクが発生するルールがあると考えられる．より多くのデータを集めれば，メンバの MBTI のタイプがわかった時点でリスクを予測することが出来ると考える．

6. 結論

本研究では，グループワークからメンバの MBTI，発生したリスクをアンケートを用いて集め，どのようなリスクがあるか調べた．その結果，特定の MBTI のタイプが揃うとリスクが発生するルールがあることがわかった．

今後もデータを集めていけば，より多くの閾値を越えたルールが増えるだろう，そして，リスクが最も少ないグループ分けの方法の提案につながることが期待される．

参考文献

[1] 中澤清, 田淵純一郎. 24 MBTI に関する研究 (1) : MBTI の概略について．日本性格心理学会大会発表論文集, No. 6, p. 52, Dec 1997.

[2] Otto Kroeger and Janet M. Thuesen. 性格学入門—運命のカギをにぎる 16 のタイプ別性格判断. 飛鳥新社, Aug 1994.

ディープラーニングを用いた Web サイトデザインの年代解析

プロジェクトマネジメントコース　矢吹研究室　1442104　増田準

1. 序論

Web サイトのデザインは，時代に合ったものが求められる [1]．スマートフォンの爆発的な普及により，Web サイトは急速に発展を遂げた．Web サイトをデザインするということは，視覚的な良し悪しを求めるだけでなく使いやすさなど様々な要素を含む．その為 Web サイトを閲覧するデバイスによってデザインを変える事もあり，現代における Web デザインの多様化は著しい．以上のことから，本研究では時代によって進化する Web デザインの解析を対象とする．

2. 目的

この研究の目的は，年代ごとの Web デザインの変化を解析することである．デザインとは数値などで表すことができるものではなく，漠然としたものである場合が多い．その為，解析の際はページに映る要素を総合的に判断させることが重要だ．

3. 手法

この研究は以下の手法を用いて行う．

3.1 画像解析

機械学習による画像解析を利用する．この研究における画像解析とは，多数の教師画像を学習させ判別モデルを作成し，別の画像を判別させることで画像の特徴を解析する処理を指す．

3.2 画像の収集

Fortune Global 500[2] にリストされた企業の，過去のホームページを Internet Archive で閲覧し，そのスクリーンショットを撮る．

4. 結果

上述の手法で画像 14422 枚を取得した．そのうち 144322 枚を訓練データ，100 枚をテストデータとし，数式処理ソフト Mathematica で教師あり学習を行った（教師データはウェブサイトの公開年）．学習後のモデルのテスト結果は次の図 1 の通りである．図 1 の横軸が実際の公開年，縦軸が予測された公開年である．

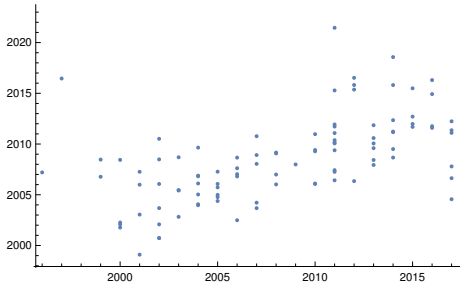

図 1　Mathematica による解析結果

また，ディープラーニング用のツール Neural Network Console にて, Web サイトの公開年を 1996 から 2002，2003 から 2009，2010 から 2017 という 3 世代に分類し画像解析した．結果は，正解率が 40.75 パーセントとなった．

5. 考察

正解のばらつきが発生した原因は教師画像が不足していたことと，学習方法が最適でなかったことが考えられる．Mathematica による解析では，最終的に教師画像 14322 枚で学習させたが，枚数を増やすごとにばらつきは少なくなった．また，Neural Network Console では学習モデルを最適化する機能によって同じ教師画像で正解率を上げることもできた．

6. 結論

機械学習を用いて年代ごとの Web デザインの変化を解析した結果，14322 枚の教師画像では満足のいく解析結果は得られなかった．正解率の向上を図るために，より多くの教師画像とより最適な学習モデルが必要となる．

参考文献

[1] こもりまさあき, 赤間公太郎. Web デザインの新しい教科書. エムディエヌコーポレーション, 改訂新版, 2016 年.

[2] Fortune Global 500. Fortune global 500 list 2017. http://fortune.com/global500/list/（2017.09.15 閲覧）.