

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. 背景

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では、Web サービスを使うことがある。例えば、チーム内でファイルのバージョンを管理することのできる「GitHub」というサービスがある [1]。その GitHub のサーバーが 2016 年 1 月 28 日にダウンした。GitHub 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らる。実際、GitHub ダウン時には、そのことに関する多くのツイートが Twitter 上に投稿された [2]。

2. 目的

GitHub のサーバーがダウンしてしまった事例から、Web サービスの障害がソフトウェア開発プロジェクトの進捗に影響を及ぼすのではないかと考えた。したがって、本研究ではソフトウェア開発プロジェクトで利用される Web サービスに障害が発生した場合、どのような影響が発生し、その影響がどれほどの人に及ぶのか調査する。また、調査結果から Web サービスの障害発生に対するリスク対策案を検討する。

3. 手法

調査方法は Twitter で投稿されている 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 では API が提供されているが、仕様によって過去のデータが取りきれないため以下の手順で行う。

1. GitHub に関連するすべてのサービスを継続的に状況監視している「GitHub Status」を参照し、2016 年で主要なサービスが停止、復旧したと記録されている時間を調べる。その時間前後を Twitter で検索し、プログラムを使って時間と本文のみを抽出する。これを各障害の発生時間ごとに繰り返す [3]。
2. サービスの停止から復旧までの GitHub に対するツイート数と、どのくらいの時間で復旧が完了するのかを調べる。調査結果をもとに、リス

ク対策案を検討する。

4. 想定される成果物

GitHub におけるサービスの停止から復旧までに投稿されたツイート数及び、サービスの停止から復旧までの時間間隔でデータを集める。これらのデータから考察することにより、障害発生に対するリスク対策案を検討する。

5. 進捗状況

GitHub Status を参照に調べたところ、2016 年の GitHub における障害発生回数は 14 回であった（ただし、10 月 21 日の障害は同時に Twitter も障害が発生していたため対象に加えない）。それらの障害についてはデータの収集を完了した。

集めたデータから考察すると、サービスの停止した時間帯や時間間隔が様々であることがわかった。そして、停止から復旧までの間隔が同じくらいでも、時間帯によってツイート数に違いがあることがわかった。

6. 今後の計画

現段階では 2016 年のデータしか集めていないため、更に過去のツイートや 2017 年のデータを集めることを検討する。

参考文献

- [1] 千葉祥太郎. 開発現場でホントに役立つツール (第 1 回) バージョン管理ツール「git」ソースコードは個人で管理 柔軟にワークフローを構築. 日経 SYSTEMS, No. 252, pp. 92–97, apr 2014.
- [2] 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)。
- [3] 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 においてデマを拡散するユーザーの特徴抽出

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. 背景

近年 Facebook や Twitter を始めとしたマイクロブログがスマートフォンなどの普及と共に急激に普及している。特に Twitter は手軽にリアルタイムな情報を多くのユーザーに伝搬出来るため社会に大きな影響を与えている。例えば Twitter は 2011 年 3 月 11 日に発生した東日本大震災時に、携帯電話がつか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]。

本研究では、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ーの特徴抽出を行う。デマツイートがリツイートされる原因として、デマをデマと見抜けないユーザー、面白半分でリツイートしているユーザーの 2 種類がいると考えた。この 2 種類のユーザーと、それ以外のユーザーには Twitter の使い方に違いがあるのではないかと考えた。

2. 目的

デマツイートをリツイートするユーザーと、それ以外のユーザーの Twitter の使い方に違いを見つける。違いからデマツイートをリツイートするユーザーなのかを判別できるようにする。

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ーの違いを見つけ、その違いが偶然生じたものではないことを示すために以下の手法で研究する。

1. TwitterAPI を用いて日本人ユーザー 50 人をランダムサンプリングする。
2. TwitterAPI を用いてデマツイートをリツイートしたユーザー 50 人を取得する。

3. TwitterAPI を用いて集めた各ユーザーの最新 100 ツイートに含まれるリツイートの数を調べて、平均を計算する。

4. 日本人ユーザー 50 人とデマツイートをリツイートしたユーザー 50 人の最新 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. 想定される成果物

デマツイートをリツイートするユーザーと、それ以外のユーザーの Twitter の使い方に違いが見つかる。例として 2 標本 T 検定の結果、日本人ユーザー群とデマツイートをリツイートしたユーザー群の最新 100 ツイートに含まれるリツイートの平均値に有意差が確認できる。

5. 進捗状況

TwitterAPI を用いて日本人ユーザー 50 人をランダムサンプリングした。3 つのデマツイートからそれぞれ 50 人のリツイートユーザーを取得した。日本人ユーザー 50 人の平均リツイート数は 20.04 人、デマツイート 1 は 56.68 人、デマツイート 2 は 62.64 人、デマツイート 3 は 58.46 人となった。デマツイート 1 と日本人ユーザー、デマツイート 2 と日本人ユーザー、デマツイート 3 と日本人ユーザーのそれぞれ 3 組で 2 標本 T 検定を行った結果、全て有意差が確認できた。

6. 今後の計画

データ量が少ないのでデータ量をさらに増やすために新しいデマツイートを発見し分析を行う。

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, aug 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. 背景

ビットコインを始めとする仮想通貨の存在が広く知られるようになってから、ブロックチェーン技術にも注目が集まっている。ブロックチェーンとは分散型のコンピューター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 2008 年にサトシ・ナカモトを名乗る人物が基本理論を提唱した。記録データを「ブロック」と呼ぶ小分けしたデータに加工し、順番に関連付けして鎖(チェーン)のように連なる構造を取る。ブロックが連なる同一の記録データを複数のコンピューターが管理・保存するというもので、コンピューター同士がブロックを比べてデータを更新する。

ブロックチェーンにはデータの改ざんが困難という特性がある。理由は 2 つある。1 つは仮にあるコンピューター上に存在するブロックを改ざんしても、他のコンピューター上に正しい記録を持ったブロックがある場合、前後のブロックと内容が異なる場合、多数決で不正に書き換えられたデータを排除するからだ。全体の 50 % 以上のコンピューター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ーが必要があり事実上不可能なためだ。

ブロックチェーンの応用分野は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。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とみなす。私はソーシャルゲームに備わっている有料アイテム抽選装置であるガチャの確率についてブロックチェーンの改ざんできないという特徴に焦点を当て研究を進める。

2. 目的

サイコロを振った結果をブロックに書き込み、集計して個別にサイコロを振った結果のデータを取り出せるアプリケーションのプロトタイプを実装する。最終的には課金によって入手できるガチャの確率と、サービスを提供している会社が公に発表している確率が同じかどうか検証するシステムを実装する。

3. 手法

ブロックチェーン上のプログラム開発に対応している HyperledgerFabric を用いる。HyperledgerFabric とは「ブロックチェーン技術推進コミュニティ」のオープンソースであり許可制ネットワーク対応のプラットフォームである。

4. 想定される成果物

サイコロを振った結果をブロックに書き込み、複数の端末機器の間で結果を共有できるアプリケーションを作成する事を目標に研究を進める。

5. 進捗状況

Bitcoin Core という仮想通貨のアプリケーションでローカル PC 上にテストネットワークを構築し、複数のアカウント間でビットコインを送金できた。

6. 今後の計画

ブロックチェーンを用いたアプリケーションのプロトタイプを作成し、複数の PC 上にあるそれぞれ異なるアカウント間で同一アプリケーションを起動させてデータをブロックに書き込む事を目標とする。

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 氏名 小山隆太郎

1. 背景

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多い。特に卒業論文は分量も多く、形式も指定されるため、文書校正にかかる労力は大きい。また、自分以外が読んでわかりやすい文書を書く必要があり、文が長いほど理解が難しくなってしまう場合や、「だから」、「かなり」といった口語が混じり、文書の質が落ちてしまうことがある。

このような状況に RedPen[1] を執筆環境に導入することで、文書の質が向上することが期待されている。RedPen は技術文書をターゲットにした文書自動添削ツールであり、現在もコードの追加、改変が行われている。

2. 目的

RedPen は学校や会社等の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。マシンを用いた文書添削を繰り返し行い、論文向けの添削システムを確立し、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

添削システムに必要な要素を以下の手法で調査する。

1. 執筆中の文書の添削に、CI（継続的インテグレーション）サーバを導入する [2]。
2. GitHub に文書をアップロードすると Wercker が自動で動作するようにし、添削結果を出力する。
3. 添削結果を記録し、添削システムに必要な要素を考察し、RedPen のコードを追加、改変する。

4. 想定される成果物

個人、複数人プロジェクトで活用できる文書添削システムを構築する。

5. 進捗状況

矢吹研究室に所属する 3 年生の課題文を添削を行った際の、エラー数の推移は図 1 のとおりである。エラー数が減った文書の修正は以下のように行われた。

1. 文頭からマルの文字数を 120 字以内に収め、カンマや同一単語の複数回利用を修正し、文長を短くした。ですます調、数値や専門用語の記述を統一した文書は右下がりの推移を得た
2. 3 回目の添削でエラー数が 30 個ある文書は、「が」、「と」等の接続詞が連続し、読みにくくなった文書。同一の専門用語が複数回利用され、文長が長くなった文書である

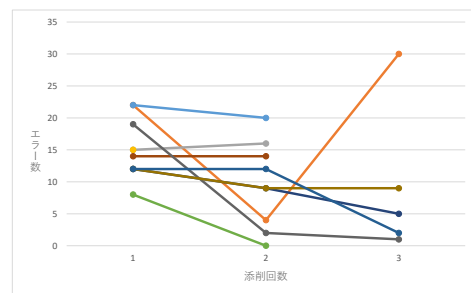


図 1 添削項数の推移

6. 今後の計画

添削機能が不十分なため、エラー数が減らなかった文書があった。卒論や課題文等を書く学生がより質の良い文書が書けるよう添削マシンを構築する。

参考文献

- [1] TakahikoIto. RedPen1.9 ドキュメント. http://redpen.cc/docs/latest/index_ja.html(2017.9.20 閲覧)。
- [2] 伊藤敬彦. RedPen を使って技術文書を手軽に校正しよう. <http://gihyo.jp/lifestyle/serial/01/redpen>(2017.9.20 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. 背景

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 年に公開されたフリーソフトウェアであり、サーバを立てることが出来れば誰でも Mastodon を自由に運用することが可能である。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 では管理者も設置場所も異なるサーバで運用できる。したがって利用者は自分自身でサーバを選びアカウントを作成してログインする。Mastodon ではこのサーバのことを「インスタンス」と呼び、その中で利用者がつぶやきを投稿することを「トゥート」と呼ぶ [1]。

2. 目的

Mastodon のインスタンスを定量的に調査し、その結果から分散型 SNS はどのような目的で使われているのか調査する。

3. 手法

Mastodon の複数のインスタンスについてのデータを集め、データマイニングにより、主成分分析を行う。また、データマイニングする。また、Mastodon と Twitter のつぶやきをベクトル化し、自然言語処理を行う。

4. 想定される成果物

Mastodon と Twitter を比較し、分散型 SNS を定量的に分類する。その結果をまとめたものを成果物とする。

5. 進捗状況

試験的に Mastodon の 30 個のインスタンスからユーザー数、トゥート数、接続中のインスタンス数

のデータを集めた。それらのデータをもとに主成分分析を行い、主成分得点を求めた。

主成分分析結果を図示すると図 1 のようになる。第 1 主成分得点（横軸）はインスタンスのコミュニティの広さ、第 2 主成分得点（縦軸）は他のインスタンスとのつながりを表していると解釈できる。

図 1 で大多数のグループからははずれたところにある 3 個は、日本最大の Mastodon のインスタンス「mstdn.jp」と pixiv が運営しているインスタンス「pawoo.net」、ダウンゴが運営しているインスタンス「friends.nico」である。それ以外の個人が運営している Mastodon のインスタンスで、現時点では大きな違いが見つかっていない。

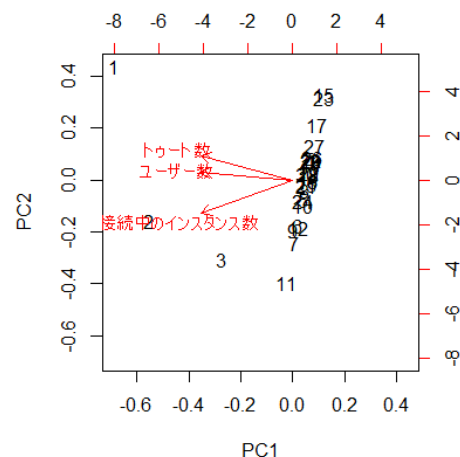


図 1 ユーザー数、トゥート数、接続中のインスタンス数を主成分分析した結果

6. 今後の計画

今後の計画は以下のとおりに行う。

- データ数を増やし、再度主成分分析を行う。
- Twitter と Mastodon のつぶやきをベクトル化し、定量的に違いを調査する。

参考文献

- [1] 武者良太. ツイッターはもう古い!? 仲間内で楽しむ SNS が人気. 日経 PC21, 2017.

Minecraft を利用する PM トレーニング法の提案と実践

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. 背景

Minecraft とは Mojang に開発されたオープンワールドゲームである。特徴としては立体系のブロックで構成された世界でブロックを設置したり、アイテムやブロックを組み合わせる新しいアイテムやブロックを製作するクラフトを行うことができる。

Minecraft は日本のプログラミング授業に使用されたことで、教育使用への関心が高まっている。現在では Minecraft を教育特化型に変更した、Minecraft Education Edition が販売されている。また Minecraft Education Edition のホームページには、年齢や教科に合わせた課題の紹介も行っている [1]。

Minecraft は教育特化型が出る前から教育活用例はある。例えば、獨協大学経済学部 2 年生のゼミにおいて全員が参加するプロジェクトとして、仮想の 3D 獨協大学が製作されている [2]。

2. 目的

本研究では、Minecraft を教育に活用している事例を調べ、実際に PM のトレーニングに活用することができる課題を製作する。

製作した課題を提案・実際に PM 学科の学生に課題を行ってもらい、製作した課題が PM としての能力向上に活用できているのかを小テストの結果から定量的に判断する。

3. 手法

Minecraft の教育活用例の調査し、PM の学習に使用できるものを参考に課題を製作する。

製作した課題を PM 学科の 4 年生に体験してもらいコストに対する資材の量のバランスがとれているかをアンケートで確認する。課題をやった被験者とやってない被験者に製作した課題に関連する 10 問ほどの小テストを行い、その結果を比較し判断する。

4. 想定される成果物

PM 的な学習を Minecraft 内でできる課題の製作をする。実際に製作している課題は、倉庫を資金と時間と品質の 3 つの要素から一番利益が出る方法を考案してもらい、実行するというものである。10 問の小テストの結果より間違いの傾向などをまとめた一覧を製作する。

5. 進捗状況

教育活用例を調査し、課題に活用するのはトレードオフの関係を使うことに決め、課題の製作を行っている。具体的には初めに考案した課題をゲームプランナーの新しい教科書やレベルアップのゲームデザイン、おもしろいのゲームデザインなどの文献を読み調整、ゲームデザインの改善を行っている。

ほかに Minecraft で再現できそうな法則等がないか調査をしている。

6. 今後の計画

資金と時間と品質の 3 つによりできるトレードオフの関係を考えた課題の製作を終わらせ、出来上がった課題を体験できる状態にする。

製作した課題を評価するのに必要な小テストの内容の考案・製作を行い、課題の体験と評価をしてもらうために協力をお願いする。

課題のバランス調整のためのアンケートを製作・実施・集計を行い、課題の修正を図る。

参考文献

- [1] MOJANG. Minecraft education edition. <https://education.minecraft.net/class-resources/lessons> (2017.9.26 閲覧)。
- [2] 堀江郁美. Minecraft の利用によるコンピュータの苦手意識への影響について. 研究報告コンピュータと教育 (CE), Vol. 2014, No. 18, pp. 1-8, mar 2014.

顔認証を用いた出席管理システムの提案

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. 背景

大学の出席システムを、機械学習という技術を利用し、顔認識という点からシステムをより良くする。

現在、大学での出席システムでは、主に IC カードを機器に通したり、出欠表、出席カードといった用紙に記入や昔からある点呼といったシステムがある。千葉工業大学では 2016 年から、iPad を使い無線で出席を自動的に確認することのできるシステムが導入された。これによって従来からあった IC カードを使ったシステムが、千葉工業大学では廃止された。このシステムの利点は、IC カードを利用する場合に発生していた、各教室の入口付近に 1, 2 個ある読み取り機器に、学生が長い列をなし、渋滞するといった問題の解消である。

このテーマで研究する以前は、主にディープラーニングを使用する画像処理について、学んでいた。今回は画像処理を、人間の神経を真似て、コンピューターによるデータの分析と学習を強力にし、「目の付けどころ」（「特徴量」）を、自動的に見つけ出すことのできる「ニューラルネットワーク」はなく、「特徴量」をあらかじめ指示し、自動的に学習する機械学習を利用して、研究を進める。なぜ特徴量を自動的に見つけ出す「ニューラルネットワーク」を利用した学習ではなく、機械学習なのかと言えば、私が機械学習の画像処理技術だけで、この研究テーマは事足りると判断したからである。

なぜこのような判断をしたかという点、実際に、ニコンの細胞培養観察装置「BioStation CT」では「ニューラルネットワーク」を利用した学習ではなく、機械学習での培養細胞の画像処理を行い、治療に適しているか否かを判定している。

このことから、機械学習でも十分に顔の特徴を掴み、ラベル分けし学習させた顔を判定させることが出来るのではないかと考えた。

2. 目的

画像処理を実際に使用することで、現在、在籍している千葉工業大学で不満のある、出席システム

の改良の余地を探ることである。

3. 手法

下記に研究の方法を記す。

1. 集合写真から顔だけを切り出す。
2. 集めた顔の画像に各個人の名前をラベル付けをする。
3. ラベル付けしたデータを学習データとして、OpenCV を使用し機械学習させる。
4. テストデータとして、上記の集合写真とは違う写真を使用し、ラベル付けごとに顔を判定させる。

4. 想定される成果物

下記に想定される成果物を記す。

- 研究のために作成したデータセット
- ラベル付けごとに顔を判定させるシステム

5. 進捗状況

OpenCV（コンピューターで画像や動画を処理するのに必要な、さまざま機能が実装されており、BSD ライセンスで配布されていることから学術用途だけでなく商用目的でも利用できます。）[1] を使用し、個人一人を対象とした顔写真の取り込みを行い、顔の判定をさせた。

6. 今後の計画

複数人の顔の判別が成功し、実用的に使えると判断できた場合は、少し難易度を上げ、人の表情を読み取り感情の識別まで、できるのかを検証する。

参考文献

- [1] Opencv とは？ 最新 3.0 の新機能概要とモジュール構成. <http://www.buildinsider.net/small/opencv/001> (2017.7.29 閲覧) .

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. 背景

「ファイナンス」と「テクノロジー」を掛け合わせた「フィンテック」が、日本でも注目を集めている。金融と最先端の技術が様々な形で融合して作り出す新しい金融サービス、さらにその潮流や動向を総称してこのように呼称する [1]。中でも世界的に流通が拡大している、ビットコインを始めとした「仮想通貨」がある。その「仮想通貨」が通貨として機能し、サービスが成り立つ上で非常に重要な技術と言われているのがブロックチェーンだ。

ブロックチェーンは、取引の履歴を記録するデータベースを分散的に管理する手法のことで、「分散型台帳」とも呼ばれる。ブロックチェーンは「仮想通貨」だけでなく他の分野への応用が始まっている。投資や土地管理、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

様々な分野で活用が広がる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プロジェクトマネジメント（以下、PM）学科において同技術を利用した際の利点・難点を研究したいと考える。複数人でのソフトウェア開発プロジェクトにおいて開発システムの分散型管理を行った際の利点と難点を調査し、より効率の高いマネジメントの遂行とプロジェクトの成功につながるよう役立てる。

3. 手法

オープンソース「ブロックチェーン技術推進コミュニティ」である Hyperledger を利用する。Hyperledger によって提供されているオープンソースのブロックチェーン実装 Hyperledger Fabric を利用し、プロトタイプを開発する。

4. 想定される成果物

ブロックチェーン技術を PM に応用した際に得られる利点・難点をまとめ、推奨される使用法を一覧とする。ブロックチェーン技術を用い複数人でソフトウェア開発を行える環境を実装する。

5. 進捗状況

ブロックチェーンで生成されるブロックの構造を理解した。チェーンの安全性を維持するため、1つ前のブロックのハッシュ値が新規生成されたブロックに含まれていることを確認し、図1の結果を得た。

```
[{"index":0,"previousHash":"","timestamp":1465154705,
"data":"my genesis
block!!","hash":"816534932c2b7154836da6afc367695e633
7db8a921823784c14378abed4f7d7"},
{"index":1,"previousHash":"816534932c2b7154836da6afc
367695e6337db8a921823784c14378abed4f7d7","timestamp":1505904883.755,"data":"Some data to the first
block","hash":"75deb75c4631b562cd1a2f88f2ea553f3fa8
8f27d3bedcaf92b645e1546c46f7"},
{"index":2,"previousHash":"75deb75c4631b562cd1a2f88f
2ea553f3fa88f27d3bedcaf92b645e1546c46f7","timestamp":1505904920.496,"data":"Some data to the second
block","hash":"2f4695c7f0a702eb85eb773089f9c3bba2ea
c8ef4454642ca421da182c782407"}]
```

図1 NaiveChain を用いたブロック作成の実行結果

この他に Hyperledger Fabric に付属する6つのサンプルを実行し、アカウントの作成とログイン、仮想通貨の送金と受取を実行した。

6. 今後の計画

Hyperledger を用いて想定される成果物のプロトタイプ作成を開始する。利用が行えるようになった段階で研究室内にて稼働させ利点・難点をまとめたいと考える。

参考文献

- [1] FUJITSU. フィンテック (fintech) とは？ 急成長を続けるファイナンスとテクノロジーの融合. <http://journal.jp.fujitsu.com/2016/10/21/02/> (2017.09.20 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.

機械学習を用いた文書検査法の提案と評価

プロジェクトマネジメントコース 矢吹研究室 1442069 須山武弘

1. 背景

人間とロボットの対話や、音声認識でのデバイス操作、機械翻訳などが増えており、コンピュータが自然言語を理解し、処理、出力することが要求されている。この際、言語を理解させ、人間とコンピュータの意思疎通を可能とさせる手段が自然言語処理である。

自然言語とは、人間がコミュニケーションを取る際、自然に発している言語のことである。自然に発している言語が故に、プログラム言語や数学的表現と違い、発言者自身の文化や曖昧さが含まれている。この曖昧な表現や発言者の文化を含んだ言語を直接コンピュータに認識させる事はできず、認識をさせるには自然言語処理が必要である。

自然言語処理とは、自然言語に含まれる発言者自身の文化や、曖昧さを含め、言葉の意味をコンピュータに処理させる一連の技術である。コンピュータと人間の意思疎通の必要性向上により、高度な自然言語の理解をする必要があるため、言語処理技術の向上が研究されている [1]。

本研究で用いる IBM Watson は、認知した情報に基づいて理解、推論し、学習するコグニティブ・コンピューティングのプラットフォームである [2]。13 種類の API があり、自然言語処理、画像処理、音声認識、データ分析など多岐にわたる分野で利用できるプラットフォームである。API が公開されており、多くの企業で質問応答システムや、自然言語のデータ化などのシステム構築がなされている [3]。

2. 目的

機械学習を用いた文書検査ツールを実装するため、自然言語処理に長けたプラットフォームである IBM Watson を使用する。この文書検査ツールにおいて、RedPen などの文書検査ツールより正確な文書の検査結果が得られるかを検証する。

3. 手法

本研究は、IBM WatsonAPI を用い、文書検査ツールを実装する。実装した文書検査ツールで文書検査を行う。次に、出力結果の有用性について検証、考察する。さらに、改善が必要な場合には改善を行い、検証、考察する。

その後、他の文書検査ツールを用い、出力結果の違いを検証、考察する。

4. 想定される成果物

文書検査ツールを実装し、その有用性や、他の文書検査ツールの出力結果の違いなどを検証し、考察する。その結果を踏まえて機械学習における文書検査の有用性の評価や改善をまとめたものを成果物とする予定である。

5. 進捗状況

IBM WatsonAPI の 1 つである Natural Language Classifier (NLC) を用い、入力した文の分類を認識させた。現在、IBM Watson の他の API を用いて文書の間違いを検査できるようにするように実装している。

6. 今後の計画

研究室内の文書データを用い、正確に文書の検査ができるかを試し、論文を執筆する。

参考文献

- [1] MahoTakara. 自然言語処理 (NLP) ってなんだろう? <http://qiita.com/MahoTakara/items/b3d719ed1a3665730826> (2017.09.22 閲覧)。
- [2] 井上研一. 初めての Watson API 用例と実践プログラミング. 株式会社リックテレコム, 2016.
- [3] 日本 IBM. IBM Watson 活用事例. <https://www.ibm.com/watson/jp-ja/use-cases/> (2017.09.28 閲覧)。

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. 背景

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]。

この MBTI を使い, プロジェクトの開始時点からメンバの性格を理解し, メンバの相互作用が原因となつて起こる事象を予測したい。本研究では MBTI を用いて, グループワークでの事象とメンバの性格との相関関係について研究する。

2. 目的

本研究の目的は, グループメンバの MBTI の 16 タイプの相互作用がプロジェクトにどのような影響をもたらしているのかを調べ, MBTI のタイプから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 [2] と失敗マンダラ [3] に基づいた事象についてのアンケートを行う
3. メンバの性格とアンケートの結果から仮説を立てる
4. 被験者に MBTI の性格検査を行う
5. 仮説に基づき数人のグループを作り, グループワークを行ってもらう
6. 課題提出時にアンケートを行う
7. タイプと事象についての仮説を実証する

4. 想定される成果物

想定される成果物はプロジェクト開始時から使用できる MBTI を用いたメンバ間のリスク予測リストである。

5. 進捗状況

千葉工業大学で開講されている講義「プログラミング言語とプログラミング」と「データマイニング入門」, 「PM 実験・演習」のグループワークで, データを収集する。

現在, 「データマイニング入門」の指導教員である矢吹太朗准教授に講義で行うグループワークの提案し, 手法と課題について話し合っている。グループ分けは「プログラミング言語とプログラミング」で収集したデータから仮説を立て思索している。

6. 今後の計画

「データマイニング入門」と「PM 実験」のグループワークで, MBTI の性格検査とアンケートを実施する。

そのデータから課題研究などのデータから立てた仮説を実証する。

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.
- [3] 亀倉正彦. 失敗マンダラを活用したアクティブラーニング授業の失敗事例分析とその知識化-学生の「やる気」を引き出す観点から-. *NUCB journal of economics and information science*, Vol. 59, No. 2, pp. 123-143, Mar 2015.
- [4] 杉浦友佳. 1-5 レスキュー・マネージメントのためのストラテジック・チーム・ビルディング論. プロジェクトマネジメント学会研究発表大会予稿集, Vol. 2001, pp. 20-25, 2001.

ディープラーニングを用いた Web サイトデザインの年代推定

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. 背景

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などもあり、現代における Web デザインの多様化は著しい。更に、人々の生活に密接に影響していることから、Web サイトに対する研究は学際的に取り組まれている [2]。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

この研究では、年代ごとの Web デザインの変化を解析することを目的とする。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 Web サイトに使用されている文字のフォントやその配置、色彩や画像など、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 機械学習

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

Internet Archive にて閲覧できる Web ページを、Selenium によるスクリーンショットを用いて多数保存する。保存の対象となる Web サイトは 2017 年度版の Fortune Global 500 [3] にリストされた企業のホームページとする。

3.3 画像解析

保存する画像は、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 世代に分類しタグ付けする。そしてこれらの画像を用いて、機械学習による画像解析を行い Web デザインのみで年代を判別可能にする。

4. 想定される成果物

想定される成果物は以下のとおりである。

- Web サイトの年代解析結果及び予測モデル
- Web サイトの年代が判別可能となる教師画像のデータセット
- 機械学習による年代解析用のコード

5. 進捗状況

上記した画像の収集にある手順で教師画像を集めた。スクリーンショットを実行するコードにおいてサイズの指定をしているが、キャプチャのサイズにばらつきが出ることや、Internet Archive を読み込む際の処理落ちによりエラー表示のまま保存されているものが多かった。エラー表示のまま保存された画像を除き、7000 枚強ほどの教師画像を収集した。

6. 今後の計画

機械学習による画像解析用のコードの書き方を習得し、解析を実行する。解析の結果によって教師画像の追加や、リサイズを洗練させるなどして正解率の向上を図る。

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] 酒巻隆治, 染矢聡, 岡本考司. Web デザインに対する印象と記憶される情報量との関係性分析. 日本デザイン学会研究論文集, Vol. 55, No. 6, pp. 59–66, 2008 年.
- [3] Fortune Global 500. Fortune global 500 list 2017. <http://fortune.com/global500/list/> (2017.09.15 閲覧).