

第3回リクルート自然言語処理ハッカソン 開催報告 速報版

株式会社リクルートテクノロジーズ

目次

1. はじめに
2. 開催要項
3. 成果報告会
4. ハッカソン風景

1. はじめに

この度は、第3回リクルート自然言語処理ハッカソンにご参加いただきまして、誠にありがとうございました。

リクルート社のサービスが保持する実データに触れていただいたことで、ビジネスで用いられているデータがどのようなものなのか、またリクルート社のデータの面白さ等を実際に感じ取っていただくことが出来ましたら幸いです。

以降にハッカソンの開催報告を速報版としてまとめましたので、ぜひご覧ください。
また、後日発表スライドや弊社からのコメントも含めたものをお送りできればと思っております。

作成いただいた発表スライドにつきましては、広報観点からの社内チェックが入るため、場合によっては全てお送りできないことがあります。あらかじめご了承ください。

引き続き、何卒よろしくお願い申し上げます。

株式会社リクルートテクノロジーズ
APソリューショングループ 検索改善チーム
大杉 直也

2. 開催要項(1/2)

- 名称: 第3回リクルート自然言語処理ハッカソン
- 主催: 株式会社リクルートテクノロジーズ
- 企画・実施支援受託: アカデミック・リソース・ガイド株式会社
- 時期: 2017年3月27日(月)～31日(金)
- 会場: 株式会社リクルートテクノロジーズ オフィス
参考) <http://recruit-tech.co.jp/>
- 参加人数: 8名
- 参加費: 無料

2. 開催要項(2/2)

- 内容

- 使用データセット: リクルートマーケティングパートナーズ社のメディアデータ
 1. 記事の文章データ
 2. ユーザーのアクセスログ
 3. TwitterなどのSNSデータ
- 分析環境: Jupyter notebook + BigQuery を使った環境
- 使用言語: Python 2.7 系または 3.5 系
- 評価: メディアデータからの知識発見を技術観点とビジネス観点の両方から評価。
優劣は予測精度ではなく社員へのプレゼンで決定

3. 成果報告会(1/3)

発表内容と発表概要を下記に記載いたします

1. 朴 炳宣(関西大学)「記事の内容と著者を考慮したタグの生成」

Webサイトに投稿されている記事は時間が経つにつれて読まれなくなる傾向があることに着目した。記事をもういちど探し当てるために記事に付与されている「タグ」はとても重要な枠割を持っているが、現状では人手で書かれていることから、作業量の多さと観点の違いによる食い違いが生まれる。そこで、記事のデータから記事の内容や著者の情報に基づき、自動で特徴的なキーワードや著者の情報を推薦するシステムを作成した。

2. 金 基煜(京都大学)「ビッグデータ&機械学習くんからのお知らせ:この記事がウケるよ！」

→優秀賞(テクノロジー部門)

Webサイトの記事データと1年間のアクセスログを統計と機械学習を用いて分析した。その結果を用いて、どのような記事がPV数を集めるか、新たなユーザーを集めるかななどを予測するアプリケーションを構築した。

3. 篠田 一聡(東京大学)「word2vecを用いた記事推薦システムの設計」

WebサイトのPV数を底上げするために、推薦システムを構築した。手法にword2vecを用いることで、ユーザーの過去の行動に応じて推薦記事を選定できるようになった。推測精度はまだ低いものの、扱うデータを増やせば今後精度を上げていけると考えている。

3. 成果報告会(2/3)

発表内容と発表概要を下記に記載いたします

4. 久保 大輝(奈良先端科学技術大学院大学)「対話を用いた記事推薦システム」

→優秀賞(ビジネス部門)

現状の検索システムでは、検索結果をカテゴリにより絞り込むことができない。そして現在、スマートフォンからのアクセスが増えており、スマートフォンからでも使いやすいUI設計を行う必要がある。そこで本発表では、対話を用いた記事の推薦システムを提案する。

5. 森 雄喜(電気通信大学大学院)「タイトルからわかるユーザーが寄ってきそうな言葉の推測」

PVを上昇させるためにタイトルに含まれる単語を考える。これをウィキペディアから作成した辞書を利用することで、PVを上げたい。時と場合で使用する単語は変わるが、これをNLPを使って解決したい。

6. 松岡 佑知(大阪大学大学院)「PV数を15%UP! ～因果推論とトピックモデルによる記事タイトルの統計解析～」

→最優秀賞

WebサイトのPV数を向上させる要因とその影響の大きさを『統計的因果推論』、『トピックモデル』を用いて分析した。これによりクリック率が向上するタイトルの付け方を特定することができ、今後のサービス改善の判断材料を得ることができたと思われる。

3. 成果報告会(3/3)

発表内容と発表概要を下記に記載いたします

7. 飯田 健司(埼玉大学)「利益率UPのためのユーザーの行動別分類」

収益につながるユーザーのアクセスの特徴を抽出することで、収益につながるユーザーにむけたサービスを改善することに失敗した。データの扱いが不慣れであった。

8. 野本 英梨子(大阪大学)「サービスアカウントがツイートすべき記事」

→優秀賞(プレゼン部門)

Twitterで注目されそうな記事をタグから予測した。これらをツイートすることで、サービスをたくさんの人に知ってもらえると考えられる。今後はタグだけでなく本文や画像を考慮した予測により、より効果的な予測ができると考えられる。

4. ハッカソン風景



4. ハッカソン風景 ～オリエンテーション～



4. ハッカソン風景 ～社内見学～



4. ハッカソン風景

～第1回ハッカソンに参加した先輩社員のスピーチ～



4. ハッカソン風景 ～環境設定～



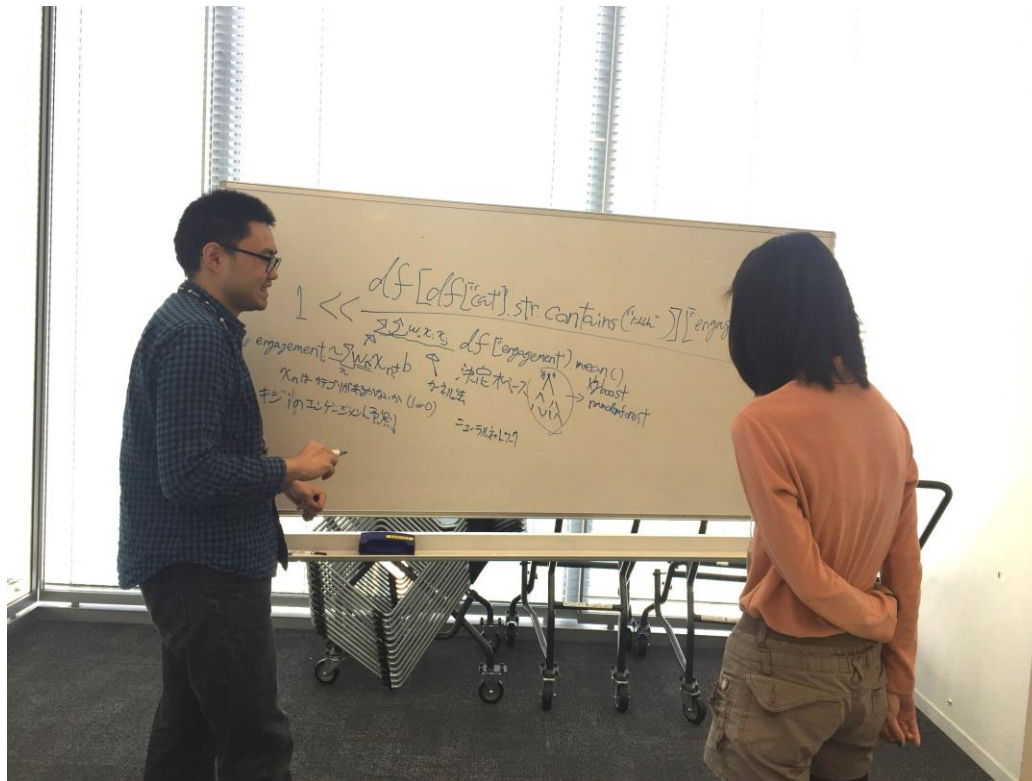
4. ハッカソン風景 ～キックオフパーティー～



4. ハッカソン風景 ～開発作業風景～



4. ハッカソン風景 ～開発作業風景～



4. ハッカソン風景 ～成果報告会～



4. ハッカソン風景 ～懇親会～

