



## vol.12 田村みゆさん

---

みなさんは、世界の第一線で活躍されている女性プログラマーの方々をご存知でしょうか？ JOI情報オリンピック日本委員会が実施する「先輩に聞く！プログラマーへの道しるべ」では、プログラミングやその周辺の技術や知識を使って活動している女性の先輩方に、お仕事内容や学生時代についてのお話を伺っていきます。

第12回目に登場いただくのは、[日本女子大学大学院理学研究科 数理・物性構造科学専攻](#)の修士課程に在籍している田村みゆさんです。聞き手は JOI 情報オリンピック日本委員会理事で東京大学の山口利恵が務めます。ぜひみなさんの進路の参考にしてみてくださいね。

---



日本女子大学大学院理学研究科 田村みゆさん

## 行政文書の解析やAIモデルの構築に取り組む

山口 日本女子大学のキャンパスはどのあたりにあるのですか？

田村さん 東京・豊島区の目白というところにあります。JR山手線の目白駅から徒歩15分くらいの場所にあり、都心に位置しているのですが、周辺には公園もあったりして、緑豊かなエリアです。

山口 数理・物性構造科学を専攻とのことですが、どんな学科ですか？

田村さん 数学と物理と情報の3つの分野を学べます。私は情報科学を専攻しています。機械学習や人工知能を使って、自然言語処理を行う研究です。どちらかというソフトウェアが専門になります。


山口 言語処理というのは、私たち人間の言葉を、いかに機械にわかってもらうかという研究だと思いますが、そのなかでも、行政文書のような特殊な文書の研究をしているそうですね。

田村さん はい。2021年から、内閣府のエビデンスシステムというところで、研究補助のアルバイトをさせていただいていたのですが、2022年から共同研究というかたちが変わって、エビデンスシステムのデータ研究をしています。行政文書の解析や、AIモデルの構築などに取り組んでいます。

### 専門用語に基づいた文書分類モデルの説明可能性の向上に向けて

田村 みゆ<sup>1</sup> 沼澤 翠<sup>2</sup> 堀浦 照乃<sup>1</sup> 倉光 君郎<sup>2</sup> 七丈 直弘<sup>3</sup>

<sup>1</sup>日本女子大学大学院理学研究科  
<sup>2</sup>日本女子大学理学部  
<sup>3</sup>一橋大学工学部



日本女子大学  
JAPAN WOMEN'S UNIVERSITY

#### 背景・課題

近年、科学技術イノベーション(科技)は重要な政策課題となっている。そこで、各府省庁の進める行政事業に関する文書から科技に関連する2値分類モデル(科技判定モデル)の構築に取り組んでいる。




Figure 1: 科技判定モデル

自然言語処理モデルmT5[1]をベースとし、2021年度行政事業レビューシート上の事業に関する文書と科技に関するラベルのペアでFine-tuningした。しかし、予測結果の根拠が不透明であるため信頼性に欠けてしまい、実用化に向けた課題となっている。

#### 目的

科技判定モデルが注目している専門用語を可視化することによって、説明可能性の向上を目指す。

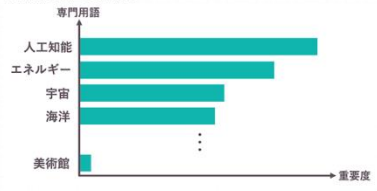


Figure 2: 科技判定モデル全体が注目している専門用語の可視化

#### アイデア

図3のように文書内の専門用語にマスクをかけることで予測結果への影響について調査する。

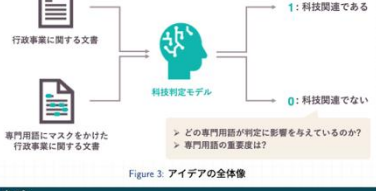


Figure 3: アイデアの全体像

#### 実験

予備実験として、専門用語が予測精度に与える影響について調査した。モデル評価用のデータ829件の内、科技に関連するデータ234件から専門用語を抽出し、専門用語をマスクトークンに置換することで作成したデータ(マスク)と、元のデータ(オリジナル)をモデルに入力した際の科技判定モデルの予測精度を比較する。




Figure 4: 専門用語抽出の流れ

#### 実験結果

科技判定モデルの予測結果について以下に示す。表1では、オリジナルとマスクでの正解率と、ラベルが1(科技に関連する)のデータにおける適合率/再現率/F値をまとめた。

データ	正解率	適合率	再現率	F値
オリジナル	0.84	0.73	0.70	0.71
マスク	0.72	0.50	0.27	0.35

Table 1: 科技判定モデルの予測結果

また、オリジナルとマスクが意味的にどの程度離れているか確認するため、それぞれのデータから一部抜粋した文書を、SentenceTransformer[2]で特徴ベクトルに変換後、Cos類似度を測定した。

オリジナル

南海トラフ地震の想定震源域におけるプレート境界の固着状態の変化を観測・監視し、最新の科学的知見に基づく解析を行い、適時適切に南海トラフ沿いの地震に関連する情報を発表することにより、南海トラフ地震による被害の防止・軽減を目指す。南海トラフ地震とその周辺に発生した地震動観測情報、ひずみ計等。により、南海トラフ地震につながる可能性がある現象を24時間体制で観測・監視し、最新の科学的知見に基づく解析を行うとともに、観測データに異常が検出された場合は、その原因について、南海トラフ沿いの地震に関する緊急討議会、及び、地震防災対策強化地域判定会、により緊急的な評価を行う。また、その評価結果を防災対応に活かすことができるよう、適時適切に南海トラフ沿いの地震に関連する情報を国民・防災関係機関・報道機関等に発表する。

Figure 5: Cos類似度測定に使用したオリジナルの文書

オリジナル同士でのCos類似度を測定したところ、0.99となった。一方、オリジナルとマスクとのCos類似度は0.58であった。これらの結果から、専門用語が科技判定モデルに影響を与えていることがわかった。

#### まとめと今後の展望

本研究では、専門用語の観点から科技判定モデルの説明可能性の向上を目指している。予備実験として、専門用語が予測精度に与える影響について調査し、科技判定モデルの予測精度の低下とCos類似度が著しく低下していることが確認できた。

- マスクをかける専門用語の選択
- 科技判定モデルに対して専門用語が及ぼす影響の評価指標の調査

#### 参考文献

[1] Linting Xun, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. *mT5: A massively multilingual pre-trained text-to-text transformer*. *arXiv preprint arXiv:2010.11934*, 2020.

[2] Nikl Remers and Inna Gurevych. *Sentence-bert: Sentence embeddings using siamese bert-networks*. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing: Association for Computational Linguistics*, 11 2019.

CLP 若手の会 (YANS) 第17回シンポジウム (2022)

山口 行政文書の研究のなかで、苦労したことや楽しかったことなどありますか？

田村さん 行政事業の文章って、なかなか日常生活では見ることがありませんよね。難しい言い回しもありますし、書き方もそれぞれなので、まず人間の側で読み解かなければい

けないのですが、そこが難しかったですね。記号の処理なども必要になってくるので、少し大変だった思い出があります。

## この夏、オフラインでの国際会議出席も経験

山口 国際会議にも参加したと伺ったのですが。

田村さん 国際会議には出ましたが、行政文書の研究ではなく、また別の内容で参加させていただきました。

山口 今年では日本で行われましたね。理系の修士課程の学生さんは、年に1、2回は海外で発表するチャンスがあるのですが、最近ではコロナ禍もあり、国外へ行くのが難しかったですよね。

田村さん 2020年に学部4年生だったのですが、ちょうどコロナ禍で、学会も完全なオンライン開催でした。2022年の夏の学会が初めての対面形式だったので、普段とは違う緊張感を持って臨めました。

山口 国際会議でのエピソードなどありますか？

田村さん 会場が広島だったのですが、フリーの日には、研究室のメンバーで宮島に観光に行くなどして楽しかったです。緊張感もありつつ、仲間と過ごす時間もあっさりして、いい経験になりました。

山口 私は教員の立場なので、ぜひ学生のみなさんには外の世界へ出て、いろいろな経験を積んでもらいたいなと思っています。初めての海外が国際会議への参加という方もいるので、外国での苦労などいい思い出になるとと思っています。

田村さん ハイブリッド開催もあり、オンラインとオフラインの両方を選べるとなると、どうしても気軽なオンラインを選びがちで、海外へ行く機会が減ってしまいます。卒業修了前までには、一度でもいいので海外での発表などを経験してみたいですね。そのためにも研究をがんばります。

## 「JOI入門講座」のチューターも務める

山口 研究以外にもいろいろな取り組みに参加されていますよね。私たち情報オリンピック日本委員会がやっている、「JOI入門講座」のチューターなども務めていただいたり。他にも何か経験はありますか？

田村さん アルバイトですと、学部1年生の時から電話オペレーターを4年間やっていました。あとは出版社の雑務なども経験しています。

山口 JOI入門講座での田村さんの説明はとてもお上手ですよ。講座のチューターをやってみていかがですか？

田村さん 何度か参加させていただいていますが、毎回ものすごく緊張しています。参加者には、プログラミングをはじめたばかりの方も多いので、ここでつまづいてほしくないと思って、毎回気合を入れて臨んでいます。

山口 趣味もいろいろお持ちだと聞きました。

田村さん 今がんばって勉強していることが、グラフィックデザインです。修士1年のときに4ヵ月間、グラフィックデザインの専門学校に通っていて、そこから自分で手を動かしてデザインをつくったりしています。



制作したロゴ

## 中高大と12年間女子校生活

山口 田村さんは千葉県の成田市出身だそうですね。

田村さん 成田の小学校を卒業して、東京に引っ越してからは、都内の中高一貫校の女子校に通っていました。

山口 女子校から女子大に行かれたのですね。

田村さん 女子だけの環境で育ってきたという、めずらしいケースかもしれません。

山口 私も女子大出身ですよ。田村さんは、そのなかでも理系を目指しましたが、きっかけはありますか？

田村さん 小学生の頃に実験教室に通っていて、そこから理系に興味を持ち始めて、そのまま進路も理系に。

山口 実験教室ではどんなことをやっていたのですか？

田村さん 生物、化学、物理問わずいろいろな分野の実験を行っていました。例えば生物だと動物の解剖や、科学だとガラスを溶かしてとんぼ玉をつくったり、物理は衝突の実験などです。月に一度通っていたのですが、本当に大事な経験をさせてもらったと感じています。



小学生時代に通っていた実験教室で参加したハワイ研修（写真の左下手前が田村さん）

山口 そこから情報系に興味を持つきっかけはなんでしたか？

田村さん 実験教室ではプログラミングには触れていなかったのですが、中高の夏休みに実験プログラムに参加しました。ロボットを自分たちで組み立てて動かしてみるという内

容で、そのときに初めてプログラミング言語を学び、おもしろいなと感じたのがきっかけです。

山口 そこで興味を持って、そのまま大学まで継続していくわけですね。

田村さん 中高でプログラミングがおもしろいとは思ったのですが、それを大学でも学びたいかというとは実はそうではありませんでした。ただ、漠然と、理系に進むとしても、自分は生命系ではなく、非生命系だろうと思っていました。大学も数学・物理・情報の3つの分野が学べる学科にそのまま進みましたが、学部時代の授業で改めてプログラミングに触れる機会があって、もうちょっと深く学んでみたいと思うようになり、情報科学を専攻しました。



高校では琴部に所属。

山口 学生時代のお写真もいくつかいただいています。





小学6年生の時に所属していた集会委員会として参加した音楽集会（写真左の仮面を被っているのが田村さん）

田村さん 小学生のときは、集会委員会という委員会に所属していて、そこで月に1回開催される音楽集会の企画運営と司会をしていました。これはそのときの写真です。ヒーローに扮して司会をするという催しでした（笑）。人を楽しませたいとか、人の心を動かしたいという気持ちが強かったように思います。

山口 当時、夢などありましたか？

田村さん 夢などはなかったのですが、ただ漠然と大人になった自分はスーツを着ていないだろうと思っていました。

山口 最後にもう少しプログラミングについて聞きたいのですが、プログラミングのどんなところに楽しさを感じていますか？

田村さん この記事を読まれている方は、プログラミング自体が好きだという方が多いかもしれないのですが、私はどちらかというと、プログラミングを手段として捉えています。プログラミング自体が楽しいと思う機会は少ないのですが、いろいろ試行錯誤したうえで、自分が期待したものができたり、思いがけないかたちで成果が得られたときにおもしろさを感じます。

山口 今後の目標などあれば教えてください。

田村さん 好奇心を頼りに進んできたので、この先も自分の好奇心に触れるものがあつたら、それに対して取り組んで行けたらいいなと思っています。そこで今やっているプログラミングや研究の知識が活かせるといいですね。

山口 最後に、未来のプログラマーへメッセージをお願いします。

田村さん 「好き」という感情はとても大事だと思っています。好きという気持ちが何よりも大きな原動力になると思うので、プログラミングにかかわらず、いろんなことに挑戦し、自分の好きなことを見つけたら、それを起点に突き進んでいっていただければと思います。

山口 本日はありがとうございました。

#### 【インタビューを終えて】

田村さんは、JOI 入門講座のチューターを長く務めてくれていて、受講している方とのコミュニケーションもとても上手で穏やかで素敵な先生でもあります。彼女がどうしてこの道に来たのかを伺ったのは初めてでしたが、小さい頃から実験などに参加する機会があったという話を聞いて、ちょっと羨ましくもありました。彼女は4月から就職してまた新しい別の道に進んでいくことになるのですが、彼女の柔らかい人当たりで、前向きにすすんでいくんだろうなと思っています。（山口）

次回もお楽しみに。