

ビッグデータ時代を担うデータサイエンティスト育成を目指す日本初のコンテスト

特集2

データ分析のアイデアとスキルを競う



ビッグデータは膨大で多彩な情報のつぼに例えられる。そこからいかに有用な知識や社会・経済のニーズをすくい上げるかは、専門家の腕前やセンスにかかっている。JSTは3月8日、大規模情報を活用した統計解析のさらなる普及とこの分野の人材育成に向けて、分析アイデアやスキルを競い合うコンテスト「データサイエンス・アドベンチャー杯」を、ソフト・サービス会社SAS Institute Japanと共同開催した。全国から寄せられたきらめくアイデアや効果的な分析の数々を紹介する。

企業のデータサイエンティスト発掘を促進する



予選審査を通過した8チームによる本選プレゼンテーション審査。

データの専門家が求められる時代

ビッグデータから有用な情報を解析する「データサイエンティスト」が脚光を浴びている。統計解析やデータ処理に関する高度な知識を持ち、情報を的確かつ効果的に分析して、研究やビジネスに役立つ予測や成果を見いだすことができる専門家である。2009年、「これからの10年で最もセクシーな職業は統計解析家」というグーグル社チーフエコノミストの言葉が注目を集め、データサイエンティストの魅力は世界中に広まった。さまざまな分野で注目を浴びるようになったものの、有能な専門家はまだまだ少なく、人材育成の場も限られている。

「データサイエンス・アドベンチャー杯」は、実務で統計・分析を担うビジネスマンや統計学・情報科学に関心を持つ大学生・高校生らを対象にしたコンテストだ。JSTが持つ科学技術データやウェブ上のオープンデータとSAS社が提供する統計ソフトを材料に、柔軟な発想でテーマを自由に設定して統計分析した作品を募集した。未来を切りひらくデータサイエンティストたちに、新しい着想や分析技術を切磋琢磨する機会を提供することを目的としている。科学技術分野の分析コンテストは、日本で初めての試みであり、世界でもほとんど例がない。

科学技術文献データの新たな活用法にも期待

昨年10月から12月まで、データやソフトを貸与するエントリー期間が設けられ、86チームの申し込みがあった。そのうち、今年1月末までに応募したのは34作品で、書類審査を通過した8チームが3月のコンテスト本選に臨んだ。

各チームは応募作品の概要や解析手法、結果などを発表した。内容についての予選結果と、この日のプレゼンテーション審査を経て各賞を決める。当日、会場には約200名の観客や関係者が詰めかけ、この分野への関心の高さがうかがわれた。その中に、制服姿で熱心に発表に聞き入る数多くの高校生の姿もあった。18歳以下のU-18部門に応募した8作品の生徒とその応援団だ。

「U-18部門で本選に勝ち上ったのは2組ですが、他のチームも一般部門にひけ

を取らない作品ばかりでした。将来を担う若い才能に本選の雰囲気を経験して刺激を受けてほしかったので、全員を招待しました」とコンテストを企画したJST情報企画部の伊藤祥主査は話す。

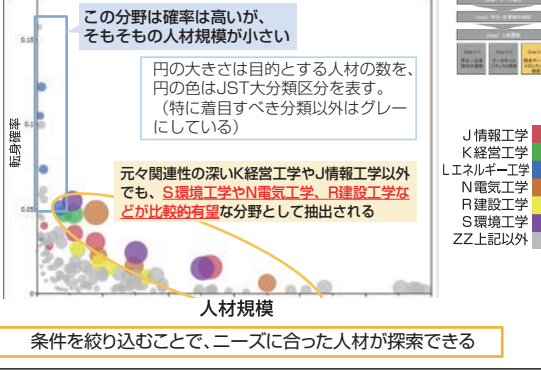
開会あいさつで、JSTの大竹暁総括担当理事は、「JSTの科学技術情報から、オープンイノベーションが生まれることを期待しています」と話した。

このコンテストには、JSTが収集提供し、日本の研究を支えてきた膨大な科学技術データの新たな活用法を開拓するねらいもある。今回、課題として出されたのは、国内資料12,000誌、国外資料4,700誌を収集している科学技術文献データベースなどJSTの各種データだ。研究論文のタイトルや著者名などの書誌情報だけでなく、データベース編纂時に整理された所属機関や化学物質の情報、科学技術・特許用語の体系なども含まれる。論文検索サービスを提供するために50年



分析結果【STEP3-3】特定データサイエンティスト探索 (エネルギー会社向けデータサイエンティストの探索結果)

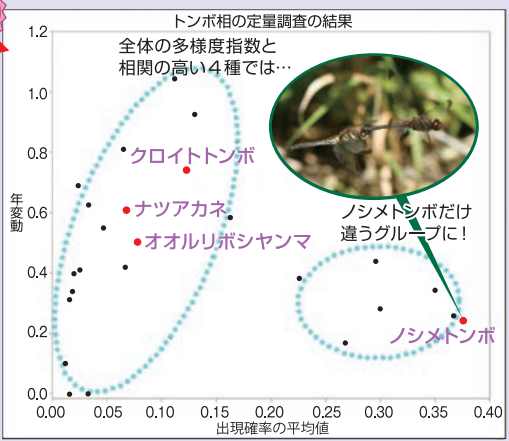
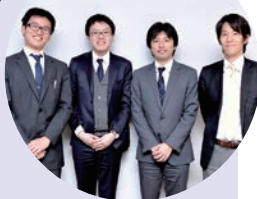
特定(S)確率(学生時代にこの分野
だった人が求める企業時代に求める
条件に合致する確率)



一般 UNAGI (東京ガス株式会社
技術開発本部 技術戦略部 戦略研究グループ)

未来のデータサイエンティストを探せ! ～研究分野遷移から見た人材マッチング～

論文データに付与された著者IDを活用し、研究者それぞれの学生時代の論文と企業就職後の論文を調べることで専門分野の変遷をたどれることに着目。企業が望む分野の人材は、学生時代の専門分野を探索することで見つけられると提案した。



U-18 北海道札幌旭丘高等学校生物部

生物多様性を探るために ～トンボの統計解析からわかった ノシメトンボと生物多様性について～

生物部で蓄積したトンボの生息数調査の結果を統計分析し、全体の多様性指数と相関の高い種を探った。さらに論文タイトルのデータを解析して、トンボに関する研究の動向を調査。多様性に大きく関係するノシメトンボの研究が少ないことを見いだした。



以上かけて蓄積したものを、新たに独創的な視点でまとめて解析することで、予想外の発見が生まれるのではないかとこの期待もあった。

コンテスト参加で 知識やスキルを試す

一般部門の金賞には、東京ガスの「UNAGI」チームの「未来のデータサイエンティストを探せ!～研究分野遷移から見た人材マッチング～」が輝いた。論文情報を分析することにより、どんな専門分野の学生が情報工学や経営工学の分野に多く進んだのかの傾向を調べ、未来のデータサイエンティスト候補を多く発掘できそうな「金脈」を統計的に明らかにするユニークな成果だ。学生と企業の縁組みを支援し、さまざまな分野の企業ニーズに合った人材を探す仕組みも提案した。「分析手法はシンプルだが、とても効果的」、「アイデアが卓抜」など、審査委員から高い評価を受けた。

リーダーを務めた釋宏介さんの呼びかけに、技術開発本部で各種データ分析を担当する同僚たちがすぐに賛同し、参加を決めたという。「テーマを決めるのにいちばん時間がかかりました」と釋さん。メンバーの藤本剛志さんは、「社会のニ

ズにいかに応えるかに重点をおきました」と続ける。日常の仕事が終わった後にとりかかり、ときには終電まで粘ったそうだ。

「解析は大変でしたが、シンプルな方法でデータを活用することに意義があるという考えで進めました」とメンバーの笹谷俊徳さん。中井洋平さんは、「改めてデータの大切さを実感しました」と話す。金賞受賞に、「自分の経験や知識を会社以外の場所で試せてよかった。今後の自信につながりました」とそれぞれ笑顔で語っていた。

高校生が予想を上回る 大活躍

アイデア賞は、一般部門と高校生部門を共通に審査した結果、高校生チームに贈られた。受賞作品は、北海道札幌旭丘

高等学校生物部の「生物多様性を探るために～トンボの統計解析からわかったノシメトンボと生物多様性について～」。

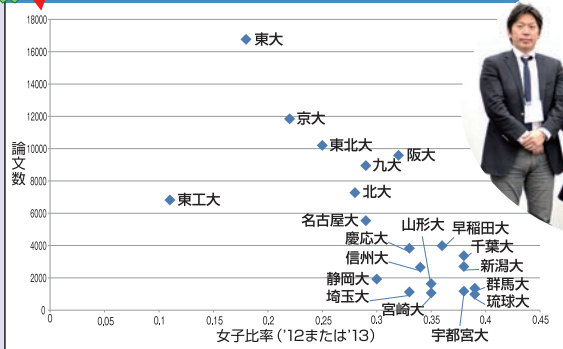
生物部が5年間にわたって採取した近隣地域のトンボ相(生息するトンボの種類組成)の調査結果と、JSTのトンボに関する論文のデータを組み合わせた。トンボ相統計の結果から、その多様性の変化に大きく関係している種がノシメトンボであることを見つけた。その上で、日本の論文タイトルを解析。生物多様性の研究は増えているものの、ノシメトンボの研究は数少ないことから、今後、研究を増やすべきだと提案した。1年生部員ながらしっかりとした説明力と、「言葉で表しきれない自然界のことを、何とかして客観的に表現したい」という熱意に満ちた内容は、審査委員や会場を驚かせた。



コンテストの審査員と受賞者たち。高校生から社会人まで幅広い参加者がデータサイエンスの可能性に挑戦した。

U-18賞

女子比率('12 or '13)と論文数



科研費の獲得総額や教授数、学生の学力、女子比率などの条件が、各大学の論文数にどう関わっているかを統計的に調査し、国立や私立での傾向の違いなど、大学の実態について考察。論文数と女子比率が反比例することを示した。

U-18 埼玉県立熊谷女子高等学校 大学のそこところ ～おカネと人と論文と～

審査委員からは、「自分たちの調査データを組み合わせたユニークな提案であり、素晴らしい発想」、「予備的な分析によってターゲットを絞った上で、文献の分析につなげる流れは、研究のプロセスそのものといえる。応募作品の中で、このような提案はほとんどなかった」などの高評価を得た。

受賞した宇久村三世さんは、「本選に出られるとは思っていませんでした」と驚きの表情。「限られた時間で、たくさんのデータを分析してまとめることに苦労しました。途中でくじけそうになりました

が、何とかやり遂げました」と喜んだ。生物部顧問の綿路昌史先生は、「私たちのやってきたことが理解されて嬉しく思います。生き物を大事にしようという気持ちで調査を指導してきましたが、生息数の

増減を単純に並べるだけでなく、データを解析することで数倍も説得力が増すことを知りました」と話す。

高校生を対象にしたU-18賞は、埼玉県立熊谷女子高等学校チームが受賞した。テーマは、研究費や教授数、学生の学力、女子比率などの大学の実態を、論文数と比較しながら考察した「大学のそこところ～おカネと人と論文と～」。

「私たち、もうすぐ大学生だけど、論文をたくさん出す大学は何が違うの?」「やっぱり学力?」「それともおカネ?」とかけ合いを交えながら、高校生らしい興味を掘り下げた意欲的な発表だ。「理系女子としての今後の活躍を期待しています」というコメントが審査委員から寄せられた。

コンテストでは、ほかにも多彩なアイデアが花開いた。例えば、入賞した慶應

銀賞



一般 技術動向観測隊 (株式会社金融エンジニアリング・グループ)

企業に着目した共同研究ネットワーク構造の解析と非連続的成長の予測

情報工学分野の企業に着目し、論文の共著関係から共同研究機関数の推移を解析。共同研究機関数の分布には非連続的な構造があることを発見するとともに、過去の論文発表履歴から将来的な共同研究の急拡大を予測できると結論した。

銅賞



一般 HSE研開部 (株式会社日立ソリューションズ東日本)

研究力の向上と実社会の発展の関係分析

分野別の文献数、特許出願件数、法人企業統計の売上高などを使って、実社会で基礎研究や応用研究がどのように社会に還元されるかをモデル化。研究力向上が社会還元につながることを、還元力は特許出願件数に表れることを見いだした。

入賞



共著関係にある研究者(四角)同士を線で結んだ。赤は中心性が高い研究者の例。



一般 Terano Lab. (東京工業大学)

論文の共著関係ネットワークの中心性分析

論文の共著関係から研究者ネットワーク(図)を描き、つながりの多さや他のグループとの媒介の度合いなどを測る「中心性分析」により、各分野内や異分野間の研究活動をけん引する研究者を洗い出した。

入賞



一般 健マネ (慶應義塾大学 大学院 健康マネジメント研究科)

研究活動の年次推移および人々の生活実感への影響に関する分析

「主成分分析」という手法で、工学的か医学的か、など動向の似た分野をまとめる観点(主成分)を見つけ、年代別の研究動向を把握。世論調査の生活満足度を各主成分で表す式をつくる「重回帰分析」で研究の傾向と生活満足度が連動することを示した。

入賞



一般 広島市立大学 ニュース記事と特許を利用した科学技術の重要性の評価

科学技術の重要性を社会的、経済的側面から評価する取り組み。特許分類の自動付与技術などで分野を分けることで、「学術論文」や「特許」、「技術開発のニュース記事」の多い分野を比較した。ニュースは日用品開発の話題が多いこと、医学特許や物理系の研究論文が増えたことを見いだした。



審査委員長の長尾真京都大学名誉教授（写真左）をはじめ、情報科学分野に詳しい各界の有識者8名が審査員を務めた。本選の発表では、成長への期待を込めて鋭い指摘や質問が投げかけられた。

義塾大学大学院健康マネジメント研究科の「健マネ」チームは、「健康」の指標として内閣府世論調査の生活満足度・充実度データを選び、国内文献の分野別統計データとの相関を調べることで、科学技術の研究が時代の変遷とともにどう人々の健やかさに寄与してきたかを分析した。

また分野データの解析で得られた研究動向の転換期について同時期の主な出来事を調べ、医学研究は技術進歩によって増え、環境に影響を及ぼす研究は社会の要請により増減したなど、科学技術と社会の関わりも考察した。

東京工業大学の「Terano Lab.」チー

ムは、論文の共著関係からわかる研究者ネットワークを解析し、各研究分野内や異分野間の研究をけん引する研究者を統計的に特定しようと取り組んだ。エネルギー関連の各分野については、それぞれ中心的研究者がごく少数存在し、研究コミュニティをつなぎ合わせていることを見いだした。

情報は大切な資源

審査委員の鈴木良介さん（野村総合研究所主任コンサルタント）は、「JSTから提供されたデータは、論文の書誌情報などかなり特殊なものなので、相当苦労されたはず。それを、どのチームもうまく使いこなしていて、アイデアが実におもしろいと思いました」と振り返った。

また、データそのものを重視する考え方は高まっているものの、それがいかに大きな可能性を秘めているかとの認識がまだ低いことを指摘。「情報は出せば終わりではありません。情報を資源ととらえ、それを分析して社会に役立てることが大切。そのために、一緒に有効な方法を考えていくことが必要です」と、今回のコンテストを「データの使い方を考える先鋭的な取り組み」と高く評価した。

審査委員長の長尾真さん（JST科学技術情報特別主監、京都大学名誉教授）は、「ビッグデータ分析のための人材育成と、JSTの今後の活動にとっても有意義でした」と総括。「JSTの科学技術情報を思わぬところに関連付けていくことが非常に面白い。私自身も、この貴重なデータをもっと有効に使わなければもったいないと思っていました。今回いろいろな形で使い、関心を持ってもらえたことを喜ばしく感じました」と評価した。

ビッグデータはこれからの社会に革命をもたらすとされている。今回の作品は、まさにそれを予感させてくれるものばかり。コンテストを通して、高校生を含む多くの若者たちがデータサイエンスに関心を持っていることもわかった。これをきっかけに、多くの若者たちが未来を担うデータサイエンティストとして羽ばたくことを期待したい。



データサイエンスへの関心の高さを実感

JST情報企画部 新規事業推進担当 伊藤 祥

コンテストを企画したきっかけは、JSTが築いてきた膨大な量のデータの新たな価値を見出すために外部の知を積極的に取り入れたいと考えたからです。また、コンテスト形式にすることによって、ビッグデータ時代に活躍できる人材が切磋琢磨する機会も提供できると考えました。これまでに科学技術データを使ってアイデア創出から分析・考察までのスキルを一貫して競い合うコンテストはほとんどなかったため、オープンデータやデータサイエンスに関心が高まりつつあるこの時期にどこよりも先がけてやろうということになりました。幸い、データの活用や人材育成の必要性に対する思いが、

SAS社をはじめとする企業、団体、審査員の先生方とも合致して、実現することができました。

本当に参加者が集まるのか不安もありました。若い世代にも積極的に参加してもらいたかったので、全国の高校などに声をかけて説明に奔走しました。最終的に8つの高校生チームから応募があったことを、審査員の先生方も大変喜ばれていました。企業からの応募も多数あり、実社会にデータサイエンスへの関心が広がっていることを実感しました。

より多くの人にデータサイエンスへの興味をもってもらい、未来のデータサイエンティストを育成するためにも、今後もこの取り組みを続けていきたいと考えています。