第2回データサイエンス・アドベンチャー杯

日程 2015年3月7日(土)

場所 独立行政法人科学技術振興機構

東京本部別館(K's五番町)1階ホール

主催 独立行政法人科学技術振興機構

■プラチナスポンサー : SAS Institute Japan株式会社, 株式会社ジー・サーチ

■コンテンツプロバイダー : 株式会社日刊工業新聞社, 株式会社日経BP

■後援 : 経済産業省, 文部科学省, データサイエンティスト協会

■協賛 : NRIサイバーパテント株式会社, ネイチャーインサイト株式会社

情報管理, 2015, vol. 58, no. 2, p. 148-153. doi: <http://dx.doi.org/10.1241/johokanri.58.148>

1. はじめに

近年、欧米を中心とした各国において、公的機関によるデータがオープン化され、その利用が可能になっている。わが国でも、経済産業省が2014年にOpen DATA METIというWebサイトを開設し、政府系データのオープン化を開始したところである。さらに、そうした利用可能なデータやソフトウェア等を活用し、かつさまざまな人々が参加可能となるオープンサイエンスやオープンイノベーションといった概念が一般化しつつある。その意味で、多様な分野の科学者、企業に所属する方々、大学生、高校生などさまざまな人々が本イベントに参加し、その創り上げる知見をもって競い合うことには大きな意義があると考えている。

独立行政法人科学技術振興機構（以下、JST）は、このたび、科学技術データ等を利用した分析アイデアや分析スキルを競い合うコンテスト「All Analytics Championship ～第2回データサイエンス・アドベンチャー杯～」(以下、アドベンチャー杯)を開催した。

今回は、前回¹⁾同様、データサイエンティストの発掘・育成機会の提供、ツールやコンテンツの利用開発を開催目的とした。前回からさまざまな形で協力いただいたSAS Institute Japan株式会社、株式会社ジー・サーチに加えて、今回はアドベンチャー杯の

開催目的に賛同いただいた株式会社日経BP、株式会社日刊工業新聞社にコンテンツプロバイダーとして新たに協力いただいた。その結果、コンテスト参加者に対して前回同様にJSTとジー・サーチからは科学技術データを、SAS Institute Japanからは分析ツールを利用可能としたことに加え、日経BPからは雑誌記事データ、日刊工業新聞社からは新聞記事データが利用可能となり、多様なデータを活用し、さまざまな分野でのデータ分析に関心のある方々が参加できる環境を整えた。

2. 研究課題と応募作品

アドベンチャー杯の研究課題に対する基本的な方向性は、「JST科学技術データ等を使用して、柔軟な発想で研究課題を自由に設定し、統計・データ分析を行う」と設定した。応募作品は一般部門と言語部門の2部門で募集を行うこととした。一般部門は、分析手法を中心とした作品を想定するが、今回新設した言語部門は、自然言語処理等を用いてデータの分析を行う作品を想定した。また、U-18賞を設け、高校生のチームのトップが受賞することとした。

応募作品は、分析テーマ、分析手法・プロセスならびに分析結果を1) 分析概要（指定書式有り）および2) プレゼンテーション資料（PowerPoint形式）の

2点にまとめて提出することとした。なお、2) のプレゼンテーション資料については、本選に進出した際にそのまま発表資料となることを想定し、15分間の内容にまとめることを義務付けた。ただし、分析プロセスなどデータ手法の正当性を判断するために必要な情報は、本選の発表では省略することも可能とし、できる限り詳しく記述してもらうようにした。

3. エントリー開始から本選までの流れ

アドベンチャー杯は、2014年9月16日からエントリーを開始し、2015年1月30日に応募作品を締め切った。2月16日に審査委員による予選(書類選考・非公開)を行い、予選審査を通過した9チームが3月7日の本選にてプレゼンテーションを行い、各賞が決定した。

エントリー数は81チーム、実際に提出された応募作品数は25にのぼった。提出された25作品のセクター別内訳(代表者所属)は、一般企業が3、公的研究機関が2、高校が6、大学が14となっている(図表1)。都道府県別でみると、首都圏が多いものの全国各地から作品が寄せられていることがわかる(図表2)。

次に、分析に使用されたデータについてみていく。アドベンチャー杯のために提供したJST科学技術データ等の利用状況は、表1のとおりであり、JSTのデータでは科学技術文献データと分野分類のデータが多く使用されている。また、新聞記事データを利用したのは4チーム、雑誌記事データを利用したのは6チームであった。なお、各チームのオープンデータ利用

率は4割であった。

4. 審査方法と審査基準

アドベンチャー杯における審査基準は、1) JST科学技術データを使用した最良の活用アイデアが示されているか、2) 実用化、産業化につながる視点が織り交ぜられているか、3) 分析が適切に行われ、分析結果が効果的なプレゼンテーション資料としてまとめられているか、の3点である。これらは、エントリー開始と同時に公開した。本選の審査では、予選審査の結果と発表者のプレゼンテーション内容を総合的に判断し審査委員会で協議のうえ各賞を決定した。審査委員会は、表2に示す各界の有識者により構成され、予選・本選ともに厳正なる審査が行われた。

5. 本選の様子および結果

開催当日は休日であいにくの雨天だったにもかかわらず、本選会場には一般観覧者と発表者・関係者を合わせて約200名が集まり、熱のこもった発表や質疑応答に注意深く耳を傾けていた(事前登録は216名)(図1)。

本選進出者は表3に示す9チームであり、それぞれ15分間の発表後に審査委員およびアドバイザーによる5分間の質疑応答が行われた。9チームの発表が終了した後、審査委員会による審査が行われ各賞が決定した。

今回のアドベンチャー杯ではJSTの提供するデータ

セクター	提出数
一般企業	3
公的研究機関	2
高校	6
大学	14
total	25

図表1 所属セクター別作品提出数

都道府県	提出数
神奈川県	8
東京都	5
福岡県	3
京都府	3
兵庫県	2
宮城県	1
埼玉県	1
千葉県	1
茨城県	1
total	25

図表2 地域別作品提出数

表1 貸与データ，ツールと利用件数

データ，ツール名	U-18	言語	一般	合計
■日刊工業新聞社 新聞記事データ		1	3	4
■日経BP社 雑誌記事データ	3	1	2	6
■JSTデータ	-	-	-	-
◎JST一次加工データ（国内・海外文献別）	-	-	-	-
・分野分類の統計データ			2	2
・シソーラス用語の統計データ	2		1	3
・化学物質の統計データ				0
◎ 科学技術文献データ：約1,000万件 （抄録は2010-2012年の300万件のみ）		3	9	12
◎ 資料名データ：約20万件	1	1		2
◎ 分野分類データ：約4,000件		1	6	7
◎ 人名名寄せデータ：約4,000万件			4	4
◎ 機関名データ：約30万件			3	3
◎ 科学技術用語シソーラス：約4万件			3	3
◎ 同義語・異表記辞書データ：約70万件		1		1
◎ 化学物質データ：約300万件			1	1
■ オープンデータ等（利用チーム数）	3		7	10
■ SASツール	1		1	2
合計利用件数	10	8	42	60
（参考）作品提出チーム数	6	5	14	25

に加えオープンデータや新聞・雑誌のデータを複雑に駆使して解析したハイレベルな作品が数多くみられた。今回の応募作品の特徴は、実用化につながるものが多かったことである。また、受賞者の方々は、データサイエンティストとして今後有望である等の寸評も聞かれた。

次に各部門の受賞作品（図2）について、概略を述べる。

5.1 U-18の部

チーム“埼玉県立熊谷女子高等学校”は、高校3年生の3人組で「今年の流行語を先取り!!!～論文数から見る流行研究の推測～」と題して、詳細なデータ分析を通じて流行語予測を行い、U-18賞を受賞した。

チーム“柏陽小町”（神奈川県立柏陽高等学校）は、「節電に対する人々の意識～東日本大震災が与えた影響～」をテーマに身近なエネルギー問題について分析を行い、データ活用につき提言を示し、日経ビッ

グデータ賞を受賞した。

5.2 言語部門

チーム“キュープラス”（九州大学と株式会社Lafla

図1 本選会場の様子

の混成チーム)は、「大学や研究所の研究活動がひと目で分る研究活動マップ生成～だれが、どこで、どんな活動をしてるの?～」をテーマに、関連する単語1つから研究活動マップを生成し、研究サーベイにかかる労力を軽減できるシステムについて研究を行い発表した。このシステムは、図書館の実運用でも利用できることを示し、研究内容および実用化の点から言語部門最優秀賞を受賞した。

チーム“T-linkage”（国立情報学研究所）は、「TermLink：言語横断論文推薦のための専門用語処理」をテーマに、複数の言語解析を行い、専門用語データベースの性能改善に有用な言語リソースを導き出した。このチームの作品は、この研究内容が評価され、言語部門優秀賞を受賞した。

表2 第2回アドベンチャー杯審査委員会

	氏名	所属・役職
審査委員長	長尾 真	JST科学技術情報特別主監, 京都大学名誉教授
審査委員	大向 一輝	国立情報学研究所 コンテンツ科学研究系 准教授
審査委員	小町 守	首都大学東京 システムデザイン学部 准教授
審査委員	福本 文代	山梨大学大学院 医学工学総合研究部 教授
審査委員	宿久 洋	同志社大学 文化情報学部 教授
特別審査委員	杉本 昭彦	日経デジタルマーケティング, 日経ビッグデータ 編集長
特別審査委員	渡瀬 博文	株式会社 ジー・サーチ 代表取締役社長

※審査委員, 特別審査委員は五十音順

表3 本選進出者一覧および各賞

発表順	部門	各賞	チーム名	代表者所属	テーマ
1	U-18	U-18賞	埼玉県立熊谷女子高等学校	埼玉県立熊谷女子高等学校	今年の流行語を先取り!!! ～論文数から見る流行研究の推測～
2		日経ビッグデータ賞	柏陽小町	神奈川県立柏陽高等学校	節電に対する人々の意識 ～東日本大震災が与えた影響～
3	言語	最優秀賞	キュープラス	九州大学付属図書館 など	大学や研究所の研究活動がひと目で分る研究活動マップ生成～だれが、どこで、どんな活動をしてるの?～
4		優秀賞	T-linkage	国立情報学研究所	TermLink：言語横断論文推薦のための専門用語処理
9	一般	最優秀賞	アズマー	東北大学	雑誌・新聞・JSTデータから見る「よりよい企業選択へのヒント」
8		優秀賞	都産技研チーム	東京都立産業技術 研究センター (開発本部)	専門家を探せ！一社内に眠る人材を活用しようー
5		SAS賞	TEAM KK	慶應義塾大学大学院 など	医学論文数から見る、政策誘導と研究動向～「褥瘡」から見える動き～
6		ジー・サーチ賞	TeranoLab	東京工業大学	研究を主導する次世代のリーダーを探る
7		入賞	若鶏の醤油揚げ	京都大学	JST科学技術文献データのネットワーク分析-シソーラス, 組織, 分野の分析を通じて

図2 受賞者と審査委員の皆さま

5.3 一般部門

チーム“アズマー”（東北大学）は、「雑誌・新聞・JSTデータから見る『よりよい企業選択へのヒント』」と題して、就職活動という現実の壁を乗り越えるためにステップごとに多様なデータを適切に組み合わせて分析し、一般部門最優秀賞を受賞した。

チーム“都産技研チーム”（東京都立産業技術研究センター 開発本部）は、「専門家を探せ！～社内に眠る人材を活用しよう～」と題して、社会的な要請が高い専門家マッチングシステムを開発し、実際の相談に活用するなど実用化し、一般部門優秀賞を受賞した。

チーム“TEAM KK”（慶應義塾大学大学院 健康マネジメント研究科と卒業生の混成チーム）は、「医学論文数から見る、政策誘導と研究動向～『^{じよくそう}痔瘡』から見える動き～」というテーマで、社会的関心の高いテーマの論文数増加を、研究内容と著者タイプを切り口に分析し、SAS賞を受賞した。

チーム“TeranoLab”（東京工業大学）は、「研究を主導する次世代のリーダーを探る」をテーマに、研究成果と競争的資金の関係を多面的に分析し、次世代リーダーの特徴を明らかにし、ジー・サーチ賞を受賞した。

チーム“若鶏の醤油揚げ”（京都大学工学部 情報学科数理コース）は、「JST科学技術文献データのネットワーク分析ーシソーラス、組織、分野の分析を通

じて」と題して、分類、組織、意味ネットワークの構造を見いだした興味深い研究を行い、入賞を果たした。

6. おわりに

今回のアドベンチャー杯では、審査基準に「実用化、産業化につながる視点が織り交ぜられているか」を加えたため、具体的に出口がわかりやすく、地に足のついた応募作品が多かったと思われる。その中で、深く正確な分析が行われているものが本選に出場し、各賞を受賞したように思える。

また、若手が数多く参加し、データサイエンティストの発掘・育成機会を提供するという目的を達成することができた。さらに、JST情報事業やコンテンツ提供者にとっては、認知度向上や他のデータの親和性を知る機会、また活用に関する外部の知の取り込み、今後強力なサポーターとなりうる機関・人物との出会いなど、さまざまな効果を得ることができた。

最後に、アドベンチャー杯の開催にあたって、コンテンツをご提供いただいた日経BP、日刊工業新聞社、スポンサーとして多大なご支援をいただいた協賛各社の皆さま（SAS Institute Japan株式会社、株式会社ジー・サーチ、NRIサイバーパテント株式会社、ネイチャーインサイト株式会社）、この取り組みに賛同いただき応援をしてくださった後援団体の皆さま（経済産業省、文部科学省、データサイエンティスト

協会), すべての応募作品に目を通し, 眞摯^{しんし}に審査を行ってくださった審査委員の先生方, エントリーをしてくださった全国の皆さま, 本選会場にお越しいただいたすべての皆さまに心より御礼申し上げます。

アドベンチャー杯公式Webサイト :
<http://www.sascom.jp/AAC/>

(科学技術振興機構 火口正芳, 木下健太郎)

参考文献

- 1) 伊藤祥. データサイエンス・アドベンチャー杯. 情報管理. 2014, vol. 57, no. 1, p. 57-61. <http://dx.doi.org/10.1241/johokanri.57.57>, (accessed 2015-04-06).