データサイエンス・アドベンチャー杯

日程 2014年3月8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 SAS Institute Japan株式会社
独立行政法人科学技術振興機構

情報管理 57(1), 057-061, doi: 10.1241/johokanri.57.57 (<http://dx.doi.org/10.1241/johokanri.57.57>)

1. はじめに

独立行政法人科学技術振興機構（以下、JST）は、このたびSAS Institute Japan株式会社（以下、SAS社）との共催で、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 Powered by SAS 〜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。

このアドベンチャー杯は、データサイエンティストを志向する人々が分析アイデアや分析スキルを切磋琢磨できる機会を提供することを目的に高校生以上を対象に実施した。

データサイエンティストとは、「統計解析やデータ分析に関する高度な知識を持ち、情報を正確かつ効果的に分析することで、ビジネスに役立つ洞察や課題解決のための有効な知見を見出すことができる人」を指し、Google チーフエコノミストのHal R. Varian氏の言葉にもあるように、“今後10年間で最もセクシーな職業”として世界中で注目されている。

しかし、日本ではこのデータサイエンティストの人材不足が指摘されており、アドベンチャー杯はその現状を打開する契機となることを目指した。

今回、データサイエンティストの発掘・育成機会の提供という開催目的で合意した民間企業のSAS社とタッグを組み、コンテスト参加者に対してJSTからは分析のためのデータを、SAS社からは分析ツールを無償で提供することにより、データ分析に関心のある誰もが参加できる環境を整えた。

2. 課題と応募作品

アドベンチャー杯の課題は、「JST科学技術データを使用して、柔軟な発想でテーマを自由に設定し、統計・データ分析を行う」と設定した。なお、JST科学技術データを使用していれば他のオープンデータ等を組み合わせた分析も歓迎するということを明記した。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。

JST科学技術データとは、JSTが長年にわたり収集・整理体系化してきた科学技術分野の論文に由来するデータ群をいう。今回それらのデータの一部を、統計処理がしやすいようJST側で一次加工を行ったデータ（タイプA）と応募者が最初から自由にデータを扱えるローデータ（タイプB）という形で提供した（表1）。

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、2013年10月8日からエントリーを開始し、2014年1月31日に応募作品を締め切った。応募作品は一般部門と18歳以下のチームが応募できるU-18部門の2部門で募集を行った。その後、2

表1 提供したJST科学技術データ

<p>■一次加工データ（タイプA）</p> <ul style="list-style-type: none"> ・対象年：1981年-2012年（文献が収録された資料の発行年） ◎ 分野分類の年次別統計データ（国内・海外文献別） <ul style="list-style-type: none"> - 科学技術文献データに索引されたJST分類コード24分野の年次別索引頻度 ◎ シソーラス用語（統制語）の年次別統計データ（国内・海外文献別） <ul style="list-style-type: none"> - 科学技術文献データに索引されたシソーラス用語の年次別索引頻度 ◎ 化学物質の年次別統計データ（国内・海外文献別） <ul style="list-style-type: none"> - 科学技術文献データに索引された化学物質の年次別索引頻度と最初に索引された年
<p>■ローデータ（タイプB）</p> <ul style="list-style-type: none"> ・対象年：2003年-2012年（文献が収録された資料の発行年） ◎ 科学技術文献データ（書誌情報）：約1,000万件 ◎ 資料名データ：約20万件 ◎ 分野分類データ：約4,000件 ◎ 人名名寄せデータ：約4,000万件 ◎ 機関名データ：約30万件 ◎ 科学技術用語シソーラス：約4万件 ◎ 同義語・異表記辞書データ：約70万件 ◎ 化学物質データ：約300万件

月17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8作品が3月8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86チーム、実際に提出された応募作品数は34にのぼった。

提出された34作品のセクター別内訳は、企業が18、大学が8、高校が8となっている（図1）。都道府県別でみると、東京都が多いものの北海道から徳島県まで全国各地から作品が寄せられていることがわかる（図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科学技術データの利用状況は、表2のとおりとなっており、一次加工データでは、分野分類とシソーラスの統計データが、ローデータでは、科学技術文献データ（書誌データ）と分野分類のデータが多く使用されている。また、今回のアドベンチャー杯ではJST科学技術データにオープンデータを組み合わせて分析することを推奨したため、各チームのオープンデータ利用率を見てみると、全体として6割を超える作品が何らかのオープンデータを使用していることがわかる（表3）。

図1 応募作品内訳（セクター別）

図2 応募作品内訳（都道府県別）

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、①JST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データ分析の中心となる部分でSASのソフトウェアが使用されているか、③データ分析手法

表2 使用されたJST科学技術データ

タイプA（一次加工データ）	チーム数
化学物質の統計データ（国内文献）	2
化学物質の統計データ（海外文献）	1
分野分類の統計データ（国内文献）	14
分野分類の統計データ（海外文献）	6
シソーラス用語の統計データ（国内文献）	13
シソーラス用語の統計データ（海外文献）	6
タイプB（ローデータ）	チーム数
科学技術文献データ（書誌情報）	13
人名名寄せデータ	3
分野分類データ	10
資料名データ	1
機関名データ	4
同義語・異表記辞書データ	3
科学技術用語シソーラス	3

表3 オープンデータ利用率

セクター	企業	大学	高校	全体
チーム数	18	8	8	34
OpenData利用有り	12	4	5	21
OpenData利用率	66.7%	50.0%	62.5%	61.8%

に正当性があるか、④分析結果が明確に示され、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4点である。これらは、エントリー開始と同時に公開した。

予選審査では、基準①、③、④について審査委員が審査を実施した。基準②については、SAS社にてSASソフトウェアの使用有無を確認したところ不適合作品はなかったため、全員が基礎点を得た。さらに、基準③については審査委員による審査と並行して、統計学的見地からみた正当性をアドバイザーが確認し審査の参考情報とした。

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4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

休日の開催にもかかわらず、本選会場には一般観覧者と発表者・関係者を併せて200名を超える人々が集まり、終始和やかな雰囲気で行進した（図3）。

本選進出者は表5に示す8チームであり、それぞれ15分間の発表の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。

最初に発表を行ったチーム“埼玉県立熊谷女子高等学校”は、高校3年生の4人組で「大学のそこところ～おカネと人と論文と～」というタイトルの作品を鋭くユーモアに富んだ考察を交えながら発表した。2番目のチーム“Terano Lab.”は東京工業大学の大学院生のチームで、「論文の共著関係ネットワークの中心性分析」をテーマにJSTの戦略立案等のヒントになるような共著関係のネットワーク分析について発表した。3番目のチーム“HSE研開部”は、株式会社日立ソリューションズ東日本のメンバーによるもので、「研究力の向上と実社会の発展の関係分析」を、仮説設定から考察までをぶれずにかつ丁寧に行った結果について発表した。4番目のチーム“健マネ”は、慶應義塾大学大学院で健康マネジメント学を専攻し

表4 アドベンチャー杯 審査委員会

審査委員会メンバー：

- 審査委員長 長尾 真 氏（JST科学技術情報特別主監、京都大学名誉教授）
- 審査委員（50音順）
 - 大向 一輝 氏（国立情報学研究所 コンテンツ科学研究系 准教授）
 - 小野寺 夏生 氏（一般社団法人情報科学技術協会（INFOSTA）会長）
 - 鈴木 良介 氏（株式会社野村総合研究所 主任コンサルタント）
 - 林 雅之 氏（NTTコミュニケーションズ株式会社 主査、国際大学GLOCOM 客員研究員）
 - 南川 敦宣 氏（KDDI株式会社 新規ビジネス推進本部 課長補佐）
 - 森川 富昭 氏（慶應義塾大学 大学院政策・メディア研究科／環境情報学部 准教授）
- アドバイザー（50音順）
 - 国友 直人 氏（日本統計学会 会長、東京大学 経済学部 教授）
 - 吉田 靖 氏（東京経済大学 経営学部 教授）

図3 本選会場の様子

ている学生によるもので、「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」について、JSTのデータと内閣府のオープンデータをうまく組み合わせた分析結果を発表した。

5番目に発表したチーム“UNAGI”は、東京ガス株式会社で社内の業務効率化や組織改善のためにデータ分析を行っている部署のメンバーによるもので、「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～」というテーマで、社会ニーズを意識し具体的な活用イメージを盛り込んだ発表を行った。6番目に発表したチーム“広島市立大学”は、「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」と題して、文書分類や情報抽出など今後のJSTにとっても非常に参考になる技術が盛り込まれた発表を行った。7番目に発表したチーム“北海道札幌旭丘高等学校 生物部”は、高校1年生による

ものであり、「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～」と題して、自校で5年間にわたり集積したオリジナルのデータを丁寧に分析・考察した後にJSTのデータを組み合わせるというアイデアに富んだ発表を行った。最後に発表したチーム“技術動向観測隊”は、株式会社金融エンジニアリング・グループ、国立情報学研究所など、3機関のメンバーから構成されるチームで、「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」について、分析結果を示すだけでなくその成果によって開発したWebアプリも併せて発表した。

6. 受賞者紹介

8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。

アドベンチャー杯における各賞の詳細と該当する応募区分は、表6に示すとおりである。

金賞はチーム“UNAGI”、銀賞はチーム“技術動向観測隊”、銅賞はチーム“HSE研開部”、アイデア賞はチーム“北海道札幌旭丘高等学校生物部”、U-18賞はチーム“埼玉県立熊谷女子高等学校”が獲得した（図4）。

審査委員会では、予選同様に審査委員とアドバイザーによる活発なディスカッションが行われた。また、今回のアドベンチャー杯でアイデアだけに留まらず、各自が設定した仮説を丁寧に検証し、それを

表5 本選進出者一覧と作品名

チーム名	作品名
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～
Terano Lab.	論文の共著関係ネットワークの中心性分析
HSE研開部	研究力の向上と実社会の発展の関係分析
健マネ	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
UNAGI	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～
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
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～
技術動向観測隊	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

表6 アドベンチャー杯 各賞の詳細

各賞と受賞者名	授与	副賞・記念品	対象区分
金賞 "UNAGI"	表彰状・トロフィー	・賞金15万円 ・未来館チケット ・長尾委員長サイン入り著書 ・SAS/バッグ	一般区分の 1チーム
銀賞 "技術動向観測隊"	表彰状・トロフィー	・賞金10万円 ・未来館チケット ・長尾委員長サイン入り著書 ・SAS/バッグ	一般区分の 1チーム
銅賞 "HSE研開部"	表彰状・トロフィー	・賞金8万円 ・未来館チケット ・長尾委員長サイン入り著書 ・SAS/バッグ	一般区分の 1チーム
アイデア賞 "北海道札幌旭丘高等学校生物部"	表彰状・トロフィー	・賞金5万円 ・未来館チケット ・長尾委員長サイン入り著書 ・SAS/バッグ	一般/U-18 のいずれか 1チーム
U-18賞 "埼玉県立熊谷女子高等学校"	表彰状・トロフィー	・賞金5万円 ・未来館チケット ・長尾委員長サイン入り著書 ・SAS/バッグ	U-18の 1チーム
入賞 "Terano Lab." "広島市立大学" "健でネ"	表彰状	・未来館チケット ・長尾委員長サイン入り著書 ・SAS/バッグ ・SAS水筒	本選出場者 のうち 上記以外の 3チーム

提言にまとめあげたハイレベルな作品がこぞって発表されたことを喜ぶ声、受賞者たちへの未来のデータサイエンティストとしての活躍を期待する声が聞かれた。

7. おわりに

今回のアドベンチャー杯の開催を通じて、ビッグデータ、オープンデータ時代において活躍する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。また、JST情報事業にとっては、JSTの情報資産である科学技術データの認知度向上やJST科学技術データとオープンデータの親和性を知る機会、JST科学技術データのオープン化に向けた課題の抽出・整理、JST科学技術データの活用に関する外部の知の取り込み、今後事業を推進するにあたって強力なサポーターとなりうる機関・人物とのネットワーク強化など、さまざまな良い効果を得ることができた。

今後もこのような機会を継続することにより、科学技術分野のデータから新たな価値の創造と、高度

ICT利活用人材の育成を通じた日本のイノベーション促進への貢献につなげていきたいと考えている。

最後に、アドベンチャー杯の開催にあたって共催者として多大なご支援・ご協力をいただいたSAS社、ならびに運営事務局として全面的なフォローをしてくださった株式会社エムエム総研、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（KDDI株式会社、株式会社ジー・サーチ、情報処理学会、日本計算機統計学会、日本統計学会、ネイチャーインサイト株式会社、株式会社マイナビ）ならびに来賓の皆様、すべての応募作品に目を通し、^{しんし}真摯に審査を行ってくださった審査委員・アドバイザーの先生方、エントリーをしてくださった全国の皆様、本選会場にお越しいただいたすべての皆様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/>

((独) 科学技術振興機構 伊藤 祥)

図4 受賞者と審査委員の皆様