

自然言語処理技術を用いたテキストマイニングによる 受講生の意識の分析

— 津山高専ジュニアドクター育成塾での学びについて —

吉富 秀樹* 宮下 卓也** 谷口 圭輔*** 山田 貴史****
中村 直人***** 寺元 貴幸**

Analysis of Student Learning Awareness by Text Mining using Natural Language Processing Technology

— Learning at Junior Doctor Science School of Tsuyama College —

YOSHITOMI Hideki, MIYASHITA Takuya, TANIGUCHI Keisuke, YAMADA Takafumi,
NAKAMURA Naoto and TERAMOTO Takayuki

The Junior Doctor Science School (hereafter, JDSS) is a project developed by the Japan Science and Technology Agency (hereafter, JST). Tsuyama College was adopted in 2018 and is working on this project. JDSS seeks out elementary and junior high school students with high motivation and outstanding abilities and trains their abilities through learning in the fields of science and mathematics and information, with the aim of encouraging outstanding human resources who will lead future science and technology innovation. Since the purpose of JDSS is to develop and implement human resource development programs, we have set goals for human resource development. It is necessary to measure and evaluate the growth of abilities. We are working on diverse learning evaluations using portfolios and worksheets at each stage of diagnostic, formative, and comprehensive stages. Since this project is the final year to receive support from JST, what did the students who have learned at our JDSS so far feel from what they have learned and what will they aim for in the future. In this report, we report on the results of analysis from a descriptive questionnaire on the changes of learning awareness budding in the minds of students.

Key Words: Junior Doctor Science School, Learning Awareness, Natural Language Processing, Text Mining

1. はじめに

ジュニアドクター育成塾は、国立研究開発法人科学技術振興機構（以下、JSTと略称する）が展開する公募型の次世代人材育成事業であり、小学校5・6年生および中学生を対象とし、将来の科学技術イノベーションを牽引する傑出した人材の育成に向けて、高い意欲や突出した能力を持つ小中学生を発掘

し、理数・情報分野の学習などを通じてその能力を伸長させる体系的な取り組みである¹⁾。2022年度現在、全国で高専、大学、NPO法人など30の機関で実施されている。津山工業高等専門学校（以下、津山高専または本校と称す）は、2018年度に採択され、JSTの支援のもと、この事業に取り組んでいる。津山高専ジュニアドクター育成塾（以下、本企画と称す）では、将来、イノベーターとして科学技術分野を牽引する100万人に一人の才能を見出し育成する体系的育成プログラムを開発・実施することを目的として、企画名を『「5 σ 」の逸材へ、発掘して育てるジュニアドクター育成塾』とした²⁾。ここに“「5 σ 」の逸材”とは、“100万人に一人の逸材”のことを統計学の標準偏差 σ を用いて表した言葉である。受講生は、毎年40名程度を新規に募集している。JSTの支

原稿受付 令和4年9月2日

* 津山高専ジュニアドクター育成塾 シニアメンター

** 総合理工学科 情報システム系

*** 総合理工学科 先進科学系

**** 総合理工学科 機械システム系

***** 総合理工学科 電気電子システム系

援期間は最長5年間となっており、本企画では本年度(2022年度)が支援を受ける最終年度となっている。

ジュニアドクター育成塾は、人材育成プログラムの開発・実施を目的としていることから、人材育成の目標を定め、受講生の成長を測定・評価することが求められる。本企画では、診断的、形成的、総括的な各段階において、ポートフォリオやワークシートおよび受講生アンケートなどを活用して多様な学習評価に取り組んでいる³⁾。

本報では、本企画が本年度(2022年度)でJSTの支援を受ける最終年度に当たることから、これまでに本校ジュニアドクター育成塾で学んだ受講生が、学びの中から何を感じたのか、そして、これから何を目指していくのか、受講生の心の中に芽生えている意識について自由記述式アンケートから分析した結果を報告する。

2. 育成課程と学びの概要

本企画の体系的育成プログラムの内容は既報⁴⁾にて報告しているため、ここでは、育成課程と学びの概要について説明する。育成課程は、第一段階プログラムと第二段階プログラムの2つの課程に分かれている。第一段階プログラムは、1年間だけの単年度課程であり、新規受講生は全員が受講する。第一段階のねらいは、さまざまな内容の講座を受講することで子どもたちの興味や適性に気付かせること、探究活動を通して研究倫理や情報リテラシーおよび実験スキルなどの第二段階へ進むための基礎的素養を学ぶことである。第一段階プログラム修了生の中で第二段階へ進級を希望する者の中から数名を選抜し第二段階プログラムへ進ませる。第二段階プログラムは、研究活動とロボコン活動から成っている。研究活動では、受講生は本校の研究室に配属され、教員や学生メンターの指導を受けながら各自の研究テーマに取り組む。ロボコン活動は、地元津山市で開催される“津山ロボコン”への参加を目指す活動であり、指導はロボコン指導の経験が豊富な本校退職教員に依頼している。なお、第二段階プログラムは、複数年度の在籍が可能で、本人が希望すれば中学3年まで在籍できる。

3. 自由記述式アンケートと分析対象者

本報で分析の対象となるデータは、各年度の講座の最終回に実施している自由記述式アンケートである。このアンケートは、JSTが作成し津山高専が実施しているものであるが、データの利用についてはJSTで了承されている。ここに、データの利用につ

いてJSTに謝意を表す。このアンケートは、選択式アンケートと自由記述式アンケートから構成されているが、本報で使うのは自由記述式アンケートの部分である。自由記述式アンケートの設問は、次の2つである。

設問1:「ジュニアドクター育成塾に参加してよかったことを自由にお書きください」

設問2:「これから、ジュニアドクター育成塾で取り組みたいことや目指したいことをお書きください」

分析の対象者は、2018～2021年度にアンケートに回答した第一段階プログラムおよび第二段階プログラムの受講生99名である。なお、第二段階の受講生は、ほとんどの者が中学3年まで在籍しているので、毎年データを使うと、同一人物が毎年回答する状況となって回答内容が重複する可能性がある。そこで、2018～2020年度の第二段階データについては中学3年生のみをピックアップした。アンケートの最終年度に当たる2021年度については、第二段階についても受講生全員のデータを使った。

分析の対象となった受講生の学年、性別、育成課程(第一段階か第二段階か)の各属性に該当する人数を表1に示す。本報では、アンケートの回答内容と属性との対応関係についても分析する。なお、学年および育成課程はアンケート実施当時のものである。

表1 分析対象者の属性と人数

学年	人数	性別	人数	育成課程	人数
小5	36	男子	78	第一段階	83
小6	20	女子	21	第二段階	16
中1	19	計	99	計	99
中2	15				
中3	9				
計	99				

4. 分析方法

自由記述式アンケートを分析する場合、分析者がクロス集計表にまとめるなどの方法をとる場合があるが、データ量が多いと作業が大変であるし、属性との関係も分析するとなると人間の手作業ではなかなか難しい。そこで、本報では、近年注目されている“AIによる自然言語処理技術を用いたテキストマイニング”による分析を行う。テキストマイニングは、大量のテキストデータを統計的に扱うことができるとともに、解析用ソフトウェアを用いて機械的に処理するので分析者の主観が介在し難しく客観性が担保できるという特徴もある。

表2 設問1の抽出語と出現回数

抽出語	出現回数	抽出語	出現回数
知る	29	数学	5
色々	27	増える	5
プログラミング	11	3Dプリンタ	4
ロボット	11	たくさん	4
実験	11	嬉しい	4
興味	10	機械	4
高まる	10	協力	4
体験	10	考える	4
楽しい	9	高専	4
自分	9	今まで	4
分かる	9	作る	4
パソコン	8	授業	4
使う	8	色々	4
人	8	前より	4
学ぶ	7	勉強	4
学べる	7	歩行	4
思う	7	力	4
良い	7	ジュニアドクター	3
学校	6	ロボコン	3
教える	6	家	3
研究	6	楽しみ	3
知識	6	好き	3
普段	6	行く	3
理科	6	実際	3
科学	5	将来	3
参加	5	先生	3
出来る	5	全く	3
触れる	5	発電	3
新しい	5	友達	3

4. 1 分析ツール

テキストマイニングは、アンケート調査などの大量のテキストデータを、AIによる自然言語処理技術を用いて品詞単位の単語に分解し（これを形態素解析という）、多変量解析等の技術を用いて単語の出現頻度や単語間の結びつきの強さ（例えば、共起関係）などを分析することでデータ全体を理解しようとする手法である⁵⁾。本報で用いたテキストマイニングのツールは、樋口耕一氏によって開発された“KH Coder⁶⁾”を使用した。テキストマイニングのツールは従前は高価であったが、KH Coderはフリーソフトウェアとなっており誰でも利用できる。また、解析結果が図表などの視覚化された形でアウトプットされるので、統計学や多変量解析の専門的知識のない者でも分析が可能であり、近年、さまざまな研究に利用されている⁷⁾。

4. 2 分析条件

使用したKH CoderのバージョンはWindows版のKH Coder-3である。形態素解析には、ChaSen（茶筌）とMeCab（メカブ）が選択可能であるが、ここではChaSen（茶筌）を使った。予備解析を実行したところ、「ジュニアドクター」のような本来複合語として現れる語が「ジュニア」と「ドクター」のように分解された形で抽出されるので、このような語は複合語として強制抽出されるように設定した。

5. 分析結果

テキストマイニングによる分析は、前記の“設問1「ジュニアドクター育成塾に参加してよかったことを自由にお書きください」と“設問2「これから、ジュニアドクター育成塾で取り組みたいことや目指したいことをお書きください」”の2つについて、それぞれ別々に分析した。設問1からは、受講生が“学びの中から何を感じたのか”を分析する。また、設問2からは、“これから何を目指していくのか”を分析し、受講生の心の中に芽生えている意識について考察する。なお、以降では、テキストマイニングによる抽出語をカッコ付で表す。

5. 1 設問1の分析結果と考察

(1) 抽出語と出現回数 設問1について、テキストマイニングで抽出された語の総数は809語で、うち異なり語は268語であった。出現回数3回以上の抽出語を表2に示す。出現回数が10回以上の語に着目すると、「プログラミング」や「ロボット」などの「体験」を通して「色々」「知る」ことができ「実験」への「興味」が「高まる」と多くの

受講生が感じていることを示唆している。

(2) 抽出語の共起関係 語と語の関連性の強さについて、KH Coderの機能を使って共起ネットワーク図を出力して解析結果の“見える化”を図り分析を進めた。抽出語の共起ネットワークを図1に示す。語を囲む円が大きいほど出現回数が多いことを表している。共起関係はJaccard係数（語と語の出現状況の類似度を示す指標⁸⁾）で評価しており、語と語が線で結ばれているものが共起関係にある。線が太いほど共起性が強いことを示している。また、互いに線で結ばれている語の集団が、共起関係などの関連性が認められるクラスターとなっている。なお、図のタイトルにある“(出現4回以上)”の意味は、出現回数が4回以上の抽出語を使って解析しているという意味である（以下の図も同様）。

図に示されている抽出語の具体的な意味や使用状況を調べる場合は、KH CoderのKWICコンコー

から何を感じたのか”について要約すると，“普段の学校生活では体験できないロボットやプログラミングなどのいろいろなことを体験できて良かった”との意識を持っていることが確認できた。

5. 2 設問2の分析結果と考察

(1) 抽出語と出現回数 設問2について、テキストマイニングで抽出された語の総数は 548 語で、うち異なり語は 226 語であった。出現回数3回以上の抽出語を表3に示す。出現回数が10回以上の語に着目すると、「自分」は「プログラミング」や「ロボット」を「作る」「研究」や「実験」に「取り組む」ことを「思う」との意思表示が認められ、多くの受講生が「ロボット」への関心が高いことを示唆している。

(2) 設問2の抽出語の共起関係 設問2の抽出語の共起ネットワークを図3に示す。右端付近のクラスターでは、「自分」で「ロボット」の「プログラミング」に「取り組む」との意欲が認められる。中央上部では、「色々」な「実験」をして「科学」への「興味」を「高める」との意識が確認できる。中心付近では、「探究活動」をさらに「深める」、「今」行っている「研究」をさらに「深める」との研究への積極性が認められる。左端下部では、「不思議」に「思う」ことに挑戦し、もっと「知る」ことを目指したいとの意志が表明されている。

表3 設問2の抽出語と出現回数

抽出語	出現回数	抽出語	出現回数
自分	15	生物	4
ロボット	14	積極	4
実験	13	探究活動	4
プログラミング	11	挑戦	4
研究	11	発表	4
作る	11	ジュニアドクター	3
思う	11	活動	3
取り組む	11	興味	3
知る	11	好き	3
色々	9	考える	3
使う	7	行く	3
ロボコン	6	今年	3
将来	6	参加	3
第二段階	6	詳しい	3
数学	5	生かす	3
化学	4	探求	3
科学	4	知識	3
学ぶ	4	調べる	3
高める	4	不思議	3
今	4	理科	3
深める	4		

(3) 設問2の抽出語と学年の共起関係

設問2の抽出語と“学年”との共起関係を図4に

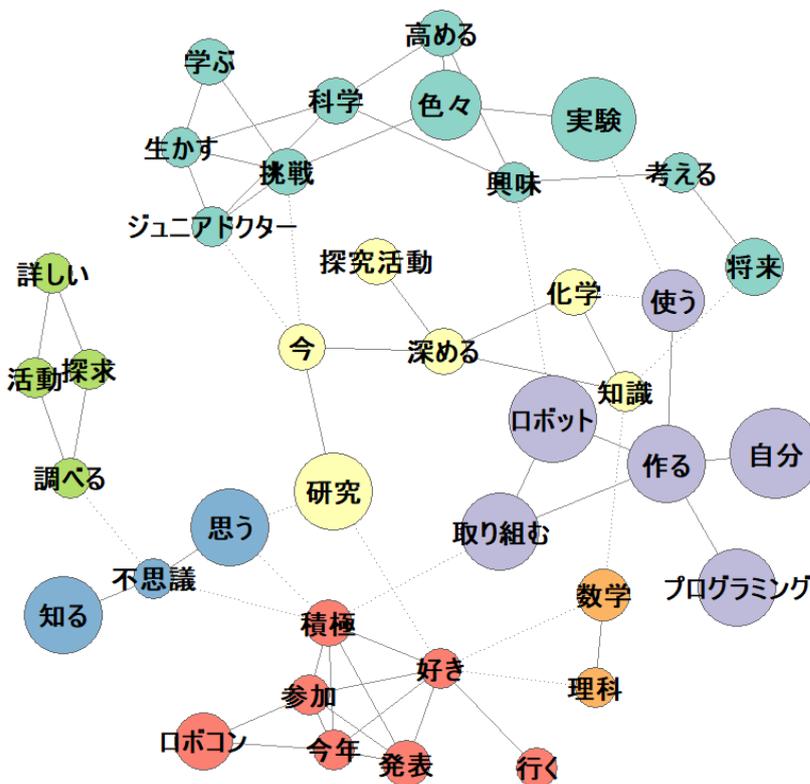


図3 設問2の抽出語の共起ネットワーク（出現3回以上）

示す。図の中心付近の「ロボット」や「研究」および「取り組み」などの抽出語は、すべての学年につながっており、多くの受講生が「ロボット」や「研究」に取り組む意志を持っていることがわかる。中1の周辺には「生物」の語が認められ、今後の活動として、「生物」の「知識」を「深める」との意欲が見て取れる。小6の周辺では、いろいろなことを「学ぶ」ことで将来「科学」の分野で活躍したいとの意志表明がある。また、小5の周辺では、「探究活動」の「発表」で頑張りたいこと、「数学」に興味を持つようになったことなどが示されている。

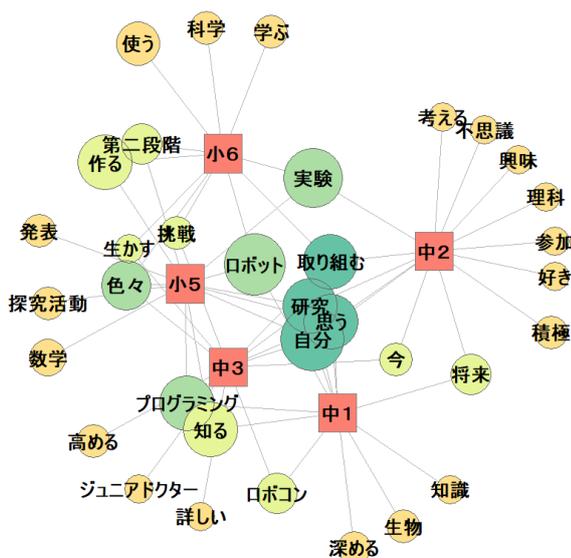


図4 設問2の抽出語と学年との共起関係（出現3回以上）

（4）抽出語と性別の対応分析 近年、次世代人材育成では、ジェンダーバイアスの解消が政策として挙げられている⁹⁾。そこで、表1に示した属性“性別”と抽出語の対応分析（コレスポネンス分析）を行った。対応分析の結果を図5に示す。この図は横軸と縦軸をもった散布図の形態をしているが、一般的な散布図とは仕組みが異なる。そこで、この図の見方を説明する。ポイントは以下の5点である。

- ①特徴のない語は原点付近に布置される
 - ②特徴の大きい語は原点から遠くに布置される
 - ③関連性の強い語どうしは、原点から見て同一方向に布置される
 - ④関連性の強い語どうしは近くに布置される
 - ⑤横軸と縦軸の意味を厳密に定義することは難しいが、事後的に意味付けすることが推奨される
- さて、図5では、円は抽出語を表し、円の大きさは出現回数に対応している。外部変数（性別）は四角で表示されており、“男子”と“女子”の2つである。軸の意味を事後的に考察すると、横軸は右に行くほど男子との関連が強く、縦軸は上に行くほど

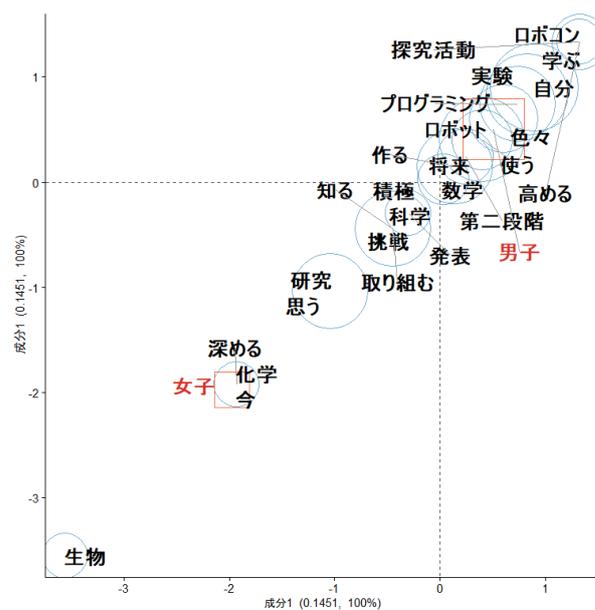


図5 設問2の抽出語と性別の対応分析（出現4回以上）

男子との関連が強い関係になっている。

図5によると、男子は「プログラミング」や「探究活動」および「実験」において「自分」を「高める」ことへの意識が強いことが特徴となっている。「ロボット」などの“ものづくり”への関心が高いこともわかる。一方、女子は「生物」や「化学」などの基礎的な理科教科への興味が高いことが特徴的である。また、「研究」を「深める」という意志も表明されている。

したがって、本分析結果からは、男子は「プログラミング」や「ロボット」などの応用的な分野、女子は「生物」や「化学」などの基礎科学的な分野に関心が高い傾向が認められるといえよう。

（5）抽出語と育成課程の対応分析 設問2の抽出語と育成課程（第一段階か第二段階か）との対応分析の結果を図6に示す。図の見方は、前記の図5と同様である。第二段階受講生は、研究活動とロボコン活動に取り組んでいるので、この2つの活動に関連した抽出語である「研究」と「ロボコン」が第二段階の特徴語として示されており、活動内容に対する興味や関心を反映していることが認められる。また、「今」が右上端に布置されており、第二段階のもっとも特徴的な抽出語となっている。多くの第二段階受講生が「今」やっている活動について言及していることがわかる。具体的には、「今」やっている「研究」や「ロボコン」を「深める」ように「積極」的に取り組みたいとの意志表明である。また、ロボコン活動に取り組んでいる者については、「ロボコン」に「挑戦」して優勝を目指したいとの記述も見られる。

