

10周年記念企画

授業実践報告

共同作業による文理融合教育の試み

—ジョイント・リサーチ「データから文化・社会を読む」の紹介—

鄭 躍軍・金 明哲・村上 征勝

1. はじめに

近年、容量 (Volume) の巨大さ、種類 (Variety) の多さ、速度 (Velocity) の早さ、価値 (Value) の多岐さを特徴とし、IT 技術の進展と利用拡大によって生まれたビッグデータが各界に注目されています。特に、ビッグデータの処理と活用に必要な理論と方法の開発が活発になっており、同分野の人材育成が強く求められています。

この時代の流れの中で、ビッグデータの複雑な管理・処理が扱えるコンピュータ技術の発達により、統計学が最も重視してきたデータの無作為性 (randomness) がそれほど重要なファクターではなくなる場合が増えていきます (鄭他, 2007)。データを中心に現象を解析するデータサイエンスは、まさに統計学から脱皮して誕生した学問分野です。データサイエンスという用語は、1960年代から様々な意味で使われてきましたが、統計数理研究所元所長故林知己夫氏は「データサイエンス (データの科学) が統計学、データ解析、分類、その他の関連諸方法を総括する概念」と先駆的に提案しました (林, 2001)。

今日では、データは科学研究だけでなく、産業界でも事実の記録やエビデンスの源として広く使われるようになってきました。それにともない、データサイエンスの専門知識をもつデータサイエンティストは最も魅力ある職業として、近々企業が重要視する人材となると予測されています。

文化情報学部は、情報化社会のニーズに応え、日本初の「データサイエンス」と「文化」の文理融合型学部として 2005 年に設置され、今年 10 歳の誕生日を迎えることになりました。文化情報

学部では、これまで自然科学の分析対象ではなかった人間の多様な営みを、データにより探究する理論と方法の開拓を目標に、教育活動に取り組んでいます。目標に合わせた教育理念は、データサイエンスを活用して文化現象を解析し、新たな社会価値を創出する人材を育成することにあります。

本学部は、ユニークなカリキュラムの導入により、「文」と「理」の枠を越え、文系得意な高校生も理系得意な高校生も育つ教育環境の提供を試みています。特に、データを適切な方法で収集、解析を行うことによって、新しい事実を浮かび上がらせるというデータサイエンス教育に一貫して取り組んでいます。図 1 に示しているように、学生は文化クラスター科目 (文化、言語、人間行動、人間社会) で文化の深みと広がりをおぼるとともに、データサイエンス科目 (基礎数理、データ分析、情報コンピュータ) では高度な知識探究の手法を身につけます。さらに、それらの力を総合して、「ジョイント・リサーチ」「卒業研究」で具体的な研究に取り組みます。

横糸と縦糸から構成された 7 つのクラスター科

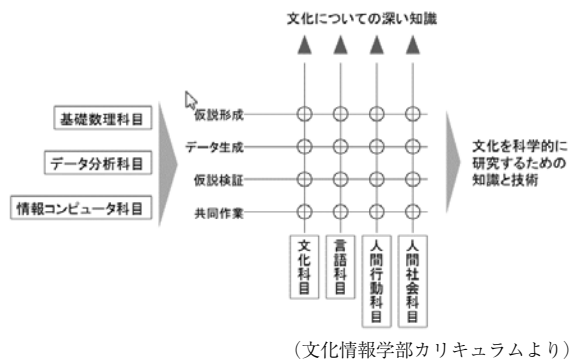


図 1. 文理融合型科目クラスター群

目群のうち、横系に位置する4つの文化クラスター科目は様々な文化領域について専門的な知識を学ぶための科目群、縦系に置かれている3つのデータサイエンス科目は文化を科学的な方法で研究するための技術と知識を学ぶための科目群です。横系と縦系が折り込むノードは「文理融合」による新たな知識を生み出すことを表しています。

カリキュラムのもう一つの特徴は、3年生を対象に「ジョイント・リサーチ」と名付ける科目群を設けていることにあります。ジョイント・リサーチ科目群は、小人数の共同作業で切磋琢磨する演習形式により文化を科学し、データサイエンスの諸技法を理論的かつ実践的に理解することを共通の目標としています。履修生は、文献レビュー、データの収集・分析、レポート作成など学問探究の基礎技法を学びながら、グループ単位での共同作業を行うことにより、あらかじめ答えがあるという保証のない問題に対し、自分の創意工夫で取り組む能力の形成に培います。また、自分とちがう考え・視点をもった人と課題に取り組むことで、視野を広げてゆくこともできます。

本報では、先般学会で報告した内容（鄭・金・村上，2013）を基に、データの設計、収集、解析をキーワードとするジョイント・リサーチ科目「データから文化・社会を読む」（以下、JR授業と省略する）の試みを紹介し、授業の狙い、内容概要、進め方及び教育効果について報告します。

2. JR 授業の目指すこと

JR 授業は、通年履修の必修科目として、グループ単位で創意・工夫を凝らした研究取り組みの実践を能動的に行わせることを目指します。その最大のねらいは、履修生が主役で、1年次と2年次の座学講義と文化情報学演習シリーズ（統計分析、文化、言語、人間行動、人間社会など）で学んだ文化現象や社会事象などの知識とデータサイエンスの実践力を活用し、問題を発見・解決する能力を鍛えることにあります。JR 授業は、グループ共同作業によるコミュニケーション、研究成果発表・ディスカッション、レポート作成などを通じて研究への理解を深め、4年生の「卒業研究」への布石を打つと位置づけています。

「データから文化・社会を読む」という JR 授業では、履修生に統計学的思考と実用性の高いデータ分析手法を理解させることと、それらの手

法を文化現象・社会現象の研究に適切に活用させ、独自に問題の本質を解明するプロセスと技能を修得させることを格段に重視します。

春学期の JR 授業（JR I）では、4～6人を1つのグループに分け、各グループが代表的なデータ分析手法を3つずつ選定した上で、メンバー間の連携により文献サーベイ、意見交換、データ分析、成果発表などの一連の共同作業を実践的に行うことになっています。これらの取り組みを行うことによって、履修者に「問題意識」「データ解析手順」及び「結果考察」を理解させます。最終的には、グループごとに人文社会科学分野によく用いられるデータサイエンスの諸手法の「数理的基礎」「適用可能なデータ」「活用要件」を、テキスト（鄭・金，2011，村上，2002）の活用、ディスカッションなどにより身につけさせます。

秋学期の JR 授業（JR II）では、履修生が春学期に身につけたデータサイエンスの諸方法を用いて、実際の文化現象・社会現象から問題意識を生み出させ、データから現象の本質を解明するための分析技能を実践的に修得させます。まず、各グループが関心のある研究テーマを選定した上で、関連する文献を調べ、閲読することによって、研究方法を確立します。次に、利用可能なデータを入手し、研究の目的に合わせて解析方法を選定します。次いでデータの収集・分析を実施し、得られた分析結果を考察します。最後に、問題意識、研究方法、データの収集・解析、結果考察を中心に、各グループが成果を発表することになります。

教員は、このような取り組みを通して、履修生全員に研究の問題意識、データの収集・活用の具体的なテクニックを実践的に理解させ、共同作業の進め方を修得させることを念頭に置いています。全体的な目標は、履修生全員がジョイント・リサーチの意義を理解した上で、メンバー間の連携のもとで自主的な発想による共同作業を展開する技能を身につけさせることです。

具体的に言えば、教員の主な役割は、履修生に授業で取りあげるデータサイエンスの方法論と文化現象・社会現象の専門知識を説明することと、グループ作業を指導することにあります。

3. 授業の内容と進め方

3.1 春学期の JR I

JR I では、文化現象・社会現象に関わるデー

タ分析に必要とする諸方法を修得させるための内容を組みました。授業では、グループごとに典型的なデータ分析方法を与え、関連する文献のサーベイや共同作業により、諸方法の数理的な考え方、適用データの特徴、分析手順及び分析結果の読み方などを理解させることを中心とします。

図2に示しているように、本授業では順を追って「関連分析」「予測」「分類」のために用いられる方法を取りあげています。

第1ステージでは、助走段階として「相関分析・クロス表における独立性検定・対応分析」を取り上げ、R言語を復習しながら、授業形式に慣れるよう工夫します。第2ステージでは、履修者が主役で「重回帰分析・数量化Ⅰ類・ロジスティック回帰」、そして第3ステージでは、「主成分分析・クラスター分析・数量化Ⅲ類」を中心に「予測」と「分類」の諸方法をリサーチします。

15回の授業では、各グループが「関連分析」「予測」「分類」に適用できるデータ分析方法から、1つずつ担当することになっています。なお、履修生にこれらの方法を着実に理解させるためには、データ分析方法を各グループに無作為に振り分けることにしています。それぞれの分析方法について、履修生が5回の授業時間を使い、数理的な基礎を理解した上で、応用例を用いて分析手順を実践的にリサーチし、関連の成果を発表します。

1回目は、所与の方法に関するテキストや授業配付資料の活用、グループ内のディスカッションなどにより、担当するデータ分析方法の統計的な考え方及び具体的な役割を理解します。

2回目は、各グループが意識疎通を行いながら、関心のあるデータ事例を調べ、分析用データを手入して整理します。続いて3回目は、R言語を用いてデータを分析する作業手順を実際に身につ



写真1. ディスカッション風景

け、データ分析のプロセスとアウトプットを実践的に理解します。4回目は、アウトプットの読み方、読み取れることから、分析方法の特徴を考察します。さらに、データ分析方法の概要、分析手順及び結果の読み方、データ例の説明、分析結果と所見を中心に、発表用資料を作成します。

5回目は、履修者が全員一堂に集まり、あらかじめ決められたメンバーの役割分担のもとで、グループごとのジョイント・リサーチ成果を発表し、リサーチの基礎力を着実に養います。発表では、他のグループの発表内容を聞いたり、質問したりすることによって、自分が担当しなかったデータ分析方法を理解することを中心とします。教員が発表を講評することによって共同作業のテクニックを履修生が互いに勉強し合うことも可能です。

図3は、「予測」用データ分析方法の1つである数量化Ⅰ類を担当したグループが考え出したデータ例を示すものです。酒専門店の1日当たりのビール販売量（量的被説明変数）を予測するには、天気状況、曜日、来店者の性別などの質的説明変数を選定しました。このような例を通して、

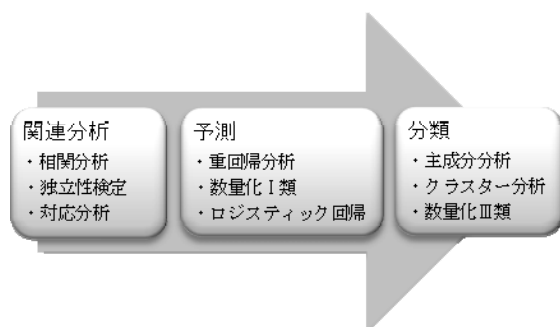


図2. JR Iで扱った統計方法と順序



写真2. データ分析作業風景



写真 3. 分析結果に喜ぶ学生たち

履修者が担当したデータ分析方法に関心を寄せて、真剣に方法の役割、使い方、特徴などについてしっかりと覚えることができます。

このような演習作業は、各グループが3回繰り返すことによって、「関連分析」「予測」「分類」のデータ分析方法の理解を徐々に深めます。



写真 4. クラスター分析グループの発表風景

説明変数：天気、曜日、来店者の性別



予測

被説明変数：
ビール販売量



図 3. 数量化 I 類のデータ例

3.2 秋学期の JR II

JR II は、典型的な文化現象または社会事象を計量的に解析するアプローチを修得させることを目的とします。グループごとに関心のある問題を選定した上で、メンバー間の連携により独自のリサーチを展開するアプローチを身につけさせることを中心に、授業内容を組みました。授業の進め方は、文化・社会を理解するためのリサーチをプロジェクト形式で展開していく仕組みです。

文化現象は、歴史的に成り立った象徴的、習得的な側面を示す言語、習慣、慣行などを意味するもので、多様性と変容性をもつことが特徴的です。一方、社会現象は、社会全体ないし特定の社会・集団で見られる行動・行為・状況などの営みであるため、複雑さと変わりやすさという特徴を有します。したがって、文化現象・社会現象の探究にあたって、データサイエンス手法の活用が効果的です。JR II の主な授業内容は、以下の2つの範疇から構成されています。

「意識分析」：調査実施方法、ワーディング、配列などが調査データの収集に与える影響を、実験調査及びデータ分析により理解させる内容

「テキスト分析」：文学作品、ブログ、口コミ、SNN (Social News Network) のコンテンツなどの統計的分析方法について学習・研究させる内容

2つの内容について、それぞれ7回の授業時間をかけて、問題意識の明確化から、データ獲得方法の確立、データ収集・分析、結果考察、成果発表までの一連の共同作業をこなすことになっています。なお、各グループが同じ期間に同一課題範疇を扱いますが、具体的なリサーチ中身が異なります。意識分析について、調査モード、ワーディング、配列などの影響を別々のグループが質問作成及び調査実施を行うことになります。また、テキスト分析について、ともにテキストマイニング手法を活用しますが、文学作品、ブログ、口コミなどの内容分析を異なるグループが担当することになります。

意識分析では、調査モード、質問の構文などが調査データの質にどのような影響を与えうるかをリサーチします。調査実験とデータ分析により、各ファクターの影響を解明することによって、調査に役に立つ知見を履修生に身につけさせます。

具体的な授業内容は、調査モード、質問の聞き方の影響分析、実験調査の企画・実施、調査データの分析などから構成されています。調査モードについて個別面接聴取法と留置法を、質問の聞き方

について中間的な選択肢の有無及び配列、ダブル・バーレルの陳述、一般的質問と個人的質問、間接質問と直接質問を実験調査の対象としています。

調査の素材として、履修生が相談しながら決めることになってはいますが、やはり関心度の高い大学教育を選定したことが多い。図4は、大学生の授業取り組み態度に関する質問紙の一部です。

実験調査は、現役の大学生を対象に、2種類の調査モードと2通りの質問の聞き方に分けて作成した4種類の調査票を用いて行うことになっています。調査票ごとに性別の割合に配慮して約50～100名分の回答を集めることを目標としています。データ分析では、調査モードによる同一質問の比較分析と質問の聞き方による結果の変化(同一調査モード)に焦点を絞ると同時に、クロスにおける独立性検定、対応分析及び数量化Ⅲ類による質問間の関連性分析を活用します。

図5は、「履修科目を登録するとき、前年度の成績評価を重視しますか?」という質問に対して、中間的な選択肢の有無による影響を示すものです。両方とも留置法による調査結果ですが、中

F1. あなたの性別をご記入ください。	1. 男	2. 女				
F2. あなたの得意教科をご記入ください。	1. 理系教科	3. 文系教科				
問 G1.1 あなたは、講義を履修する際に出発点を気にしますが、それとも気にしませんか。	1. 気にする	2. やや気にする	3. どちらでもない	4. あまり気にしない	5. 気にしない	6. その他(記入)
問 G2.2 あなたは、履修科目を決めるときに、前年度の成績評価を重視しますか。	1. 重視する	2. やや重視する	3. どちらでもない	4. あまり重視しない	5. 重視しない	6. その他(記入)
問 G3.2 あなたは、大学の講義をレジュメ配布で進めるべきだと思いますか。	1. そう思う	2. ややそう思う	3. ややそう思わない	4. そう思わない	5. その他(記入)	6. その他(記入)
問 G3.3 あなたは、大学の講義をパワーポイントで進めるべきだと思いますか。	1. そう思う	2. ややそう思う	3. ややそう思わない	4. そう思わない	5. その他(記入)	6. その他(記入)
問 G4.2 あなたは、卒業研究配属先を選ぶ際に、研究家の研究内容を重視しますか。	1. 重視する	2. やや重視する	3. あまり重視しない	4. 重視しない	5. その他(記入)	6. その他(記入)
問 G4.3 あなたは、卒業研究配属先を選ぶ際に、先生の人間性を重視しますか。	1. 重視する	2. やや重視する	3. あまり重視しない	4. 重視しない	5. その他(記入)	6. その他(記入)

調査票 A: 中間的な選択肢がある質問例

F1. あなたの性別をご記入ください。	1. 男	2. 女			
F2. あなたの得意教科をご記入ください。	1. 理系教科	3. 文系教科			
問 G1.1 あなたは、講義を履修する際に出発点を気にしますが、それとも気にしませんか。	1. 気にする	2. やや気にする	3. あまり気にしない	4. 気にしない	5. その他(記入)
問 G2.1 あなたは、履修科目を決めるときに、前年度の成績評価を重視しますか。	1. 重視する	2. やや重視する	3. あまり重視しない	4. 重視しない	5. その他(記入)
問 G3.1 あなたは、大学の講義をレジュメ配布もしくはパワーポイントで進めるべきだと思いますか。	1. そう思う	2. ややそう思う	3. ややそう思わない	4. そう思わない	5. その他(記入)
問 G4.1 あなたは、卒業研究配属先を選ぶ際に、研究家の研究内容や先生の人間性を重視しますか。	1. 重視する	2. やや重視する	3. あまり重視しない	4. 重視しない	5. その他(記入)

調査票 B: 中間的な選択肢がない質問例

図4. 実験用調査票の一例

間的な選択肢の有無による影響があることが明らかになりました。履修生は、このような実験を通じて意識調査にとって重要なファクターが何かを1つひとつ理解することが期待できます。

一方、テキスト分析は、電子化した様々なジャンルの文章(文字と記号)からパターンを見つけ、有用な情報を抽出するプロセスです。テキストマイニングの一般手順は、図6のようになっています。

まず、テキストに対して形態素解析、構文解析などを行い、分析用データを生成します。次に、分析用データから各構成要素の出現頻度などを集計します。最後に、作品ごとの集計データを用いて、主成分分析、対応分析、クラスター分析、判別分析、機械学習法などを用いて、作品の成立、特徴などに関する有用な情報を抽出します。

各グループは、関心のあるジャンルのテキストを選定した上で、7回の授業時間をかけ、テキストマイニングへの理解を実践的に行います。たとえば、はじめて分析用ソフトウェアを触る学生も、グループのメンバー間の協力し合いによって徐々に使いこなすよう指導しています。

図7は、4名の代表的なプロレタリア文学作家の43作品に現れる「名詞」の使用特徴を分析した例を示しています。梶井基次郎、林芙美子の作品に関しては、それぞれほぼ特徴的に集まっていますが、小林多喜二と宮本百合子の作品に似ている特徴があり、名詞だけで上手く分類できないこ

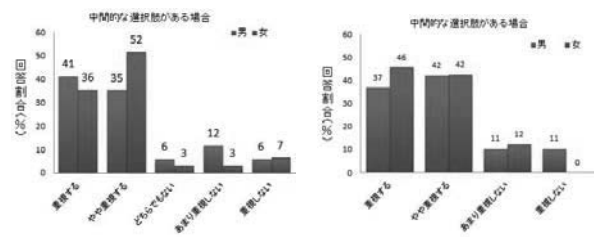


図5. 中間的な選択肢の有無による実験結果

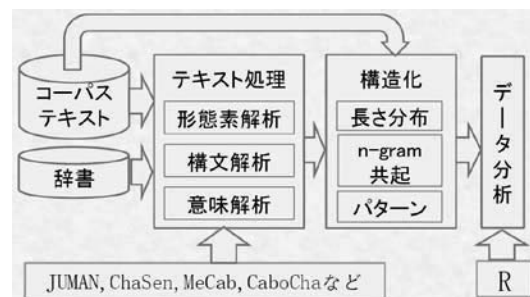


図6. テキストマイニングの一般手順

とが分かりました。この結果を受け、動詞や形容詞などの情報を新たに加え、試行錯誤しながら作品の特徴抽出を試みることを繰り返します。

図8は、価格.comで公開した売れ筋ランキング上位の掃除機に対する口コミ情報を分析するワード・クラウド (word cloud) です。これは、形態素解析により集計した名詞と形容詞の頻度に応じた大きさで図示する手法です。例えば、X3Dyson (1位) は、キャニスタータイプなどの機能に関する口コミが多く書かれていますが、ルンバ (3位) についてココロボやスケジュールなどの機能に関する言葉が目立ちます。このように評価用言葉の特徴を基に、掃除機の売れ筋の特徴を抽出する試み試みができます。

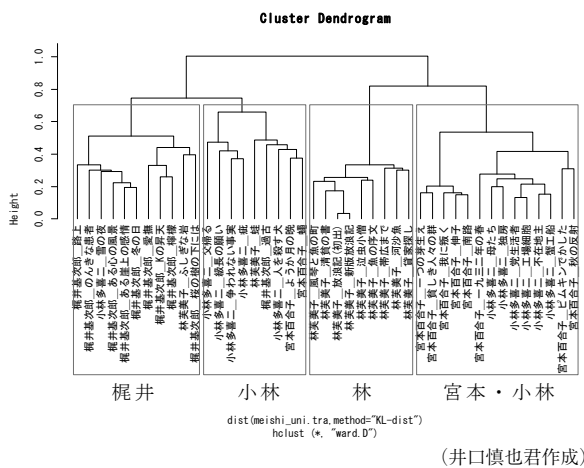


図7. 「名詞」のクラスター分析結果



図8. 「名詞」のクラスター分析結果

4. おわりに

通年授業を通して、履修生が互いに助け合いながら親しくなると同時に、共同作業型演習授業の教育効果が多く得られていると考えます。

第一に、グループの共同作業に参加することによって、履修生は知識吸収の効率を高め、座学授業で学んだ「文化」知識とデータサイエンス手法への理解を深めながら、文化現象・社会現象の研究に活用する技能を身につけることができます。

第二に、グループの組換え、授業中発表を行うことによって、履修生が個人の潜在力を引き出し、与えられた役割を果たすのに各々のデータ分析方法の詳細を最大限に理解する努力を払う学生が多く現れています。

第三に、繰り返しの発表資料作成、プレゼンテーション及び質問への対応を通じて、履修生の学術的コミュニケーション能力を向上させる効果が見られています。

第四に、グループのリーダー (班長・副班長) を担当する経験と得られた成果を総括することによって、履修生はリーダーの役割を十分に理解することができます。

以上のような修得知識と基本技能を基に、履修生は卒業研究に取り組む意識の向上が期待できると実感しています。今後も、学部のカリキュラムに合わせ、クラスの運営方法を改善しながら、学生の潜在力を引き出すよう、工夫していく予定です。

参考文献

- [1] 鄭躍軍・金明哲・村上征勝 (2007) : データサイエンス入門, 勉誠出版.
- [2] 鄭躍軍・金明哲 (2011) : 社会調査データ解析, 共立出版.
- [3] 鄭躍軍・金明哲・村上征勝 (2013) : 共同作業によるデータサイエンス教育の新しい試み—ジョイント・リサーチ授業の事例を中心に—, 第41回日本行動計量学会大会発表抄録集: 124-125.
- [4] 林知己夫 (2001) : データの科学, 朝倉書店.
- [5] 村上征勝 (2002) : 文化を計る—文化計量学序説, 朝倉書店.