

社会科学系大学生を対象とする モデリング科目の設計



千葉商科大学政策情報学部 准教授

長尾 雄行
NAGAO Takeyuki

プロフィール

東京大学大学院数理科学研究科数理科学専攻博士課程単位取得退学修士（数理科学）

2006年～2015年 産業技術大学院大学研究員・助教

2016年～2017年 静岡理工科大学専任講師

2018年～2019年 千葉商科大学専任講師

2020年より現職。数理科学、e-Learningシステムに興味を持つ。

1 はじめに

千葉商科大学政策情報学部では、専門科目「モデル・シミュレーション」において、社会科学系の大学生を対象とするモデリングとシミュレーションの講義を実施している。この科目は、社会における様々な現象を数式や図式等のモデルで表現する技能を修得するためのものであり、今後社会で文系学生にも要求されると思われる、コンピュータを駆使したデータやモデルの活用に関する基礎知識を提供する機会となっている。

本稿では、2020年度において実施している科目「モデル・シミュレーション」の内容にもとづいて、確率統計や微分積分等の数学的な知識や、プログラミングの経験を前提とせず、文系の学生にとって理解しやすいソフトウェアやオンラインサービスを活用して、モデリングとシミュレーションの授業を実施するための科目設計を事例として報告する。また、当科目をデータサイエンス分野の知識項目に対応させるための課題

と対策を記述する。

2 データサイエンス教育との関連

2017年に滋賀大学データサイエンス学部が創設され、文理融合方式の教育や Project Based Learning (PBL) 演習を取り入れ、企業との連携による実践的な教育が実施されている（竹村 2018）。その後、データサイエンス系学部の設置が各大学で進むなど、データサイエンス教育への取り組みが盛んに行われている。

社会科学系の大学においても、一般的に従来からコンピュータを活用した統計学の教育が行われており、多変量解析やデータマイニング等の授業が SPSS 等の専用ツールを用いて実施されている。今後はこれまで以上に広い分野で各種統計データや実世界のセンサーから取得したデータをコンピュータで活用し、近年急速に開発が進んでいる人工知能や機械学習の成果を応用する教育が重視されると予想される。

特に、データサイエンスの分野ではプログラミング言語の一つである Python 言語が最も重要な言語の一つとなっており、学生は早い段階で Python に触れる機会を得ておく必要がある。また、従来であればコンピュータの実習環境を整えなければ Python 言語等を利用できなかったが、近年ではクラウドサービス上で Python 言語を無料で活用できるプラットフォームが Google Colaboratory¹をはじめとして複数提供されている。このようなプラットフォームの利用技術も学生にとっては今後必須となっていくと予想できる。

1 <https://colab.research.google.com/>

3 科目「モデル・シミュレーション」

専門科目である「モデル・シミュレーション」の目的は、社会科学系の様々な分野を学ぶ文系の学生が、現実世界の対象をモデリングし、コンピュータ等を活用してシミュレートすることの重要性を体験的に学んで理解することにある。2019年度は90分の授業を全15回実施する形で、2020年度は105分の授業を全13回実施する形で実施している。

この科目の履修者は政策情報学部の2年生以上の学生である。デザインや映像制作に興味がある学生、プログラミング・3Dモデリング・都市計画・環境問題を学んでいる学生、さらに、公務員を目指す学生等が履修者として想定される。数学の知識についても、高等学校時代に数学Iと数学Aのみを学習した学生もいれば、微分積分まで学習した学生も混在している。物理学を学んだ者もいればそうでない学生も多い。これらの想定履修者の興味を引きやすくするために、事例は視覚的に把握しやすい図で扱える内容に絞り、実社会での応用例を紹介する方針で科目を設計した。

モデル・シミュレーションに関する教材の先行事例として(正司 2006)が挙げられる。本講義の設計に当たっては、利用するツールや想定する学生像は異なるが、この教材における講義の流れが社会科学の学生にも受け入れやすいと考え、大まかな授業の流れの設計を参考にした。そして、履修者がデータサイエンスの分野への下地を整えるために、Python言語やGoogle Colaboratoryを活用する形で学修環境を設計した。以下では2020年度の内容にもとづき、全13回の各授業回で扱っている内容を記述する。

3.1. 初回授業におけるモデル紹介

一般的にモデルという言葉は多様な分野において利用されるため、文脈によって意味が大きく異なる。そこで、初回の授業においては、モデルとは現実世界の対象を単純化して表現する為の数式・図・文章・データのことであると定義して、モデルを多数例示している。定義を明確にせずに授業を進行した場合には、モデルという多義語によって誤解を招く恐れがあり、また、具体例を利用しない場合には、人によってモデルという言葉でイメージする内容が異なる可能性がある

ため、定義と例示によりモデルという概念を具体的に説明している。授業の導入部分での具体例として取り上げている項目を以下では個別に列挙する。

世界地図

実世界の立体を平面に投影して表現する手法の例として世界地図を紹介している。現実には球体である地球を平面に投影する方式で図示する役割を持つのが世界地図であり、正距円筒図法やメルカトル図法のような投影図法が用いられていることを説明している。球体と平面図である世界地図との対応を用いることで、距離あるいは面積といった性質が正確には表現できない場合があることを述べ、モデルには表現できる性質と表現できない性質があることを紹介している。身近なところでは、360度動画でも正距円筒図法によって撮影者の周囲の映像が平面へ変換されて記録・再生されていると伝え、立体の平面的表現についての現実感を持たせている。さらに、インターネット上で公開地図を共有管理するシステムであるOpenStreetMapを紹介し、地図を構成する建物等のデータを誰もが編集可能であると伝え、モデルとして完成したものを利用するだけでなく、必要に応じてモデル自体を作成することが可能であるという事実を実例で紹介している。

路線図

グラフの概念を説明する前準備として、日常生活の中で見られるグラフの例である路線図を紹介している。ある駅から別の駅へ電車を用いて長距離移動する場合には、詳細な建物や道路の様子が描かれた住宅地図は不要であり、むしろ詳細情報が経路や乗換地点を見つけづらくするため経路検討の妨げとなることを説明した上で、地図の詳細情報を大胆に省略して駅だけを丸印等の記号で図示し、駅と駅の接続を線で結んで表現する路線図が有用であることを紹介している。実際の地図で見られる路線の形と路線図上に見られる線の形には多少のずれがあったとしても、つながり方が正しく記述できることが重要であるというグラフの概念を、専門用語をあえて省略して説明し、その後別の授業回において抽象的なグラフの定義へと進んでいる。

台風の進路予想図

自然現象に関する数値モデルをコンピュータ上でシ

ミュレートすることにより近未来に生じる事象を予測する身近な例として、気象庁の気象予報モデル(山口2018)や気象庁のWebサイトを紹介している。日常生活で頻繁に見る天気図を実現するためにコンピュータを用いて動的にシミュレーションが行われていることを意識させることを意図している。

自動車の3Dモデル

自動車メーカー各社のWebサイトにおいては、3Dモデルを活用して自動車の配色を試し、検討している自動車を好きな角度から閲覧することができるシステムが提供されている。このようなシステムでは自動車の3Dモデルのデータが利用されており、高価な自動車を購入する前に、消費者が検討している商品を実物に非常に近い視覚的表現で体験することができるメリットがあることを紹介している。

古典力学の運動方程式

古典力学では運動方程式によって物体の運動を記述すること解説し、身近なゲームにおいて物体が重力の影響を受けて落下することを表現するために、ゲームエンジン(ゲーム制作用のソフトウェア)内で物理エンジンを用いて古典力学に従った運動のシミュレーションが実施されていることを紹介している。数理モデルにおいて方程式を用いることが多いが、方程式がゲームなどの日常生活で利用する対象に適用されていることを意識させることで、方程式を解く手法や手順ではなく、それを日常の対象に適用することの重要性を理解させることを意図している。また、Web上でも物理シミュレーションを視覚的に行うことができる作品(遠藤2016)を紹介し、シミュレーションには多数のパラメタが用意されており、パラメタの値が変化すると結果もそれに応じて変化することを体験させている。

3.2. 第2回のテーマ：グラフ

グラフ理論は基本的なモデリング手段の一つであるが、高等学校までの教育では扱いがないため、図示の作法を把握し、著名な定理やアルゴリズムの応用を行うことに焦点を当てつつ講義を実施している。

講義では、頂点・辺等の概念やグラフの図示方法を解説したのち、作業の前後関係をグラフで表現する手法であるPERT図の描き方とクリティカルパスの算出

方法を紹介している。PERT図を採用している理由は、ITパスポート試験等のIT系の資格において必要となる概念であり、IT系の資格取得を希望する者にとっては学修のインセンティブとなるからである。

PERT図を取り扱った後に、一筆書きの可能性を判定するオイラーの定理を紹介している。準備として、グラフの連結性の概念、さらに、頂点の次数(頂点につながる辺の個数)を説明したのち、オイラーの定理では、連結グラフの全頂点の次数を求めておけば、一筆書きできる場合(奇数次数の頂点が0個または2個の場合が該当)と一筆書きできない場合(奇数次数の頂点が4個以上の場合が該当)の両者が容易に判定できると説明し、例題を用いて幾つかの例で判定方法を練習させている。

一筆書きは身近な遊びであり、多くの学生が幼少時に実施したと思われるため、グラフ理論への敷居を下げる効果が期待できる。また、一筆書きに関するオイラーの定理を紹介することで、18世紀の数学が現代でも普遍的な道具として応用可能であることに注目させ、普遍的な数学の理論の魅力を意識させる効果を期待している。

オイラーの定理に引き続いて、一筆書きの経路を探索するアルゴリズムである、フラリーのアプローチのアルゴリズムを紹介している。オイラーの定理を用いると一筆書きができるかどうかを判定することは可能であるが、一筆書きの経路を求める手順は定理に含まれていないという事実に注目させた後、一筆書きの解となる経路を求めるフラリーのアプローチのアルゴリズムを紹介している。このアルゴリズムを学ぶ中で、橋・多重辺・ループ・孤立点という概念を修得させている。

一筆書きの解が存在するかどうかを述べているオイラーの定理と、一筆書きの解を求めるためのフラリーのアプローチのアルゴリズムを対比させることで、解の存在証明と解を求めるアルゴリズムは別のものであることを意識させる効果を期待している。

また、レポート作成や卒業論文での利用を想定してWeb上でグラフを描画するためのツールであるViz.js²を紹介している。これは、グラフを描画する古典的なツールであるGraphvizをWeb上で実装したものである。

² <http://viz-js.com/>

3.3. 第3回のテーマ：線形計画法

線形計画法は連立不等式によって表現された領域における特定の線型関数を最小化・最大化する問題であるため、2変数問題の場合には高等学校でも数学Bや数学IIの範囲で実質的に取り扱われている。一方で、実社会での応用やコンピュータでの活用方法については大学以降の教育で取り扱う必要がある。そこで、本講義では、複数の材料を混合して製品を生産する際に利益が最大となる材料の配合を求める配合問題を題材としている。

線形計画法のコンピュータ実習環境を学生自身で構築する必要が無いようにするために、Webブラウザ上で利用可能なPython言語の実行環境であるGoogle Colaboratoryと線形計画法のモデリングツールであるPuLP³を利用している。本講義では、2変数に限定して連立不等式による制約条件の定式化、及び、目的関数の設定方法等の最適化問題の記述手法を解説し、その後、PuLPを用いて最適化問題をPython言語で実装する例題を実施し、類似問題の演習を行っている。

授業回数が一回のみしか割り当てられないため、線形計画法で用いるシンプレックス法や内点法等のアルゴリズムを解説することはできていない。また、ソルバー等のソフトウェアにおいて数値計算がどのように実施されるのかについても触れることはできていない。

3.4. 第4回のテーマ：確率・統計

確率的手法を用いてシミュレーションを行うことが必要な場合も多いため、確率及び統計についての基礎知識をExcel又はGoogle スプレッドシート等の表計算ソフトウェアを用いて演習を行いながら修得する機会を提供している。高等学校の数学Aにおけるような数学的確率の概念と、観測したデータから算出する統計的確率の概念を紹介している。

履修者は1年次にExcelの基本操作を修得済みのため、表計算ソフトの関数利用には抵抗がない。そこで、表計算ソフトを用いて観測データや統計データを集計して度数・相対度数・ヒストグラム・平均・標準偏差を求める過程を体験する機会を提供している。具体的には、整数乱数生成用のrandbetween関数を用いてサイコロを模した乱数を生成し、観測された乱数についてその度数表・相対度数表・ヒストグラムの作成を順に取り扱い、試行回数を増やすと理論的に予想され

る数学的確率へ収束する様子を観察する内容となっている。

その後、現実のデータに関する統計処理を体験するために、気象庁が公開している過去の年間台風発生件数データ⁴を取得して、台風発生件数のヒストグラムを図示し、平均値・標準偏差を算出し、サイコロの場合は一様な分布となっていたヒストグラムが、台風発生件数については釣り鐘状の分布となっていることを確認する内容となっている。これらの学修内容によって、確率密度の概念を間接的に体験することができるかと期待される。一方で、集合論を用いた確率の定式化や、確率変数や確率密度・正規分布等を導入及び解説することまではできていない。

3.5. 第5回のテーマ：待ち行列

この時点までに学修している確率・統計の内容と表計算ソフトの活用を発展させる形で、刻一刻と窓口へ並び列をなす顧客の列をシミュレートする技法を取り扱っている。この講義では、井家(2015)の手法にもとづいて、離散事象シミュレーションを用いている。この手法は、顧客の到着とサービスの提供という2つの現象を指数乱数によりシミュレートする方法であり、表計算ソフト上で学生自身が関数を組み合わせながら構築できることから、シミュレーションがどのように実現されているのかを体験するために有用と思われる。実際の講義では、はじめに指数乱数の生成方法とその値の意味を説明し、引き続いて顧客の到着のみをシミュレートするシートを作成している。顧客の到着が事前に設定した到着率の値に従うように指数乱数の発生方法を調整している。その後、サービスの提供も同様の方法で一定のサービス率に従って行われるようにシートを拡張する形を取っている。サービス率と到着率を変化させた場合の挙動の違いを考察させることで、これらのパラメタの意味をシミュレーション経由で考えさせる内容となっている。ここでの授業内容は待ち行列理論への入門ではなく、乱数を応用したシミュレーションを学生自身が構築する機会を提供するものになっている。

3.6. 第6回のテーマ：モデル化

前半部分のまとめとなる授業回においては、初回に示したモデルの具体例に加えグラフ・線形計画問題・

3 <https://pypi.org/project/PuLP/>

4 <https://www.data.jma.go.jp/fcd/yoho/typhoon/statistics/generation/generation.html>

待ち行列等のモデルが出そろっている。この時点で、正司 (2006) に従って表現形式によるモデルの分類 (物理的モデル・図的モデル・数式モデル) を説明し、各分類項目の具体例をインターネット上で公開されている資料を紹介するなどの方法により提示している。モデルの特性については、静的モデルと動的モデルの違い、連続時間モデルと離散時間モデルの対比、確率的モデルと確定的モデルの差異を解説している。

先行する複数の授業回において事前にモデルを多数紹介したのちに、抽象的な分類やモデル化の流れを説明することで、目的ごとに多様なモデルが存在しており、モデルには必ず単純化や理想化が行われるため、表現の限界が存在していることを具体例により体験的に学ぶことができると期待される。

モデル化の説明と併せて、シミュレーションの手法として、プログラミング言語による方法、表計算ソフトによる方法、及び、シミュレーション用のソフトウェアによる方法を紹介している。プログラミング言語としては、Python を採用し、離散事象シミュレーションのための SimPy⁵ を Google Colaboratory 上で利用することで、待ち行列の授業で扱ったモデルを Python で実装する内容を実施している。表計算ソフトによるシミュレーションについては、確率・統計及び待ち行列の授業回に実施した内容である。シミュレーション用のソフトウェアについては、Vensim PLE⁶ を利用している。このソフトウェアでは、システムダイナミクスと呼ばれる手法に従ってモデルを図と数式で表現する。

3.7. 第7回以降：システムダイナミクス

システムダイナミクスは J. W. フォレスターによって 20 世紀中頃に提案されたモデリング手法である。システムダイナミクスでは、モデリングの対象に含まれる要素を、ストック (預金口座の残高のように変化する値を記述する要素)、フロー (要素間の値の流入及び流出を表現する要素)、及びコンバータ (定数値や外部入力を表す要素) を組み合わせて表現する。要素と要素の間の因果関係を図示することで検討しているシステムの全体像をトップダウン方式でモデリングすることができる。

システムダイナミクスでは、数学・物理学・工学の分野で用いられる微分・積分の概念を微分記号や積分記号を用いずに図で表現することができるので、常

微分方程式に相当するモデルを取り扱うことが可能である。この講義で採用している Vensim PLE では、Windows または macOS の環境で日本語による図式の作成にも対応しているため、利用方法を習熟すれば学生が日本語でモデルを作成してシミュレーションを実施することができる。Linux 環境でも Windows のエミュレータ Wine を用いると動作させることが可能である。

システムダイナミクスの記法を説明した後、マルサスモデルに従った人口の指数増大、温度センサーと暖房をテーマとしたフィードバック制御 (遅延がない場合と遅延がある場合の両者)、ロジスティックモデルによる S 字型の成長曲線、そして、ヴォルテラ捕食系モデルによる周期的な個体数の変動を扱っている。

物理現象のシミュレーションをテーマとする授業回では、古典力学の運動方程式を用いた等速直線運動、自由落下、及び、単振動を扱っている。物理学や微分概念を知らない学生でも、位置の変化が速度によって引き起こされるという因果関係をストックとフローを用いて図示することで視覚的にモデルを表現できるため、理論的準備がなくても常微分方程式に相当する表現が可能である点がシステムダイナミクスの利点である。

社会現象のシミュレーションを扱う授業回では、ロジスティックモデルを用いて口コミの広がりをも簡易的にシミュレートし、更に、感染症の広がりを表現する SIR モデル (人の状態を未感染状態、感染状態、免疫獲得状態の 3 状態で表現) を用いて感染症の拡大の様子をシミュレートしている。

カオス現象をテーマとする授業回では、ローレンツ方程式をシステムダイナミクスで表現してシミュレートし、結果を Google Colaboratory 上の Python スクリプトにより動画で可視化することで、ローレンツアトラクタへ吸い込まれる非周期的な軌道を体験する内容となっている。

4 科目設計の課題と対策

ACM Data Science Task Force (2019) によると、学部教育におけるデータサイエンスの知識領域は (1)

⁵ <https://simpy.readthedocs.io/en/latest/>

⁶ <https://vensim.com/>

分析と表現、(2) 人工知能、(3) ビッグデータシステム、(4) 計算と計算機の基礎、(5) データの取得・管理・ガバナンス、(6) データマイニング、(7) データのプライバシー・セキュリティ・完全性・セキュリティ分析、(8) 機械学習、(9) プロフェッショナルリズム、(10) プログラミング・データ構造・アルゴリズム、そして、(11) ソフトウェア開発とメンテナンス、という11領域に分けられている。

科目「モデル・シミュレーション」で扱う内容は、実世界の対象をモデリングするという点では、上記の知識領域(1)、(2)、(5)及び(8)におけるモデリングに関連している。一方で、後半部分のシステムダイナミクスに関する授業で紹介したモデルは確定的なモデルであるため、データサイエンスの分野で活用が期待される確率モデルとは大きく異なっている。よりデータサイエンスの分野に当科目の内容を近づけるには、ベイズモデル等を扱うようにして機械学習にも触れられる機会を設ける方法が考えられる。

知識項目(2)、(5)、(8)は主に人工知能や機械学習に関連する領域であり、一般的に画像や音声認識アルゴリズム、自然言語処理等が関連する。科目「モデル・シミュレーション」の目的を大きく変更せずにデータサイエンス分野へ近づけるために、当科目では特に自然言語処理を実施するためのモデルに焦点を当てた上で、授業回1回を割り当て、単語出現頻度・TF-IDF (Term Frequency-Inverse Document Frequency)・形態素解析を取り扱うことがよいと思われる。そのための道具として、機械学習用のPythonライブラリであるscikit-learn⁷を用い、データセットとして青空文庫形態素解析データ集⁸を活用する方法が考えられる。

機械学習については、現状では当科目で扱うことができていないが、scikit-learnを用いると回帰分析や、手書き文字認識が比較的簡単な手順で体験できる

ため、授業で取り入れることが可能である。知識項目(1)の分析と表現においては、機械学習によって得られたデータを可視化し、結果を視覚的に表現する手法を紹介することが効果的と思われる。その道具として、機械学習の結果を可視化するためのライブラリであるYellowbrick⁹が有用と思われ、当科目でも取り入れることが可能な技術的難易度と考えられる。

以上の議論から、2回程度の授業を機械学習や人工知能周辺技術を取り扱うように設計を変更することで、データサイエンス分野へ当科目をより近づけることが可能と思われる。

5 おわりに

本稿では、モデリング科目「モデル・シミュレーション」の現状を記述し、授業の手法及び授業内で活用している学修環境やツールを紹介した。ACMのデータサイエンス分野の知識領域と知識項目に照らし合わせた結果として、確定的モデルへの科目設計の偏りが見られた。モデリングとシミュレーションを実施するという科目としての目的を逸脱しない範囲で実施可能な対策として、自然言語処理を中心とする機械学習分野への入門的教育を一部取り入れることが良いと思われる。

一方で、それらのアイデアが全て実施できた場合においても、数学・情報科学等の分野の数学的・理論的側面を当科目の中で取り扱うことは困難であるため、ツールを使ってモデルやその原理をよく理解せずに活用する傾向を学生に持たせてしまう懸念がある。原理や理論に興味を持つ学生には、より進んだ文献・資料・オンライン教材を紹介するなど、科目外で学べる教材や情報源とのつながりを構築していくことも今後の課題である。

⁷ <https://scikit-learn.org/stable/>

⁸ <http://aozora-word.hahasoha.net/>

⁹ <https://www.scikit-yb.org/>

参考文献

- ACM Data Science Task Force. (2019) Computing Competencies for Undergraduate Data Science Curricula Draft 2. 2021年1月10日閲覧 <http://dstf.acm.org/DSReportDraft2Full.pdf>
- Dantzig, G. B. (2002) Linear Programming. *Operations Research* 50 (1) :42-47. 2021年1月6日閲覧 <https://doi.org/10.1287/opre.50.1.42.17798>
- Forrester, J. W. (1996) The Beginning of System Dynamics. 2021年1月9日閲覧 <http://web.mit.edu/sysdyn/sd-intro/D-4165-1.pdf>
- 井家敦, 岸康人, 佐久間大 (2015) 表計算ソフトで待ち行列を再現してみよう. *オペレーションズ・リサーチ*, 60 (9), 526-531.
- 石村園子 (2001) *すぐわかる確率・統計*, 東京図書株式会社
- 遠藤理平 (2016) HTML5による仮想物理実験室(物理シミュレーション). 2021年1月6日閲覧 <http://www.natural-science.or.jp/PHYSICS/>
- 正司和彦, 高橋参吉他 (2006) *基礎シリーズ 最新 モデル化とシミュレーション*. 実教出版
- 竹村彰通, 和泉志津恵, 齋藤邦彦, 姫野哲人, 松井秀俊, 伊達平和 (2018) データサイエンス教育の滋賀大学モデル. *統計数理*, 66 (1), 63-78.
- 橋本洋志, 石井千春 (2008) Scilab/Scicosで学ぶシミュレーションの基礎—自然・社会現象から、経済・金融、システム制御まで. オーム社
- 山口春季他 (2018) 気象庁「全球アンサンブル予報システム」について. 平成29年度「異常気象と長期変動」研究会報告.