

高等学校における機械学習についての指導の可能性と 授業実践

春日井優^{†1}

概要: 社会においては、人工知能を謳った技術が広がりを見せ、深層学習や機械学習などを扱った書籍が増えている。高等学校においては、学習指導要領の改訂で新設される科目「情報」においてデータサイエンスが学習内容に加わり、データサイエンスの一分野として機械学習について生徒が学ぶ可能性がある。現行の学習指導要領の範囲を基に、高校生に指導可能な学習内容を検討した。また、機械学習の一手法である単純ベイズ分類器を指導した授業実践について報告する。

キーワード: 機械学習, 単純ベイズ分類器, 高等学校, 情報科, 授業実践

Study and Practice of Machine Learning in Senior High School

YU KASUGAI^{†1}

1. はじめに

近年、人工知能の技術が発展し、将棋や囲碁ではトップ棋士にも勝利するなど、めざましい成果を挙げている。また、スマートスピーカーなどの AI を謳う製品も販売されるなど世の中の関心が高くなっている。このようななか、人工知能などの技術と関わっていくためには、仕組みの理解も不可欠になる。

また、初等中等教育の場面では、人工知能が広がりを見せていることを背景に、必要となる資質・能力が検討され、学習指導要領が改定されつつある。そのなかで、高等学校情報科においてはデータサイエンスの分野を学習する単元が新設される。

そこで、本稿では、人工知能やデータサイエンスと関わりが深い機械学習に焦点を当て、高等学校で指導可能な機械学習のアルゴリズムについて検討する。また、検討した結果をもとに、機械学習のアルゴリズムの 1 つとして、単純ベイズ分類器についての授業実践を行った。その授業実践について述べる。

さらに、授業実践をとおして明らかになった学習の成果や授業における課題についても述べ、高等学校における機械学習の可能性について検討する。

2. 高等学校で機械学習を指導する背景

2.1 社会の状況

近年、人工知能技術のめざましい発展が、一般の新聞、テレビ、書籍などにおいて紹介され、人工知能の専門家だけではない人にも関心を持たれている。

2011 年には、IBM 社が開発した Watson が、クイズ番組で歴代のクイズチャンピオン 2 名を相手に勝利した[1]。また、将棋においては、2013 年には第 2 回将棋電王戦において、5 本の将棋ソフトと 5 人のプロ棋士が対戦し、将棋ソ

フトが 3 勝 1 敗 1 引き分けとプロ棋士に勝ち越した[2]。

将棋では将棋ソフトが勝利したが、将棋は局面の数が 10 の 220 乗、囲碁は 10 の 360 乗と推測され、囲碁でコンピュータがトップ棋士に勝利するには 10 年はかかると予想されていた[3]。

しかし、囲碁においても 2016 年に、Google 社の Alpha Go が世界トップクラスの棋士イ・セドルに 4 勝 1 敗で勝利し、一般紙においても衝撃として掲載された[4]。

また、2011 年から国立情報学研究所においては「ロボットは東大に入れるか」(東ロボ)というプロジェクトが行われ、人工知能の可能性と技術限界について研究された。そのプロジェクトで開発された「東ロボくん」はセンター模試において、5 教科 6 科目で偏差値 57 以上を達成し、70% 以上の大学で合格可能性 80% 以上を達成した[5]。

オックスフォード大学のマイケル・オズボーン准教授が発表した「雇用の未来」において 10~20 年程度のうちに機械に代替可能な仕事が労働人口の 47% であるという推計値を発表し、人工知能技術の進展とともに一般紙などでも引用されている[6]。

また、近年スマートスピーカーが AI スピーカーと呼ばれ販売されるなど、人工知能技術を応用した家電製品が注目されている[7]。

以上のように、人工知能技術が発達するなか、専門家以外でも仕組みを知ることができるような書籍が出版されたり、新聞の記事などでも大まかな仕組みが掲載されたりしている。しかし、一般には十分理解されているとはいえない。単に人工知能などの技術を盲信したり、逆に拒否したりというような態度を取るのではなく、仕組みを理解した上で適切に扱うことが必要であると考えられる。

2.2 初等中等教育における人工知能の扱い

現行の学習指導要領は、小中学校については 2008 年に、

高等学校については2009年に改定され、この時点ではまだ人工知能について記述はない。

現在、学習指導要領改定に向けた動きが進んでいる。小中学校については、2017年に次期学習指導要領が示され、高等学校については2017年度中に次期学習指導要領が示される予定になっている。

2016年に、次期学習指導要領編成に向けてまとめられた中教審答申では、人工知能の発達による社会の変化予想を取り上げている[8]。また、「人工知能がいかに進化しようとも、それが行っているのは与えられた目的の中での処理である。一方で人間は、完成を豊かに働かせながら、どのような未来を創っていくのか、どのように社会や人生をよりよいものにしていくのかという目的を自ら考え出すことができる」と人間の役割について述べられている。

次に、既に次期学習指導要領が公示されている小中学校について述べる。小学校・中学校ともに改訂の経緯として、人工知能による社会の変化および人工知能との比較により「思考の目的を与えたり、目的のよさ・正しさ・美しさを判断したりできるのは人間の最も大きな強み」であることを挙げている[9][10]。さらに、中学校では教科の内容として、技術科では、「情報の技術」の単元において「生活や社会における人工知能の活用について、人間の労働環境や安全性、経済性の視点から、その利用方法を検討する」という利用面について学習することが例示されている[11]。さらに、社会科では、公民分野において「情報化については、人工知能の急速な進化などによる社会の構造的な変化など関連付けたり、災害時における防災情報の発信・活用などの具体的な事例を取り上げたりすることとした」と、社会における活用事例として取り上げられている[12]。また、理科では第1分野の「科学技術と人間」の単元において、科学技術の発展の1つとして調べさせることが挙げられている[13]。中学校では、人工知能は活用についての学習が主であり、仕組みについて踏み込んで学習する見込みは低い。そのため、コンピュータが人間を超える存在として盲信してしまったり、逆に拒絶する姿勢を見せてしまったりする可能性は否定できない。人間にとって人工知能の仕組みを理解する。

高等学校については、本稿執筆時点では学習指導要領が示されていないため、中教審答申をもとに議論する。

中教審答申では、具体的な内容として人工知能についての記述はないが、情報活用能力に統計が含まれることが明記された。統計的な内容の充実が求められ、情報科と数学科で連携して指導することとなった[14]。

また、情報科では、大幅に科目構成が変更になり、単元構成案についても示されている(表1)[15]。新科目「情報」において、「情報社会の進展と情報技術」、「情報とデータサイエンス」という単元が新設される。

「情報社会の進展と情報技術」の内容から、社会におけ

る人工知能等の技術を展望するものになり、技術面について扱う可能性は薄いと推測される。

データサイエンスの分野には、統計、データマイニング、機械学習、人工知能などがある[16]。「情報とデータサイエンス」の単元では、「データサイエンスの手法を活用して情報を精査する力を育む」ことが内容案に示され、情報を精査するための仕組みの理解が期待される。

次期学習指導要領は、高等学校については2022年から年次進行で実施されるが、ますます人工知能技術が進展し社会に浸透することが見込まれることから、現行の学習指導要領の範囲内で指導することを検討した。ただ、人工知能技術という範囲が広範にわたることもあり、定義もまちまちであることが見受けられることから、データを解析しその結果から学習して判断や予測を行うためのアルゴリズムである機械学習に対象を絞って検討した。

表1 次期学習指導要領における情報科科目構成案

科目名	単元名	内容
情報	(1)情報社会の問題解決	中学校までに経験した問題解決の手法や情報モラルなどを振り返り、これまでに情報社会の問題の発見と解決を適用して、情報社会への参画について考える。
	(2)コミュニケーションと情報デザイン	情報デザインに配慮した的確なコミュニケーションのちからを育む。
	(3)コンピュータとプログラミング	プログラミングによりコンピュータを活用する力、事象をモデル化して問題を発見したりシミュレーションを通してモデルを評価したりする力を育む。
	(4)情報通信ネットワークとデータの活用	情報通信ネットワークを用いて、データを活用する力を育む。
情報	(1)情報社会の進展と情報技術	情報社会の進展と情報技術の進展について歴史的に捉え、AI等の技術も含め将来を展望する。
	(2)コミュニケーションと情報コンテンツ	画像や音、動画を含む情報コンテンツを用いた豊かなコミュニケーションの力を育む。
	(3)情報とデータサイエンス	データサイエンスの手法を活用して情報を精査する力を育む。
	(4)情報システムとプログラミング	情報システムを活用するためのプログラミングの力を育む。
	課題研究	情報 および情報 (1)~(4)における学習を総合し進化させ、問題の発見・解決に取り組み、新たな価値を創造する。

3. 高等学校の授業における指導の可能性

3.1 検討する範囲の決定

2.2 節において、現行の学習指導要領は人工知能や機械学習の進展を前提にしておらず、学習内容についても含まれていない。

しかし、現行の学習指導要領での学習内容の範囲内でも機械学習の仕組みと関連する内容が含まれている。高等学校で実施可能な内容を調べるために、現行の学習指導要領の中で、機械学習の仕組みと結びつく可能性がある教科として、情報科・数学科・理科を対象に調査する。

また、機械学習についても様々なアルゴリズムがあり、現在もアルゴリズムが研究されているが、高等学校の授業では専門的に理解するのではなく、機械学習が何であるかを知り、その仕組みの概要をつかむことができれば十分であると考えられる。そこで、ある程度一般向けに解説された書籍でも紹介されているアルゴリズムを対象とする。

3.2 高等学校学習指導要領における関連内容

共通教科の情報科には、「社会と情報」・「情報の科学」という2科目により編成されている。

「社会と情報」は、「情報の活用と表現」・「情報通信ネットワークとコミュニケーション」・「情報社会の課題と情報モラル」・「望ましい情報社会の構築」という4つの単元から構成されている。「社会と情報」の科目の目標が、主に情報社会に積極的に参画する態度を育てることが柱としていることから、情報に関する仕組みについては、あまり多く扱われていない。そのため、特に機械学習の仕組みにつながる学習内容は含まれていない。

「情報の科学」は、「コンピュータと情報通信ネットワーク」・「問題解決とコンピュータの活用」・「情報の管理と問題解決」・「情報技術の進展と情報モラル」の4つの単元から構成されている。これらのうち、「問題解決とアルゴリズム」では、モデル化や処理手順の自動化が扱われている[17]。モデル化については、機械学習のアルゴリズムは、それぞれのアルゴリズムの考え方となるモデルがあることから、モデル化の発展として学習できる。また「処理手順が込み入ったものであっても、適切なアルゴリズムでコンピュータに自動実行させることによって、誤りなく繰り返す使用することができるなど、自動実行の有用性について考えさせる」ことが挙げられており、機械学習のアルゴリズムの学習をし、自動実行させて動作を確認するというような授業展開をすることが可能である。

数学科では、機械学習と関連する内容として、数学の「データの分析」、数学の「図形と方程式」、数学Aの「場合の数と確率」、数学Bの「確率分布と統計的な推測」が挙げられる[18]。

理科では、生物の「生物の環境応答」の単元において「刺激の受容と反応」が扱われ、脳の仕組みを学習する[19]。脳の仕組みは後述するニューラルネットワークのモデルになっている。機械学習の指導においては、生物学の体系的に理解する必要はなく、脳における情報の伝達の仕組みが概念的に理解できれば十分であることから、トピック的に指導することも考えられる。

3.3 対象とする機械学習

本稿で対象とする機械学習として、「回帰分析（線形回帰）」・「k近傍法」・「サポートベクターマシン」・「決定木」・「単純ベイズ分類器」・「k平均法」・「ニューラルネットワーク」とした。これらの機械学習の手法の概要を示し、高等学校での学習内容との関連について述べる。

(1) 回帰分析（線形回帰）

単純な単変量の線形回帰は、単一の特徴量(説明変数 x)と連続値の応用(目的変数 y)との関係をモデルとして表現したものである。説明変数が1つだけの線形モデルの方程式は次の式で定義される[20]。

$$y = ax + b$$

この式は1次関数であり、この直線に近似することにより値を求めることができることは、中学校の数学で十分理解可能である。一方、この方程式を求めるには、最小二乗法が用いられる。最小二乗法により係数は

$$a = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - a\bar{x}$$

により求められる。ここで、 s_{xy} は x と y の共分散、 s_x は x の分散、 \bar{x} は x の平均、 \bar{y} は y の平均である。分散や共分散は、数学の知識が必要となる。

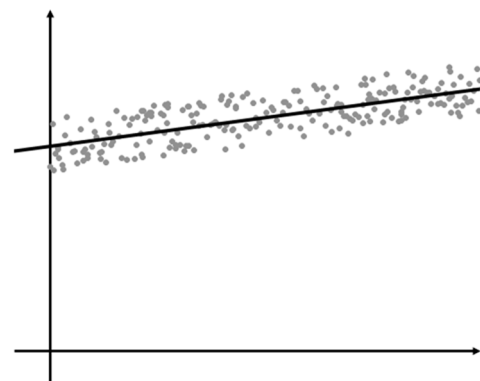


図1 線形回帰の概念図

(2) k近傍法

k近傍法による分類は、分類したいサンプルからk個の最近傍のデータを見つけ出し、多数決により分類する方法である[21]。近傍であること、すなわち距離の概念が理解できれば十分なので、高等学校では十分指導可能である。実際にアルゴリズムを記述するには距離の求め方が必要になる。2点間の距離については、数学の「図形と方程式」において座標平面の距離の求め方を学習する。距離を求めるもとなる考え方として、三平方の定理は中学校で学習しているので、アルゴリズムについては十分理解可能である。

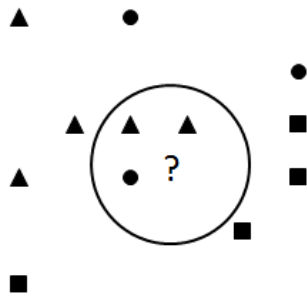


図 2 k 近傍法(k=3 の場合)の概念図

(3) サポートベクターマシン

サポートベクターマシンによる分類は、決定境界という超平面によって 2 つのクラスを分類するものである[22]。概念的には 2 次元平面上に直線を引くことで、2 つのクラスを分類できることを理解できれば十分で、特別な知識は不要である。もう少し深く理解するには、数学で学習する「図形と方程式」の単元にある「不等式が表す領域」が必要となる。不等式により領域を示すことができることを理解することにより、数式として表現できることまで理解が深まる。サポートベクターマシンを実装するためのアルゴリズムを学習するには、決定境界とデータのマージンを最大化する必要がある。この決定境界を求めるには、ラグランジュの未定乗数法が必要であり、偏微分など高等学校の学習範囲を超える。

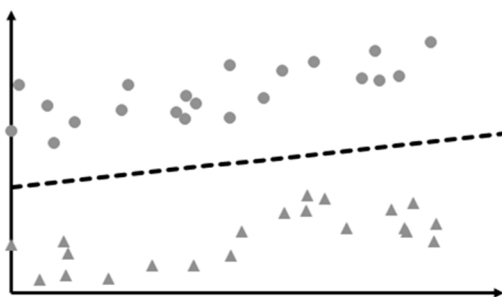


図 3 サポートベクターマシンの概念図

(4) 決定木

決定木による分類は、複数の質問によってどのクラスに属するかを絞り込んでいくものである[23]。これは、単に条件分岐に過ぎず、分類の考え方自体は難しくない。決定木を構成するにはジニ不純度などの指標を求める必要がある。ジニ不純度は

$$I_G(t) = 1 - \sum_{j=1}^c p(j|t)^2$$

により求められる。ここで、 $I_G(t)$ はジニ不純度、 $p(j|t)$ はノード t におけるクラス j の相対頻度である。数学 A の確率を学習すれば、 $p(j|t)$ は求められる。



図 4 決定木の概念図

(5) 単純ベイズ分類器

ベイズの定理に基づいて分類する手法である。ベイズの定理は数学 A の「場合の数と確率」の条件付き確率をもとにして学習する。ベイズの定理は、

$$P(Y = k|X) = \frac{P(X|Y = k)P(Y = k)}{P(X)}$$

である[24]。ベイズの定理により確率を求め、確率が最大になったクラスに属すると分類するものである。アルゴリズムを実装するには、値を求める際にアンダーフローが起きる可能性が高いため、対数関数が用いられる。対数関数は、数学の「指数関数・対数関数」で学習する。

(6) k 平均法

k 平均法は教師なし学習の一種で、クラスタリングに用いられるアルゴリズムである。k 平均法のアルゴリズムは次のとおりである[25]。

1. k 個のクラスタの中心をランダムに割り振る。
2. 各データと k 個のクラスタの中心との距離を求め、最も近いクラスタに割り当て直す。
3. 各クラスタの算術平均を求め、新たなクラスタの中心とする。
4. 収束するまで 2,3 を繰り返す。

このように、距離と平均を求めるだけであるので、概念は指導可能である。距離については、(2)k 近傍法で述べたように、三平方の定理をもとにすれば、数学の知識を前提にしなくても指導可能である。

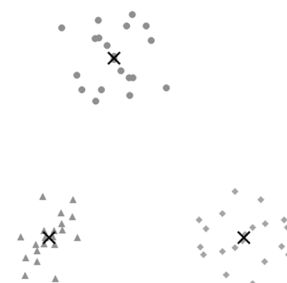


図 5 k 平均法の概念図

(7) ニューラルネットワーク

ニューラルネットワークは、脳内の神経細胞において別の神経細胞から電気信号を入力として受け取り、ある閾値を超えると受け取った神経細胞も次の神経細胞に伝える様

子を，数理モデル(McCulloch-Pitts モデル)をもとに考案されたものである[26]．概念的には，生物で学習する神経細胞での信号の伝達が理解できれば十分である．アルゴリズムを考えると，閾値を超えた際にどの程度伝達するかを示す活性化関数が，ステップ関数やシグモイド関数などが用いられ，高校での範囲を超える．

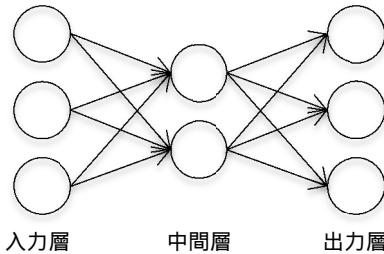


図 6 ニューラルネットワークの概念図

以上の 7 つの機械学習のアルゴリズムについて，高等学校における指導の可能性について整理すると表 2 となり，高等学校でも十分指導可能であるといえる．

表 2 機械学習のアルゴリズムと高等学校の学習内容

アルゴリズム	概念の理解に必要な知識	アルゴリズムを理解するのに必要な知識
回帰分析	1 次関数(中学校数学)	分散・共分散(数学 A)
k 近傍法	距離の概念	2 点間の距離(数学)
サポートベクターマシン	不等式が表す領域(数学)	高等学校の範囲を超える
決定木	条件分岐	確率(数学 A)
単純ベイズ分類器	条件付き確率(数学 A)	対数関数(数学)
k 平均法	距離の概念・平均	2 点間の距離(数学)
ニューラルネットワーク	神経細胞での信号の伝達(生物)	高等学校の範囲を超える

4. 高等学校における授業実践

4.1 授業内容の検討

著者の勤務校では，3 学年で「情報の科学」を 2 単位で開講しており，「モデル化とシミュレーション」について教科の内容として含まれている．その単元の教材として機械学習を指導することとした．対象となる生徒は，全員が数学 および数学 A を履修しており，前節で述べたアルゴリズムについて，ニューラルネットワーク以外の概念は前提となる知識を持っている．機械学習によりコンピュータが回答する答えについて，生徒が仕組みを理解する上ではアルゴリズムの指導も必要であると考え，網羅的に機械学習を取り扱うのではなく，一つのアルゴリズムを深く指導することとした．

また，高等学校の情報科の目標に「社会の中で情報および情報技術が果たしている役割や影響を理解させ，社会の

情報化の進展に主体的に対応できる能力と態度を育てる」ことが示されており，単に機械学習の仕組みを理解させるのでは不十分であると考えた．

そこで，機械学習の技術が応用される場面についても指導する必要があると考え，生徒にとって仕組みとあわせて理解できる教材の検討が必要であった．

以上を検討した結果，自然言語処理とあわせて単純ベイズ分類器を指導することとした．また，単純ベイズ分類器の仕組みの理解が深まるよう，TF-IDF 法による特徴語抽出およびワードクラウドによる出現頻度が高い単語の可視化を行った．



図 7 ワードクラウド

4.2 授業内容

(1) 形態素解析

自然言語処理を対象にするにあたり，情報科の授業内容との関連付けを検討した．自然言語は形態素の出現回数だけに着目してモデルとする Bag-of-Words モデルで扱われることがある．このモデルをもとに機械学習をすることを検討することにより，モデル化の内容とプログラミング言語での処理というコンピュータによる自動実行として学習することにした．

自然言語処理を指導するにあたり，文章が形態素に分解されていることを理解させる必要がある．そのため，形態素解析をはじめに指導した．Web 上で形態素解析するサイトを用い，形態素に分解できることを体験させた．また，プログラムで文章を扱うことができることも理解できるように，Python のプログラムを配布し，プログラムを用いての形態素解析も体験させた．

(2) TF-IDF

3 つの果物の特徴語を抽出することを題材に指導を行った．それぞれの果物について 5 単語程度ずつ与え，手作業で単語の出現頻度を数え，値を求めさせた．指導の際に用いた式は下記のものである[27]．

$$tfidf_{i,j} = tf_{i,j} \cdot idf_i$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$idf_i = \log \frac{|D|}{|\{d: d \ni t_i\}|}$$

ここで $n_{i,j}$ は文書 d_j における単語 t_i の出現回数， $\sum_k n_{k,j}$ は

文書 d_j におけるすべての単語の出現回数の和, $|D|$ は総文書数, $|\{d: d \ni t_i\}|$ は単語 t_i を含む文の数である.

TF は単語の出現頻度であり, 生徒に理解しやすい値になるが, IDF はわかりにくい. 生徒が理解しやすくなる工夫として, IDF を求める式の真数の逆数

$$\frac{1}{|\{d: d \ni t_i\}|}$$

は単語 t_i を含む文の割合であり, 0 以上 1 以下の値を取ることから, この割合を横軸, IDF の値を縦軸に描いたグラフ (図 7) を示して, IDF の値の説明をした. なお, このグラフでは, すべての生徒が学習している底が 10 の自然対数としたため, 値としては正確ではない.

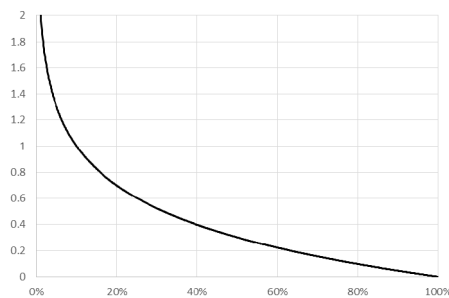


図 7 単語を含む割合と IDF の関係

(3) 単純ベイズ分類器

著者の勤務校では, 1 学年で数学 A を履修しているが, 確率の学習がそれ以来となる生徒もいるため, 条件付き確率とベイズの定理について, 簡単な例を用いて指導した.

ベイズの定理については, 「振り込め」という言葉が含まれることにより絞り込むことで迷惑メールである確率が高まることをもとに説明した.

単純ベイズ分類器については, TF-IDF 法の項と同様に, 3 つの果物について 5 単語程度ずつ与え, 分類対象の文がどの果物に分類されるかを説明した. 説明の際にベイズの定理の分子だけに着目して

$$P(Cat|Doc) \propto P(Doc|Cat)P(Cat)$$

により, この $P(Cat|Doc)$ が最大になるカテゴリーに分類できることを用いた. ここで, $P(Cat|Doc)$ は文書 Doc が与えられた際にカテゴリー Cat に含まれる確率, $P(Doc|Cat)$ はカテゴリー Cat が与えられた際に文書 Doc が得られる確率, $P(Cat)$ はカテゴリー Cat が得られる確率である. $P(Cat)$ はベイズの定理において, 事前確率とよばれており, 情報が多いカテゴリー (授業においては文章の行数が多いカテゴリー) は確率が高くなるようにした.

この際, ゼロ頻度問題が生じる可能性とその対策としてのスムージングの必要性にも触れた.

(4) コンピュータの活用

生徒には Python のプログラムを配布し, アルゴリズムのうち, 手計算で用いたものとプログラムが同じ式になって

いることを確認させた. 現状では高等学校情報科で初めてプログラミングを経験する生徒が多数のため, それ以上のプログラミングには踏み込まなかった.

また, ワードクラウドの作成も配布した Python のプログラムにより実行させた. 一般的な「これ」などの不要な単語を表示しないように, ストップワードを追加する程度のプログラムの改変はさせた.

(5) 単純ベイズ分類器を問題解決に活用

(1)~(4)の説明だけでも, 単純ベイズ分類器の仕組みを理解できるが, 機械学習の有用性の理解としては不十分であると考え, 身のまわりや社会の問題を解決する場面に活用させる学習をさせた. 3~4 名程度のグループで問題を発見し, 機械学習に必要なデータを収集させ, 実際にプログラムを実行させる流れである.

問題は生徒に発見させたが, 単純ベイズ分類器の適用場面として多様な問題を発見した. その一部を表 3 に示す.

生徒は, TF-IDF 法やワードクラウドをもとにどのカテゴリーに分類されるかを予想して文を作成し, プログラム実行時に入力し結果を得る. 予想どおりに結果を得られる場合もあるが, 事前確率 $P(Cat)$ の違いにより, 予想と異なる結果になってしまう場合があり, 実際に活用する経験を通して単純ベイズ分類器についての理解を深めた.

表 3 生徒が単純ベイズ分類器を活用した題材

● 客の希望をもとに, 希望に合った観光地やテーマパークを示す
● 好みの性格や特徴をもとに, おすすめの異性をタレントに例えて示す
● 知りたい情報をもとに, 朝のテレビ番組の中からおすすめの番組を示す
● アレルギーの症状をもとに, 原因となっている花粉を示す
● 好みの犬の性格をもとに, おすすめの犬種を示す
● 食べたい料理その調理法をもとに, その料理にあうサツマイモの品種を示す
● ショッピングモールで買いたい商品をもとに, 販売している店名を示す

(6) 授業実施による生徒の理解

生徒が単純ベイズ分類器の仕組みを理解し, 実際にプログラムを実行することを通して, 生徒が理解したことを示す.

- 分類に必要な単語が多く含まれる文章を集める必要がある.
- 必要でない単語を消す必要がある.
- キーワードになっている言葉が多く含まれるものに分類される.
- 似た単語が使われているものの分類は難しい.
- データの量(事前確率に使われている行数)の差で, 結果が大きく異なる.
- アンケートなどをもとに分類するとき, ふざけた回答があると影響を受ける.

教員の観点から、生徒の理解について感じた点を述べる。単純ベイズ分類器とあわせて TF-IDF 法やワードクラウドについて教材としたが、複数の考え方を同時に扱うと生徒にとって区別が付きにくい。そのため、生徒が理解するための時間が必要である。

しかし、生徒が機械学習の特徴をつかむことができれば、特徴語との一致の程度、事前確率の影響といった単純ベイズ分類器が返す回答の理由の理解が深まった。

本授業において、自然言語を対象とした単純ベイズ分類器が題材だったため、単純ベイズ分類器を動作させずに生徒自身が文章を分類している例が見られた。数値計算の結果を記録させて示すようにさせていなかったため、手間がかかるプログラムの活用をしなかったものと考えられる。今後、検討すべき課題である。

5. 今後の課題

5.1 情報科における学習内容について

今回の授業において、生徒の手作業によるデータの収集を行わせたが、アンケートのような多数収集されるデータやネット上からプログラム等により収集したデータが対象となる。機械学習に与えるデータが、ある意図を持ったデータが多数を占めるとノイズとなり、そのデータに結果が引きずられることになる。これまで、情報モラルについての題材は、個人情報やプライバシー、他者を攻撃したりするものが主であった。今後、ビッグデータの利用が進むことを想定し、データを生成する立場とデータを利用する立場の両面から、情報モラルについて理解するための学習が必要と考える。

また、今回の授業では触れることができなかったが、著作権法 47 条 7 にある「情報解析のための複製等」についての学習も必要となる。単に機械学習を扱うだけでなく、法整備等とも関連付けて学習できるように検討したい。

5.2 情報科以外の学習内容について

高等学校における機械学習の指導の可能性について、情報科以外の教科についても内容を精査した。その際に、検討したことを課題として示す。

情報科以外の教科で学習していないと指導できない内容ばかりではない。生物における神経細胞による伝達や数学における距離などは、それぞれの教科で学習していなくても概念が理解できれば十分なので、機械学習を扱う際に説明することは可能と考える。

また、距離について数学では平面上(2次元)の2点間の距離しか学習していないが、空間上(3次元)、さらには4次元以上についても、計算だけでもできるようにすると、k近傍法やk平均法が多次元で扱うことができる。

さらに、高等学校の数学においてベクトルは、平面上もしくは空間における有向線分として幾何的な要素だけ学習するが、複数の値をまとめて処理できる量として代数的な

面についての学習もなされれば、内積とあわせて cos 類似度といった考え方にも発展できる。

6. おわりに

本稿では、高等学校における人工知能等の新しい技術についての学習の必要性、機械学習の指導の可能性、実践した授業について述べた。また、今後の課題として、情報科で教材して検討が必要な内容およびそれ以外の教科の指導について検討すべき点を挙げた。

今後も、単純ベイズ分類器などによる機械学習の授業実践を続け、他のアルゴリズムを学習することによる効果や課題等について検討をしていきたい。

参考文献

- [1] “クイズで勝ったコンピュータ「ワトソン」の素顔”。
https://www.nikkei.com/article/DGXNASDD2305K_T20C11A30000/, (参照 2018-01-24).
- [2] “将棋ソフト、プロ軍団を3勝1敗で破る 電王戦”。
https://www.nikkei.com/article/DGXNASDG2002L_Q3A420C1CR8000/, (参照 2018-01-24).
- [3] “囲碁 AI、人知迫る一石 トップ棋士に勝ち越し 「アルファ碁」3連勝”。
<https://mainichi.jp/articles/20160313/ddn/001/040/003000c>, (参照 2018-01-25).
- [4] “囲碁 AI プロに4勝1敗 最終局も熱戦制す”。
https://www.nikkei.com/article/DGXLASDG15H7R_V10C16A3CR8000/, (参照 2018-01-25).
- [5] 新井紀子。「ロボットは東大に入れるか」という企て。情報処理, 2017, vol. 58, no. 7, p. 598-599.
- [6] “AIの雇用への影響を考える(1)「雇用の未来」を世界中で研究”。
<https://www.nikkei.com/article/DGXMZO23054500S7A101C100000/>, (参照 2018-01-25).
- [7] “AIスピーカー 続々発売 グーグル6日 LINEは値下げ”。
<https://www.nikkei.com/article/DGXMZO21913680V01C17A0M00000/>, (参照 2018-01-25).
- [8] 中央教育審議会. 幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について(答申). 2016, p.9-12.
- [9] 文部科学省. 小学校学習指導要領解説 総則編. 2017, 1p.
- [10] 文部科学省. 中学校学習指導要領解説 総則編. 2017, 1p.
- [11] 文部科学省. 中学校学習指導要領解説 技術・家庭編. 2017, p.57-58.
- [12] 文部科学省. 中学校学習指導要領解説 社会編. 2017, p.19-22.
- [13] 文部科学省. 中学校学習指導要領解説 理科編. 2017, p.64-65.
- [14] 中央教育審議会. 幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について(答申). 2016, p.142-143.
- [15] 中央教育審議会. 幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について(答申)別添資料. 2016, 87p.
- [16] “人工知能とデータサイエンティストの役回り”。
<https://www.msi.co.jp/userconf/2015/pdf/muc15.pdf>, (参照 2018-01-25).
- [17] 文部科学省. 高等学校学習指導要領解説 情報編. 2010, p.30-32.
- [18] 文部科学省. 高等学校学習指導要領解説 数学編.

2009,p.19-58.

- [19] 文部科学省. 高等学校学習指導要領解説 理科編. 2009,89p.
- [20] “単回帰分析”.
<http://www.sist.ac.jp/~kanakubo/research/statistic/tankaiki.html>,
(参照 2018-01-25).
- [21] Sebastian Raschka. Python 機械学習プログラミング. インプレス,2016,p.89-92.
- [22] 多田智史. 新しい人工知能の教科書. 翔泳社,2016,p.175-177.
- [23] Sebastian Raschka. Python 機械学習プログラミング. インプレス,2016,p.77-86.
- [24] Rachel Schutt,Cathy O’Neil. データサイエンス講義. オライリー・ジャパン,2014,p.104-111.
- [25] Rachel Schutt,Cathy O’Neil. データサイエンス講義. オライリー・ジャパン,2014,p.87-91.
- [26] Sebastian Raschka. Python 機械学習プログラミング. インプレス,2016,p.17-23,325-332.