

問題解決に自然言語処理と機械学習を用いた協働学習の実践

埼玉県立川越南高等学校 春日井 優

次期学習指導要領の情報Ⅱでデータサイエンスが内容に含まれることになった。また、近年人工知能の発展がめざましく、機械学習などデータサイエンスの手法が用いられている。そこで、人工知能を支える技術のうち、自然言語処理（形態素解析・tf-idf）と機械学習（単純ベイズ分類器）について授業を行った。さらに、グループでの協働学習により、それらの手法を問題解決での活用法を生徒が考え、インターネット上の文章を用いて処理し、解決方法の提案をする授業を行った。

1. はじめに

昨今、人工知能が話題に上がり、学習指導要領改訂の理由の1つにも挙げられている^①。人工知能を支える技術には、データサイエンスの手法も用いられており、次期学習指導要領で設置される「情報Ⅱ」においてもデータサイエンスが学習内容に含まれることになった^②。

そこで、データサイエンスのうち機械学習の仕組みを説明し、生徒は学習した知識・技能を活用して問題解決に取り組む授業を実践した。その授業及び生徒の成果物について報告する。

2. 授業計画

授業の目的として、人工知能などのコンピュータが算出する結果を、我々人間が適切に捉え有効に活用するための知識を習得し、適切に活用し、その経験を通して望ましい情報社会の在り方を考えることとした。

そのための計画として、初めに知識を理解することを計画した。情報のデジタル表現やプログラムによる情報処理の仕組み、自然言語処理、機械学習の仕組みが知識として必要である。これらの内容は、高等学校においても十分指導可能である^③。次に、これらの知識を用いて、身のまわりや社会の中で活用できる場面を考えることにより、自然言語処理や機械学習の適切な活用を考えさせた。これらの学習を振り返り、情報社会の在り方について考えさせた。本実践に関連する授業計画については、表1に年間計画、表2に授業計画として整理した。

3. 一斉授業による知識の理解

3.1 知識の理解を活用・探究につなげる工夫

本授業において、データサイエンスのうち特に機械学習の内容を理解するための授業において工夫する点について検討を行った。本授業後にテストで知識の理解について評価することも行った。しかし、それだけでは十分でなく、問題解決を生徒が考える際に内容の理解が不十分であると、問

表1 年間計画(本内容関連の概略)

前期	文字のデジタル表現 基本的なプログラミング(Python)
後期	機械学習の仕組み 機械学習を用いた問題解決

表2 本授業の計画

1 時間	形態素解析
2～3 時間	tf-idf による特徴語抽出 Word Cloud による特徴語の可視化
4～6 時間	ベイズの定理・単純ベイズ分類器
7～14 時間	グループでの問題解決・発表・振り返り

表3 活用・探究につなげるために配慮した観点

- ① ブラックボックス化せず仕組みがわかる。
- ② 他教科の内容は高等学校の範囲内に抑える。
- ③ 対象とするデータの収集が容易なものとする。
- ④ 問題となる事項とデータの関連がわかる。
- ⑤ 多様な問題への適用が可能である。

題解決のために知識を活用することができない。そこで、授業内容を検討する際に表3に示す観点④を踏まえて、授業内容を決定した。また、単に数式を示すだけでなく、具体例を用いて生徒が内容の理解を深めることができるようにした。

3.2 形態素解析

自然言語処理をするにあたり、コンピュータは言葉の意味を理解しているのではなく、単に辞書との対応をもとに処理していることを説明した。形態素解析をするサイトを利用して、生徒が考えた短文を形態素解析させた。また、Python のプログラミングを配布し、プログラムを用いた形態素解析も経験させた。Python での形態素解析にはライブラリ janome を用いた。

この経験を通して、生徒は「ディズニーランド」が「ディズニー」と「ランド」に分割される経験をし、形態素解析の特性についても理解を深めた。

3.3 tf-idf と Word Cloud

次に、文章を形態素解析した結果をもとに特徴語を抽出する仕組みについて説明した。

tf = 形態素の出現回数 / 全形態素数
idf = $\log(\text{全文書数} / \text{形態素を含む文の数})$
tf-idf = tf * idf

の式を、いくつかの果物を対象に形態素を与え、手計算により値を求め、特徴語を抽出した。あわせて Python を用いて計算させ、その結果を Excel に出力し、グラフとして描画させた。

また、特徴語を視覚化する方法として Word Cloud を紹介し、Python のライブラリを用いて作成させた。

3.4 ベイズの定理と単純ベイズ分類器

次に、機械学習の手法の1つである単純ベイズ分類器(Naive Bayes classifier)と、その考えのもとになるベイズの定理の説明をした。

ベイズの定理については、数学で学習している条件付き確率と関連付けて、

$P(\text{cat} | \text{view}) = P(\text{cat}) * P(\text{view} | \text{cat}) / P(\text{view})$ が導かれ、迷惑メールの分類を例に説明した。また、観測情報に影響され推測結果が変化することも説明し、定性的には観測したことを経験則に当てはめて推測する場面が日常的に行っていることも補足した。

単純ベイズ分類器については、ベイズの定理の計算結果により分類できることを、果物の分類を題材として説明した。果物の分類する際に、学習データの文章に出現しない形態素が出現した場合の扱い(ゼロ頻度問題)や Python で処理する際のアンダーフローについても具体例を用いて説明した。説明後に配布した Python のプログラムを用いて、実際に果物の分類をさせた。

4. 協働学習による問題解決

4.1 課題提示・問題発見の工夫

3章に示した一斉授業を終えた後、3~4名による問題解決への活用を協働学習として行うこととした。生徒に示した課題は「tf-idf と単純ベイズ分類器を、社会における問題の発見や解決への用い方を考えて、社会における問題解決を提案してください」というものである。

ここまでの学習は、単に計算の仕組みの説明と与えられた例題を解くにとどまっている。そのため、生徒にとってはどのように問題解決に活用することができるかイメージしにくい。

そのため、教師側から実際に問題解決に活用する例を示し、生徒が制作する成果物や発表がイメージできるようにした。

また、問題を発見する工夫として、①個人で考える、②グループで意見を交換する、③他のグループの人に説明して意見を交換する、ということ

を行った。

4.2 生徒の考えた問題

生徒はグループで様々な問題の解決に適用しようと考えた。花粉症の症状について書かれた文章を学習させ原因となる花粉を推測させたり、犬種について書かれた文章を学習させペットを飼いたい人に提案したりさせるなど、様々な問題を発見したり解決したりする場面に適用していた。

生徒の成果物の一例を図1に示す。問題解決を通して、単純ベイズ分類器の特徴である「ない」という否定語が加わったときの挙動、学習データの質に依存すること、与える文章が何に分類されるか予想するなど仕組みについて理解が深まった。

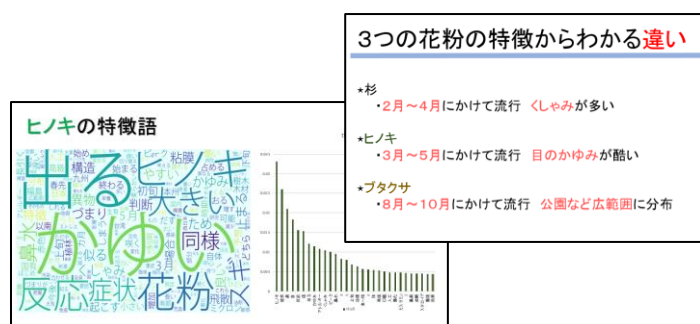


図1 生徒が作成したスライド

5. おわりに

本稿執筆時には、高等学校学習指導要領解説編は公表されていない。そのため、本内容が次期学習指導要領の内容と適合するか不明である。しかし、授業を実践する際に考慮する観点など参考になる点があると考えている。

本授業での単純ベイズ分類器を用いて、秋田県の日本酒を分類する bot を作成した。Twitter および LINE のそれぞれの bot は、下記からアクセスできる。動作の保証はできないが、お土産選びに試していただきたい。



Twitter bot
@NaiveBayesSake



LINE bot
@ozj1358q

参考文献

- (1) 小学校学習指導要領解説総編、文部科学省(2017)
- (2) 高等学校学習指導要領、文部科学省 (2018)
- (3) 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、情報処理学会研究報告 Vol2018-CE-143No19、春日井優 (2018)
- (4) 活用・探究につなげる授業の検討-機械学習を題材とした授業-、日本情報科教育学会第11回全国大会、春日井優・森本康彦 (2018)