

アソシエーション分析による学生の社会・消費行動にみる傾向の分析

2019SS048 長田ゆうか 2019SS049 仲井諒子

指導教員：小市俊悟

1 はじめに

私たち人間はそれぞれ行動や好み、感覚が異なる。しかし、相手との共通点が多ければ会話やイベントなど一緒に過ごす時間を楽しむことができる。相手の情報の一部から、趣味や行動などの傾向を推測することができ、自分との共通点を事前に知ることができたら人間関係の構築に役立つのではないかと考える。本研究では、大学生を対象にこのような人間関係に役立つ趣味・行動などを推測することを目的とし、そのために大学生・専門学生の傾向を探る。より具体的には、大学生と専門学生を対象としたアンケート学生がどのように実施し、学生が持つどのような条件が、お金の使い方・コンビニエンスストアの使い方に関連があるのかをアソシエーション分析という手法を用いて明らかにする。

2 調査と回答データ

調査方法として、Google Forms を用いたオンラインアンケートを行った。全国の学生を対象とし、SNS を使ってアンケート画面が表示される URL を送信した上で回答を依頼したところ、全国の 124 人の学生 (大学生・専門学生) から回答を得た。アンケートの回答データ (dataA) と、アンケートの回答データを一部加工して作ったデータ (dataB) を分析対象とする。

2.1 dataA

dataA は次に挙げる項目からなる。

1. 学生について学年、学部、性別、出身地、兄弟構成、部活 (サークル)、アルバイト、暮らし方、留学経験
2. コンビニエンスストアの使い方について使用頻度、利用する日、1 回の利用の滞在時間、多く利用するコンビニエンスストア、多く利用する理由、よく買うもの、好きなコンビニエンスストア、好きな理由

2.2 dataB

dataB は dataA の項目の一部を大まかにまとめることで得られるものである。新たに追加される項目とカッコ内に値ごとの該当データ数を以下に記す。

1. 専攻分野の文理 (文系：72 人、理系：52 人)
2. サークル、部活の所属有無 (有：73 人、無：51 人)
3. サークル、部活の種類 (スポーツ系：46 人、文化系、その他：78 人) ※兼部含む
4. アルバイトの有無 (有：111 人、無：13 人)
5. 兄弟姉妹構成においての自分の立場 (長子：44 人、間の子：17 人、末っ子：47 人、一人っ子：16 人)

3 分析方法

3.1 アソシエーション分析

アソシエーション分析とは、データマイニングの手法の一つで、データから「 $X \Rightarrow Y$ 」というような関係 (連想・アソシエーション) を見つけ出すものである。そのような関係は、必ずしもすべてのデータで成り立つとは限らないので、信頼度 (confidence) や支持度 (support)、リフト (lift) などの指標で評価し、評価値が高いパターンを導き出す。次に各指標について説明する。関係「 $X \Rightarrow Y$ 」について、データのうち条件 X と Y を満たすものを同じく X と Y を用いて、それぞれ集合 X と Y と書く。また、データ全体を集合 A と書く。

信頼度 (confidence) 集合 X に含まれるデータのうち、集合 Y にも含まれるもの、すなわち、集合 $X \cap Y$ が、集合 X に対してどれくらいの割合を占めるかを表すが、信頼度 $\text{Conf}(X \Rightarrow Y)$ であり、下記の数式 (1) で求められる。

$$\text{Conf}(X \Rightarrow Y) = \frac{|X \cap Y|}{|X|} \quad (1)$$

信頼度が高いということは、条件 X を満たすものの多くが条件 Y も満たすことになるので、関係「 $X \Rightarrow Y$ 」が成り立ち易いということを意味する。

支持度 (support) データ全体に対して、集合 $X \cap Y$ が占める割合を表すものが支持度 $\text{Support}(X \Rightarrow Y)$ であり、下記の数式 (2) で求められる。

$$\text{Support}(X \Rightarrow Y) = \frac{|X \cap Y|}{|A|} \quad (2)$$

これは、データ中に条件 X と Y を同時に満たすものが、どれくらい現れるかを表すものであるから、その値が小さい場合、関係「 $X \Rightarrow Y$ 」の信頼度が高いとしても、それを適用できるデータがあまり多くないことを意味する。したがって、データに頻出する関係を見出すという観点では、この支持度は高い方が良いということになる。

リフト (lift) データ全体に対して、集合 Y が占める割合を考え、さらに、信頼度をそれによって割ることで求められるのが、リフト $\text{Lift}(X \Rightarrow Y)$ である。下記の数式 (3) で求められる。

$$\text{Lift}(X \Rightarrow Y) = \frac{\frac{|X \cap Y|}{|X|}}{\frac{|Y|}{|A|}} \quad (3)$$

これは、集合 X に含まれる集合 Y の割合と、データ全体に含まれる Y の割合を比較するものであるから、それが

高いということは、集合 X の中では条件 Y を満たす傾向が、全体と比較して強いことを意味する。したがって、基本的には、リフトの値が大きい関係「 $X \Rightarrow Y$ 」がデータから導くものとしては望ましくなるが、そもそもデータ全体に含まれる Y の割合が小さいと自ずとその値が大きくなるので、この点に注意する必要がある。

実際のアソシエーション分析には、python により実装されたアプリアリアルゴリズムを用いた。関係「 $X \Rightarrow Y$ 」の抽出には、支持度、信頼度、リフトのそれぞれについて抽出するか否かのしきい値を設定する必要がある。

4 データの項目群の設定

分析に先立って、対象とするデータに対して下記のようなデータ項目群をいくつか設定することにした。要旨ではそのうちの2つの項目群を取り上げる。

兄弟姉妹・社会性に関する項目群

- (i) 部活・サークルに所属しているか
- (ii) 友人知人とのコミュニケーションにお金を使うのか
- (iii) 兄弟姉妹の立場（一人っ子か一番年上か一番年下か、長男長女と末っ子の間か）

項目群の1つ目は、上記の3項目である。これらは、兄弟姉妹の立場・社会性に関する項目であるので、以降では、兄弟姉妹・社会性に関する項目群と呼ぶ。

コンビニ利用状況項目群

- (i) 1週間で何日コンビニを利用するかどうか
- (ii) 1回の利用での滞在時間
- (iii) コンビニを利用するのは平日か休日かどうか

項目群の2つ目は、上記の3項目である。これらは、コンビニの利用状況に関する項目なので、以降では、コンビニ利用状況項目群と呼ぶ。これらは、好きなコンビニとよく利用するコンビニの違いに関する分析結果なので、以降ではコンビニ利用の理由項目群と呼ぶ。

兄弟姉妹・社会性に関する項目群は全体を1つのデータセットとして扱う。兄弟姉妹の有無や立場によって、社交性の高さに違いがあるのかを分析するが、社交性の高さを検証する項目は部活・サークル活動の所属の有無と、友人知人とのコミュニケーションにお金を使うかの項目である。コンビニの利用状況項目群については、データをまず、一人暮らしか実家暮らしかに分け、さらに実家暮らしについては、男女で分ける。一人暮らしの学生データは27.4%ほどしかないので、男女で分けると該当のデータ数が少なくなるため男女で分けることはしない。

以上のような項目の限定とデータの分類により、合計4個のデータセットが得られる。アソシエーション分析は、それらのデータセットのそれぞれに対して行った [1], [2].

5 分析結果

5.1 兄弟姉妹・社会性に関する項目群に対する分析結果

- (1) 部活・サークルに所属している \Rightarrow 友人知人とのコミュニケーションにお金を使わない
- (2) 友人知人とのコミュニケーションにお金を使わない \Rightarrow 部活・サークルに所属している
- (3) 部活・サークルに所属している \Rightarrow 長子である
- (4) 長子である \Rightarrow 部活・サークルに所属している
- (5) 部活・サークルに所属していない \Rightarrow 末っ子である
- (6) 末っ子である \Rightarrow 部活・サークルに所属していない
- (7) 部活・サークルに所属しており、友人知人とのコミュニケーションにお金を使わない \Rightarrow 長子である
- (8) 長子であり、友人知人とのコミュニケーションにお金を使わない \Rightarrow 部活・サークルに所属している
- (9) 長子である \Rightarrow 部活・サークルに所属しており、友人知人とのコミュニケーションにお金を使わない
- (10) 知人・友人とのコミュニケーションにお金を使わず、部活・サークルに所属していない \Rightarrow 末っ子である
- (11) 末っ子であり、部活サークルに所属していない \Rightarrow 知人・友人とのコミュニケーションにお金を使わない
- (12) 末っ子であり、知人・友人とのコミュニケーションにお金を使わない \Rightarrow 部活サークルに所属していない
- (13) 末っ子であり、知人友人とのコミュニケーションにお金を使わない \Rightarrow 部活・サークルに所属していない
- (14) 部活・サークルに所属していない \Rightarrow 末っ子であり、知人友人とのコミュニケーションにお金を使わない
- (15) 末っ子である \Rightarrow 部活・サークルに所属しておらず、知人友人とのコミュニケーションにお金を使わない
- (16) 友人知人とのコミュニケーションにお金を使う \Rightarrow 部活・サークルに所属していない
- (17) 部活・サークルに所属していない \Rightarrow 友人知人とのコミュニケーションにお金を使う
- (18) 友人知人とのコミュニケーションにお金を使う \Rightarrow 末っ子である
- (19) 末っ子である \Rightarrow 友人知人とのコミュニケーションに

お金を使う

- (20) 兄弟姉妹で、長子と末っ子の間の立場である
⇒ 友人知人とのコミュニケーションにお金を使わない
- (21) 友人知人とのコミュニケーションにお金を使う ⇒ 長子である
- (22) 長子である ⇒ 友人知人とのコミュニケーションにお金を使う

表1 兄弟姉妹・社会性に関する項目に対して抽出された関係の各指標値

関係	支持度	信頼度	リフト
(1)	0.40	0.70	1.02
(2)	0.40	0.58	1.02
(3)	0.25	0.44	1.26
(4)	0.25	0.72	1.26
(5)	0.19	0.45	1.19
(6)	0.19	0.51	1.19
(7)	0.17	0.42	1.21
(8)	0.17	0.72	1.26
(9)	0.17	0.49	1.20
(10)	0.14	0.47	1.25
(11)	0.14	0.70	1.02
(12)	0.13	0.55	1.29
(13)	0.14	0.55	1.28
(14)	0.14	0.36	1.25
(15)	0.14	0.32	1.05
(16)	0.14	0.45	1.05
(17)	0.14	0.32	1.05
(18)	0.13	0.42	1.11
(19)	0.12	0.34	1.11
(20)	0.11	0.78	1.21
(21)	0.11	0.37	1.06
(22)	0.11	0.33	1.06

5.2 兄弟姉妹・社会性に関する項目群に対する分析結果の考察

得られた関係を兄弟・姉妹における立場別に考察する。

長子に関する関係 長子に関して、関係(3), (4), (7), (9), (21), (22) が得られた。

1. 関係(3)の信頼度は0.44であるので、部活・サークルに所属している学生のうち、約4.4割は長子である。
2. 関係(4)の信頼度は0.72であるので、長子の学生の約7割は部活・サークルに所属している。
3. 関係(7)の信頼度は0.42であるので、部活・サークルに所属しており、友人知人とのコミュニケーションにお金を使わない学生の約4.2割は長子である。

4. 関係(9)の信頼度は0.49であるので、長子の学生の約4.9割は部活・サークルに所属しており、友人知人とのコミュニケーションにお金を使わない。
5. 関係(21)の信頼度は0.37であるので、友人・知人とのコミュニケーションにお金を使う学生の約3.7割は長子である。
6. 関係(22)の信頼度は0.33であるので、長子の学生の約3.3割は友人・知人とのコミュニケーションにお金を使う。

長子と末っ子の子に関する関係 長子と末っ子の子に関して関係(3-20) が得られた。

1. 関係(20)の信頼度は0.78で、長子と末っ子の間の学生の約7.8割は友人知人とのコミュニケーションにお金を使わない。

末っ子に関する関係 末っ子に関して、関係(5), (6), (13), (14), (18), (19) が得られた。

1. 関係(5)の信頼度は0.45であるので、部活・サークルに所属していない学生の約4.5割は末っ子である。
2. 関係(6)の信頼度は0.51であるので、末っ子の学生の約5.1割は部活・サークルに所属していない。
3. 関係(13)の信頼度は0.55であるので、末っ子であり、知人友人とのコミュニケーションにお金を使わない学生の約5.5割は部活・サークルに所属していない。
4. 関係(14)の信頼度は0.50であるので、部活・サークルに所属していない学生の約5割は末っ子であり、知人友人とのコミュニケーションにお金を使わない。
5. 関係(18)の信頼度は0.42であるので、友人知人とのコミュニケーションにお金を使う学生の約4.2割は末っ子である。
6. 関係(19)の信頼度は0.34であるので、末っ子の学生の約3.4割は友人知人とのコミュニケーションにお金を使う。

以上の結果から、長子と末っ子の学生は比較的、部活・サークルや友人知人とのコミュニケーションに時間を使っていると考えられる。部活・サークルの所属について長子と末っ子の学生を比較すると後者の方が所属率が低いことがわかった。また、長子、長子と末っ子の間、末っ子のそれぞれの立場について友人知人とのコミュニケーションにお金を使う割合はいずれも3割程度で差異がない。一人っ子の学生はサンプル数が少なく、傾向が得られなかったので兄弟姉妹のいる学生との比較ができるものではなかった。この分析結果によれば、長子であることが部活やサークルを通じた交流に対して比較的、積極的であると言えるだろう。

5.3 コンビニの利用状況項目群に関する分析結果

一人暮らし、実家暮らしの男子学生、実家暮らしの女子学生のそれぞれについてアソシエーション分析で得られた

主な関係は次である。

一人暮らしの学生

- (23) 文系学生 ⇒ 1週間で1日コンビニを利用する
- (24) 1週間で1日コンビニを利用する ⇒ 文系学生
- (25) 1週間で2～3日コンビニを利用する ⇒ 理系学生
- (26) 理系学生 ⇒ 1週間で2～3日コンビニを利用する

表2 一人暮らしの学生のコンビニの利用状況項目群に関して抽出された関係の各指標値

関係	支持度	信頼度	リフト
(23)	0.21	0.39	1.07
(24)	0.21	0.58	1.07
(25)	0.18	0.50	1.10
(26)	0.18	0.40	1.10

実家暮らしの男子学生

- (27) 1週間で2～3日コンビニを利用する ⇒ 文系学生
- (28) 文系学生 ⇒ 1週間で2～3日コンビニを利用する

表3 実家暮らしの男子学生のコンビニの利用状況項目群に関する分析に対して抽出された関係の各指標値

関係	支持度	信頼度	リフト
(27)	0.26	0.62	1.27
(28)	0.26	0.53	1.27

実家暮らしの女子学生

- (29) 1週間で2～3日コンビニを利用する ⇒ 文系学生
- (30) 文系学生 ⇒ 1週間で2～3日コンビニを利用する
- (31) 理系学生 ⇒ ほとんどコンビニを利用しない
- (32) ほとんどコンビニを利用しない ⇒ 理系学生
- (33) 1週間で1日コンビニを利用する ⇒ 理系学生
- (34) 理系学生 ⇒ 1週間で1日コンビニを利用する

表4 実家暮らしの女子学生のコンビニの利用状況項目群に関する分析に対して抽出された関係の各指標値

関係	支持度	信頼度	リフト
(29)	0.19	0.69	1.06
(30)	0.19	0.30	1.06
(31)	0.12	0.35	1.08
(32)	0.12	0.39	1.08
(33)	0.11	0.46	1.32
(34)	0.11	0.30	1.32

5.4 コンビニの利用状況項目群に関する分析結果の考察

一人暮らしの学生は、関係(24)、(25)を見てみると、文系学生の方が1週間で1日コンビニを利用するに対し、理

系学生は1週間で2～3日コンビニを利用するという結果が得られたことから、利用頻度は僅かではあるが理系学生の方が高い。実家暮らしの学生は、関係(27)、(29)を見てみると、男女共に文系学生の利用頻度は1週間で2～3日コンビニを利用する。実家暮らしの女子学生は、関係(31)、(32)を見てみると理系学生はほとんど利用しない、もしくは1週間で1日コンビニを利用するという結果が得られた。男子学生の理系学生に関しては関連する関係が得られなかったが、実家暮らしのコンビニ利用頻度は文系学生の方が低く、一人暮らしとは反対の結果となり、興味深い。

6 おわりに

兄弟姉妹・社会性に関する項目に対する分析から、兄弟姉妹の立場別で部活・サークルに関する社会性やお金の利用による友人・知人コミュニケーションの積極性に違いがあることがわかった。しかし、これらは兄弟姉妹の立場とお金を友人・知人とのコミュニケーションに使うのか、部活やサークルに所属しているかという項目のみで分析しているので、お金を使わずに友人・知人とのコミュニケーションに時間を使っている場合もあることを考慮すると、これらの分析は学生の社会活動の活発性を明らかにするには、分析する項目が不十分だと言えるだろう。今後の課題として、兄弟姉妹の立場で学生生活での社会活動の活発性を明らかにするには、より多くの学生に対して再度アンケートを実施し、これまで実施したアンケートの項目に加えて、友人・知人とのコミュニケーションに時間を使う頻度も調査して分析する必要がある。

コンビニの利用事情について、一人暮らしか実家暮らし、男子学生か女子学生、理系学生か文系学生かでそれぞれ違う傾向が得られた。どの時間帯にコンビニを利用するか分かれれば、昼食目的で利用するのか、夕食目的で利用するのかなど利用目的で傾向が掴めたり、コンビニの1回の利用で使う利用金額が分かれば他の傾向を掴めたかもしれない。しかし、一定の回答数のアンケート収集に時間がかかるため、本研究ではアンケートの再調査に取り組めなかったが、分析を進めていく中でこの項目があれば、より関連性が見つかりそうだと思うこともあり、それを今後の課題にしたい。

参考文献

- [1] 赤石雅典：『Pythonで儲かるAIをつくる』。日経BP、東京、2020。
- [2] 赤石雅典：2020、『Pythonで儲かるAIをつくる』、Qiita。
(<https://qiita.com/makaishi2/items/c5f06f844cdb8454b6c3>, 2022年6月5日取得), 2020年9月22日。