

機械学習を用いた口コミサイトの名店選定に対する分析

2019SS067 白川日南子 2019SS091 横山未来

指導教員：小市俊悟

1 はじめに

食べログは、日本全国のお店を無料掲載し、実際に食事したユーザーによる感想や写真を口コミとして公開することで、信頼できるレストランガイドとして活用できる口コミサイトである。この口コミサイトでの評価を見て、お店を決める人も多い。そのサイトでは、食べログ百名店として、ジャンルごとに高い評価を集めた100店を選出している。本研究の目的は、機械学習を用いて、食べログ百名店に選出されるような人気店の特徴を分析することである。はじめに、東京のラーメン店に着目し検証を行う。そのために、お店のいくつかの属性について、データを収集し、百名店に選ばれた店と、そうでない店が持つ特徴を、決定木分析やランダムフォレストを用いて分析する。次に、地域やジャンルによって人気店の特徴に違いがあるのかを検証するために、西日本のラーメン店、全国のそば店、全国のフレンチ店について同様の分析を行い、東京のお店に対して得られた結果と比較・考察を行う。

2 使用するデータ

東京と西日本のそれぞれについて、人気ラーメン店とその他のラーメン店の違いを調べるために、食べログ百名店100店舗のデータとその他ラーメン店100店舗について、下記に挙げる属性に関するデータを集めた。ただし、2021年のラーメン百名店は99店舗の選出であったため、その他ラーメン店も99店舗とした[1, 2, 3].

- 最寄り駅からの距離
- テーブル席の数
- 総営業時間
- 価格
- 定休日の有無

このうち、定休日の有無については、定休日がある場合は1、無い場合は0で表す。また、後の分析において、これらはこの順に属性 $X[0]$ から $X[4]$ に対応する。

3 解析方法の説明

解析方法として、決定木分析とランダムフォレストを用いる。決定木分析で百名店とその他の店の特徴を分析し、ランダムフォレストで各属性の重要度を算出することを目的とする。

3.1 決定木分析

決定木では、木構造の各頂点にはデータが持つ属性に対する条件が対応し、それらの条件を木構造によって表現される順番で、根から階層的に適用することでデータを分類

する。本研究では、Pythonにおけるscikit-learnパッケージを用いて決定木を求めた。図1は階層の深さを5とした決定木の1例である。一つ目の分割では、属性 $X[2]$ 、すなわち、総営業時間の特微量について、9.25を閾値にデータを分割している。属性 $X[2]$ が9.25以下に該当するデータに関する情報が、一番上の四角につながっている2つの四角のうち、左側の四角に示されている。その四角では、“sample=97”となっているので、学習に用いた171個のデータのうち、97個のデータが該当していることがわかる。また $value=[15, 82]$ となっているので、その97個のデータのうち、15個はその他のラーメン店であり、残り82個は百名店である。このように決定木では、木構造に従って属性に対する条件を適用することでデータを分類する。末端の四角に該当するデータが百名店とその他のラーメン店のどちらか一方で占められるようになれば、良い決定木が得られたことになり、データの百名店とその他のラーメン店を分ける条件が明らかになったと言える。

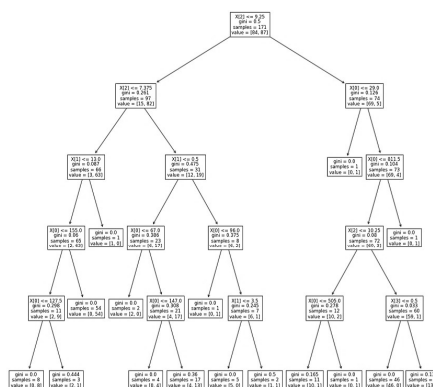


図1 東京のラーメン店を対象とした決定木

3.2 ランダムフォレスト

ランダムフォレストは過学習を起し易いと言われる決定木の欠点を補いつつ、その高い分類能力を維持向上させることを目的として考案された手法である。決定木が過学習を起す要因の一つは、データに誤りのデータが含まれているからであり、誤りのデータを分類するためだけの条件が選ばれてしまうからである。そこで、ランダムフォレストでは、学習データをすべて使うのではなく、ランダムにその一部を抽出して利用し、誤りのデータが除かれた学習を行うことが期待できる。しかし、学習データの数が減る分、学習の精度が落ちる可能性があるため、学習データのランダムな抽出を適当な数だけ繰り返し、それぞれに対

して決定木の作成を行う。その結果として、決定木が複数得られ、さらに一つのデータに対して、複数の判別結果が得られるので、最終的な判別結果は多数決で決め、分類能力の精度を維持向上させる。

決定木を構成している条件は、データが持つ属性のいずれかを対象とする。ランダムフォレストでは、決定木を一つ作る際に、使用できる属性をランダムに選抜することも行っており、これにより多様な決定木を作成している。そうして得られた多様な決定木において、よく採用される属性が存在する。そのような属性は、分類に際して重要な属性であると考えられる。このような考えに基づき、それを定量的に評価したものが属性(特徴量)の重要度である。

4 分析結果

4.1 分析結果の予想

分析を行う前に、百名店に選ばれる店の特徴を、最寄り駅からの距離が近い、テーブル席の数が多い、総営業時間が長い、価格が安い、定休日が少ないという予想を立てた。

4.2 決定木分析を用いた結果

はじめに、東京のラーメン百名店 99 店舗と、その他のラーメン店 99 店舗を scikit-learn によって提供される決定木分析を用いて分析した。その結果、得られた決定木が図 1 である。図 1 から読みとれる百名店の結果は、最寄り駅からの距離が 155m 以上、テーブル席の数が 13 席以下、総営業時間が 7.375 時間以下となった。一方、その他のラーメン店の特徴は、最寄り駅からの距離が 29m 以上 811.5m 以下、総営業時間が 10.25 時間以上、定休日が 0.5 日以下となった。

次に、西日本のラーメン店も同じ特徴がみられるかを確認するために、西日本のラーメン店を対象に同様の分析を行った。西日本のラーメン店の百名店は、大阪府、京都府、奈良県、兵庫県、福岡県、滋賀県、石川県、富山県、三重県の 9 の地域から選出されている。その他のラーメン店も同じ地域の 100 店を、データとして集めた。収集対象とした属性は、東京と同じである。結果として得られたのは、図 2 に示す決定木である。図 2 から読みとれる百名店の特徴は、最寄り駅からの距離が 180.5m 以上、テーブル席の数が 5 席以下、総営業時間が 8.75 時間以下となった。

その他のラーメン店の特徴は、総営業時間が 8.75 時間以上、価格が 790 円以下、定休日が 0.5 日以下、テーブル席が 0.5 席以上となった。

4.3 ランダムフォレストを用いた結果

東京のラーメン百名店とその他のラーメン店を分類するのに、最も重要だと考えられる属性を定量的に明らかにするために、ランダムフォレストを用いて、各属性の重要度を求めた。その結果、総営業時間の値が飛びぬけて高いため、総営業時間が百名店か否かの分類に重要であることが分かる。

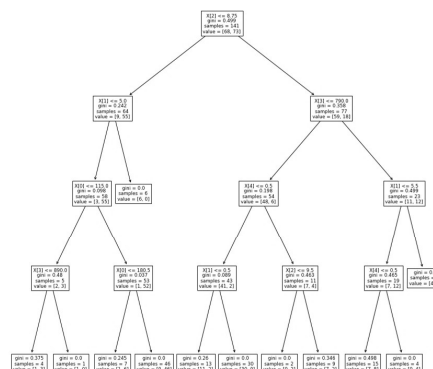


図 2 西日本のラーメン店を対象とした決定木

次に西日本のラーメン百名店とその他のラーメン店を分類するのに、最も重要だと考えられる属性を定量的に明らかにするために、ランダムフォレストを用いて、各属性の重要度を求めた。その結果、東京と同様に、総営業時間の値が飛びぬけて高いため、総営業時間が百名店か否かの分類に重要であることが分かる。

5 考察

5.1 東京の百名店

分析結果より、東京の百名店が持つ特徴は、最寄り駅からの距離が遠い、テーブル席の数が少ない、総営業時間が短い、であった。百名店か否かを分ける特徴について明らかとなったが、決定木分析やランダムフォレストだけでは、特徴と分類の間の因果関係まではわからない。あくまで百名店に広く共通し、その他の店にはあまり共通しない特徴を明らかにしただけなので、そのことにも注意して、得られた結果について順に考察する。

東京の百名店とその他のラーメン店の各属性の平均を表 1 に示す。表 1 から、最寄り駅からの距離、テーブル席の数、総営業時間、定休日の有無に大きな差があり、価格についてはそこまで大きな差がないことが読みとれる。

表 1 東京の百名店とその他のラーメン店の各属性の平均

属性名	百名店	その他のラーメン店
最寄り駅からの距離	349	245
テーブル席の数	0.71	4.51
総営業時間	6.3	11.7
価格	850	795
定休日の有無	0.72	0.28

5.1.1 最寄り駅からの距離について

一般には、最寄り駅からの距離が遠いということは、立地があまりよくなく、通いづらい場所であると考えられる。そこで、百名店の最寄り駅について、もう少し詳細に調べ

たところ、その最寄り駅は都心から離れたところが多く、ラーメン店も地下鉄以外の公共交通機関を利用しなければ行けない場所にあることが多いことが分かった。したがって、百名店の立地は、一般には集客が難しいと思われる場所が多かった。それにも関わらず、百名店である、すなわち、人気を集める理由を考えた。一つ考えられるのは、立地が悪い分、家賃を押さえることができるため、その分を、こだわった食材に当てることができる。これにより味を高め、ビル街や街中ではないものの、住宅街にあることが多いため、地元の人や近所の人がたくさん通うと考えられる。

5.1.2 テーブル席について

テーブル席の数が少ないということは、店が狭いと考えられる。これも家賃を押さえることにつながる。また、店が小規模であることから、同時に対応が必要な客数も限定されるため、従業員も少なく済む。これらは、経費の削減に繋がることから、最寄り駅からの距離と同様に、食材の調達などに有利に働くと考えられる。ただし、総売上を考えると十分な客の回転数も必要となることから、テーブル席の少なさが百名店となる直接的な要件とはならないであろう。むしろ、百名店となるような店は少ないテーブル席でも十分な利益を上げられる店であり、そのような店は上述のような好循環の中で経営ができていていると考えられる。

5.1.3 総営業時間について

総営業時間が短いということは、仕込みや準備に時間をかけているとも考えられる。これは味の向上に繋がると考えられるが、テーブル席と同様に、総営業時間が短い分、十分な客の回転数がないと経営が難しいのではないかと考えられる。したがって、経営できていることを前提に総営業時間が短いことと百名店であることに強い関係があるのだと考える。

5.2 その他のラーメン店

解析結果より、その他のラーメン店が持つ特徴は、最寄り駅からの距離が近い、総営業時間が長い、定休日が少ない、という結果を得た。この結果について順に考察する。なお、その他のラーメン店はランダムに選択したが、それでもチェーン店と思われる店が、百名店に比べるとやや多い。その他のラーメン店が持つ特徴はチェーン店にも共通するような特徴とも考えられるので、これが結果に反映されていると考えられるところもある。

5.2.1 最寄り駅からの距離について

最寄り駅からの距離が近いということは、一般的には立地が良いことを意味し、また、より詳細な調査で多くの店の最寄り駅が街中や繁華街にあることが多いことが分かった。したがって家賃が高いのではないかと考えられる。そのため、食材のコストや人件費など、その他の経費を考えると、食材にそこまでこだわることができなると考えられ

る。これは味について少し妥協を迫る可能性がある。

5.2.2 総営業時間について

総営業時間が長いということは、アルバイトを雇いながら、交代して営業していると考えられる。多くのアルバイトを雇うには、相応の人件費がかかる。最寄り駅からの距離と同様に、食材仕入れに影響があるかもしれない。また、アルバイトに調理を任せている店は、味に影響が出る可能性もある。営業時間が長いとノーグエスの時間ができる可能性も高まるので、光熱費や人件費の効率性に苦慮しなければならない。

5.2.3 定休日について

定休日が少ないということは、閑散日でも人件費や光熱費がかかるので、店を開けているとかえって赤字になることがある。効率性が悪いと食材仕入れに影響があるかもしれない。

5.3 西日本について

分析結果より、西日本の百名店による特徴は、最寄り駅からの距離が遠い、テーブル席の数が少ない、総営業時間が短い、価格が高いという結果を得た。西日本のその他のラーメン店による特徴は、テーブル席の数が多く、総営業時間が短い、価格が低い、定休日が少ないという結果を得た。価格に関する条件以外は東京のラーメン店にも共通するので、東京の特徴と異なる価格に関する条件についてのみ考察する。

西日本の百名店とその他のラーメン店の各属性の平均を表2に示す。

表2 西日本の百名店とその他のラーメン店の各属性の平均

属性名	百名店	その他のラーメン店
最寄り駅からの距離	534	243
テーブル席の数	1.41	5.43
総営業時間	7.18	11.78
価格	786	743
定休日の有無	0.67	0.18

表2から、最寄り駅からの距離、テーブル席の数、総営業時間、定休日の有無に大きな差があり、価格についてはそこまで大きな差がないことが読みとれる。

百名店は、人件費や家賃が低い分、材料費に当てることができるため、味重視の商品を提供できる。また、一食当たりの価格の平均の差は約40円であり、そこまで大きな差ではないことから、多少通いづらい場所であるとしても味の良い店に足を運ぶ人が多いと考えられる。

6 その他の飲食店に対する分析

他の飲食店も同じ特徴がみられるかを確認するために、全国のそば店を対象に同様の分析を行ったところ、ラーメ

ン店とほぼ同じ結果を得た。

全国のそば百名店とその他のそば店を分類するのに、最も重要だと考えられる属性を定量的に明らかにするために、ランダムフォレストを用いて、各属性の重要度を求めた。その結果、総営業時間と定休日の有無の値が飛びぬけて高いため、総営業時間と定休日の有無が百名店か否かの分類に重要であることが分かる。

全国のそば百名店とその他のそば店の各属性の平均を表3に示す [4, 5].

表3 全国のそば百名店とその他のそば店の各属性の平均

属性名	百名店	その他のそば店
最寄り駅からの距離	2142	367
テーブル席の数	5.5	11.16
総営業時間	4.97	9.3
価格	991	798
定休日の有無	0.96	0.21

表3から、百名店の最寄り駅からの距離が飛びぬけて高いことが分かる。そば百名店は、山奥や田舎に位置している店が多く交通の便が悪いため、表5のような結果を得た。その他の各属性の平均は、ラーメン店とほぼ同じ結果であった。

さらに、他の飲食店も同じ特徴がみられるかを確認するために、東京のフレンチ店を対象に同様の分析を行ったところ、ラーメン店とは異なる結果を得た。

東京のフレンチ百名店とその他のフレンチ店を分類するのに、最も重要だと考えられる属性を定量的に明らかにするために、ランダムフォレストを用いて、各属性の重要度を求めた。その結果、価格の値が飛びぬけて高いため、価格が百名店か否かの分類に重要であることが分かる。

東京のフレンチ百名店とその他のフレンチ店の各属性の平均を表4に示す [6, 7].

表4 東京のフレンチ百名店とその他のフレンチ店の各属性の平均

属性名	百名店	その他のフレンチ店
最寄り駅からの距離	416.2	261.45
テーブル席の数	7.75	5.25
総営業時間	5.6	8.625
価格	22093	8694.5
定休日の有無	0.8	0.65

表4から、百名店の価格が飛びぬけて高いことが分かる。その他の属性については、百名店とその他のフレンチ店でそこまで大きな差は見られず、ラーメン店とほぼ異なる結果を得た。ラーメン店は、一人当たりの単価が低く客の回転率が高いのに対して、フレンチ店は、一人当たりの単価が高く客の回転率が低い。そのため、総営業時間に差がないと考える。また、ラーメン店はチェーン店や姉妹店

が多くそれらの店はほぼ定休日が無いのに対して、フレンチ店はほぼ個人で経営している店が多いため、定休日の有無に差がないと考える。

7 おわりに

本研究では、機械学習を用いて、食ベログ百名店に選出されるような人気店の特徴分析を行った。はじめに、東京のラーメン店に着目し分析を行ったところ、百名店に選出されるような店の特徴は、「最寄り駅からの距離が遠い」、「テーブル席の数が少ない」、「総営業時間が短い」と考えられる結果となった。また、地域によって百名店の特徴が異なるのかを分析するため、西日本でも同様の分析を行ったところほぼ同じ特徴を得た。同じ食べ物であれば、百名店に選出される店の特徴は、どこの地域でも共通していることが分かった。さらに、他の飲食店でも同じ分析を行ったところ食べ物のジャンルによって、結果が異なることが分かった。そば店で分析を行ったところ、ラーメン店とほぼ同じ結果を得た。一方、フレンチ店で分析を行ったところ、ラーメン店とは異なる結果を得た。これは、一人当たりの単価や客の回転率の違いが関係していると考えられる。ラーメン店やそば店のような、一人当たりの単価が安く客の回転率が高いジャンルの店は、「最寄り駅からの距離が遠い」、「テーブル席の数が少ない」、「総営業時間が短い」ことが百名店か否かの分類に重要であると考えられる。

参考文献

- [1] 食ベログラーメン TOKYO 百名店 2021(2022/07/22 アクセス): 『https://award.tabelog.com/hyakumeiten/ramen_tokyo/2021/』.
- [2] 食ベログラーメン WEST 百名店 2021(2022/07/22 アクセス): 『https://award.tabelog.com/hyakumeiten/ramen_west/2021/』.
- [3] おすすめの美味しいラーメンをご紹介!(2022/07/22 アクセス): 『<https://tabelog.com/rstLst/ramen/>』.
- [4] 食ベログそば百名店 2021(2022/10/27 アクセス): 『<https://award.tabelog.com/hyakumeiten/soba/>』.
- [5] おすすめの美味しいそばをご紹介!(2022/10/27 アクセス): 『<https://tabelog.com/rstLst/soba/>』.
- [6] 食ベログフレンチ TOKYO 百名店 2021(2022/11/24 アクセス): 『https://award.tabelog.com/hyakumeiten/french_tokyo/2021/』.
- [7] 東京でおすすめの美味しいフレンチをご紹介!(2022/11/24 アクセス): 『<https://tabelog.com/tokyo/rstLst/RC021101/>』.