

卒業論文

トレーディングカードゲームのゲーム性能と希少性を考慮した中古価格の要因分析

公立はこだて未来大学
システム情報科学部 情報アーキテクチャ学科
情報システムコース 1017154

穴戸 建元

指導教員 新美 礼彦

提出日 2021年1月26日

BA Thesis

Analysis of Price Factors for Used Trading Card Games Considering Scarcity and Game Performance

by

Takeru Shishido

School of Systems Information Science, Future University Hakodate
Information Systems Course, Department of Media Architecture
Supervisor: Ayahiko Niimi

Submitted on January 26th, 2021

Abstract— The trading card game (TCG) market has expanded significantly in recent years. One of the major factors contributing to this expansion is the sale of used cards. Several factors affect the selling price of these cards. However, there are few examples wherein the degree of impact of each factor is categorized or degree of scarcity is considered. In this study, we analyze the factors that contribute to the price formation of used TCG cards using machine learning for the used sales market of “Magic: The Gathering.” The cards’ characteristics are classified into two categories: (1) cards performance attributes within the game; and (2) card scarcity attributes. We then analyze what attributes contribute to the card price formation. The results of regression analysis and decision tree (CART) analysis using categorical and numerical attributes showed that the scarcity attribute contributed more to price determination than the performance attribute. The extreme bias of the price distribution around 0.1 USD and accuracy of the price prediction using the results showed that only a more complex model could predict the price.

Keywords: Data Mining, Machine Learning, Price Forecasting, Regression Analysis, Decision Tree, Trading Card Game

概要： 近年，トレーディングカードゲーム（以下，TCG）市場は大きく広がりを見せている．TCG を形成する大きな要因のうち，ゲーム部分やコミュニティ以外の大きな要因として，中古カードの販売が存在する．この中古市場において，カードの販売価格に影響する要因は複数存在するが，要因それぞれがもたらす影響度合いが明確に分類されている例や，希少度を考慮した例は少ない．そこで，本研究では，TCG の1つである「Magic: The Gathering」における中古販売市場を対象とし，機械学習を利用した TCG の中古カードの価格形成要因の分析を行った．特に，カード自体が持つ属性を (1) 『カードゲームにおける性能属性』と (2) 『カードの希少度を表す属性』に該当するものに分類し，それぞれどのような属性が価格決定に寄与しているのか分析した．カテゴリ属性，数値属性を利用し，回帰分析と決定木 (CART) による分析を行った結果，性能属性よりも希少度属性の方が価格決定に寄与していることを示した．しかし，価格の分布が 0.1USD 付近に極端に偏っていることや，その結果を利用した価格予測の精度から，より複雑なモデルでなければ価格の予測が困難であることが判明した．

キーワード： データマイニング，機械学習，価格予測，回帰分析，決定木，トレーディングカードゲーム

目次

第 1 章	序論	1
1.1	背景	1
1.2	TCG について	1
1.3	TCG の中古販売市場について	2
1.4	TCG 価格予測システムの構築	2
1.5	問題点	3
1.6	目的	3
1.7	論文構成	4
第 2 章	関連研究	5
2.1	中古販売市場における先行研究と本研究の位置づけ	5
2.2	TCG に関する先行研究と本研究の位置づけ	6
2.3	TCG の中古販売に関する先行研究と本研究の位置づけ	6
第 3 章	提案手法・前提知識	9
3.1	Magic: The Gathering のルールについて	9
3.2	既存技術の紹介	10
3.3	分析手法の提案	14
第 4 章	実験と評価および考察	15
4.1	実験環境	15
4.2	カード性能データセット・中古価格データセットの作成	15
4.3	回帰分析による価格要因分析	18
4.4	決定木学習 (CART) による分類実験	25
4.5	価格帯ごとの属性傾向の確認実験	30
第 5 章	結論	34
5.1	まとめ	34
5.2	今後の展望	35

参考文献	38
.1 実験で利用した中古価格データセットの構造	40
.2 実験で利用したカードデータの属性とその一例	41
.3 実験 4.3 における回帰係数の一覧	42
.4 実験 4.5 における価格帯ごとの属性傾向の一覧	46

第 1 章

序論

本章では，研究の背景と目的について述べる

1.1 背景

近年，トレーディングカードゲーム (以下，TCG) 市場は大きく広がりを見せている．TCG とは，「トレード等によりカードの収集を行い，その中からルールに従って自由にカードを選び，カードを使用して遊ぶことのできるゲーム」のことを指す．TCG は目的の一つはカード自体の収集・コレクションである．鑑賞目的などでそのカードを集めることを目的とするが，もう一つの目的として，そのカードを利用したゲームが存在していることがある．

1.2 TCG について

TCG とは，各自がコレクションしたカードを用いて遊ぶことのできるアナログゲームである．元来コレクションして楽しむアイテムとして，「トレーディングカード」という商品が存在した．これは，写真やイラスト等が描かれたカードであり，主に鑑賞や収集等のコレクションが目的であった．海外においては，現在でもゲームの要素を含むものと，単にコレクションするだけのものを一緒にたにし，「Collectible Card Game」(CCG) という名称でカテゴライズされている例も存在する．その後，コレクション要素と合わせて，カード自体にゲーム中の性能を記述しそれを複数枚集めて遊ぶという，ゲームの要素を含んだものとして TCG が開発された．

「Magic: The Gathering」[1] は，世界で初めて制作されたトレーディングカードゲームである．このゲームは，当時数学教授であった Richard Garfield 氏が発案し，Wizards of the Coast 社によって 1993 年に発売された．[2] 当時は TRPG などが主流であったアナログゲーム業界の中で爆発的人気となり，その後アナログゲーム内外を問わず大きく知名度を上げ，現在では世界中で多くのプレイヤーによって親しまれるゲームに成長している．

「Magic: The Gathering」の登場以降，TCG のタイトルとして様々なタイトルが開発・販

売されており、「ポケモンカードゲーム」[3] や「遊戯王 OCG デュエルモンスターズ」[4] , 「カードファイト!!ヴァンガード」[5] などがその例である。これらのタイトルは日本の会社で作られた製品も存在するが、現在では世界中で翻訳がなされ販売されるタイトルも存在し、世界的なマーケットとなる事例も存在している。実際に日本において、玩具市場における TCG の売り上げは 2019 年で 1100 億円規模となっており、2018 年から 2019 年の間では成長率は 104.9% と、20 年以上たった今なお成長を続けている大規模な市場である。[6]

近年では、そのゲーム部分をさらに掘り下げ、特定タイトルの世界大会が開催される事例や、企業スポンサーが出資し、実際にプロプレイヤーとして世界中の大会で活躍する事例も存在する。

1.3 TCG の中古販売市場について

TCG 市場を形成する要因のうち、ゲーム要素やコミュニティ以外のものとして、カードの中古販売が存在する。一般に TCG を遊ぶユーザは、ゲームで遊ぶために複数枚のカードを用意する必要がある。そのためユーザは、商品を購入する必要があるが、それらの商品の多くがランダムな確率でカードが封入されているため、本当に目当てのカードのみを手に入りたい場合には、トレードや中古カード単品販売など別の手段を通じ、自身のコレクションを増やしていくこととなる。

TCG は特に中古カードの販売市場が発達し、「中古 TCG 専門店」としてカードのみを販売する例もある。現在では、EC サイト・アプリケーションを活用した電子商取引も活発になっており、中古 TCG 市場の内部で情報技術が活用される例も増加している。

また、TCG の中古市場ではコレクションとしての価値が認められ、高額な値段で取引されるものも存在する。特に希少なカードについてはそのコレクション的な価値の高さから、PSA[7] などの専門機関による真贋鑑定がなされるほどである。実際に鑑定がなされた高額なカードが数千万円で市場に出回る例も存在し、実店舗販売のみならずオークションが用いられる事例も存在する。

以上のように、TCG はそのゲームに関する側面だけではなく、中古販売市場に関する側面も非常に重要な要素となっている。

1.4 TCG 価格予測システムの構築

上述の中古販売市場の活発化から、本研究では TCG の中古カード自動値付けシステムの構築を最終目標としている。

中古 TCG 市場では、日々新しいカードが追加されるため、需要と供給が変化し、その価格も大きく変化する。また、カードゲームにおける性能は、実際にゲームについて深く理解しなければ判断が難しく、初心者や門外漢の人々にとって、その是非を判断しながら価格をつけることは難しい。そのため、このシステムを実現することができれば、カードに関する

知識が乏しい場合でも、カードの情報を入れることで簡便に値付けを行えるようになると考えられる。

現時点で提案しているシステムでは、新たに販売されたカードの特性情報を入力とし、そのカード自体の性能や希少度から自動的に値付けするシステムである。例えば、新しく発売されたカードの持つ数値やテキスト・レアリティなどをシステムに入力すると、過去のカードの情報と価格を照らし合わせて、新しいカードの中古予測価格を出力するシステムである。

また発展として、入力として与えられたカードと関連するような既存のカードを探し出し、それらの高騰予測までを扱い、市場の変化を予測するシステムも展望としている。

1.5 問題点

しかし、このシステム構築の前段階において、どのような要因がカード価格に影響しているのか詳細に解明されていない、という問題がある。

TCG の中古販売市場において、カード 1 枚の販売価格は様々な要因によって決定されている。例えば、特定の年数を経た商品に高額な価格がつく例や、イラストや言語の違い、カード光沢の有無などで価格が変わる例が存在する。また、ゲーム上の性能が価格と直結する例も存在する。例えば、前週に 1 枚 30 円の最低価格で取引が行われていたカードが、相性の良い新しいカードの発売により需要が増加し、次週では 1 枚 2000 円以上で取引される事例も存在する。このように TCG 市場では、価格に影響する要因が複数存在し、実際に価格が高騰した事例は存在する。しかし、要因それぞれが価格にもたらす影響が明確に分析された例は少ない。

この要因を明らかにしなければ、価格予測システムを構築する際にも、予測の際にカードのどの属性に注目すべきかわからず、本当に価格予測に必要な属性情報を入力できず、影響度の強い属性を利用できない可能性が存在する。また、必要以上に多くの属性を入力するようなシステムを作成した場合、結果的にカードの知識がなければ利用できないシステムとなり、利用が容易ではなくなる。そのため、システムの前段階としてカードの中古価格要因の分析が不可欠となる。

1.6 目的

前述の背景を踏まえ本研究では、TCG の 1 つである「Magic: The Gathering」における中古販売市場を対象とし、機械学習を利用した TCG 中古カードの価格要因分析を行う。

本研究では特に、その価格がカードの持つゲーム上の性能とそのカードの希少度によって説明できると仮定する。その上でカード自身の持つ属性を、(1)『カードゲームにおける性能属性』と、(2)『カードの希少度を表す属性』に分類し、それぞれの属性がどのように価格要因として作用するのかを、機械学習を用いて明らかにする。

1.7 論文構成

本節では、本論文の論文構成について述べる。第1章では、本研究の背景と目的を述べた。第2章では関連研究について述べる。第3章では提案手法や前提知識について述べる。第4章では本研究で行った実験とその評価・考察について述べる。第5章では本研究のまとめと今後の展望について述べる。

第 2 章

関連研究

この章では本研究を行うにあたって参考にした既存研究，また本研究の立ち位置について述べる．

2.1 中古販売市場における先行研究と本研究の位置づけ

商品の中古販売市場における価格要因分析の研究は，様々な先行研究が存在する．仁ノ平ら [8] の研究では，中古ファッションアイテムの価格予測を，ファッションブランドの出品履歴 67,211 件を対象とし，季節ごとの値下がり率を潜在クラスとして考慮して混合回帰モデルによる分析を行った．

また，越川ら [9] の研究では，中古車販売価格に対する既存研究を踏まえて，中古車の性能情報と価格情報を含んだデータセット 157,611 件を利用し，ランダムフォレストによる価格予測モデルを構築した．

表 2.1 は価格要因分析に関する先行研究と本研究との違いをまとめたものである．本研究の目指す価格予測システムの構築や価格要因分析は，他分野における価格要因分析の研究と同様の統計的手法による価格の分析から成る．しかし本研究は，TCG の中古市場を対象としたものであり，先行研究ではあまり扱われてこなかった商品を対象としている点で，独自性があるといえる．

表 2.1 価格要因分析に関する研究における本研究の位置づけ

論文	対象とする商品	データ数	研究内容
仁ノ平ら	ファッションブランドの 出品履歴データ	67,211 件	潜在クラスを用いた混合回帰モデルによる分析
越川ら	中古車の出品履歴データセット	157,611 件	ランダムフォレストによる価格予測モデル構築
本研究	TCG の販売履歴データセット	41,236 件	線形回帰・学習木による価格要因分析

2.2 TCG に関する先行研究と本研究の位置づけ

背景で示した TCG の普及により，本研究に限らず，TCG を題材として様々な観点からの学術的な研究が行われている．

Bjørke ら [10] は，ゲームのために用意されるカードの束 (デッキ) にどのカードを使用するのかについて，最適化問題の例としてとらえ，遺伝的アルゴリズムを使用したカード選択システムの提案・実装をした．この研究は，主に TCG のゲーム部分についてアプローチを行った研究といえる．

また，TCG を行う人々によって形成されるコミュニティに関する研究も存在する．Felipe ら [11] は，TCG プレイヤー 272 名に対してアンケート調査を行い，それらの回答の相関を見ることで，TCG のタイトルごとのコミュニティの意識の違いや，タイトルによらないプレイヤー共通の嗜好について論じている．

このように，TCG に関する研究にはその内部における様々な領域の研究がある．表 2.2 は，前述した TCG に関する先行研究の内容とその研究が扱う TCG 内の領域についての対象を示したものである．

表 2.2 TCG に関する研究における本研究の位置づけ

論文	論文の内容	扱う領域
Bjørke ら	TCG のアルゴリズムに関する研究	ゲーム部分に関する領域
Felipe ら	TCG コミュニティの意識や趣向の違いを論じる	TCG のコミュニティに関する領域
本研究	中古 TCG の価格形成要因をとらえる	ゲーム部分 + コレクション部分に関する領域

本研究は，前述した研究と同じ領域ではないが，カードの性能やカードの希少性からの説明を行おうとすることから，ゲーム部分に関する領域と一部重複している部分があるといえる．また，コレクション部分に関する領域も関係しているため，TCG に関する研究の中でも他の既存研究と比べ，独自の位置づけにあるといえる．

2.3 TCG の中古販売に関する先行研究と本研究の位置づけ

TCG の中古販売価格の分析・予測については，いくつかの先行研究が存在する．

Loh[12] の研究では，「カードファイト!!ヴァンガード」のプレイヤー対象としたアンケート調査を行い，カードイラスト，ゲーム中の性能，カードの希少性，コミュニティの 4 つについて，中古販売市場における価格に影響を及ぼしていることを示した．本研究は，ここで示された「ゲーム中の性能」と「カードの希少性」について，カードの持つどのような属性が特に寄与しているのかを明らかにしようとする試みである．また本実験では対象とするタイトルも異なり，「Magic: The Gathering」のカードにおいて，特にどのような属性が寄与しているのかを明らかにするのが目標である．加えて，この研究はアンケートインタビュー

である質的研究であり，本研究のデータや実価格から価格を分析する量的研究とは異なっている．

また，機械学習を利用した TCG の価格要因の分析には，Matthew ら [13] の研究が存在する．この研究では，TCG の中古市場に見られる価格の高騰について，ロジスティクス回帰や SVM を利用し価格の変動を予測している．しかし，この予測は価格推移の履歴や大会での使用率に大きく依存しているため，カード価格履歴やトーナメントのプレイタイムがない新しいカードの価格への適用には向かない．この点については，カード持つ性能や希少性からの説明を目的としている本研究ともアプローチが異なっている．また特に，カード全体としての傾向を確認する本研究と違い，数日間中の価格変化の方がより重要視されている点も本研究とは異なる．

加えて，Dustin ら [14] の研究においても，機械学習を利用した TCG の分析が行われている．この研究では，中古カード価格の予測において，カードのテキスト性能を n-grams による解析を行って利用し，カードゲーム上の性能から価格を予測している．この研究で扱われている予測は，本研究の目指している『カードゲームにおける性能属性』からの説明と合致している．しかし，『カードの希少度を表す属性』についてはあまり考慮されていないため，本研究では両方の属性について説明が可能であるか検証する．

このように，TCG の中古販売価格の分析はいくつかの参考研究が見られる．表 2.3 は，TCG の中古販売価格の先行研究についてまとめたものである．

表 2.3 TCG の価格要因分析に関する研究における本研究の位置づけ

	Loh ら	Matthew ら	Dustin ら	本研究
解析手法	アンケート調査	ロジスティクス回帰	n-grams によるカード テキスト解析，回帰分析	線形回帰，学習木
対象データ	アンケート	2012 年～2014 年までの カードデータ，トーナメント 使用カードデータ，価格データ	2015 年までのカード データ，価格データ	2020 年 11 月 9 日までの カードデータ，価格データ
データ件数	107 件	13,608 件	14,352 件	カードデータ 41,236 件 価格データ 14,143 件
対象タイトル	『カードファイト！ ヴァンガード』	『Magic: The Gathering』	『Magic: The Gathering』	『Magic: The Gathering』
研究内容	中古価格に関わる 要因の分析	トーナメント利用履歴， 価格履歴を用いた 価格の変動要因を分析	テキストを含めた， カードの性能からの 価格予測	カード価格の全体傾向と， 価格に影響を及ぼす属性 を明らかにする

この表からもわかるように，既存の TCG の中古販売の研究において対象にしているのは Loh らのアンケート分析による質的研究か，Matthew ら，Dustin らのような 14,000 件程度のデータによる量的研究である．しかし，本研究では「Magic: The Gathering」において，公式ゲームフォーマット上で利用できるすべてのカード 41,236 件を対象とし，既存研究を上回るデータ数における解析を行う．さらに，過去の量的研究とは異なり，カードの性能に関

する属性のみではなく、カードの希少度も考慮に入れ分析を行う点が、既存研究と異なる部分である。また、その価格の要因についても Matthew らの研究とは違い、価格の変動に特に注目するのではなく、カード価格全体の傾向について分析する点も異なる点である。

第3章

提案手法・前提知識

この章では実験の際に用いる TCG のルール，分析に用いる既存技術の紹介，提案する分析手法について述べる．

3.1 Magic: The Gathering のルールについて

分析に用いられるカードゲームの性能属性について，特にそのゲームにおけるルールを理解していなければ把握することが難しい．そのため，前提知識として，「Magic: The Gathering」のルールを記述する．

「Magic: The Gathering」は2名で行う対戦型のゲームである．プレイヤーは60枚のカードを持ち寄ってゲームを開始する．ゲーム開始前，各プレイヤーは7枚の手札と20点のライフを持ってゲームを始める．

プレイヤーはお互いに自身のターンを繰り返して，その中で行動する．毎ターン使用できるカードを1枚山札から引き，手札を増やしていく．プレイヤーは各ターンの間，「土地」カードの設置，カードや能力の使用，戦闘を行うことができる．

プレイヤーは，カードやその能力を使用して，戦闘を行うことで相手のライフを0にすることが目的となる．戦闘を行うカードは「クリーチャー」と呼ばれるタイプを持ち，戦闘の際，自身の「パワー」の値をダメージとしてを相手や相手のクリーチャーに与えることができる．クリーチャーはそれぞれ体力のような「タフネス」という値を持ち，クリーチャー同士の戦闘ではお互いにパワーの数値をタフネスから引き算する．タフネスが0になったクリーチャーは，戦闘を行うクリーチャーが存在できる「フィールド」から，墓地という使用済みカードの置き場に置かれる．そのため，「タフネス」や「パワー」といった数値は最終的にゲームを決めるために重要な指標の一つといえる．

また，「クリーチャー」カード以外にも，使用后フィールドに存在し常に効果を発揮する「エンチャント」「アーティファクト」「プレインズウォーカー」が存在する．そのほか，「クリーチャー」のように戦闘ができず，使い切りの効果のみを持つ「インスタント」「ソーサリー」等が存在する．

しかし、これらのカードを使用するためには、カードに記載された「マナコスト」という資源を支払わなければいけない。この資源の支払いには、毎ターン1度だけ自身のフィールドに置くことができる「土地」というカードが必要になる。各プレイヤーは、毎ターン自身の土地の枚数分の資源しか利用できないため、プレイヤーはいかに手札のカードを効率よく使うのか、選択しカードを使用する必要がある。

そのため、一般的に「マナコスト」の小さいものはすぐに使いやすく利用しやすいが「パワー」や「タフネス」の数値が小さく、その能力もあまり強力ではない。逆に、「マナコスト」が大きいものは、使用した際にゲームに及ぼす影響が大きい、使用することが、難しいといった傾向がある。また、「マナコスト」には特定の色が付いたものと数値だけのものが存在し、特定の色を持つものは同じ色を持つ「土地」カードからのみしか資源を支払うことができない。この色は赤・青・緑・白・黒の5色が存在し、それぞれに対応した色の「土地」カードが存在する。

プレイヤーは、このゲームで使用する60枚のカードを集める際、使用するカードや、それを利用するための土地カードの種類、またそれらの配分を自身で調整し、各自がゲーム中で使用するカードを自由に組み替えることができる。

このようなルールに基づいて、「Magic: The Gathering」のゲームが成り立っている。

3.2 既存技術の紹介

ここでは、本研究の分析において使用する既存技術について述べる。

3.2.1 線形回帰

線形回帰とは、学習データから目的変数 (Y) と説明変数 $X(x_1, x_2, \dots, x_n)$ を置く時、 β_0 を切片、 ϵ を誤差、 $\beta_1 \sim \beta_n$ までを回帰係数とし、目的変数と説明変数との関係を式 3.1 のようにモデルに当てはめる回帰分析手法である。

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (3.1)$$

線形回帰では、目的変数と説明変数との間の比例関係を仮定し、各説明変数が目的変数に対して独立して影響することを仮定している。また、回帰係数を決定する際、予測値 $Y'(y'_1, y'_2, \dots, y'_n)$ と目的変数 $Y(y_1, y_2, \dots, y_n)$ の乖離を最小化するために、式 3.2 のような損失関数 L を最小化させるような回帰係数を決定する。(最小二乗法)

$$L = \sum_{i=1}^n (y_i - y'_i)^2 \quad (3.2)$$

本研究では, scikit-learn[15] の LinearRegression によって線形回帰を行い, 回帰係数を求める.

3.2.2 Ridge 回帰・Lasso 回帰

Ridge 回帰・Lasso 回帰は, 線形回帰において学習データに対して過剰に適合すること (過学習) を防ぐために, 線形回帰の損失関数 L に正規化項を加えた手法である.

Ridge 回帰は, 正規化項として L2 ノルム (ユークリッド距離) の二乗を用いる手法である. Ridge 回帰では, 回帰係数の絶対値を抑え, 説明変数の影響が過大にならないようにすることが可能である. しかし, 回帰係数が完全に 0 にはならないため, 説明変数が非常に多い場合ではモデル解釈が困難となる問題点が存在する.

Lasso 回帰は, 正規化項として L1 ノルム (マンハッタン距離) を用いる手法である. Lasso 回帰では, 目的変数の説明に不要な説明変数は回帰係数が 0 になり, 不要な特徴量を削り属性選択が可能である. しかし, 説明変数同士の相関が高い場合について, 片方の変数しか扱うことができないという問題点が存在する. また, 属性選択の際に, 標本数が説明変数の数よりも小さい場合, 標本数までしか説明変数を選択できないという問題が存在する.

本研究では, scikit-learn の Ridge, Lasso によってそれぞれの回帰を行い, 回帰係数を求める.

3.2.3 Elastic Net

Elastic Net は, Ridge 回帰の L2 正規化項と Lasso 回帰の L1 正則化項, 2 つの和を正規化項として組み合わせた手法である. Elastic Net は, Lasso 回帰の欠点であった説明変数の選択数限界の問題に対応し, 相関の高い変数にも対応することが可能である.

本研究では, scikit-learn の ElasticNet によって ElasticNet での回帰係数を求める.

3.2.4 回帰分析の性能評価に用いた指標

本分析では, 回帰分析で求められたモデルに対し, 以下のような指標を用いてモデルの評価を行う.

決定係数 (R^2)

決定係数とは, 説明変数がどれだけ目的変数を説明できているか, モデルの当てはまり具合を表す指標である. 寄与率とも呼ばれ, 目的変数 $Y(y_1, y_2, \dots, y_n)$ と予測値 $Y'(y'_1, y'_2, \dots, y'_n)$ について, 決定係数は式 3.3 で表される.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3.3)$$

本研究では、scikit-learn の `r2_score` 関数を用いて算出する。

二乗平均平方根誤差 (RMSE)

二乗平均平方根誤差とは、目的変数と予測値の誤差の大きさを表す指標である。目的変数 $Y(y_1, y_2, \dots, y_i)$ と予測値 $Y'(y'_1, y'_2, \dots, y'_i)$ について、その誤差の値の二乗平均 (平均二乗誤差・MSE) を算出し、その値を平方根で補正したもので、式 3.4 のように表される。

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - y'_i)^2} \quad (3.4)$$

本研究では、scikit-learn の `mean_squared_error` 関数によって平均二乗誤差を求め、その平方根を算出することで求める。

3.2.5 決定木

決定木は、データを分類するためのルールについて、分岐過程を階層化し、樹形図などで表したものである。機械学習の過程では、実際の事例データから決定木を構築し、最終的な樹形図として学習器となる。決定木の構築は、まず根ノードと呼ばれる属性を決定し、その属性の値に応じた子ノードを作成する。その後、子ノードの中で属性を選択し、その属性の値に応じた更なる子ノードを生成する、という工程を繰り返して分類木を作成する。

決定木は、木の可視化による分類ルールの解釈が簡単であるが、剪定がなされなければデータにオーバーフィッティングする可能性がある。また、数値属性を離散値に置き換えて処理を行うため数値属性をうまく扱えないことや、少しの変化で構築される学習木が大きく異なるという特性がある。

決定木は、分岐に使用するアルゴリズムによって、様々な実装方法が存在する。ID3 は、分岐ノードを作成する際、分岐によって最大の情報利得比 (情報ゲイン) をもたらすような特徴量を見つけて利用する実装である。C4.5 は、ID3 を改良した実装方法である。ID3 の枝刈り基準を改良し、ID3 で扱うことのできなかった連続値を離散値でも扱うことができる。CART は、ID3 や C4.5 とは違い、木全体の Gini 係数の減少を最小化するような特徴量を見つけて利用する実装方法である。Gini 係数は入力データの一様分布を測る指標であり、最小化することで目的変数を最もよく分類するような説明変数を利用することができる。

本研究では、scikit-learn を利用して決定木を構築する。scikit-learn の内部の実装では、CART アルゴリズムを最適化させたものを利用しているため、本研究においても CART アルゴリズムを用いて学習木を構築している。

3.2.6 決定木の性能評価に用いた指標

本研究において，決定木学習で構築された学習器の分類性能評価には，以下のような 4 種類の評価指標を用いる．また，予測の際には以下の表 3.1 のように分類後の結果を分け，その結果について性能を評価する．

表 3.1 予測結果に対する混同行列

	実際は陽性	実際は陰性
陽性と予測	True Positive(TP)	False Positive(FP)
陰性と予測	False Negative(FN)	True Negative(TN)

Accuracy

Accuracy は，学習器の正答率を表す指標である．全体から True Positive と True Negative に分類できたものの割合を示すものであり，式 3.5 のように表される．

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3.5)$$

Precision

Precision は，分類の誤検出の少なさを評価するための指標である．陽性と予測されたもののうち，実際に陽性である事例の割合を表し，式 3.6 のように表される．

$$Precision = \frac{TP}{TP + FP} \quad (3.6)$$

Recall

Recall は，分類の検出漏れの少なさを評価するための指標である．実際に陽性であるもののうち，陽性と予測されたものの割合を表し，式 3.7 のように表される．

$$Recall = \frac{TP}{TP + FN} \quad (3.7)$$

F 値 (F1 値)

F 値 (F1 値) は Precision と Recall について調和平均をとって学習器を評価する指標である．式 3.8 のように表される．

$$F = \frac{2Recall \cdot Precision}{Recall + Precision} \quad (3.8)$$

3.3 分析手法の提案

本研究では、Loh[12]の研究において示された中古販売市場への影響を与える要因について、特に「ゲーム中の性能」を「カードの希少性」について、カード性能を表す複数の属性から表現することが可能であると仮定し、分析を行う。

分析の際には、「Magic: The Gathering」で発行されたカードについて、その性能を複数の属性で記述したデータセットを用意する。その属性を(1)『カードゲームにおける性能属性』と、(2)『カードの希少度を表す属性』に該当するものに分類し、それぞれどのような属性が価格決定に寄与しているのか、回帰分析と決定木学習によって明らかにする。

(1)『カードゲームにおける性能属性』とは、TCGのゲーム上の性能を決める要素である。例えば、カードタイプやカードテキスト、パワーなど、ゲームルームで参照されカードゲーム上での性能を表す属性を指す。

(2)『カードの希少度を表す属性』とは、TCGのゲーム上の性能にかかわらず、カードの希少性を決定する要素である。例えば、カードのレアリティや入手できる環境、再印刷の頻度など、そのカード自体の希少性の高さを表す指標となるものを指す。

この2点について、各属性の価格への影響度合いを分析していく。

また本研究では価格についても、『相場全体における価格要因の傾向』を分析するものとする。これは、Matthewら[13]の研究で確認している時系列的な価格変動とは違い、相場全体の傾向としてどのような要因が高騰につながるのかを把握することが本研究の目的のためである。そのため、本研究における価格は、一定の日時において確認された価格を参照するものとし、その前後における価格の変動性については、本研究では扱わないものとする。

第 4 章

実験と評価および考察

この章では本研究で行った実験と評価および考察について述べる。

4.1 実験環境

本研究では，機械学習を用いた分析・実験のために以下のような環境を利用した．表 4.1 に利用したツール・ライブラリとそのバージョンについて記載する．

表 4.1 利用した各種ツール・ライブラリとそのバージョン

使用したツール・ライブラリ	バージョン
Ubuntu(GoogleColab)	18.04.5(LTS)
Python	3.6.9
scikit-learn	0.22.2.post1
Pandas	1.1.4
Numpy	1.19.5
MTG JSON	5.0.1

4.2 カード性能データセット・中古価格データセットの作成

本研究の実験を行う前に，使用するカードゲームタイトルとそのデータセットについて，選定を行った．

TCG において，中古販売が広く行われているタイトルは長期間継続して販売されているタイトルが主である．そのためカードの種類が多様多様にあり，各カードを確認しながら手動でデータセットを作成するのは容易ではない．そのため，すでに作成されたデータセットを利用し，そのデータセットに修正を加えることで実験用データセットを作成した．

4.2.1 既存データセットの調査・選定

本研究で分析する TCG のタイトルとして、「Magic: The Gathering」を利用した。TCG 市場において、中古販売が行われ高額な金額での取引事例が存在しているタイトルは、長期間継続して販売されているタイトルが主である。その点を考慮し、調査の前段階においては「遊戯王」、「カードファイト!!ヴァンガード」、「ポケモンカードゲーム」等を候補としていた。データセットの調査したところ、「Magic: The Gathering」、「遊戯王」のデータセットが確認できた。しかし、「遊戯王」のカードデータセットは Kaggle 上で発見されたデータセット 1 件と有志による API が 1 件のみであった。「Magic: The Gathering」については、既存データセットが複数種類存在し、公式の API を利用し作成できる可能性があることや、各種言語におけるデータ等の追加データも存在し、情報が取得しやすいため、本研究では「Magic: The Gathering」を分析の対象とした。

既存データセットの選定にあたって、(i) 最近まで情報が更新され多くのカード情報が存在すること、(ii) 中古販売価格データの取得が容易であること、この 2 点を満たすようなデータセットの調査を行った。

その結果、今回は、「MTG JSON」[16] というサイトから「Magic: The Gathering」のすべてのカード情報が記述されたデータセット、AllPrintingsCSVFiles を利用し作成を行った。「MTG JSON」は、「Magic: The Gathering」のコンテンツ内において印刷されたすべてのカードについて、データでの記述を JSON 形式で行うオープンソースプロジェクトである。「MTG JSON」で公開されているデータセットは有志によって作成・管理され、毎週カード情報が更新されている。そのため、(i) 最近まで情報が更新され多くのカード情報が存在すること、を満たしている。また、(ii) 中古販売価格データの取得が容易であること、についても海外の販売サイトにおける取引データがデータセットとして集約されており、毎日更新されていることから、今回のデータセット作成に適切であると判断した。今回使用したデータセットは、公開されているもののうち、カードそのものの情報を記述した AllPrintingsCSVFiles というデータセット、その価格データを記述した AllPrices というデータセット、また公式に販売されている商品に関する情報を記述した SetList というデータセットの 3 つを利用した。

4.2.2 データセット作成・処理

前述したデータセットを用い、実験で使用する最終的なデータセットを作成した。

カード性能データセットは、AllPrintingsCSVFiles というデータセットをもとに作成した。本研究では公式ゲームフォーマット上で利用できるカードのみを利用するため、今回の実験で対象としないカードデータを削除した。フォーマット利用対象外であることを表す金枠カード、対象とするフォーマットで利用できないカードタイプを持つものや、通常のカードサイズでないことを表す isOversized という属性を持つものを削除した。また、カードと

して印刷されていないカードも存在したが、今回は中古販売価格の値段についてそれらのカードの価格予測はできないため、調査の対象外とし削除した。

カード性能データセットの各データの持つ属性は、MTG JSON 内に存在する属性を利用して作成した。前述したように対象としないカードを削除した際、今回対象としないカードのみが持つ属性やカード id 等の一部不要な属性も削除した。また、Python の pandas 内部で読み込む際に表記揺れやデータ形式が不正なものが存在したため、それらを修正・変更した。加えて、今回は元のデータセットに存在していなかった属性として year, printings_num, textColor という 3 つの属性を追加した。year は、各カードが封入されている商品の発売年度をそのカードの発行年として考慮に入れるため、カードが封入されている商品を表す setCode という属性と、SetList データセットを利用して作成した。printings_num は、同じカードの発行回数を流通量として考慮するため、同じカードが封入された商品を記す printings という属性から商品数を算出し利用した。textColor は、カード自身が持っている色ではないが、テキスト中に出てくるマナコストの持つ色を表す属性である。カードの持つ色に関する情報について、本来は colorIdentity と colors の 2 属性が存在したが、colorIdentity は colors の持つ情報にテキスト中のマナコストの色の情報を含み、colors の内容が重複していた。そこで、colorIdentity から colors の情報を取り除くことで textColor を作成し、カード自体の持つ色を表す属性 colors と、カード自身が持っていないテキストの色 textColor の 2 つの属性からカードの持つ色を定義した。

以上の様に対象としないデータの削除・表記揺れ等の修正、属性の作成を行い、最終的なカード性能データセットを作成した。作成したカード性能データセットの属性と、実際の値の対応例は付録に記載する。

中古価格データセットは、AllPrices というデータセットを元に、前処理を行って作成した。このデータセットは、JSON 形式であるが、“MTG JSON” のフォーマットにより 6 重の入れ子構造になっていた。また、入れ子の内部構造の一部が定まった形ではなく、店舗価格データと店舗価格データとの間に通貨情報データが入る非構造データとなっていた。そのため、入れ子構造をなくし構造データに変換するプログラムを作成し、すべてのデータの対応を示した CSV の形式に変換し利用した。AllPrices について JSON ファイル内部の詳細な構造については、付録にて記載する。CSV に変換したデータについては、1 枚のカードに対して複数の属性を持っていた、表 4.2 にそのデータセットの持つ属性を表す。

表 4.2 中古販売価格データセットの属性

属性	意味
uuid	カードを識別するための id
format	販売している形態 (実際に販売されている paper の形態のみ使用)
priceType	カードの販売サイト (cardKingdom, cardmarket, tcgplayer の 3 店舗)
cardType	カードの種類 (今回は nomal のデータのみ使用)
newestDay	参照したカードの価格取得日の最新日
currency_type	カードの通貨形態 (USD, EUD)
val	実際のカード価格 (USD, EUD のものは通貨相場より価格を 1.2 倍して利用 0.01USD ~ 40084.505USD)

表 4.2 のような属性を持つデータセットから、今回はカード価格についてその最新日の価格のみを利用した。実際のデータセットでは 2020 年 6 月 28 日 ~ 同年 9 月 27 日までの価格が存在しているが、今回は時系列的な傾向ではなく、相場全体における価格要因の傾向を確認するため、各カードの取得最新日の価格データのみを利用した。また、カードについても複数の店舗におけるカードの価格が存在したため、各店舗の最新価格取得日における価格同士の平均をカードの価格として利用した。その結果、価格が 0.01USD ~ 40084.505USD の範囲におけるカードを取り扱った。

最終的に、カード性能データについては 41,236 件、中古価格データについては、2020 年 6 月 28 日から同年 9 月 27 日の期間における販売価格 141,434 件を利用し、それぞれのデータセットを結合し、実験で用いる最終的なデータセットを作成した。

4.3 回帰分析による価格要因分析

4.3.1 実験概要

本実験では、(1) 『カードゲームにおける性能属性』と (2) 『カードの希少度を表す属性』について、テキスト以外の属性のみを利用し価格要因を説明が可能であるか、機械学習による回帰分析から評価を行った。

実験で使用したカードデータは、前述したカードデータセット 41,236 件とその価格データ 141,434 件を対象とした。データセットの属性のうち、ゲーム上で参照される属性を公式ゲームルールと照らし合わせ考慮し、(1) 『カードゲームにおける性能属性』と (2) 『カードの希少度を表す属性』として分類して利用した。

また、今回の実験で回帰分析に利用した属性は、それらの分類し属性のうち、カテゴリ属性と数値属性であるものと価格データを利用した。表 4.3 に (1) 『カードゲームにおける性能属性』として利用した属性と、表 4.4 に (2) 『カードの希少度を表す属性』として利用した属性を示す。また、図 4.1 に実際のカードとの対応の例を示す。

表 4.3 実験に用いたデータセットの (1) 『カードゲームにおける性能属性』

属性	意味
colors	カード自身の色 (赤:R, 白:W, 黒:B, 青:U, 緑:G)
textColor	カード自身の色以外にテキストで利用されている色 (同上)
convertedManaCost	manaコストを色を考慮せず数値に変換したもの (0~16)
hasAlternativeDeckLimit	枚数制限の有無 (4枚以上の使用が可能か否か)
layout	カードのレイアウト (normal, aftermath, split, flip, leveler, saga, transform, adventure, meld の9種)
loyalty	忠誠度 (Planeswalker タイプのみ使用, 0~7)
power	パワー (-1~16)
toughness	タフネス (-1~16)
types	カードのタイプ (Enchantment, Creature, Land, Instant, Sorcery, Artifact, Planeswalker, Tribal の8種)

表 4.4 実験に用いたデータセットの (2) 『カードの希少度を表す属性』

属性	意味
hasContentWarning	人種差別を想起させるようなカードとして禁止されているか否か
hasFoil	フォイル版 (特殊加工) の有無
hasNonFoil	ノンフォイル版 (加工なし) の有無
isAlternative	特殊バージョンのカードの有無
isFullArt	フルアート版 (特殊なイラスト) の有無
isPromo	プロモーション配布されたものか否か
isReprint	同名カードが再印刷されたか否か
isReserved	再録禁止リストにあるか否か
isStarter	スターターセット (商品) に入っているか否か
isTextless	テキストなしの特殊バージョンか否か
year	封入された商品の発売年度 (1993~2020)
printings_num	同名カードの印刷回数 (1回~174回)
borderColor	カードの枠線の種類 (black, white, borderless の3種)
frameVersion	カードレイアウトの種類 (1993, 1997, 2003, 2015, future の5種)
rarity	カードの商品内封入率 (common, uncommon, rare, mythic の順に低い)

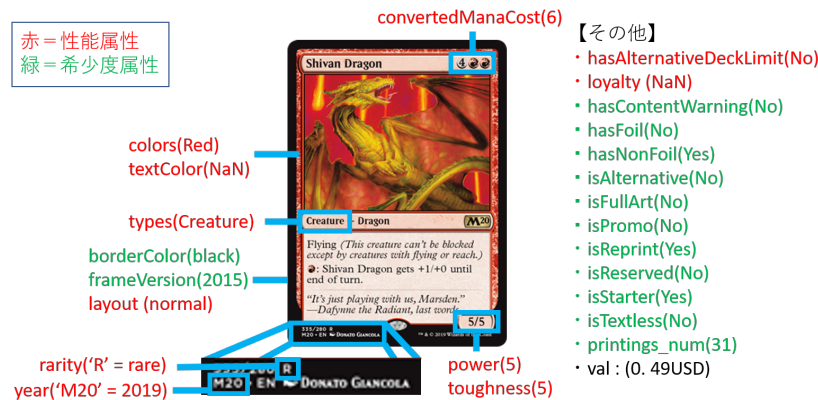


図 4.1 実験に使用した属性と実際のカード [17] との対応の例

これらの属性は回帰分析に用いる際に、カテゴリ属性についてはダミーコーディングを行い利用した。カテゴリ属性のうち、特に colors, textColor, types については 1 枚のカードが複数のカテゴリを同時に有するケースが存在した。そのため、colors, textColor は色 5 色に対応した属性 (例えば、colors が赤を含む場合 colors_R など) を作成、types はカードの種類 8 種類に対応した属性 (例えば、types がクリーチャーを含む場合 types_Creature など) を作成し、カードがそのカテゴリを含むかどうかを判定しダミーコーディングを行った。

また、数値属性については“X”や“1+*”など、初期値がゲームの状況によって変化する旨が記載されている値が存在した。そのため、数値属性である power, toughness, convertedManaCost, loyalty については可変であることを示す属性 (例えば、power_X など) を新たに作成した。また数値属性それぞれの取りうる値の範囲が異なる属性のため、回帰係数の影響を評価するためにはすべての属性のとりうる範囲を一定にする必要がある。そのため、前述の数値属性 4 つと year, printings_num について、以下のような数式を用いて 01 正規化を行い、すべての属性を 0 から 1 の間の数値として扱った。属性 X について、各インスタンスがとりうる値 X_i は、式 4.1 で 01 正規化できる。

$$x_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (4.1)$$

データの前処理を行った後、データのうち 9 割を学習データ、1 割をテストデータに分割し、学習データのうち価格 (val) を目的変数 Y, それ以外の属性を従属変数 X として、回帰分析を行った。本実験では、scikit-learn[15]に含まれている linear_model を利用し、線形回帰, ElasticNet, Ridge, Lasso の 4 つの回帰モデルによる学習器の作成を行った。

学習器の作成の際、(1) 『カードゲームにおける性能属性』と (2) 『カードの希少度を表す属性』の分類についてどのようにそれぞれが作用しているのか、(a) 『カードゲームにおける性能属性』のみ、(b) 『カードの希少度を表す属性』のみ、(c) 両方の属性を利用した場合の

3 条件について、各 4 種類の回帰モデルの作成を行った。

最終的に各回帰モデルの示した回帰係数を確認し、実際にどのような属性が価格要因として影響しているか確認を行った。また、(a)(b)(c) の条件下における回帰係数を確認し、性能・希少性のみを利用した場合と両方考慮した場合の違いについて確認を行った。

またそれらの学習器について、分割したテストデータを利用し、モデルの当てはまりを評価する決定係数 (R^2) とテストデータに対する二乗平均平方根誤差を求めることで、各属性から価格が説明可能であるか否かを評価した。

4.3.2 実験結果

それぞれの回帰モデルの決定係数と二乗平均平方根誤差について、表 4.5, 4.6, 4.7 に示す。

表 4.5 (a) 『カードゲームにおける性能属性』のみの各回帰モデルの性能

回帰モデル	決定係数 (R^2)	二乗平均平方根誤差 (RMSE)
線形回帰	11.03×10^{-3}	158.2
ElasticNet	3.637×10^{-3}	158.8
Ridge	11.09×10^{-3}	158.2
Lasso	7.069×10^{-3}	158.5

表 4.6 (b) 『カードの希少度を表す属性』のみの各回帰モデルの性能

回帰モデル	決定係数 (R^2)	二乗平均平方根誤差 (RMSE)
線形回帰	90.76×10^{-3}	151.7
ElasticNet	30.38×10^{-3}	156.4
Ridge	90.72×10^{-3}	151.7
Lasso	99.12×10^{-3}	151.0

表 4.7 (c) 両方の属性を利用した各回帰モデルの性能

回帰モデル	決定係数 (R^2)	二乗平均平方根誤差 (RMSE)
線形回帰	102.5×10^{-3}	150.7
ElasticNet	33.64×10^{-3}	156.4
Ridge	102.7×10^{-3}	150.7
Lasso	104.6×10^{-3}	150.5

また、性能属性・希少度属性・全属性の 3 条件において、Lasso 回帰において回帰係数を示した属性と、その回帰係数を表 4.8, 4.9, 4.10 に示す。なお、ここに示されている以外の属性については、属性削減を行う Lasso 回帰において回帰係数を示さず 0 だったものである。すべての属性の各モデルにおける回帰係数は付録にて記載する。

表 4.8 Lasso 回帰において回帰係数を示した属性とその回帰係数 (性能属性のみ)

属性	線形回帰	ElasticNet	Ridge	Lasso
types.Artifact	5.115	6.924	5.424	40.11
convertedManaCost	-160.4	-1.384	-159.2	-10.39
types.Creature	-81.67	-5.239	-81.08	-10.91

表 4.9 Lasso 回帰において回帰係数を示した属性とその回帰係数 (希少度属性のみ)

属性	線形回帰	ElasticNet	Ridge	Lasso
isReprint	15.60	0	16.52	2.098
isReserved	460.2	16.40	461.5	414.5
frameVersion_1993	-4.960×10^{15}	14.82	41.00	53.22
rarity_rare	-2.020×10^{15}	9.833	5.979	8.558
frameVersion_1997	-4.960×10^{15}	-3.859	-26.26	-7.260

表 4.10 Lasso 回帰において回帰係数を示した属性とその回帰係数 (全属性)

属性	線形回帰	ElasticNet	Ridge	Lasso	属性分類
types.Artifact	5.337	6.372	5.471	28.03	性能属性
convertedManaCost	-171.5	-1.596	-169.8	-10.09	性能属性
types.Creature	-52.15	-4.852	-51.77	-7.928	性能属性
types.Enchantment	-50.19	-1.415	-49.82	-9.731	性能属性
isReprint	13.13	0	13.26	0.457	希少度属性
isReserved	456.6	16.36	455.2	412.1	希少度属性
frameVersion_1993	4.850×10^{14}	14.73	41.78	52.12	希少度属性
rarity_rare	-3.730×10^{14}	9.830	7.565	8.553	希少度属性
frameVersion_1997	4.850×10^{14}	-3.781	-23.31	-6.230	希少度属性

4.3.3 考察

分類性能に関する考察

表 4.5, 4.6 より, (a) 『カードゲームにおける性能属性』のみと (b) 『カードの希少度を表す属性』のみの各回帰の性能を比較すると, 希少度のみを考慮した方が, 決定係数がより高いため説明変数の当てはまりがよく, 二乗平均平方根誤差もわずかではあるが抑えられていることから, TCG における価格要因としては, 『カードゲームにおける性能属性』より 『カードの希少度を表す属性』の方が強く影響していることがわかる.

しかし, 表 4.7 について, 両方の属性を考慮した場合においても, それぞれの回帰において, その決定係数は線形回帰・Ridge・Lasso において 100×10^{-3} 程度, ElasticNet では 30×10^{-3} とモデルへの当てはまりは低いものとなった. 加えて, 二乗平均平方根誤差はど

のモデルにおいても 158 程度と、実際の価格帯が 0.01USD 単位であることから考慮すれば大きな誤差であり、予測精度は低いものとなった。このことから、性能属性と希少度属性の影響度合いを比較するという目標は達成できたが、数値属性・カテゴリ属性を用いて価格を説明が可能であるか、という点においてはうまく説明ができておらず失敗であるといえる。失敗した理由について、一つは、線形回帰を利用した点そのものに問題があると考えられる。カード価格のデータ数上位 10 件の価格とそのデータ数を確認したところ、安い金額 (約 0.1USD) への極端な偏りが判明した。表 4.11 にその結果を示す。

表 4.11 価格データ数上位 10 件の価格とその数 (0.01USD ~ 40084.505USD)

カード価格 (USD)	データ数
0.101	538
0.099	501
0.096	464
0.094	455
0.104	451
0.091	407
0.106	391
0.125	374
0.109	355
0.089	339

また、図 4.2、図 4.3 (図 4.3 は図 4.2 のうち 2.00USD 以下を拡大したもの) のように四分位範囲を確認したところ、第 3 四分位数が 0.765USD とほとんど 1.00USD 未満の数値が占めていた。

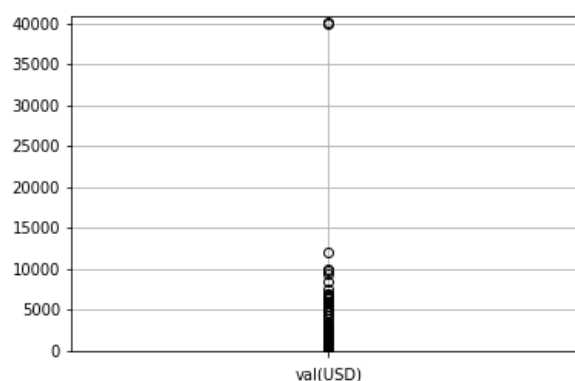


図 4.2 販売価格の箱ひげ図 (0.01USD ~ 40084.505USD)

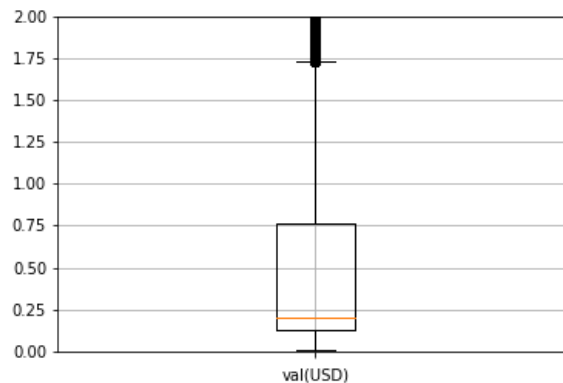


図 4.3 販売価格の箱ひげ図 (0.01USD ~ 2.00USD)

このことから，目的変数である価格が均一に分布しておらず，正規化されていない状態では線形回帰によってうまく説明ができなかったと考察できる．また，(a)(b)(c) いずれの属性を用いた場合でも，価格帯に大きく偏りがあることから，線形回帰で表現できるモデルの複雑さではなく，実験に線形回帰を適用する手法自体が適切でないといえる．そのため，今後は対数などを利用して正規化した値を用いて再度線形回帰を試みることや，価格帯を等級に分けたクラス分類問題等に変更し，より複雑なモデルを用いた分析を行うなどが課題としてあげられる．

回帰係数と属性に関する考察

前述した分類性能に関する考察から，希少度属性の方がその価格への影響度が高いことを示せたが，予測精度が乏しいため，求められた属性ごとの回帰係数についてもその妥当性の判断が難しく，どの要因が特に寄与しているのかを定量的に判断することは難しい．しかし，回帰係数を示した属性自体の妥当性については考察が可能であるため，Lasso 回帰において求められた属性を考察していく．

性能属性のみの実験においては，`types_Artifact` が正の回帰係数を示した．これは，実際の事例においてもゲーム中において強力に利用しやすいカードが多く，実際に高額カードの中では `Artifact` を持つカードが多いことから，結果通り価格要因として寄与しているといえる．また，負の回帰係数を示したものとして，`convertedManaCost`，`types_Creature` があげられる．`convertedManaCost` について，実際のゲームルールにおいては大きければ大きいほどゲーム中ではつかわれにくい．そのため，性能による需要としてカード価格が付きにくいと考えられるため，結果として負の回帰係数が付くことに妥当性があるといえる．しかし，`types_Creature` については負の回帰係数を示した．これは，`types_Artifact` が正の価格要因として出ていることから，アーティファクトではないタイプを持つカードについて，今回はクリーチャーのタイプを持つもののみが価格として安く設定されているといえる．

希少度属性のみの実験においては，`isReprint`，`isReserved`，`rarity_rare` が正の回帰係数

を示した。isReserved, rarity_rare, isReprint については, isReserved は再録禁止の有無, rarity_rare は商品における封入率が低いことを表し, isReprint は同じカードの再印刷の有無など, 実際の事例として供給側を表す指標であり価格要因として妥当であるといえる。また, frameVersion_1993 は正の価格要因, frameVersion_1997 は負の回帰係数を示した。このことから, 古いほど希少度価値が付き価格が付くと考えられるが, そのほかの frameVersion_2003, frameVersion_2015, frameVersion_future では回帰係数を示さなかったことや, frameVersion という属性の多くが年代別の分類であるため year の属性とある程度相関関係にあると考えられることから, frameVersion が示した回帰係数は価格要因として妥当ではなかったといえる。しかし, 実際の事例においても古いカードは供給が少ないため, より値段が付きやすいといった事例があり, 「年代」自体が価格要因として影響している可能性はあるといえる。

全属性を考慮に入れた実験においては, 性能属性, 希少度属性それぞれで回帰係数を示した。しかし, 性能属性のみの考慮では回帰係数を示していないかった types_Enchantment が回帰係数を示した。この属性に関しては types_Creature と同様に負の回帰係数を示した。これは, types_Artifact も正の回帰係数を示していることから, アーティファクトではないタイプを持つカードについて, 価格として安く設定されているといえる。だが, なぜ両方の属性を考慮した場合にのみ, types_Enchantment に回帰係数を示すのか把握することができず, 興味深い結果となった。

4.4 決定木学習 (CART) による分類実験

4.4.1 実験概要

前項で示した回帰分析による価格要因分析の結果から, 回帰分析以外の手法を用いて検証するため, 本実験ではカードの持つ特性について, 決定木学習を用いて各属性による説明が可能であるか分析を行った。

実験で使用したデータは, 4.3 の実験と同様に, 前述したカードデータセット 41,236 件とその価格データ 141,434 件を対象とした。使用したデータセットの属性は前述した表 4.3 と表 4.4 と同様のものを利用した。その際, 数値属性のみ前処理段階での 01 正規化をせずに学習木を構築した。

また, クラス分類問題として学習木を適用するため, 価格のうち第 3 四分位点であった 0.765USD を基準とし, それより高価なカードを “High” クラス, それ以外を “Low” クラスとする 2 クラス分類として学習木を構築した。

本実験では, 学習木の構築に scikit-learn に含まれている Decision Tree Classifier (CART) を用いた。その際には, 分類の基準にエントロピーを利用し, それ以外のいくつかのパラメータについてはデフォルトの値を用いた。この実験においても, 前述 4.3 の実験のように, (a) 『カードゲームにおける性能属性』のみ, (b) 『カードの希少度を表す属性』のみ, (c) 両方の

属性を利用した場合の3種類の学習モデルの作成を行いその比較を試みた。

最終的に、実験に使用したデータセットのうち、9割を学習データ、1割をテストデータとし、学習データを用いて決定木を作成した。また、前述した決定木の評価指標について、陽性を“High”クラスとして計算し、テストデータを利用した予測結果の評価を行った。その結果より、構築した学習木の性能からそれぞれの条件による説明が可能か否か評価した。

4.4.2 結果

構築された学習木の予測結果の評価について表 4.12, 4.13, 4.14 のように示す。

また、属性の条件 (a)(b)(c) について、それぞれ構築した決定木の4段目までを図 4.4, 4.5, 4.6 に示す。グラフの各ノードについて、一番上の表記は学習木の分岐条件を表している。entropy の項目は、各ノードにおける平均情報量を表し、samples は、データ全体におけるそのノードに分岐するサンプル数の割合を表している。value の項目は、それぞれデータ全体のうちどの程度“High”、“Low”のクラスに分類されるか割合を表している。class はそのノードに到達した際の分類結果となる。

表 4.12 (a) 『カードゲームにおける性能属性』のみを用いた際の予測結果の評価尺度とそのスコア

評価尺度	スコア
Accuracy	0.741
Precision	1.000
Recall	0.022
F 値	0.043

表 4.13 (b) 『カードの希少度を表す属性』のみを用いた際の予測結果の評価尺度とそのスコア

評価尺度	スコア
Accuracy	0.806
Precision	1.000
Recall	0.331
F 値	0.470

表 4.14 (c) 両方の属性を用いた際の予測結果の評価尺度とそのスコア

評価尺度	スコア
Accuracy	0.806
Precision	1.000
Recall	0.331
F 値	0.470

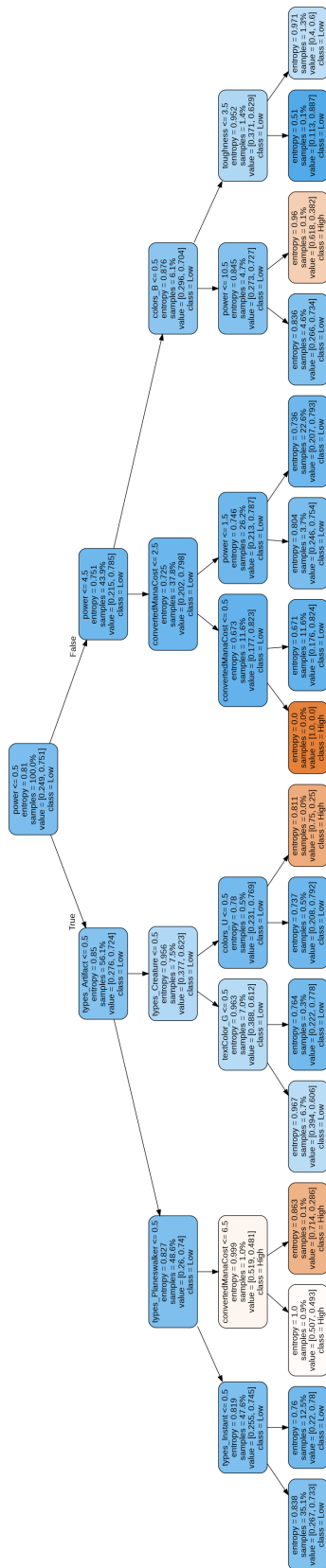


図 4.4 (a) 『カードゲームにおける性能属性』のみを用いて構築した決定木 (4 段目まで)

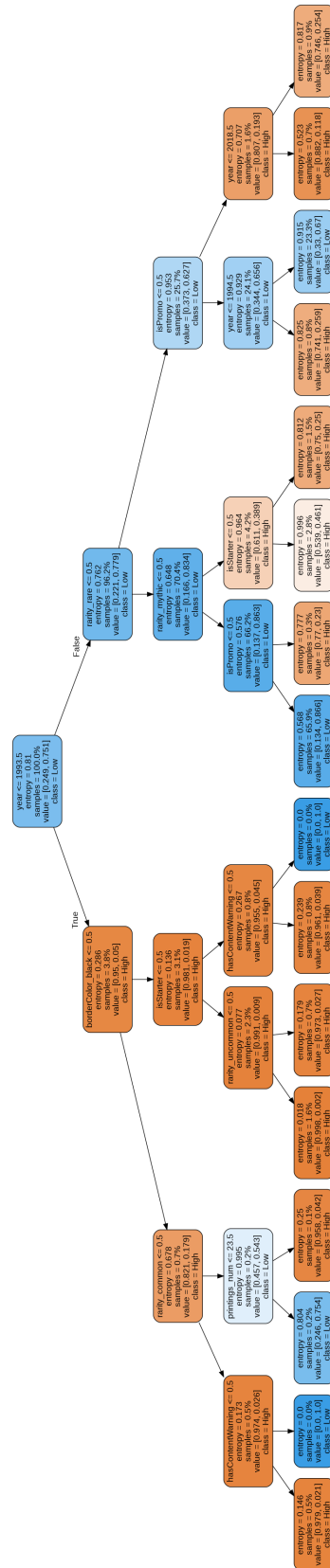


図 4.5 (b) 『カードの希少度を表す属性』のみを用いて構築した決定木 (4 段目まで)

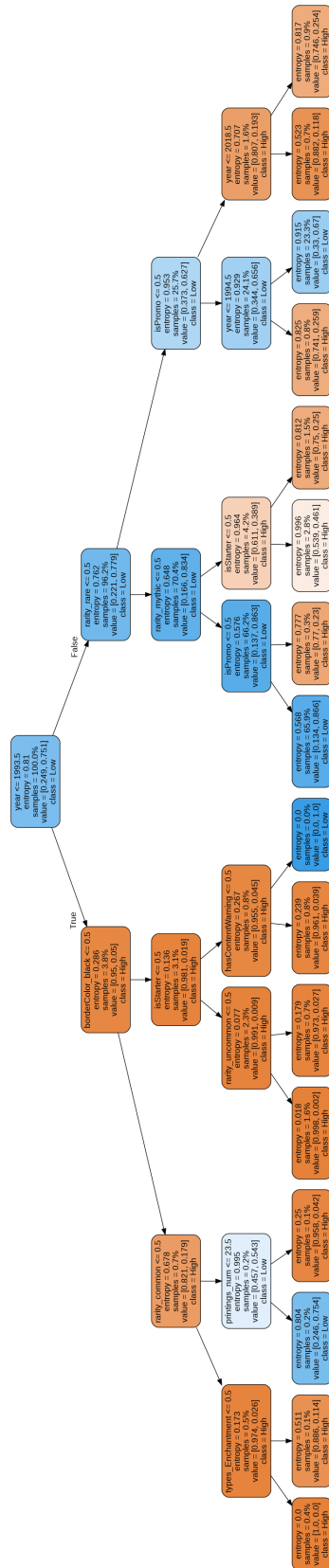


図 4.6 (c) 両方の属性を用いて構築した決定木 (4 段目まで)

4.4.3 考察

構築された学習木に関する考察

性能属性のみの場合である図 4.4 を確認すると、最終的に“High”に分類されたサンプルを確認すると 1% 程度しか残らなかった。これについて、前述したように価格の極端な偏りも原因であると考えられる。しかし、性能属性だけでは sample が全体の第 3 四分位点である 25% から 1% 程度しか残らず、カードの希少度のみの場合である図 4.5 を確認すると、最終的に“High”に分類されたサンプルが 10% 程度残っていることから、やはり (a) 『カードゲームにおける性能属性』のみではうまく分類できないと考察できる。

また、(c) 両方の属性を考慮した場合においても、図 4.6 を確認すると、左下の types_Enchantment が分類に使われていることを除き、ほとんど (b) 『カードの希少度を表す属性』と構築された学習木とほぼ同じであることから、カードゲーム全体の価格要因として、数値属性・カテゴリ属性を利用した場合においては、『カードゲームの性能属性』よりも『カードの希少度を表す属性』の方が強く影響しているといえる。

両方の属性を考慮した場合において、types_Enchantment が学習木の分類に利用されていることは、4.3 の実験で行った回帰分析の際にも同様の傾向があることから、価格要因として寄与している可能性が高いと考えられる。しかし、性能属性だけでは分類において重要視されなかった types_Enchantment が、両方の属性を考慮した場合に学習木でも現れたことについては、説明がつかず興味深い結果となった。

学習木の分類性能に関する考察

だが、決定木の分類性能については表 4.12, 4.13, 4.14 から、F 値の値がどの学習木においても 0.47 を上回らず、学習木の分類性能が低いことから、数値属性・カテゴリ属性を用いて価格を説明することが可能であるか、という点においてはうまく説明ができておらず失敗であるといえる。また、分類性能が乏しいものとなったのに対し Precision が 1 となったことから、ほとんどのテストデータについて“Low”に分類される学習木になったといえる。分類性能が低いことから、やはり分類に使われている属性それぞれについて、その影響度合いを十分に比較するのは難しいといえる。

4.5 価格帯ごとの属性傾向の確認実験

4.5.1 実験概要

実験 4.3, 4.4 より、回帰や分類に利用されてきた属性が確認できたが、どちらも精度が低いものであり、それらにおける属性の妥当性を定量的に検討することは難しいことが示された。そこで、本実験では実験 4.3, 4.4 で示された属性について、価格帯ごとのデータ割合を調べ、実際のデータ中において、その属性に関する傾向がみられるかどうか実験を行った。

実験で使用したデータは、実験 4.3 や 4.4 と同様に、前述したカードデータセット 41,236 件とその価格データ 141,434 件を対象とした。使用したデータセットの属性は前述した表 4.3 と表 4.4 と同様のものを利用した。

今回は、それらのデータについて、価格 (val) の値ごとに 4 つの価格帯ラベルに分類した。表 4.15 は価格帯とその分類クラス、該当するデータ数を表す。各価格帯の分類は、実際の相場価格を参考にラベル境界を決定した。実験では、このラベルごとに確認した属性における各値の割合を確認し、その妥当性を検討した。また、全体における割合についても同時に確認した。

表 4.15 価格 (val) の値と該当する価格帯ラベルとの対応

価格帯ラベル	価格 (val) の範囲	データ数
Low	0.01 ~ 0.74USD	30920
Middle	0.75 ~ 10.0USD	8178
High	10.01 ~ 150.0USD	1796
More_High	150.01USD ~ 40084.505USD	342

また確認を行った属性として、実験 4.3 おいて Lasso 回帰が回帰係数を示した属性である、types_Creature, types_Enchantment, types_Artifact, isReprint, isReserved, convertedManaCost, frameVersion_1993, frameVersion_1997, rarity_rare の 9 属性を利用した。また、実験 4.4 の学習木において、前述した 9 属性以外に分類に利用された、border_Color_Black, year, rarity_common, rarity_mythic, isStarter, isPromo の 6 属性についても確認を行った。

最終的に、以上の属性について価格帯ごとのユニークな値の割合を調べ、実際に実験 4.3, 実験 4.4 で示されたような属性に、価格との関係性があるのか確認した。

4.5.2 結果

属性の値と価格ごとの割合について、特に High, More_High と Low, Middle について異なる傾向を示したものを、表 4.16, 4.17, 4.18, 4.19 に示す。カテゴリ属性の割合を表した表 4.16, 4.18 ではそれぞれの価格帯における True, False の割合を表す。数値属性の割合を表した表 4.17, 4.19 では割合の高い上位 5 件について、その値の割合を表す。

また、どの価格帯においても傾向が変わらないものを含めて、確認を行った全属性の割合については付録に記載する。

4.5.3 考察

表 4.16 について、rarity_rare=True の割合が Low の段階では 17.78% と少数であったが、Middle, High, More_High では True の方が多い結果となった。さらに、More_High

表 4.16 rarity_rare におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	72.78%	82.21%	45.30%	47.32%	11.40%
True	27.21%	17.78%	54.69%	52.67%	88.59%

表 4.17 convertedManaCost(0~16) におけるデータ上位 5 つと価格帯ごとの割合

	ALL		Low		Middle		High		More_High	
値	割合	値	割合	値	割合	値	割合	値	割合	
3	21.30%	3	22.06%	3	19.61%	3	17.53%	0	29.23%	
2	19.41%	2	20.52%	4	16.39%	0	15.86%	2	20.17%	
4	16.50%	4	16.65%	2	16.05%	2	15.59%	1	15.20%	
1	11.40%	1	11.30%	0	12.37%	1	15.08%	4	12.86%	
0	11.28%	0	10.53%	5	11.21%	4	15.03%	3	12.86%	

表 4.18 frameVersion_1993 におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	89.98%	94.02%	85.14%	58.74%	4.67%
True	10.01%	5.97%	14.85%	41.25%	95.32%

表 4.19 year(1993~2020) におけるデータ上位 5 つと価格帯ごとの割合

	ALL		Low		Middle		High		More_High	
値	割合	値	割合	値	割合	値	割合	値	割合	
2019	9.77%	2019	10.37%	2019	8.96%	1993	30.40%	1993	85.96%	
2018	7.05%	2018	6.97%	2020	8.35%	2020	9.57%	1994	9.35%	
2020	6.29%	2017	6.39%	1993	8.31%	1994	9.24%	1999	2.92%	
2017	6.22%	2020	5.62%	2018	7.71%	2018	6.62%	1998	1.16%	
2016	5.24%	2016	5.35%	2017	6.45%	1999	5.40%	2020	0.29%	

では 88.59% が True を示す結果となった。このことから, rarity_rare, つまりカードのレアリティがレアに属するものについて価格との関係性が高い属性であるといえる。また, rarity_rare は実験 4.3 で回帰係数が正の値を示したことから, レアリティがレアであればより価格が高くなるといえる。これは, 実際の商品においても, レアリティは商品内でのカードの封入率を表し, レアリティがレアのカードは封入率が低いものである。そのため, レアリティはそのカード自体の供給量を表す指標と考えることも可能であり, カードの価格との影響があることには妥当性があるといえる。しかし, 回帰分析の際にはその回帰係数の大きさは特出したものではなかったが, 実際には ALL において少数属性であり, 価格ごとの割合をみた際に大きな差がみられ, 興味深い結果となった。

表 4.17 について, convertedManaCost の各価格における上位 5 つは, All や Low, Middle の際には, 3 が上位を占めていたが, More_High では 0 が 30% 近くを占めていた. このことから, 特に最上位価格層については特に軽量の convertedManaCost, つまりマナコストが軽いものがより重要視されると示された. また, convertedManaCost は実験 4.3 で回帰分析の際に負の相関係数を示したが, これは実際に最上位価格層では小さい値が重要視されており, 軽ければ軽いほど価格が上がり, 負の回帰係数を示した原因となったといえる. これについては, 実際のゲーム上の性能においても, マナコストの値が小さいほどゲーム中で利用しやすい傾向がある. そのため, 性能による需要としてマナコストが小さいほどカード価格が高くなることには妥当性があるといえる.

表 4.18 について, frameVersion_1993 の各価格における割合は, Low において True が 5.97%, More_High において True が 95.32% となった. また, 表 4.19 について, 各カードが封入された商品の発売年を表す year の割合上位 5 件が, High, More_High において 1993 年が一番高い結果となった. このことから, 価格の要因として発行年度は価格との強い関係があるといえる. 実際に, Low や Middle の場合では, 最上位の 2019 でも 11% を超えないのに対し, 1993 年について High では 30.40%, More_High では 85.96% となっていることから, 特に 1993 年が価格に対して関係しているといえる. これは, 実験 4.4 の希少度属性のみ, 両方の属性を考慮した場合の学習木において, 根ノードの分岐条件が $year \leq 1993.5$, つまり 1993 年か否かを確認していたことから, 実際の分析結果と一致するものとなっている. さらに, 1993 年は「Magic: The Gathering」が発売された最初の年であるため, 特に最初期に出たカードについて中古販売価格が高くなることが示唆された.

以上のことから, 線形回帰や決定木で確認できた属性の一部について, 実際の割合から妥当性があることが示唆され, 実験は成功したといえる.

第 5 章

結論

5.1 まとめ

本研究では、中古 TCG の価格要因分析を行うため、カード自体が持つ属性を (1) 『カードゲームにおける性能属性』と (2) 『カードの希少度を表す属性』に該当するものに分類し、それぞれどのような属性が価格決定に寄与しているのか分析を行った。

第 1 章では、本研究の背景を説明し、本研究の目標である中古販売価格予測システムや、その前提として価格要因を捕らえる研究であることを論じた。

第 2 章では、中古価格要因分析の先行研究、TCG に関する研究、TCG の中古販売価格に関する研究の 3 つの視点から、本研究の立ち位置を明らかにした。

第 3 章では、前提となる基本的な知識、また対象とした「Magic: The Gathering」のルールを説明し、改めて本研究の提案手法を論じた。

第 4 章では、実際にカテゴリ属性、数値属性を利用し、回帰分析と決定木 (CART) による分析を行い、その結果について考察を述べた。その結果、性能属性よりも希少度属性の方が価格決定に寄与していることを示した。また、価格帯ごとに属性の持つ値とその割合を確認し、レアリティ、マナコスト、封入商品の発売年が価格に影響していることが示唆された。しかし価格予測の点においては、価格の分布が 0.1USD 付近に極端に偏っていることや、回帰分析の結果を用いた価格予測の精度や分類木の性能が著しい結果となったことから、属性同士の影響を比較し精査することは達成できなかったものといえる。また、本研究の目指す TCG の価格予測システムについても、その予測精度から数値予測ができていたとはいえず、より複雑なモデルでなければ TCG の予測が困難であることが判明した。

最終的に、回帰分析や決定木の精度が著しいことから、各属性同士の正確な影響度合いを確認することができなかったが、本研究の分析手法として挙げた性能属性と希少度属性について、希少度属性の影響度が大きいと示すことができた。また、価格に影響していると考えられるの属性について興味深い結果となったことから、本研究の目的として一部は達成し、TCG の中古価格要因分析に寄与することができたといえる。

5.2 今後の展望

今後の展望について、実験で示された内容から属性に関する検討や、別の機械学習手法による分析などがあげられる。

今回の実験において、性能属性は希少度属性ほど価格要因として寄与していないと示されたため、性能属性が及ぼす傾向について再度考慮していく必要があるといえる。これは、実験 4.3 の学習木のようにほとんど寄与していない可能性も考慮できるが、`types.Enchantment` のような性能と希少性を同時に利用することで確認された属性も存在し、実例として性能による高騰が起きた例が確認されているため、性能属性と価格との影響について再検討する必要がある。

また、販売価格に極端な偏りがあり、線形回帰や学習木では精度の低いモデルしか構築できなかったため、今後は利用するカードの価格帯を絞った分析や、より複雑なモデルを利用した分析の必要性がある。価格帯を絞った分析については、現状安い価格への極端な分布が見られるため、価格を対数変換し分析する手法を検討している。さらに複雑なモデルとしては、ディープラーニングやサポートベクター回帰、アンサンブル学習による回帰木などを用い、価格要因とその学習器の予測精度について検証する必要があると考えられる。

さらに、本研究では扱っていない部分についても、展望として一考の余地がある。

性能属性については、本研究では利用していないテキスト部分を含めた価格要因分析を行う必要がある。今回使用していないテキスト部分のうち、特にカードの効果の部分についてはカードの性能を決める大きな指標の一つであるため、カードの性能属性の 1 つとして考慮していく必要があると考えられる。また、実際にテキスト属性が強く影響しているのであれば、予測精度の向上も見込まれるため、テキスト属性を分析対象とすることは検討の余地があるといえる。テキスト部分の処理については、Dustin ら [14] の研究で行われていた n-grams を用いた分析を採用し、検証していく等もあげられる。

また、本研究では取り扱っていない時系列的な価格変化についても検討の余地があるといえる。先述した通り、性能属性の評価はそのほかのカードとの相対的な強さである場合があり、新商品発売によるカードプールの増加によって性能の評価が時間的に変わると考えるのであれば、時系列的にその性能の評価が変わるため、時系列的な価格の変化と性能との連動についても検討すべきである。またその性能評価の変化要因も、カードプールの増加だけでなく実際のトーナメントでの使用率や、ゲームルールでの禁止・制限措置等によっても変化するため、それぞれの要因の価格への影響を検証することも課題である。

このように、TCG の中古価格要因の解明には多くの課題が残るが、本研究で明らかにした内容を含め、今後も研究を継続し、実験・検証を行う予定である。

謝辞

本研究を進めるにあたり，研究内容やその方針に関するご指導をいただきました公立はこだて未来大学システム情報科学部情報アーキテクチャ学科の新美礼彦准教授に心から感謝いたします。また，研究に関するアドバイスをして頂いた新美研究室の皆様にも深く感謝いたします。

The authors would like to thank all the volunteers who created, managed, and operated the dataset used in this study, “MTG JSON.” In addition, the authors would like to thank to the people of “Wizards of the Coast”, the developer of “Magic: the Gathering”, which was the subject of this research.

The authors would like to thank Enago (www.enago.jp) for the English language review.

発表・採録実績

発表等

- [1] DEIM2021 第13回データ工学と情報マネジメントに関するフォーラム（第19回日本データベース学会年次大会）（2021年3月1日～3日発表予定）

参考文献

- [1] 「Magic: The Gathering」, <https://mtg-jp.com/>, (2021/01/18 アクセス).
- [2] “First modern trading card game”, <https://www.guinnessworldrecords.com/world-records/first-modern-trading-card-game/>, (2021/01/18 アクセス).
- [3] 「ポケモンカードゲーム」, <https://www.pokemon-card.com/>, (2021/01/18 アクセス).
- [4] 「遊戯王 OCG デュエルモンスターズ」, <https://www.yugioh-card.com/japan/>, (2021/01/18 アクセス).
- [5] 「カードファイト!!ヴァンガード」, <https://cf-vanguard.com/tcg/>, (2021/01/18 アクセス).
- [6] “日本玩具協会 2019 年度国内玩具市場規模”, https://www.toys.or.jp/pdf/2020/2019_data_zenpan.pdf, (2021/01/18 アクセス).
- [7] “PSA”, <https://www.psacard.co.jp/>, (2021/01/18 アクセス).
- [8] 仁ノ平 将人, 三川 健太, 後藤 正幸, “販売履歴データに基づく中古ファッションアイテムの販売価格予測モデルに関する一考察”, 情報処理学会論文誌, Vol.60, No.4, pp.1151-1161, 2019.
- [9] 越川 翼, 林 高樹, “中古自動車の価格要因分析”, 慶應義塾大学大学院経営管理研究科修士課程学位論文, 2018.
- [10] Sverre Johann Bjørke, Knut Aron Fludal, “Deckbuilding in Magic: The Gathering Using a Genetic Algorithm”, Master thesis, NTNU, 2017.
- [11] Felipe Gomes Rufino Moura Paiva, Artur de Oliveira da Rocha Franco, Glaudiney Mendonca Junior, Jose Gilvan Rodrigues Maia, “Analyzing Player Profiles in Collectible Card Games”, SBC Proceedings of SBGames 2018, 2018.
- [12] Loh Zhen Yu, “Factors Affecting Prices Of Singles in Cardfight !! Vanguard”, PhD thesis, UTAR, 2017.
- [13] Matthew Pawlicki, Joseph Polin, Jesse Zhang, “Prediction of Price Increase for Magic: The Gathering Cards”, Stanford University CS229 Projects, 2014.
- [14] Dustin Fink, Benjamin Pastel, Neil Sapra, “Predicting the strength of Magic: The Gathering cards from card mechanics”, Stanford University CS229 Projects, 2015.

- [15] “scikit-learn”, <https://scikit-learn.org/stable/>, (2021/01/18 アクセス).
- [16] “MTG JSON”, <https://mtgjson.com/>, (2021/01/18 アクセス).
- [17] “Shivan Dragon”, <https://gatherer.wizards.com/Pages/Card/Details.aspx?multiverseid=469888>, (2021/01/18 アクセス).

付録

.1 実験で利用した中古価格データセットの構造

本節では、実験で使用した中古価格データセット AllPrices について、実際のデータ例と共にその構造を以下に示す。一部属性を除き、詳細な説明は表 4.2 に記載する。

```
1 "0001e0d0-2dcd-5640-aadc-a84765cf5fc9": //uuidカードを識別するための(id)
2 {"paper":
3   {"cardkingdom":
4     {"buylist": //店舗買取価格
5       {"normal":
6         {"2020-06-28": 1.8,
7           "2020-06-30": 1.3,
8           "2020-07-01": 1.3,
9           "2020-07-05": 1.3,
10          "2020-07-06": 1.3,
11           //中略
12          "2020-09-23": 2.2,
13          "2020-09-24": 2.2,
14          "2020-09-25": 2.2,
15          "2020-09-26": 1.65,
16          "2020-09-27": 1.65
17         }
18       },
19     "currency": "USD", //通貨情報
20     "retail": //中古販売価格
21     {"normal":
22       {"2020-06-28": 3.99,
23         "2020-06-30": 3.99,
24         "2020-07-01": 3.99,
25         "2020-07-05": 3.99,
26         "2020-07-06": 3.99,
27         //中略
28         "2020-09-23": 5.49,
29         "2020-09-24": 5.49,
30         "2020-09-25": 5.49,
31         "2020-09-26": 5.49,
32         "2020-09-27": 5.49
33       }
34     }
35   },
```

```

36 "tcgplayer":
37 {"buylist":
38 {"normal":
39 {"2020-08-13": 1.76,
40 "2020-08-14": 1.71,
41 "2020-08-15": 1.58,
42 "2020-08-16": 1.57,
43 "2020-08-18": 1.67,
44 //中略
45 "2020-09-23": 1.95,
46 "2020-09-24": 2.57,
47 "2020-09-25": 2.57,
48 "2020-09-26": 2.09,
49 "2020-09-27": 2.09
50 }
51 },
52 "currency": "USD",
53 "retail":
54 {"normal":
55 {"2020-06-28": 2.9,
56 "2020-06-30": 2.92,
57 "2020-07-01": 2.92,
58 "2020-07-05": 2.93,
59 "2020-07-06": 2.93,
60 //中略
61 "2020-09-23": 4.03,
62 "2020-09-24": 4.03,
63 "2020-09-25": 4.03,
64 "2020-09-26": 4.03,
65 "2020-09-27": 4.03
66 }
67 }
68 }
69 }
70 }

```

2 実験で利用したカードデータの属性とその一例

本節では、実験で利用したカードデータについて、その属性と実際の値との対応例を表 1 に示す。

表 1 実験で利用したカードデータの属性と実際のカード [17] の例

属性	値
uuid	2c2eeb5aaebc5e9c94b3ffe6b6789407
name	Shivan Dragon
borderColor	black
colors	R(Red)
convertedManaCost	6
frameVersion	2015
hasAlternativeDeckLimit	0(No)
hasContentWarning	0(No)
hasFoil	0(No)
hasNonFoil	1(Yes)
isAlternative	0(No)
isFullArt	0(No)
isPromo	0(No)
isReprint	1(Yes)
isReserved	0(No)
isStarter	1(Yes)
isTextless	0(No)
layout	normal
loyalty	NaN
power	5
printings_num	31
rarity	rare
textColor	NaN
toughness	5
types	creature
year	2019
val	0.49

3 実験 4.3 における回帰係数の一覧

本節では、実験 4.3 で求めた回帰係数について、全ての属性に対する回帰係数を表 2, 3, 4 に示す。

表 2 線形回帰モデルで求められた回帰係数 (性能属性のみ)

属性	線形回帰	ElasticNet	Ridge	Lasso
convertedManaCost	-160.4	-1.384	-159.2	-10.39
hasAlternativeDeckLimit	-6.553	0	-5.801	0
loyalty	41.15	0	33.14	0
power	89.72	0	88.59	0
toughness	58.11	0	57.65	0
loyalty_X	-10.61	0	-10.29	0
power_X	16.59	0	16.46	0
toughness_X	20.79	0	20.52	0
colors_B	-3.544	-1.363	-3.547	0
colors_U	6.159	0	6.129	0
colors_G	-5.305	-1.625	-5.295	0
colors_W	-5.094	-1.767	-5.097	0
colors_R	-4.296	-1.416	-4.291	0
textColor_B	18.28	0.417	18.25	0
textColor_U	31.10	0.937	31.06	0
textColor_G	15.13	0.091	15.10	0
textColor_W	12.62	0.018	12.60	0
textColor_R	14.99	0.132	14.96	0
types_Enchantment	-60.38	-0.729	-59.96	0
types_Creature	-81.67	-5.239	-81.08	-10.91
types_Land	-101.5	0	-100.9	0
types_Instant	-67.11	-0.769	-66.64	0
types_Sorcery	-48.16	0	-47.76	0
types_Artifact	5.115	6.924	5.424	40.11
types_Planeswalker	-70.15	0	-64.82	0
types_Tribal	-9.193	0	-9.109	0
layout_normal	2.048	0	2.117	0
layout_aftermath	12.64	0	12.01	0
layout_split	9.596	0	9.332	0
layout_flip	-2.810	0	-2.614	0
layout_leveler	-7.778	0	-7.468	0
layout_saga	3.815	0	3.682	0
layout_transform	-11.87	0	-11.70	0
layout_adventure	-5.898	0	-5.676	0
layout_meld	0.261	0	0.319	0

表 3 線形回帰モデルで求められた回帰係数 (希少度属性)

属性	線形回帰	ElasticNet	Ridge	Lasso
hasContentWarning	-109.0	0	-104.8	0
hasFoil	13.07	-6.938	8.404	0
hasNonFoil	8.749	0	12.42	0
isAlternative	-2.613	0	2.258	0
isFullArt	17.00	0	-4.678	0
isPromo	-1.647	0	-0.271	0
isReprint	15.60	0	16.52	2.098
isReserved	460.2	16.40	461.5	414.5
isStarter	-4.741	-2.750	-7.522	0
isTextless	-27.44	0	-1.244	0
year	-31.46	-8.524	-31.37	0
printings_num	2.756	0	3.237	0
borderColor_black	7.690×10^{14}	0.428	2.991	0
borderColor_borderless	7.690×10^{14}	0	14.57	0
borderColor_white	7.690×10^{14}	-0.498	-17.56	0
frameVersion_1993	-4.960×10^{15}	14.82	41.00	53.22
frameVersion_1997	-4.960×10^{15}	-3.859	-26.26	-7.260
frameVersion_2003	-4.960×10^{15}	-4.339	-7.726	0
frameVersion_2015	-4.960×10^{15}	-4.572	-0.400	0
frameVersion_future	-4.960×10^{15}	0	-6.619	0
rarity_common	-2.020×10^{15}	-5.188	-7.331	0
rarity_mythic	-2.020×10^{15}	0	7.073	0
rarity_rare	-2.020×10^{15}	9.833	5.979	8.558
rarity_uncommon	-2.020×10^{15}	-3.746	-5.721	0

表 4 線形回帰モデルで求められた回帰係数 (全属性)

属性	線形回帰	ElasticNet	Ridge	Lasso
convertedManaCost	-171.5	-1.596	-169.8	-10.09
hasAlternativeDeckLimit	0.634	0	1.296	0
hasContentWarning	-99.80	0	-95.85	0
hasFoil	9.057	-6.773	9.742	0
hasNonFoil	17.38	0	18.21	0
isAlternative	10.59	0	10.40	0
isFullArt	-3.304	0	-2.513	0
isPromo	-5.902	0	-6.310	0
isReprint	13.13	0	13.26	0.457
isReserved	456.6	16.36	455.2	412.1
isStarter	-1.646	-2.636	-1.559	0
isTextless	-4.382	0	-4.308	0
loyalty	53.68	0	44.47	0
power	104.3	0	103.5	0
toughness	23.89	0	23.59	0
loyalty_X	-7.888	0	-9.142	0
power_X	25.80	0	25.13	0
toughness_X	-7.403	0	-6.671	0
year	-22.96	-8.467	-22.95	0
prings_num	0.631	0	0.441	0
colors_B	-5.994	-1.292	-6.154	0
colors_U	2.947	0	2.824	0
colors_G	-6.106	-1.582	-6.212	0
colors_W	-5.725	-1.689	-5.862	0
colors_R	-6.694	-1.384	-6.764	0
textColor_B	14.78	0.518	14.61	0
textColor_U	29.51	1.059	29.88	0
textColor_G	13.21	0.204	13.26	0
textColor_W	10.51	0.116	10.36	0
textColor_R	11.56	0.236	11.68	0
types_Enchantment	-50.19	-1.415	-49.82	-9.731
types_Creature	-52.15	-4.852	-51.77	-7.928
types_Land	-80.87	0.088	-80.50	0
types_Instant	-40.03	-0.328	-39.64	0
types_Sorcery	-23.58	0	-23.28	0
types_Artifact	5.337	6.372	5.471	28.03
types_Planeswalker	-50.82	0	-45.48	0
types_Tribal	4.565	0	4.239	0

表 5 線形回帰モデルで求められた回帰係数 (全属性)(続き)

属性	線形回帰	ElasticNet	Ridge	Lasso
layout_normal	1.240×10^{13}	0	-6.899	0
layout_aftermath	1.240×10^{13}	0	11.83	0
layout_split	1.240×10^{13}	0	14.44	0
layout_flip	1.240×10^{13}	0	-10.62	0
layout_leveler	1.240×10^{13}	0	-9.744	0
layout_saga	1.240×10^{13}	0	15.61	0
layout_transform	1.240×10^{13}	0	-10.08	0
layout_adventure	1.240×10^{13}	0	-1.222	0
layout_meld	1.240×10^{13}	0	-3.320	0
borderColor_black	7.130×10^{14}	0.492	12.42	0
borderColor_borderless	7.130×10^{14}	0	-4.806	0
borderColor_white	7.130×10^{14}	-0.538	-7.616	0
frameVersion_1993	4.850×10^{14}	14.73	41.78	52.12
frameVersion_1997	4.850×10^{14}	-3.781	-23.31	-6.230
frameVersion_2003	4.850×10^{14}	-4.337	-8.729	0
frameVersion_2015	4.850×10^{14}	-4.562	-4.782	0
frameVersion_future	4.850×10^{14}	0	-4.960	0
rarity_common	-3.730×10^{14}	-5.213	-11.62	0
rarity_mythic	-3.730×10^{14}	0	12.75	0
rarity_rare	-3.730×10^{14}	9.830	7.565	8.553
rarity_uncommon	-3.730×10^{14}	-3.773	-8.691	0

4 実験 4.5 における価格帯ごとの属性傾向の一覧

本節では、実験 4.5 で確認した価格帯ごとの属性の値の割合について、以下に各表を示す。

表 6 types_Enchantment におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	88.87%	89.70%	86.69%	84.63%	88.59%
True	11.12%	10.29%	13.30%	15.36%	11.40%

表 7 types_Creature におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	52.86%	50.15%	58.96%	66.81%	78.36%
True	47.13%	49.84%	41.03%	33.18%	21.63%

表 8 types_Artifact におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	90.02%	91.45%	86.63%	83.85%	74.56%
True	9.97%	8.54%	13.36%	16.14%	25.43%

表 9 isReprint におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	50.08%	51.49%	44.83%	49.10%	53.50%
True	49.91%	48.50%	55.16%	50.89%	46.49%

表 10 isReserved におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	98.08%	99.51%	97.04%	88.53%	56.43%
True	1.91%	0.48%	2.95%	11.46%	43.56%

表 11 convertedManaCost(0 ~ 16) におけるデータ上位 5 つと価格帯ごとの割合

	ALL		Low		Middle		High		More_High	
値	割合	値	割合	値	割合	値	割合	値	割合	
3	21.30%	3	22.06%	3	19.61%	3	17.53%	0	29.23%	
2	19.41%	2	20.52%	4	16.39%	0	15.86%	2	20.17%	
4	16.50%	4	16.65%	2	16.05%	2	15.59%	1	15.20%	
1	11.40%	1	11.30%	0	12.37%	1	15.08%	4	12.86%	
0	11.28%	0	10.53%	5	11.21%	4	15.03%	3	12.86%	

表 12 frameVersion_1993 におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	89.98%	94.02%	85.14%	58.74%	4.67%
True	10.01%	5.97%	14.85%	41.25%	95.32%

表 13 frameVersion_1997 におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	82.97%	81.96%	85.47%	86.46%	95.61%
True	17.02%	18.03%	14.52%	13.53%	4.38%

表 14 rarity_rare におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	72.78%	82.21%	45.30%	47.32%	11.40%
True	27.21%	17.78%	54.69%	52.67%	88.59%

表 15 borderColor_black におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	90.21%	90.60%	90.03%	84.85%	87.71%
True	9.78%	9.39%	9.96%	15.14%	12.28%

表 16 year(1993~2020) におけるデータ上位 5 つと価格帯ごとの割合

ALL		Low		Middle		High		More_High	
値	割合	値	割合	値	割合	値	割合	値	割合
2019	9.77%	2019	10.37%	2019	8.96%	1993	30.40%	1993	85.96%
2018	7.05%	2018	6.97%	2020	8.35%	2020	9.57%	1994	9.35%
2020	6.29%	2017	6.39%	1993	8.31%	1994	9.24%	1999	2.92%
2017	6.22%	2020	5.62%	2018	7.71%	2018	6.62%	1998	1.16%
2016	5.24%	2016	5.35%	2017	6.45%	1999	5.40%	2020	0.29%

表 17 isStarter におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	81.16%	84.48%	68.90%	77.00%	96.19%
True	18.83%	15.51%	31.09%	22.99%	3.80%

表 18 isPromo におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	97.60%	99.38%	92.52%	89.92%	97.95%
True	2.39%	0.61%	7.47%	10.07%	2.04%

表 19 rarity_common におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	60.29%	50.91%	87.48%	90.59%	99.12%
True	39.71%	49.09%	12.52%	9.41%	0.88%

表 20 rarity_mythic におけるデータと価格帯ごとの割合

	ALL	Low	Middle	High	More_High
False	95.78%	99.13%	86.90%	77.73%	99.71%
True	4.22%	0.87%	13.10%	22.27%	00.29%