

コピュラ関数を用いた 検索行動と実行行動の関連性分析

小西 諒¹・桑野 将司²・森山 卓³・細江 美欧⁴

¹学生会員 鳥取大学 工学部社会システム土木系学科 (〒680-8552 鳥取県鳥取市湖山町南四丁目 101)
E-mail: b17t4044y@edu.tottori-u.ac.jp

²正会員 鳥取大学教授 工学研究科社会基盤工学専攻 (〒680-8552 鳥取県鳥取市湖山町南四丁目 101)
E-mail: kuwano@tottori-u.ac.jp

³正会員 鳥取大学助教 工学研究科社会基盤工学専攻 (〒680-8552 鳥取県鳥取市湖山町南四丁目 101)
E-mail: moriyama@tottori-u.ac.jp

⁴学生会員 鳥取大学 工学研究科社会基盤工学専攻 (〒680-8552 鳥取県鳥取市湖山町南四丁目 101)
E-mail: d19t4003b@edu.tottori-u.ac.jp

本研究は香川県で運行される鉄道路線「ことでん」を対象に、経路検索システム「NAVITIME」の検索履歴データと交通系 IC カード「IruCa」の乗降履歴データを用いて、検索履歴データから乗降件数を予測するための分析システムの構築を目的とする。分析では、コピュラ関数を用いて両者の相互依存性を表現するとともに、潜在クラスモデルを用いてサンプルの異質性を表現する。2014年1月1日から12月31日までの1年間データを用いた実証分析の結果、潜在クラス1にフランク・コピュラを、潜在クラス2にクレイトン・コピュラを適用した際のモデル適合度が最も高いことが分かった。そして、提案手法による検索履歴データによる乗車件数の期待値と分散を算出し、提案手法の有用性を確認した。

Key Words: route search service, public transit IC cards, copula function, latent class model

1. はじめに

近年、情報技術の発展により、日常生活の様々な場面でWebサービスの利用が進んでいる。Webサービスの利用例として、Amazon や楽天市場などを用いた購買、目的地までの最適な道りを把握するために行う経路検索、SNS を用いたコミュニケーションなどが挙げられ、様々な分野で Web サービスが提供されている。総務省が発表した令和2年度版情報通信白書¹⁾によると、2019年のインターネット利用率(個人)は89.8%であり、年齢別で見ると13歳~69歳までの各世代で9割を超えている。このことから、インターネットの利用は多くの個人で日常の一部となっており、Webサービスの利用も日常化されていると考えられる。

インターネット上での Web サービスの利用履歴は、単にオンライン上での活動履歴を示すだけでなく、利用者の実際の行動を知るための重要な指標となる。しかし、インターネット上での検索行動は実際の行動を必ずしも反映しない。例えば、インターネット上である商品につ

いて検索し調べたとしても、購入に至るとは限らない。そのため、インターネット上の検索行動から実際の行動を把握する研究^{2),3)}が様々な分野で行われている。このような研究は、主にマーケティング分野で盛んであるが、交通分野においても経路検索システムの検索履歴データに着目した研究などが進んでいる⁴⁾⁻⁷⁾。

経路検索システムとは目的地や出発地、公共交通を利用したい時間帯を入力すると、検索者の希望に即した最適な経路や運賃、所要時間等が提示される Web サービスの一種である。経路検索システムのサービスには「NAVITIME」や「yahoo!乗り換え案内」、 「駅すばあと」などが存在する。これらの全国版経路検索システムサービスでは新幹線や飛行機、高速バスなど全国のあらゆる地点間での移動案内が提供されており、公共交通利用者によく利用されている。2016年12月に内閣府が発表した公共交通に関する世論調査⁸⁾によると、公共交通の経路を調べる際に経路検索システムを利用すると回答した割合が40%以上であり、経路検索のために最も使用されているツールであることがわかる。

この経路検索システムには、検索者が入力した出発地や目的地、指定した日時等の情報が検索履歴データとして蓄積されている。そのため、検索履歴データはどこから、どこへ、いつ移動したいかという検索者の移動希望に関する情報を含み、実際の利用者のニーズを知る重要な情報源になる可能性がある。

経路検索システムの検索履歴データを用いた研究として、伊藤ら⁹⁾が挙げられる。伊藤らは、鳥取県の路線バスを対象とする乗換案内サービス「バスネット」のアクセス記録から、公共交通の利用状況を把握する手法を検討した。その結果、バスネットのアクセス記録から公共交通の需要や利用状況を把握できる可能性を示唆した。石村ら¹⁰⁾は、経路検索システム「乗換 NAVITIME (iOS版)」にて収集される3つのデータ(経路検索実績データ、検索経路データ、経路選択データ)に着目し、情報提供が経路検索行動に及ぼす影響を分析している。香川ら¹¹⁾は、検索履歴データのうち起終点情報に着目し、主成分分析を適用することによって、組み合わせ数が多い起終点情報をいくつかの特徴のある検索パターンに集約した。さらに、算出された主成分得点表に対して時系列分析を適用することにより、各検索パターンの曜日変動および月変動を求めることで対象地域における交通行動の特徴を明らかにした。

このように、経路検索システムの検索履歴データに着目した研究は進んでいるが、その数は少なく先行研究の蓄積は十分とはいえない。また、上述の研究では、検索件数が何件である場合に、乗降件数が具体的に何件に徙う可能性があるか、両者の関係性を分析するには至っていない。全国的に利用される経路検索システムには、通勤・通学などの日常的な利用だけでなく、出張・観光などの非日常的な利用を含めた検索履歴データが蓄積されている。そのため、検索履歴データから将来の乗降件数を予測できれば、これまでの交通需要分析が実際の行動に基づいて行われてきたことを踏まえると、実用的な意義は大きい。

本研究では、香川県で運行されている鉄道路線「ことでん」を対象に検索された全国版経路検索システムである NAVITIME の検索履歴データと、ことでんで使用されている交通系 IC カード「IruCa」の乗降履歴データから実際の交通需要を予測することを目的とする。具体的には、検索履歴データと乗降履歴データの相互依存性をコンピュータ関数により表現したモデルを用いて、検索履歴データから実際の乗降件数の予測を行う。



図-1 ことでん路線図⁸⁾

2. ことでんと使用データの概要

(1) ことでんの概要

本研究では、香川県で運行されている鉄道路線「高松琴平電気鉄道(以下、「ことでん」という)」を分析対象とする。ことでんは、図-1に示すように琴平線、長尾線、志度線の3路線全52駅が存在しており、高松市中心部から香川県の要所を放射線状に結んでいる。

(2) 使用データの概要

a) 「NAVITIME」検索履歴データ

本研究では、検索履歴データとしてことでんを対象に検索された経路検索システム「NAVITIME」のデータを用いる。検索履歴データの条件として、出発地と目的地の双方において、ことでんの駅が指定されること及び検索指定時刻がことでんの運行時間帯である5時台～24時台であることを課す。データの収集期間は2014年1月1日から2014年12月31日までの12ヶ月間であり、データ数は198,351件である。

図-2に経路検索システムの日別検索履歴件数の推移を示す。平均検索件数は約543件/日であり、最も検索件数が多かったのは8月13日の1,531件、最も検索件数が少なかったのは10月13日の193件である。

図-3に日別検索件数の分布を示す。図-3より、1日の検索件数が400～599件のときにピークが現れている。また、全365日(サンプル)のうち92.9%の検索件数は350～800件である。

b) 「IruCa」乗降履歴データ

本研究では、乗降履歴データとしてことでんで導入されている交通系 IC カード「IruCa」のデータを用いる。IruCaは、2016年時点で341,706枚発行⁹⁾されている。乗降履歴データの条件として、ことでんの運行時間帯である5時台から24時台に利用され、かつ移動所要時間が60秒以上であることを課す。2014年1月1日～2014年12月31日までの12ヶ月間に収集されたデータ数は7,265,029件である。

図-4に交通系 IC カードの日別乗降件数の推移を示す。平均乗降件数は約19,904件/日であり、最も乗降件数が多かったのは4月18日の28,378件、最も乗降件数が少なか

ったのは10月13日の1,623件である。

図-5に日別乗降件数の分布を示す。図-5より、1日の交通系 IC カードの乗降件数が 25,500~26,499 件の日（サンプル）の割合が 20.5%と最も多い。一方で7,500~8,499 件が 10.4%と多く、ピークが2つあることがわかる。

c) 検索件数と乗降件数の同時分布

図-6に日別の検索履歴件数と乗降履歴件数の同時分布を示す。図-6より乗降履歴件数約 15,000 件を境に2つの集団が存在していることがわかる。この2つの集団の特徴を把握するため、平日と休日（土日及び祝日）で区分した検索件数と乗降件数に着目する。

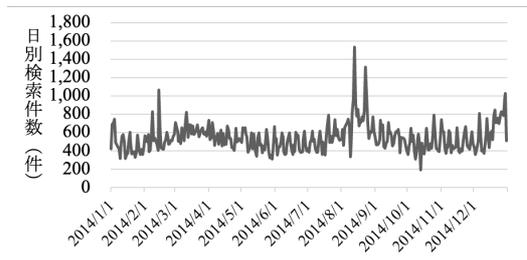


図-2 日別検索件数の推移 (NAVITIME)

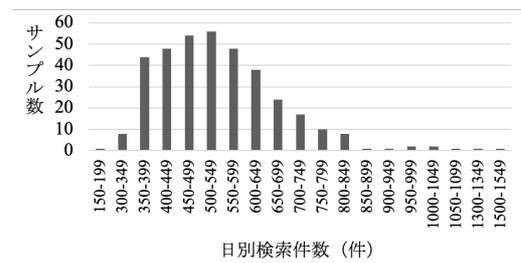


図-3 日別検索件数の分布

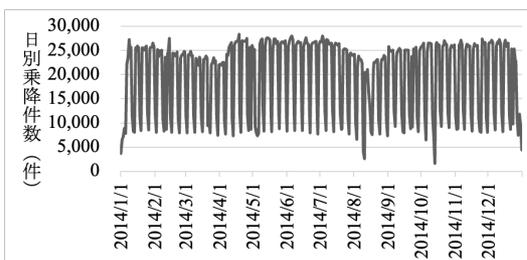


図-4 日別乗降件数の推移 (IruCa)

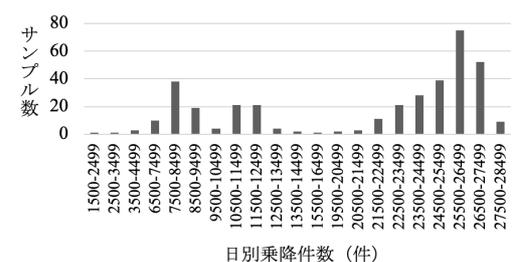


図-5 日別乗降件数の分布

図-7に平日の検索履歴件数と乗降履歴件数の同時分布を示す。平日では、検索件数が350~399件、乗降件数が25,500~26,499件の日が最も多いことがわかる。これは、全247日（サンプル）のうち24日である。また、分布図から、検索件数が少ない日ほど、乗降件数が多い傾向がみられる。平日は交通系 IC カードの利用者の通勤、通学などの日常的な利用により乗降件数も多い。一方、経路検索システムは、普段利用していない経路や時間帯の公共交通を利用するときに検索される。そのため、通勤や通学などの日常的な、パターン化された外出行動が多い平日には基本的に利用が少ないと考えられる。その結果、検索件数が少ない平日は日常的な外出行動の多い一般的な平日と考えられるため、乗降件数が多い傾向にあると考えられる。それとは反対に、検索件数が多くパターン化されていない外出行動が多い平日は、特異な日と考えられる。そのような日は平日の乗降件数の大部分を占める日常的な外出行動が少なく、乗降件数も少ないために、負の関係が検出されたと考えられる。

図-8に休日の検索履歴件と乗降履歴件数の同時分布を示す。休日では、検索件数が450~499件かつ乗降件数が7,500~8,499件の日、検索件数が550~599件かつ乗降件数が7,500~8,499件の日の割合が高いことがわかる。休日では検索件数の増加に伴い乗降件数も増加する傾向がみられ、相関係数も0.49とやや強い正の相関を持つ。経路検索システムの利用特性から、観光や出張、帰省などの非日常的な、パターン化されていない外出行動が多い休日には利用が多くなると考えられる。一方、休日は交通系 IC カードの利用者の通勤、通学などの日常的な利用が少ないため乗降件数も少ない。その結果、休日の乗降件数は非日常的な利用が多くを占めるため、検索件数が多い日に乗降件数も多くなるという正の関係が検出されたと考える。

以上の基礎分析から、検索件数の増減と乗降件数の増減の関係性は、平日と休日で異なることがわかった。そこで本研究では、正の相関が見られる休日データのみに着目し、分析を行う。

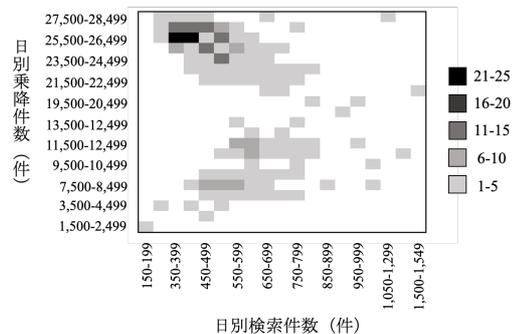


図-6 日別の検索件数と乗降件数の同時分布

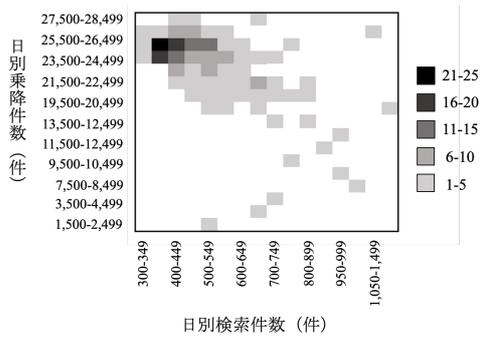


図-7 日別の検索件数と乗降件数の同時分布 (平日)

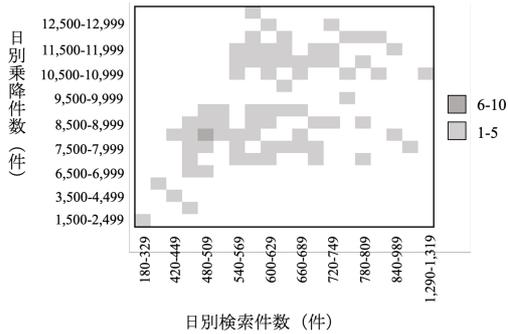


図-8 日別の検索件数と乗降件数の同時分布 (休日)

3. 潜在クラスモデルを用いた2変量コピュラ関数の定式化

(1) コピュラ関数

本研究では、経路検索履歴データと乗降履歴データの関係性を表現するために、コピュラ関数^{10,11)}を用いる。

2変量のデータ x_1, x_2 に対し、それらを確率変数 X_1, X_2 で表す。周辺分布関数を F_1, F_2 とすると、同時分布関数 $F(x_1, x_2)$ との間に、以下に示すスクラーの定理が成立する。

<スクラー (Sklar) の定理>

周辺分布関数 $F_1(x_1), F_2(x_2)$ を持つ連続な2変量同時分布関数を $F(x_1, x_2)$ とすると、次式を満たすコピュラ関数 C が一意に存在する。

$$P_r(X_1 \leq x_1, X_2 \leq x_2) = F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \quad (1)$$

スクラーの定理から、コピュラ関数 C に周辺分布関数 F_1, F_2 を適用することで生成される $C(F_1, F_2)$ は、周辺分布を区間 $[0,1]$ の一様分布とする同時分布関数である。つまり、連続な多変量分布関数は、それぞれのデータの挙動を表現する周辺分布関数 F_1, F_2 とデータ間の依存構造を表すコピュラ関数 C に分解が可能である。

スクラーの定理から2変量同時分布関数 F が周辺分布関数 F_1, F_2 とコピュラ関数 C を持つとすると、任意の

$u = (u_1, u_2)$ (ただし、 $u_i \in [0,1]$) に対し、式(2)に示す関係が成り立つ。また、この時の C は、各周辺分布が区間 $[0,1]$ の一様分布の同時分布関数となる。

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \quad (2)$$

変量間の複雑な依存関係を表現するために、様々な形でコピュラ関数の定式化が行われている¹⁰⁾。変量間の相関を行列で表現する正規コピュラとtコピュラや、変量間の相関構造を1種類のパラメータで表現する1パラメータ・アルキメディアン・コピュラ (クレイトン・コピュラ, ガンベル・コピュラ, フランク・コピュラ) などが代表的なコピュラ関数として多用されている。

本研究では、異なる依存構造を表現する4つのコピュラ、1) 正規コピュラ、2) クレイトン・コピュラ、3) ガンベル・コピュラ、4) フランク・コピュラを候補として挙げ、各コピュラについてモデルを推定し、情報量基準 (AIC) を用いて、最も当てはまりが良いコピュラを選択する。本研究で用いる4つのコピュラ関数を式(3)から式(6)に示す。

1) 正規コピュラ

正規コピュラは、変量間に多変量正規分布を仮定した場合と依存構造が一致するコピュラであり、式(3)で表される。ここでの $\phi_1(\bullet)$ は1変量標準正規分布関数、 $\phi_2(\bullet; \Sigma)$ は、相関行列 Σ を持つ2変量正規分布関数である。

$$C(u_1, u_2) = \phi_2(\phi_1^{-1}(u_1), \phi_2^{-1}(u_2); \Sigma) \quad (3)$$

ここで、相関係数が0のとき2変量は無相関である。また、正の絶対値が大きいほど2変量間の正の依存度が強く、負の絶対値が大きいほど負の依存度が強くなることを表す。

2) クレイトン・コピュラ

クレイトン・コピュラは、 $\alpha(\geq 0)$ をパラメータとして、式(4)で表される。

$$C(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}} \quad (4)$$

ここで、 α が0に近づくとき2変量は無相関となる。また、正の絶対値が大きいほど2変量間の正の依存度が強く、負の絶対値が大きいほど負の依存度が強くなることを表す。

3) ガンベル・コピュラ

ガンベル・コピュラは、 $\gamma(\geq 1)$ をパラメータとして、式(5)で表される。

$$C(u_1, u_2) = \exp\left\{-\left\{(-\ln u_1)^\gamma + (-\ln u_2)^\gamma\right\}^{\frac{1}{\gamma}}\right\} \quad (5)$$

ここで、 γ が1のとき2変量が無相関であり、正の絶対値が大きいかほど2変量間の正の依存度が強くなる。ただし、負の依存関係を表現することはできない。

4) フランク・コピュラ

フランク・コピュラは、 $\delta(\geq 0)$ をパラメータとして、式(6)で表される。

$$C(u_1, u_2) = -\frac{1}{\delta} \ln \left(1 + \frac{(e^{-\delta u_1} - 1)(e^{-\delta u_2} - 1)}{e^{-\delta} - 1} \right) \quad (6)$$

ここで、 γ が0に近づくとき2変量は無相関となる。また、正の絶対値が大きいかほど2変量間の正の依存度が強く、負の絶対値が大きいかほど負の依存度が強くなることを表す。

(2) コピュラのパラメータ推定

コピュラ・パラメータの推定手法として、最尤推定法が用いられる¹⁰⁾。独立な同時分布に従う N 個のデータに対し、第 i 変量の第 j 番目のデータを $x_i^j (i = 1, 2, j = 1, \dots, N)$ と表し、パラメータ ψ_i を持つ第 i 変量の周辺分布関数を $F_i(\cdot; \psi_i) (i = 1, 2)$ 、パラメータ α を持つコピュラを $C(\cdot; \alpha)$ とする。このとき、対数尤度関数は式(7)で表される。

$$l(\psi_1, \psi_2, \alpha) = \sum_{j=1}^N \ln C \left(\left(F_1(x_1^j; \psi_1), F_2(x_2^j; \psi_2) \right); \alpha \right) + \sum_{j=1}^N \sum_{i=1}^2 \ln f_i \left(F_i(x_i^j; \psi_i) \right) \quad (7)$$

(3) 潜在クラスを用いた2変量コピュラモデル

本研究では、検索履歴データと実際の乗降件数の関係性を表現するためにコピュラ関数を用いる。しかし、-8で示すように、休日の交通系ICカードの乗降件数には分布のピークが2つ存在しており、その形状は単峰ではない。すなわち、コピュラ関数を適用する上でサンプル全体を1つの分布系で表現することは不適切であることが予め予想できる。そこで、多峰分布を表現可能な手法として、潜在クラスモデルを適用する。

潜在クラス数が k であるとき、第 h クラスに属する帰属確率を $P_h (h = 1, \dots, k)$ 、第 h クラスを表現する2変量コピュラを $C(u_1, u_2)_h$ とすると、潜在クラスを考慮した2変量コピュラ関数は式(8)で表される。

$$C(u_1, u_2) = \sum_{h=1}^k P_h C(u_1, u_2)_h \quad (8)$$

帰属確率 P_h は、第 h クラスの対数尤度関数を式(7)を用いて $l(\psi_1, \psi_2, \alpha)_h$ とすると、式(9)で表される。

$$P_h = \frac{\exp(l(\psi_1, \psi_2, \alpha)_h)}{\sum_{h=1}^k \exp(l(\psi_1, \psi_2, \alpha)_h)} \quad (9)$$

式(9)の帰属確率 P_h を用いて、全ての潜在クラスに関わる対数尤度関数を式(10)で表す。本研究では、式(10)で表す対数尤度関数を最大化するパラメータを求めるため、EMアルゴリズムを適用する。

$$\ln L = \sum_{h=1}^k P_h l(\psi_1, \psi_2, \alpha)_h \quad (10)$$

4. 潜在クラスを用いたモデル推定結果と実行動予測

(1) 潜在クラスを用いた周辺分布の決定

本研究では、検索履歴データと乗降履歴データの依存構造を表現するための事前分析として、潜在クラスを用いてそれぞれのデータの周辺分布の適合度を比較する。周辺分布には、1) 正規分布、2) 指数分布、3) ポアソン分布、4) ワイブル分布の4つの確率分布を候補とする。モデル比較のための適合度指標には、最終対数尤度とパラメータ数から算出される Akaike's Information Criterion (AIC) を用いる。

検索履歴データの周辺分布の適合度の比較結果を表-1に示す。表-1 B) 潜在クラスを考慮する場合において、クラス1が行、クラス2が列に対応しており、 P_1 と P_2 はそれぞれのクラスに属する平均帰属確率を示している。表-1より、A) 潜在クラスを考慮しない場合と、B) 潜在クラスを考慮した場合のなかで、クラス1とクラス2の両方の周辺分布にワイブル分布を適用したときのAICが最小であった。

検索履歴データと同様に乗降履歴データの周辺分布の適合度の比較結果を表-2に示す。表-2より、A) 潜在クラスを考慮しない場合と、B) 潜在クラスを考慮した場合のなかで、クラス1とクラス2の両方の周辺分布にワイブル分布を適用したときのAICが最小であった。

適合度の比較結果から、経路検索システムの利用者の異質性を考慮する方がモデルの適合度が高くなることが明らかとなったため、本研究では両変量に対して潜在クラスを考慮し、周辺分布にワイブル分布を仮定する。

(2) 潜在クラスを用いたコピュラ関数の決定

本研究では、検索履歴データと乗降履歴データの依存構造の候補と設定した1) 正規コピュラ、2) クレイト

ン・コンピュータ, 3) ガンベル・コンピュータ, 4) フランク・コンピュータの4つのコンピュータ関数を用いたモデルの推定結果を比較する。なお, 事前分析の結果から, 周辺分布には潜在クラスを考慮したワイブル分布を仮定する。

はじめに, 潜在クラスを考慮しない場合の4種類のコンピュータを仮定したモデル推定結果を表-3に示す。表-3より, 潜在クラスを考慮しない場合では, クレイトン・コンピュータを適用したモデルのAICが最も小さくなった。

次に, 潜在クラスの数を2と設定した場合のAICと最終対数尤度, それぞれの潜在クラスに属する平均帰属確率 P_1, P_2 を表-4に示す。クラス1が行, クラス2が列にそれぞれ対応している。表-4より, クラス1にフランク・コンピュータ, クラス2にクレイトン・コンピュータを仮定するモデルのAICが最小となった。なお, クラス2にガンベル・コンピュータを仮定するモデルでは, パラメータを推計することができなかった。

潜在クラスの数を3に設定した場合のAICと最終対数

尤度, 平均帰属確率 P_1, P_2, P_3 の比較を表-5に示す。コンピュータの項目において上からそれぞれ, クラス1, クラス2, クラス3に対応し, 表-4の結果から, クラス1とクラス2には, それぞれフランク・コンピュータとクレイトン・コンピュータを仮定する。表-5より, 全ての結果で潜在クラスの数を2に設定したモデルよりもAIC値が大きくなり, モデルの適合度が低下した。なお, クラス3にガンベル・コンピュータを仮定したモデルでは, これまでと同様, パラメータを推定できなかった。以上の結果から, クラス数4以上は検証せず, 潜在クラスの数を2と設定し, フランク・コンピュータとクレイトン・コンピュータの組み合わせを最適なモデルとして選択する。

モデル適合度が最も高いと判断されたフランク・コンピュータ×クレイトン・コンピュータのモデルに対して, 推定されたパラメータ値から算出される検索履歴件数と乗降履歴件数の同時確率密度の分布を図-9に示す。

表-1 検索履歴データの周辺分布の比較

A) 潜在クラスを考慮しない場合

確率分布	AIC	lnL
1) 正規	-22.516	13.26
2) 指数	675.294	-336.67
3) ポアソン	441.405	-219.70
4) ワイブル	15.540	-5.77

B) 潜在クラスを考慮した場合

AIC lnL P_1 P_2	B) 潜在クラスを考慮した場合			
	1) 正規	2) 指数	3) ポアソン	4) ワイブル
1) 正規	-57.573 32.79 52.38% 47.62%	-36.571 21.29 93.89% 6.11%	-24.595 15.30 85.58% 14.42%	-67.691 37.85 49.11% 50.89%
2) 指数	-32.816 19.41 7.70% 92.30%	717.171 -356.59 60.57% 39.43%	506.702 -251.35 27.07% 72.93%	-16.549 11.28 7.77% 92.23%
3) ポアソン	-24.596 15.30 14.42% 85.58%	508.594 -252.30 80.53% 19.46%	443.405 -219.70 50.00% 50.00%	-9.972 7.99 15.45% 84.55%
4) ワイブル	-67.691 37.85 50.89% 49.11%	-20.972 13.49 93.95% 6.05%	-9.948 7.97 84.56% 15.44%	-79.115 43.56 46.71% 53.29%

表-2 乗降履歴データの周辺分布の比較

A) 潜在クラスを考慮しない場合

確率分布	AIC	lnL
1) 正規	46.062	-21.031
2) 指数	759.585	-378.79
3) ポアソン	482.874	-240.44
4) ワイブル	-5.870	4.94

B) 潜在クラスを考慮した場合

AIC lnL P_1 P_2	B) 潜在クラスを考慮した場合			
	1) 正規	2) 指数	3) ポアソン	4) ワイブル
1) 正規	8.044 -0.02 57.81% 42.19%	-11.387 8.69 93.54% 6.46%	2.535 1.73 87.46% 12.54%	-147.943 77.97 41.68% 58.32%
2) 指数	-11.387 8.69 6.46% 93.54%	761.585 -378.79 50.00% 50.00%	550.258 -273.13 27.07% 76.36%	-0.412 3.21 6.13% 93.87%
3) ポアソン	2.952 1.52 12.59% 87.41%	550.259 -273.13 76.36% 23.64%	484.874 -240.44 50.00% 50.00%	13.774 -3.89 12.76% 87.24%
4) ワイブル	-147.944 77.97 58.32% 41.68%	-0.412 3.21 93.87% 6.13%	13.331 -3.67 87.28% 12.72%	-170.535 89.27 43.31% 56.69%

また、図-9を見やすくするために、 u_1 (検索件数) の範囲を [6.0, 7.0] と u_2 (乗降件数) の範囲を [8.5, 9.5] に定め、図-9を拡大した図を図-10に示す。図-9、図-10にある2つの分布の内、上側の分布がクラス1に属するフランク・コピュラを、下側の分布がクラス2に属するクレイトン・コピュラを表している。

クラス1に属するフランク・コピュラの同時確率分布の形状は、分布の中心から確率密度等高線が左右に広がっている。また、中心から右側の方が等高線の間隔が狭く、左側に比べて依存度合いが高いことを示している。検索履歴件数が6.5~6.7, 乗降履歴件数が9.3~9.4あたりで確率密度の値が最も高い。

クラス2に属するクレイトン・コピュラの同時確率分布の形状は、分布の左下側と右上側で確率密度等高線が広く、分布の中心で狭い。検索履歴件数が6.4~6.5, 乗降履歴件数が9.0あたりで確率密度の値が最も高く、全体的に正の相関が見られる。

表-3 潜在クラスを考慮しないコピュラの比較

コピュラ	AIC	lnL
1) 正規	637.2149	-309.6074
2) クレイトン	-26.831	18.42
3) ガンベル	-3.109	6.55
4) フランク	-26.780	18.39

表-4 潜在クラスを考慮したコピュラの比較

AIC lnL P_1 P_2	1) 正規	2) クレイトン	3) ガンベル	4) フランク
1) 正規	712.970 -338.49 86.46% 13.54%	14.341 6.83 7.62% 92.38%	NaN	60.546 -12.27 7.40% 92.60%
2) クレイトン	3.988 12.01 87.29% 12.71%	-48.232 34.12 48.26% 51.74%	NaN	2.632 8.68 48.94% 51.06%
3) ガンベル	36.874 -4.44 92.08% 7.92%	-68.280 44.14 49.30% 50.70%	NaN	-42.745 31.37 46.33% 53.67%
4) フランク	21.247 3.38 91.86% 8.14%	-76.557 48.28 50.51% 49.49%	NaN	-6.776 13.39 49.81% 50.19%

表-5 潜在クラス別でのAICと最終対数尤度の比較

コピュラ	AIC lnL $P_1 : P_2 : P_3$
フランク クレイトン 正規	4.948 16.53 40% : 46% : 14%
フランク クレイトン クレイトン	-55.409 42.70 32% : 36% : 32%
フランク クレイトン ガンベル	NaN
フランク クレイトン フランク	-41.354 35.68 32% : 36% : 32%

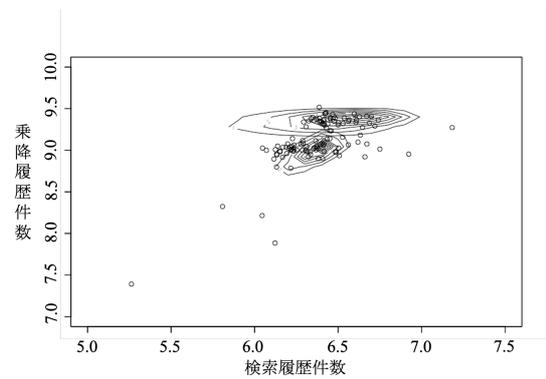


図-9 選択したコピュラの同時密度関数

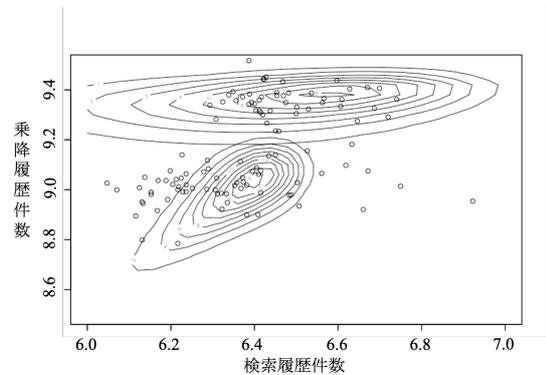


図-10 選択したコピュラの同時密度関数 (拡大図)

(3) 経路検索件数による乗降件数の予測結果

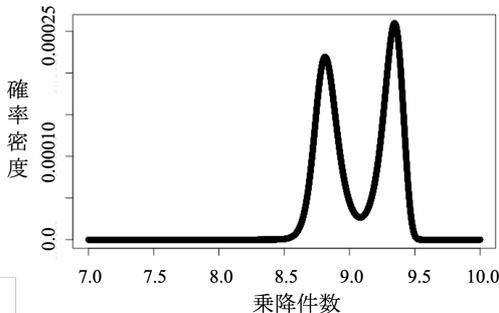
検索件数と乗降件数の間の依存関係に、フランク・コピュラとクレイトン・コピュラを仮定したモデルの推定結果を用いて、1) 検索件数を6.2とした場合、2) 検索件数を6.4とした場合、3) 検索件数を6.6とした場合の乗降件数の確率密度を図-11に示す。なお、ここでの検索件数は対数値であるため、スケールを戻すと、それぞれ検索件数が493件、602件、735件のときに対応する。

確率密度の比較結果から、検索件数によって2つの分布のピーク離れ度合いが異なることが明らかとなった。また、検索件数の増加に伴い2つの確率密度が接近する

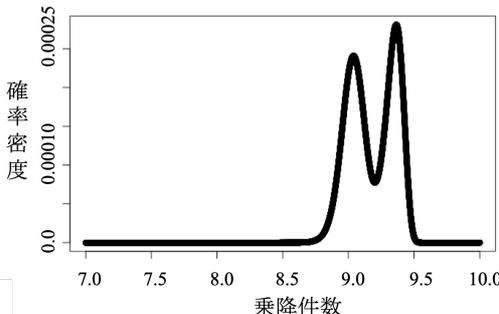
が、検索件数が 6.4 以降ではほとんど変化しないことがわかる。

図-11 に示す確率密度から得られた乗降件数の期待値と 95%信頼区間の関係性を、図-12 のグラフに示す。図-12 では、それぞれ、1) 元の検索履歴データを対数変換した場合での期待値と 95%信頼区間、2) 「1)」による推定値を指数変換し、実際の乗降件数のスケールに戻した期待値と 95%信頼区間を示している。1) に着目すると、検索履歴データの平均値である 6.38 を境に、期待値の増加の傾きが変化することがわかる。また、95%信頼区間の下側 2.5%点と上側 2.5%点に着目すると、6.3 に近づくにつれて信頼区間の幅が狭まり、それ以降ほとんど横ばいになっている。これは、図-10 で示すように、検索履歴件数の増加に伴い 2 つの確率密度が接近し、検索履歴件数が 6.3 以上の場合、確率密度の形状がほとんど変化しないためである。2) も同様に、検索履歴データの平均値である 602.63 件を境に乗降件数の期待値の傾きが変化しており、全体的に 1) とほとんど同じ形状となっている。

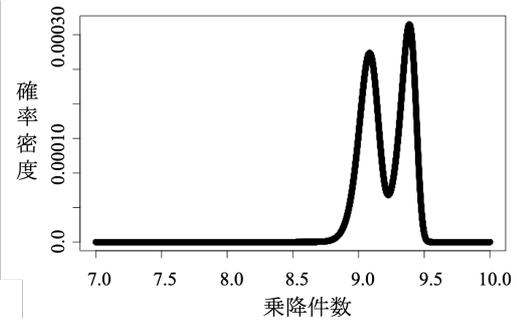
以上の結果から、検索履歴件数の増加に伴い乗降件数の期待値も増加することがわかる。また、検索履歴データの平均値である 6.38 以下では、推定される乗降件数の 95%信頼区間の幅が広く、モデルの予測精度が低い。一方で、6.38 以上では乗降件数と 95%信頼区間ともにほとんど横ばいで推移しており、モデルの予測精度が高いことがわかった。



1) 検索件数が 6.2 の乗降件数の確率密度

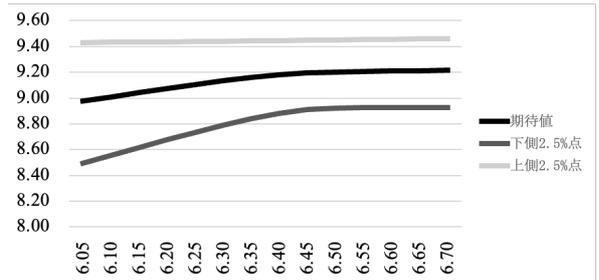


2) 検索件数が 6.4 の乗降件数の確率密度

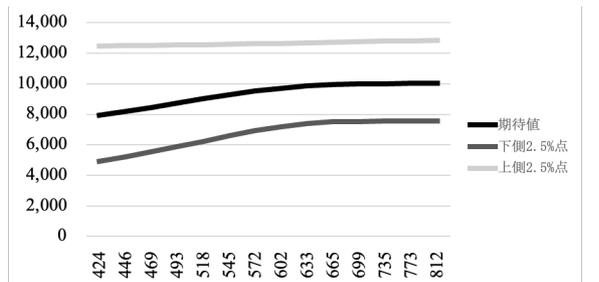


3) 検索件数が 6.6 の乗降件数の確率密度

図-11 検索件数別での乗降件数の確率密度



1) 対数変換後の期待値と 95%信頼区間



2) 指数変換後の期待値と 95%信頼区間

図-12 乗降件数の期待値と 95%信頼区間

5. 結論

本研究では、2014 年度を対象期間とし、経路検索システム「NAVITIME」の検索履歴データと鉄道路線ことでんで使用されている IC カード「IruCa」の乗降履歴データを用いて、それぞれのデータの関係性を分析した。両者の関係性を概観すると、平日と休日（土日および祝日）では異なる依存関係が存在することが明らかとなった。これは、公共交通を普段とは異なる目的で利用するときに使われるという経路検索システムの特徴が反映されているためと考えられ、平日の乗降件数を検索性から直接予測することは困難であることが示唆された。

そこで、本研究では休日にのみ着目し、検索性に基づく乗降件数の予測を試みた。関係性を分析するにあたり、データ間の依存関係を柔軟に表現できるコンピュータ関数を用い、検索履歴データと乗降履歴データが異なる周辺分布に従う場合に、両者の相互依存性を明示的に考慮した同時決定モデルを構築した。分析から得られたモデ

ルを用いて、任意の検索件数を与え、乗降件数の予測を行なったところ、検索件数の増加に伴い乗降件数の期待値も増加する結果が得られた。乗降予測件数の期待値と 95%信頼区間に着目すると、検索履歴データの平均値付近を境に期待値と 95%信頼区間の両方でグラフの傾きが変わっており、検索件数が平均値以上のときに、より良い精度で乗降件数を予測できることがわかった。

検索履歴データから交通需要を分析した研究は存在するが、具体的な乗降件数を予測した研究は少ない。本研究によって、経路検索履歴データと乗降履歴データの相互依存関係がコンピュータ関数によって表現可能である点や、検索件数が平均値に近づくにつれ 95%信頼区間の幅が狭まることが明らかになることが示唆された。これらの成果は、今後の経路検索履歴データを用いた交通需要予測手法の発展に寄与すると期待できる。

本研究では 2014 年度の土日および祝日のみを分析の対象期間として乗降件数の予測を行なったが、乗降件数の 95%信頼区間の幅が広く、十分に予測できているとはいえない。そのため、今後の重要な課題として、対象期間を延長することによってサンプル数を増やす、あるいはことடன்全体での関係性ではなく路線や起終点を限定するなど、予測精度を向上させるための方法について検討する必要がある。

参考文献

- 1) 総務省：令和 2 年度版 情報通信白書，
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/02honpen.pdf>
- 2) 大野俊典，上野山勝也，松尾豊：消費に関する検索

- 共起語から市場を観測する手法に関する研究，人工知能学会全国大会論文集，Vol.JSAI2017，4pages，2017.
- 3) 松尾豊：ウェブから実世界の観測と予測，電子情報通信学会論文誌，Vol.J96-B，No.12，pp.1309-1315，2013.
- 4) 内閣府：公共交通に関する世論調査，
<https://survey.gov-online.go.jp/h28/h28-kotsu/2-2.html>，2016.
- 5) 伊藤昌毅，見生元気，川村尚生，菅原一孔：乗換案内サービス利用履歴から探る公共交通の利用状況，第 9 回電子情報通信学会ヒューマンプロープ研究会，CD-ROM，2013
- 6) 石丸怜美，梶原康至，太田恒平：乗換検索サービスの経路選択データを用いた公共交通の経路選択行動分析，土木計画学会・講演集，Vol.49，CD-ROM，2014.
- 7) 香川喬之，桑野将司，福山敬，谷本圭志，川村尚生：バス経路検索履歴データを用いた移動特性の分析，交通工学論文集，第 2 巻，第 2 号（特集号 A），pp.A_115-A_124，2016.
- 8) ことடன்グループ：ことடன்路線図，
<http://www.kotoden.co.jp/publichtm/kotoden/time/>
- 9) 高松市民政策局交通政策課：コンパクト・プラス・ネットワークのまちづくり～ICカードを活用した公共交通利用促進施策の展開～，
http://www.estfukyu.jp/pdf/2016cyugoku/02_takamatsu.pdf
- 10) 戸坂凡展，吉羽要直：コンピュータの金融実務での具体的な活用方法の解説，日本銀行金融研究所，金融研究，pp115-162，2005.
- 11) 桑野将司，藤原章正，塚井誠人，張峻屹，岩本真由子：コンピュータを用いた自動車保有期間と走行距離の同時決定モデルの開発，土木学会論文集 D，Vol.66，No. 1，pp. 54-63，2010.

A COPULA-BASED RELEVANCE ANALYSIS OF ROUTE SEARCH LOG DATA AND PUBLIC TRANSIT BOARDING HISTORY DATA

Ryo KONISHI, Masashi KUWANO, Taku MORIYAMA and Mio HOSOE

Focusing on the route search log data of “NAVITIME” and public transit boarding history data of “IruCa” used on the Kotoden train in Kagawa prefecture, Japan, the purpose of this study was constructing model which is able to predict actual number of boarding and alighting. Analyzing these data by Copula-based Multivariate Latent Class model, the use of Flank copula in latent class 1 and Clayton copula in latent class 2 was supported. The estimated copula model indicated possibility of predicting expectation and variance from route search log data.