図書館情報メディア研究科修士論文

系列マイニングによる
スポットパス抽出に関する研究

2013 年　3 月

201121745

渡邊　直人

目次

図目次

表目次

第 1 章

序論

1.1 背景

近年，ユーザが自身の訪れたスポットに対してチェックインをする，ロケーションベース SNS (Location-Based Social Network Services, 以下 LBSNS) が普及してきている．LBSNS の代表的なサービスの 1 つである Foursquare[*1]は，2009 年に米国でリリースされ，2012 年 4 月現在で，ユーザ数は全世界で 2,500 万人を突破し，チェックイン数は累計 25 億回を超えている[*2]．日本での利用も盛んであり，東京でのチェックイン数が米国ニューヨークよりも多いという調査もある[*3]．

ユーザの利用目的は，チェックインをすることで得られるサービス内での称号獲得，訪れたスポットを記録するライフログとしての利用，友人との情報共有など様々である．また，Foursquare 上でのチェックインは，Twitter[*4]や Facebook[*5]，Mixi[*6]など既存の SNS サービスとの連携が可能なことも，Foursquare の利用を促進している要因と考えられる．

チェックインデータは，店舗名や所在地（緯度・経度情報），店舗の業種・業態を表すカテゴリ情報など，スポットに関する情報だけでなく，チェックインをしたユーザに関する情報も含んでおり，Foursquare のサーバ上に蓄積されている．チェックインデータから観測される，誰がいつ，どこにいるのか，次にどこへ行くのかという情報は，ユーザ行動をモデリングする上で有用である．例えば，雑貨店を訪れたユーザは，次にカフェに訪れる傾向があるということが分かれば，雑貨店にカフェの割引クーポンを置くなどのマーケティングが行える．また，ユーザの移動軌跡を利用した行動ナビゲーションに関する研究も盛んである．

[*1] Foursquare, https://ja.foursquare.com/
[*2] http://markezine.jp/article/detail/16431
[*3] http://scobleizer.com/2010/05/16/the-king-and-queen-of-location-based-services/
[*4] Twitter, https://twitter.com/
[*5] Facebook, https://www.facebook.com/
[*6] Mixi, http://mixi.jp/

本研究では，LBSNS のチェックインデータを対象として，系列マイニングの手法を適用し，頻出する移動軌跡をスポットパスとして抽出する．これによりユーザがどのようなスポットを訪れ，次にどこへ向かうのか，その特徴を明らかにすることを目的とする．LBSNS の代表的なサービスである Foursquare から実際のチェックインデータを収集し，24 時間を単位としたチェックインのシーケンスをユーザ毎に作成する．得られた個々人のシーケンスに対し，系列マイニングのアルゴリズムである PrefixSpan を用いることで，典型的なスポットパスを抽出する実験を行う．

1.2　本論文の構成

本論文の構成は次の通りである．2 章では，関連研究と本研究の位置づけについて述べる．3 章でスポットパス抽出のための，系列マイニングについて説明する．4 章でチェックインのシーケンス，スポットパスについて分析を行い，5 章で考察を行う．6 章で論文のまとめと今後の課題について述べる．

第 2 章

関連研究

本研究では，LBSNS の代表的なサービスである Foursquare に蓄積されたチェックインデータを対象に，系列マイニングを適用することでユーザが訪問したスポットのシーケンスであるスポットパスの抽出する．本研究に関連する，スポット推薦に関する研究，系列マイニングを利用した研究について概観することで，本研究の位置付けを明確にする．

2.1 スポット推薦に関する研究

篠田ら [1] は，ユーザが訪れたスポットと，そこに滞在している時間に着目し，これらを行動特性の素性として，類似するユーザの情報を参照することで，ユーザが興味を示すと思われるスポットを推薦する方法を提案している．zhijun ら [2] は，写真共有サービス Flickr[*1]に投稿される写真に付与されたジオタグ (位置情報) を利用し，世界の 12 都市での旅行ルートの作成を試みている．Choudhury ら [3] も Flickr の写真のジオタグを利用し，ユーザ毎のアップロードした写真の撮影時間の間隔から，２地点のルートの長さを定義して，頻出する旅行ルートの抽出を試みている．一般ユーザによる手作業で作成されたルートと，提案手法により自動抽出されたルートの比較を行い，スポット推薦をするための良好な結果が得られたと報告している．Kurashima ら [4] は，トピックモデルと隠れマルコフモデルを適用したスポット推薦方法を提案し，Flickr のデータを用いて提案法の有効性を検証している．cho ら [5] は，ユーザの行動モデルを明らかにすることを目的とし，LBSNS のデータ分析をしている．その結果，ユーザの行動モデルには個々人の行動パターンの他，友人関係も影響を及ぼしていることを報告している．

[*1] Flickr, http://www.flickr.com/

2.2　系列マイニングの利用に関する研究

旭ら [5] は，ブログから様々な人間の行動連鎖を系列マイニングによって抽出し，ユーザに抽出したシーケンスを提示する検索システムを作成している．例えば，結婚式に関するクエリを入力すると，結婚式での流れを理解できるとともに，結婚式に関係するクエリが表示される仕組みになっている．郡ら [6] は，ブログ記事からスポット名を抽出してシーケンスを生成し，スポット名とそのルート，スポット名に関連するキーワードをマップ上に表示する，旅行プラン作成補助を目指したシステムを作成している．筱島ら [7] は，商品レビューを評価するために，レビュー文章を分単位に分割し，文単位での評価の並びをシーケンスとして，商品レビューに付与されている評価値とシーケンスとの関係を調査している．山田ら [9] は，プログラミングのコードに対して系列マイニングを適用している．オブジェクト，メソッド，引数などをシーケンスとしてマイニングすることで，プログラムの欠陥を発見する手法を提案している．

2.3　本研究の位置付け

本研究では，代表的な LBSNS である Foursquare の実データに対して，系列マイニングを適用することで，多くのユーザが訪れる共通のスポットパスを抽出し，行動モデルの分析を試みる．LBSNS は，写真共有サイト Flickr に写真をアップすることと比較して簡便に利用できることから，ユーザは日々の生活での利用が多いと考えられる．そのため，抽出されるスポットパスも，日々の生活を直接的に反映しているものと思われる．

第 3 章

系列マイニング

ユーザが日々の生活の中でスポットを訪れるにあたり，スポットを訪問する順序も重要な要素であると考えられる．例えば，居酒屋でお酒を飲んだ後，締めのラーメンを食べに行く，ショッピングモールで買い物をして，歩き疲れたらカフェで休憩するなどが挙げられる．そこで本研究では，スポットの訪れる順序を考慮してシーケンスを抽出することができる系列マイニング（sequential pattern mining）を用いる．本章では，系列マイニングの定義と具体的なアルゴリズムである PrefixSpan について説明する．

3.1 系列マイニングの定義

系列マイニングとは，大量の系列データ (sequence data) を蓄積した系列データベース (sequence database) の中から，系列情報を保持したまま頻出する部分シーケンス (sequential pattern) を抽出する手法である．

表 3.1 系列データベース

	TID				
SID	1	2	3	4	5
1	A	B	C	D	B
2	A	B	C	A	C
3	B	C	A	A	
4	D	A	A	C	
5	C	A	B	C	B

系列データベースに格納される系列データは，一般に表 3.1 のような構造を持つ．系列データベースは $SDB = \{ s_1, s_2, \cdots, s_n \}$と表記される．ここで，$s_k$はシーケンスであり，それぞれ

シーケンスには識別子 (SID) が付与されている．シーケンス s は，$s = <i_1, i_2, \cdots, i_n>$と表記され，スポットとなるアイテム ($item$) の識別子 ($TID$) は，シーケンス中の何番目に出現するかを表している．

シーケンスの長さは，構成するアイテムの個数によって定義され，シーケンス α が k 個のアイテムで構成されているとき，シーケンス α の長さは k とする．また，同一シーケンスの出現回数は支持度と定義される．系列マイニングの処理開始時に，抽出するシーケンスの最小支持度 (minimum support) を設定し，系列データベース内で長さ k 以上のシーケンス α の支持度を調べ，最小支持度を満たすシーケンスを抽出する．表 3.1 に示す系列データベースに対して、最小支持度 k を 3 として，長さ 2 以上の系列を抽出した結果は、表 3.2 となる．

表 3.2　系列マイニングの結果

SID	部分シーケンス
1	A C
2	B C
3	C A

3.2　PrefixSpan

系列データベースから，与えられた長さ（支持度）以上の部分シーケンスを抽出するためのアルゴリズムを説明する．本研究では，pei[10] らによって提案された PrefixSpan アルゴリズムを用いて部分シーケンスを抽出する．系列データベースに対して Prefix projection と呼ばれる射影を行い、それによって生成される Prefix database を探索することで、候補となるシーケンスを生成することなく高速な処理を実現している．Prefix projection と Prefix database ついて述べた後、具体的な処理手順について述べる．

3.2.1　Prefix prejection と Prefix database

Prefix projection では，射影元のシーケンスから射影対象より後ろに存在するシーケンスのみを抽出する．例えば、表 3.1 に対し、アイテム B について射影した場合、表 3.3 のようにアイテム B より後に存在する部分列がシーケンスとして抽出され、生成されるデータベースが Prefix database となる．

3.2.2　PrefixSpan の処理手順

PrefixSpan の処理手順は以下の通りである．

表 3.3　アイテム B に対する Prefix database

SID			
1	C	D	B
2	C	A	C
3	C	A	A
4	C	B	

1　長さ 1 の頻出シーケンスを抽出する

2　深さ優先探索で Prefix projection と Prefix database のマイニングを繰り返す

表 3.4 のような系列データベースについて，PrefixSpan の具体的な処理手順を，図 3.1 を用いて説明する．最小支持度 k を 3 として，長さ 1 以上のシーケンスを抽出するとする．

表 3.4　系列データベース

SID			
1	A	C	D
2	A	B	C
3	A	A	B

まず，長さ 1 の頻出シーケンスを抽出する．その結果として、<A>:4，<B>:3，<C>:3，<D>:1 が抽出される．各シーケンスに付与されている数値は支持度であり，<A>:4 は，系列データベース内にシーケンス<A>が 4 つ存在していることを示している．<A>，<B>，<C>については支持度が 2 以上であるので，長さ 1 のシーケンスとして抽出される．<D>は支持度が 1 であるため不適となる．次から深さ優先探索で射影をしていく．まず，<A>について射影を行い，データベースを構築する．ここでは，<CD>，<BC>，<AB>，<B>が抽出される．このデータベースに対して支持度を確認すると，<A>:1，<B>:3，<C>:2，<D>:1 となる．ここで，支持度が 2 以上である<B>と<C>は，<A>と組み合わせた<AB>，<AC>が長さ 2 のシーケンスとして抽出される．続いて<AB>について射影するをすると，抽出されるシーケンスとその支持度は<ABC>:1 となり，最小支持度 2 を満たさないので不適である．<AC>について射影すると，抽出されるシーケンスとその支持度は<ACD>:1 となり，これも不適である．これで<A>から始まるシーケンスの抽出の処理が終わる．続いて<B>について射影を行うと，射影されたシーケンスは<A>:1，<C>:1 となり，共に支持度が 1 であるため不適である．最後に<C>について射影を行うと，<CA>:1，<CB>:1，<CD>:1 として抽出されるが，いずれも支持度が 1 であるために不適である．以上の流れから，表 3.5 が結果として得られる．

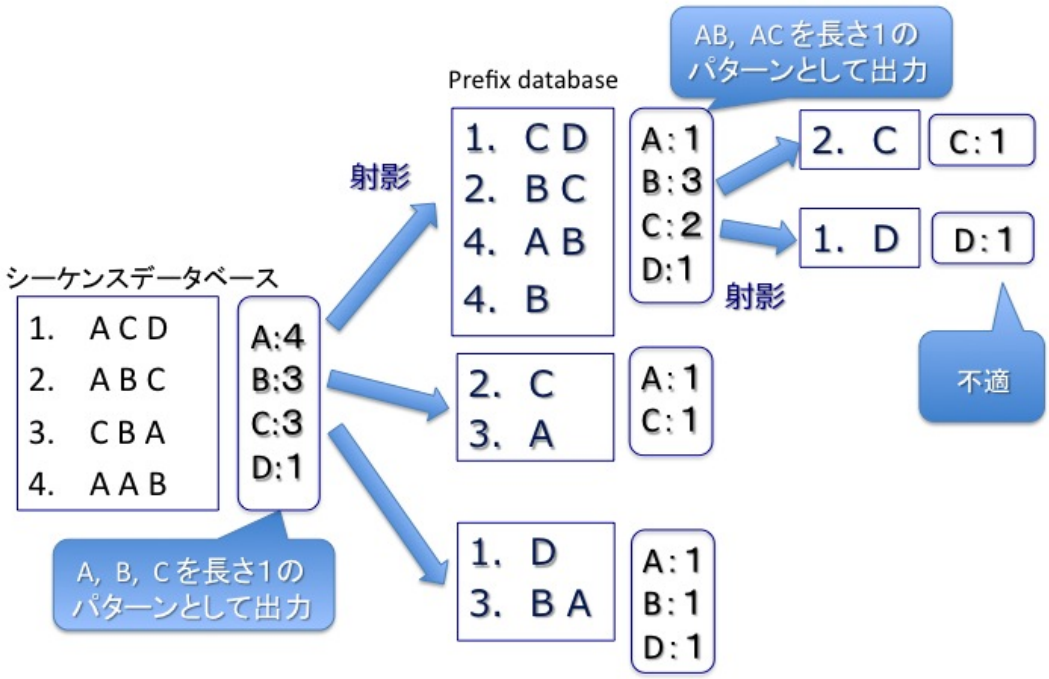

図 3.1　PrefixSpan の概要

表 3.5　PrefixSpan による抽出結果

SID	
1	A
2	A B
3	A C
4	B
5	C

第 4 章

系列データの分析

本研究では，実際の LBSNS のデータを分析する．4.1 節で分析に用いるチェックインデータについて説明し，4.2 節でスポットパスの抽出結果について分析を行う．

4.1　分析に使用するチェックインデータ

本研究ではユーザ数やサービスの普及度合いから，Foursquare のチェックインデータを分析に使用する．データセットの基本的な情報を表 4.1 に示す．

表 4.1　分析のデータセット

項目名	
期間	2012/06/15 - 2012/06/30
チェックイン数	419,349
ユーザ数	45,447
スポット数	135,655
カテゴリ数	394

Foursquare では，既に登録されているスポットに対してチェックインをするか，新たにスポットを登録してチェックインすることも可能である．スポットには，スポット名，カテゴリ，緯度経度，国名，住所などの情報項目があり，ユーザによって情報が付与されている．カテゴリは Foursquare によって候補決められており，現時点では 400 種類以上のカテゴリが存在している．

ユーザのチェックイン数

ユーザあたりのチェックイン数を示したのが図 4.1 であり，ユーザによってその利用頻度に差があることが確認できる．このデータを両対数グラフで表したのが図 4.2 であり，べき乗分布に近いことが確認できる．

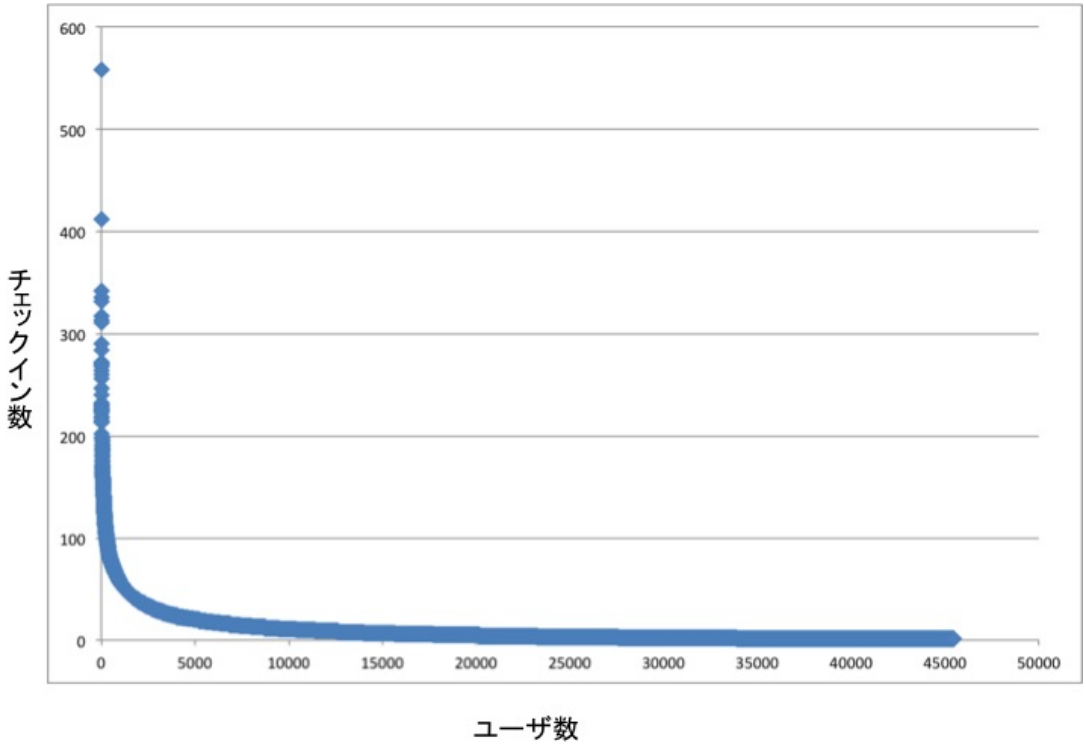

図 4.1　ユーザのチェックイン数

チェックインされるスポット

表 4.2 はチェックインの多い上位 30 位のスポットである．25/30 を都市部の鉄道駅が占めている．その他ランクインしている ”ヨドバシカメラ マルチメディア Akiba” は東京に存在する大型電器店である．”渋谷ヒカリエ” は，2012 年春に渋谷にオープンした大型商業施設兼御オフィスビル施設である．”羽田空港 第 1，第 2 ターミナル” と北海道の ”新千歳空港” と空港もランクインしている．このようにチェックインされるスポットは鉄道駅や空港などの交通機関，話題となるスポットが多いことがわかる．

スポットを個別に観測するには，データがスパースであることから，ここから先はカテゴリに着目して分析を進めていく．チェックイン の多いカテゴリ上位 30 件を示したのが表 4.3 で

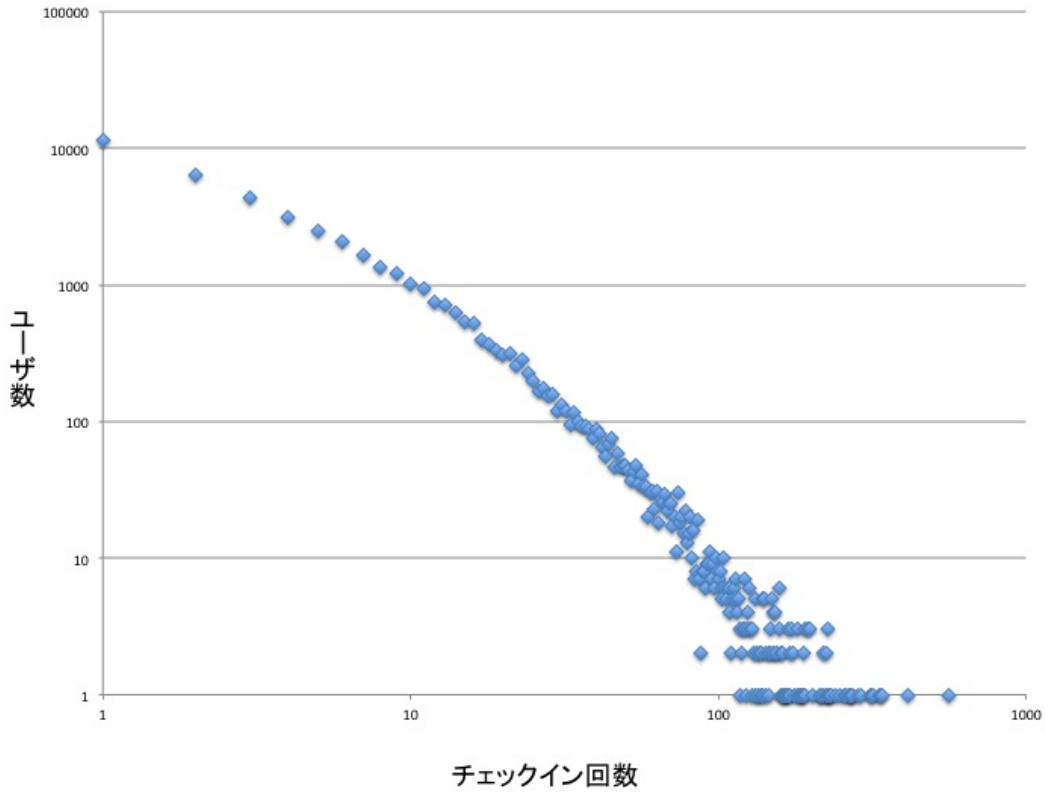

図 4.2　チェックイン数のべき乗分布

ある．鉄道駅が非常に多く，全体の約 25% を占めていることがわかる．また同様に交通機関で地下鉄，バス停，空港，ターミナルもランクインしている．飲食系のカテゴリとしてラーメン，和食，カフェ，コーヒーショップなどがランクインをしている．生活用品を買うためのコンビニエンスストアやショッピングモール，スーパーマーケットもランクインしている．オフィスや大学なども件数が多いことから，社会人や学生のユーザが多いとも考えられる．公園やゲームセンターなど他のカテゴリに比べ人気を想像し難いカテゴリもランクインしている．

チェックインの時間帯

次にチェックインがされる時間帯を調べた．データセットのチェックインデータ数を 1 時間単位として表したのが図 4.3 である．9 時前後，13 時前後，19 時前後でチェックインが多くされていることがわかる．また，カテゴリによるチェックインのピークの時間帯を調査した．チェックインの多いカテゴリ上位 3 つの鉄道駅，ラーメン，コンビニエンスストアを調べた結果が図 4.4，図 4.5，図 4.6 である．鉄道駅は 9 時前後と 19 時前後でのチェックインが多いことがわかる．一方、ラーメンでは 13 時台にチェックインが集中していて，その後 21 時前後でのチェックインが多いことがわかる．コンビニエンスストアでは，9 時前後，13 時台，19 時

表 4.2　チェックイン数上位 30 位のスポット

No	スポット名	件数
1	秋葉原駅	2,583
2	新宿駅	2,163
3	東京駅	1,946
4	渋谷駅	1,347
5	横浜駅	1,177
6	池袋駅	1,166
7	大阪駅	1,026
8	大宮駅	885
9	ヨドバシカメラ マルチメディア Akiba	800
10	川崎駅	799
11	品川駅	688
12	名古屋駅	656
13	阪急 梅田駅	614
14	京都駅	611
15	新横浜駅	581
16	仙台駅	569
17	羽田空港 第 2 旅客ターミナル	552
18	北千住駅	545
19	札幌駅	529
20	立川駅	504
21	上野駅	491
22	新大阪駅	489
23	渋谷ヒカリエ	482
24	調布駅	468
25	千葉駅	424
26	新橋駅	423
27	東急 東横線 渋谷駅	379
28	羽田空港 第 1 旅客ターミナル	372
29	新千歳空港	368
30	三鷹駅	350

表 4.3　チェックイン数上位 30 位のカテゴリ

No	カテゴリ名	件数
1	鉄道駅	102,899
2	ラーメン	16,384
3	コンビニエンスストア	16,307
4	和食	13,553
5	地下鉄	12,902
6	ショッピングモール	11,982
7	スーパーマーケット	8,329
8	カフェ	7,936
9	電器店	7,358
10	コーヒーショップ	6,975
11	サービスエリア	6,496
12	道路	5,557
13	書店	5,453
14	公園	5,135
15	ファーストフード	5,123
16	バス停	4,747
17	ゲームセンター	4,479
18	オフィス	4,371
19	橋	4,161
20	大学	4,039
21	デパート	3,998
22	中華料理	3,929
23	プラットフォーム	3,842
24	ホテル	3,679
25	居酒屋	3,408
26	バー	3,207
27	イタリア料理	3,164
28	ホビーショップ	3,076
29	空港	2,982
30	ターミナル	2,791

前後にそれぞれピークがあることがわかる．このように、スポットのカテゴリによってチェックインされる時間帯の傾向が異なることが判明した．

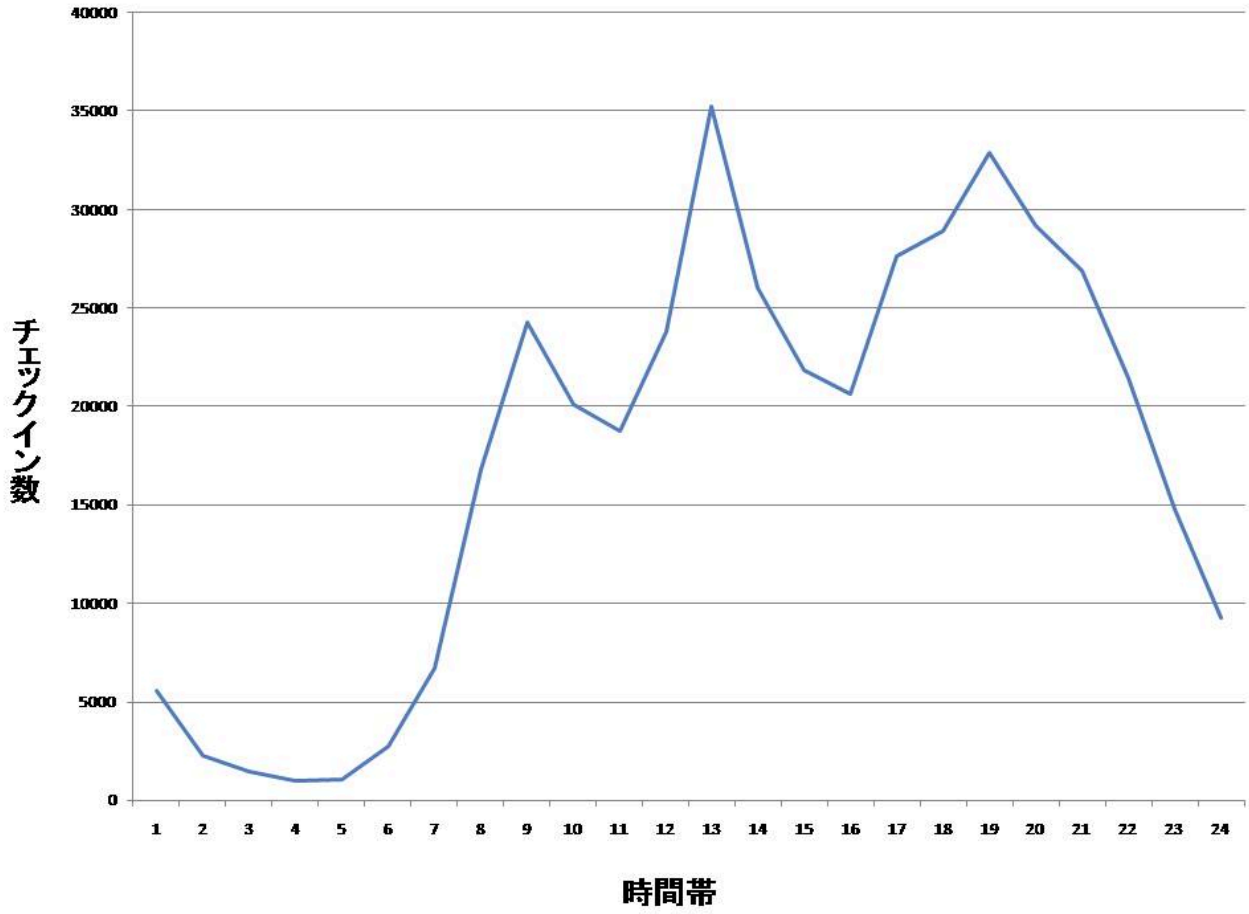

図 4.3　チェックインの時間帯

チェックインシーケンス

次にチェックインデータからシーケンスを作成し，その特徴を調べた．ユーザ毎に 00:00 から 23:59 までの時間のチェックインを，１シーケンスとして作成する. その結果，シーケンス数と平均シーケンス長は表 4.4 の通りであった．この結果から，あるユーザがチェックインを１回以上する日には，１日のうち平均は 2.29 回チェックインすることが判明した．

表 4.7 はシーケンス長の分布である．スポットパス抽出が可能となる，シーケンス長 2 以上のシーケンスの割合は半数近い．軸を対数で表した 4.8 から，シーケンス長の分布もべき乗分布に近いことがわかる．

ここで，シーケンスに含まれるカテゴリに着目し，表 4.3 のカテゴリのスポットへのチェッ

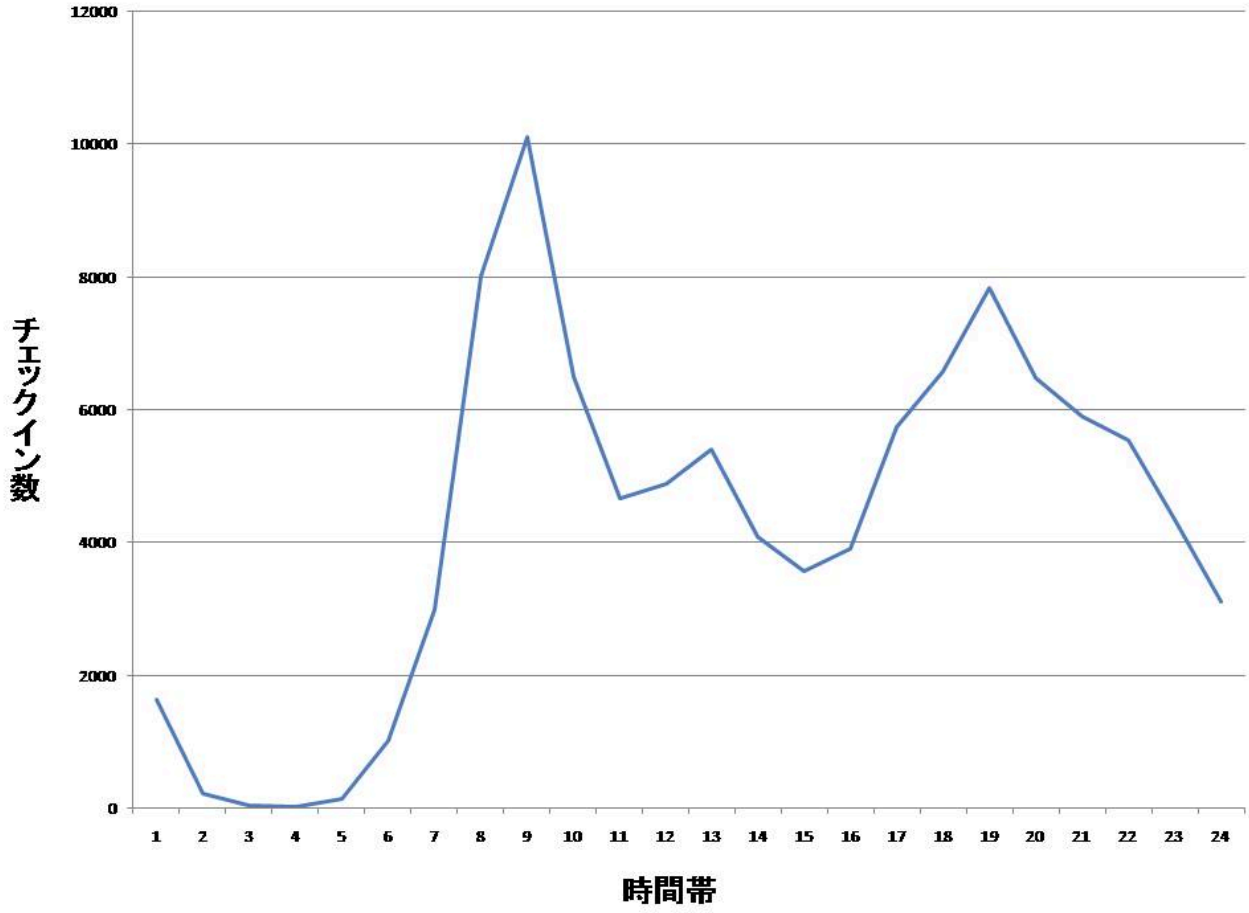

図 4.4　鉄道駅のチェックイン時間帯

表 4.4　シーケンスデータ

シーケンス種類	件数
シーケンス数	182,689
平均シーケンス長	2.29

クインが含まれるシーケンス数と，その平均シーケンス長を調べた．その結果が図 4.5 である．交通機関である鉄道駅，地下鉄へのチェックインがされた場合には，全体シーケンス長の平均よりも多い 3 回以上のチェックインがされているということが明らかとなった．また，ラーメン，和食，カフェ，コーヒーショップといったグルメ関係のカテゴリでは，シーケンス長が 2 未満となっている．

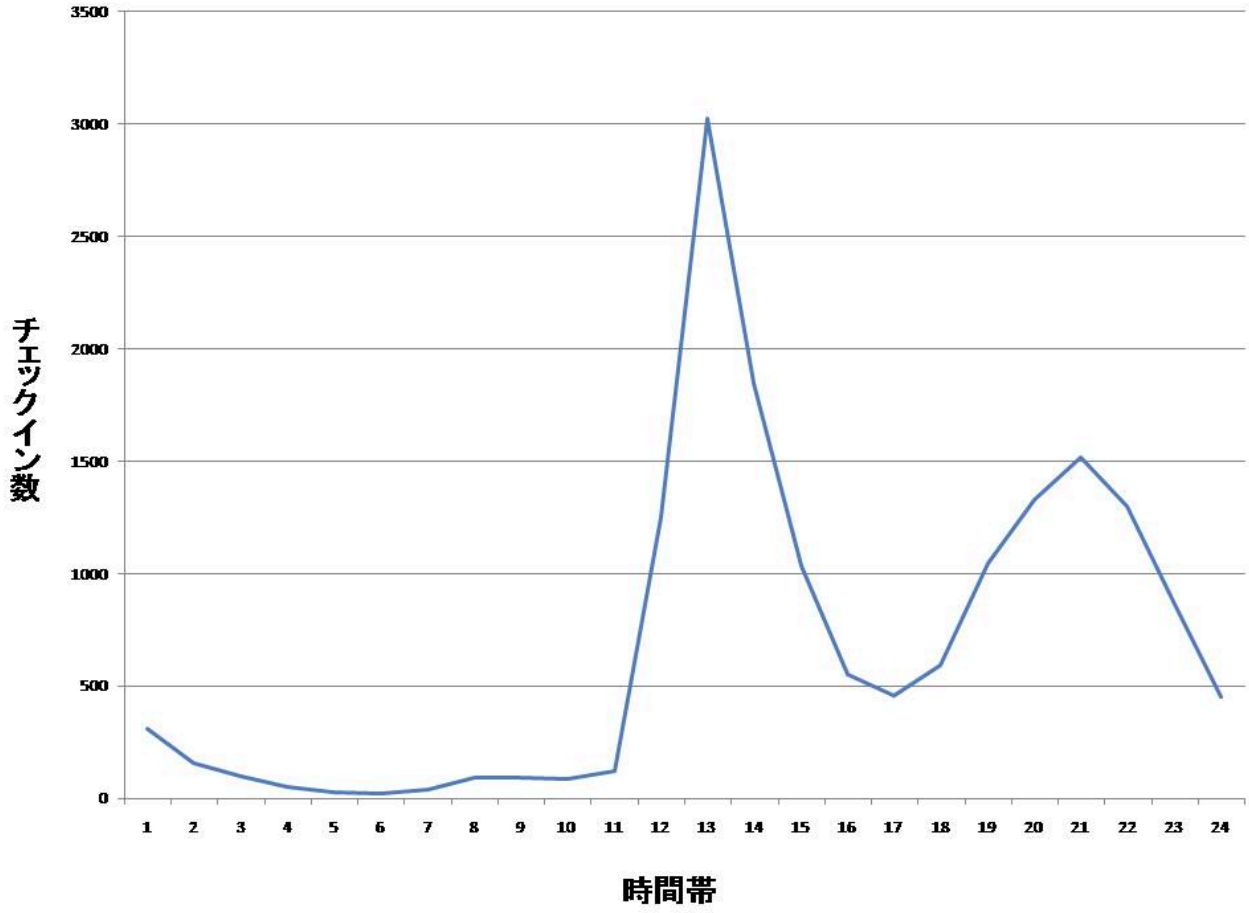

図 4.5　ラーメンのチェックイン時間帯

表 4.5　カテゴリを含むシーケンス数と平均シーケンス長

カテゴリ名	シーケンス数	シーケンス長
鉄道駅	38,136	3.17
ラーメン	9,580	1.53
コンビニエンスストア	6,677	2.83
和食	7,447	1.53
地下鉄	4,299	3.10
ショッピングモール	5,426	2.03
スーパーマーケット	3,249	1.92
カフェ	4,329	1.63
電器店	2,580	2.05
コーヒーショップ	3,714	1.88

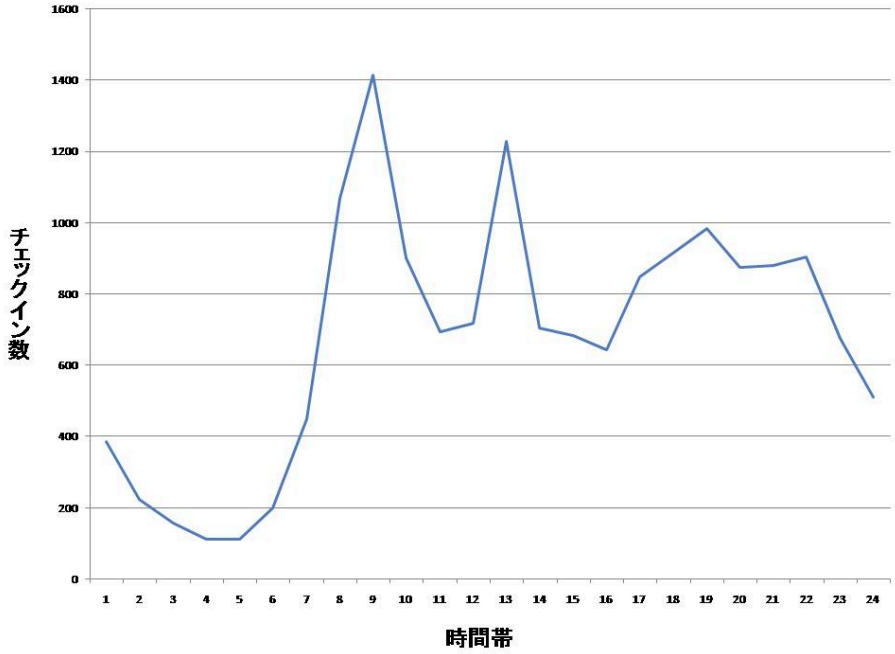

図 4.6　コンビニエンスストアのチェックイン時間帯

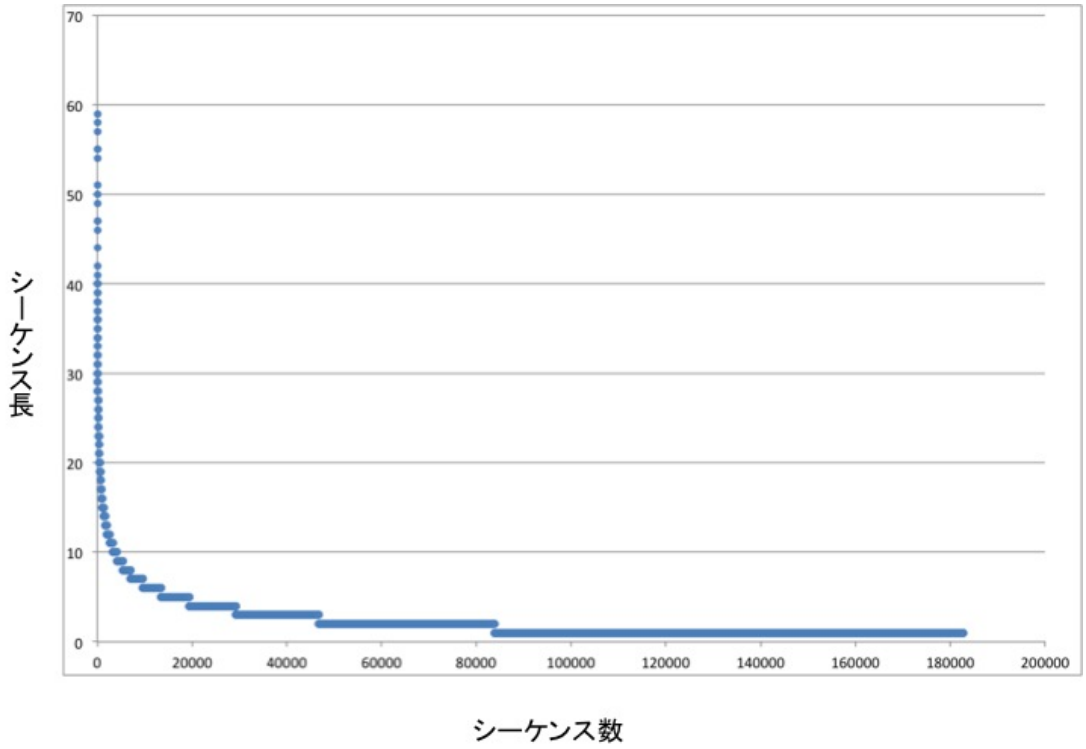

図 4.7　シーケンス長の分布

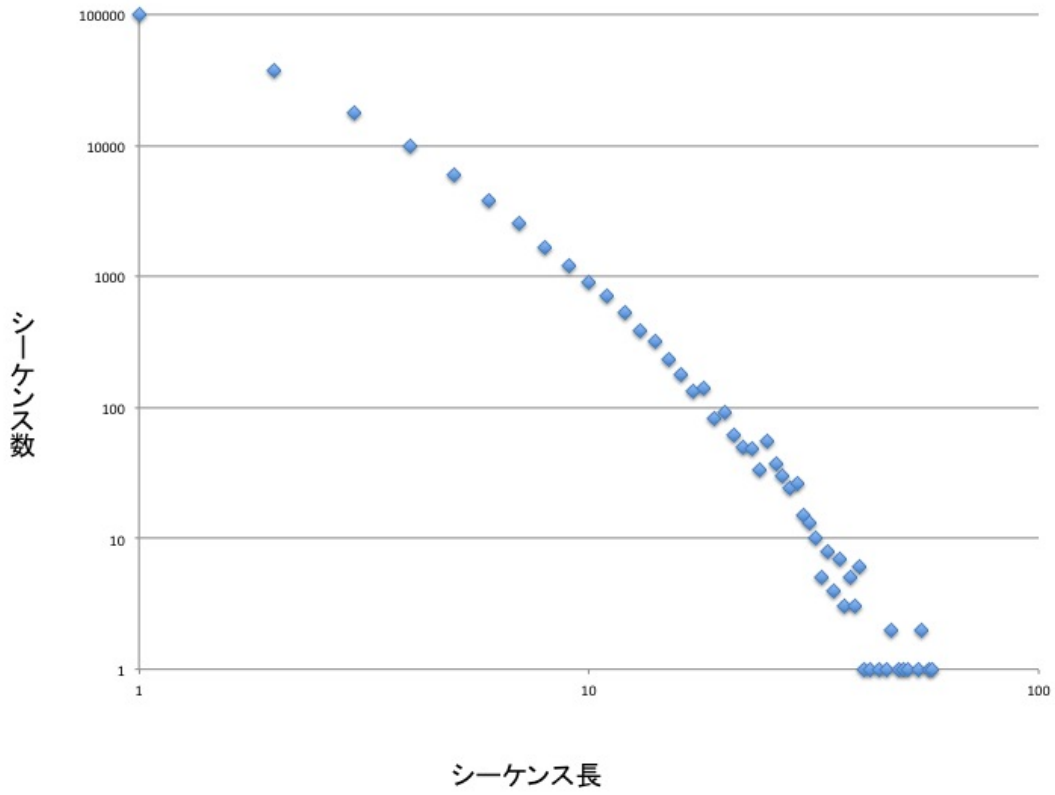

図 4.8　シーケンス長のべき乗分布

4.2　系列マイニングの結果

4.1 節で作成したシーケンスに対して，PrefixSpan を適用してスポットパスを抽出した．最低支持度を 2 と設定し，全シーケンス数の 0.1 %に当たる 180 以上存在するスポットパスを抽出した．その結果，340 種類のスポットパスが得られた．スポットパスの平均長は，2.61 であった．スポットパスの長さの分布を図 4.9 に示す．多くのスポットパスが長さ 2 であることがわかる．また，長さ 8 および 9 のスポットパスが 1 つずつ存在しているが，これはすべてカテゴリが鉄道駅のスポットによって構成されている．

スポットパスの最初のスポットのカテゴリを表 4.6 に示す．この表では，頻出したカテゴリの上位 5 件を示している．1 位の鉄道駅はチェックイン数そのものが多い．コンビニや地下鉄、ショッピングセンター，ラーメンもチェックイン数が上位のカテゴリである．一方，チェックインの多かった和食は 10 件未満となっている．

最後に抽出した 340 種類のスポットパスを，長さ 2 毎のパスに分割し，スポットをノード，次のスポットへの経路をエッジとして有向グラフを生成した．その結果を図 4.10 に示す．鉄

表 4.6　スポットパスの最初のスポットのカテゴリ

カテゴリ名	件数
鉄道駅	154
コンビニエンスストア	27
地下鉄	20
ショッピングモール	14
ラーメン	10

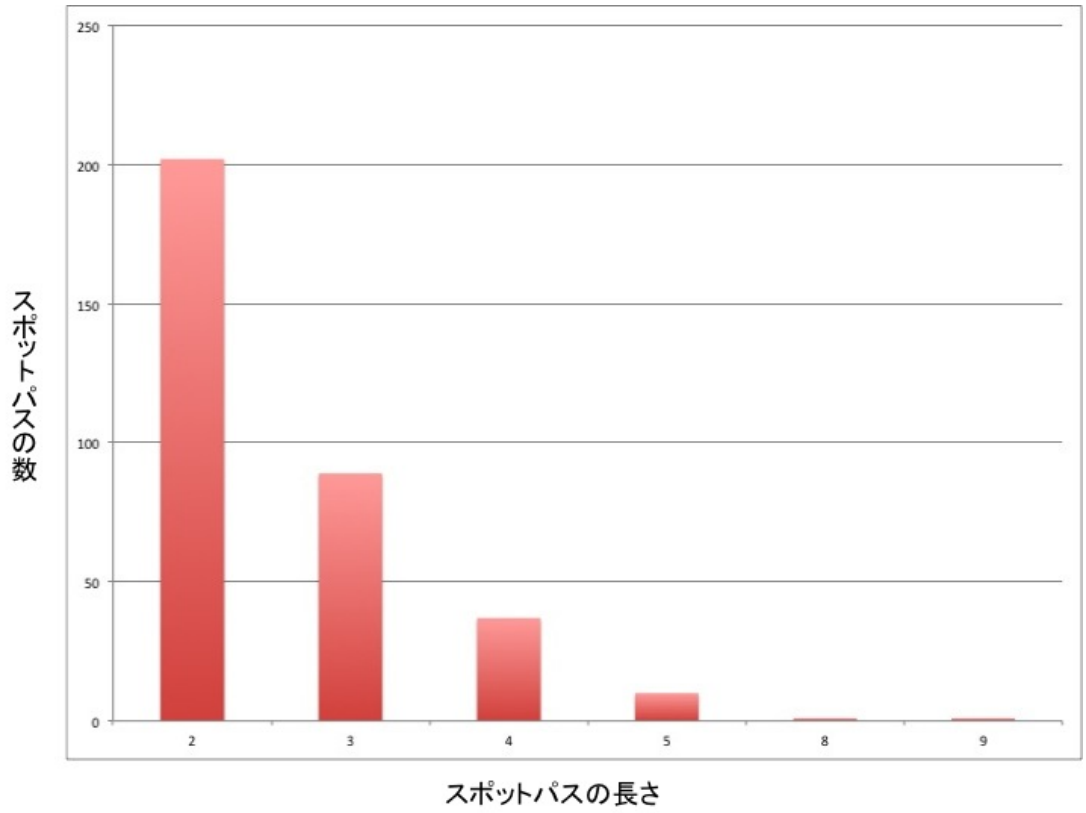

図 4.9　スポットパスの長さ

道駅の入出力のエッジ数が共に最も多いことが確認できる．中央の円の外側には，チェックイン数のトップ 30 にはランクインしていないカテゴリが並んでいて，鉄道によって様々な種類のスポットパスが形成されている．また，拡大した図 4.11 から，コンビニエンスストア，ショッピングモール，スーパーマーケットなどもエッジ数からハブとなっていることが確認できる．

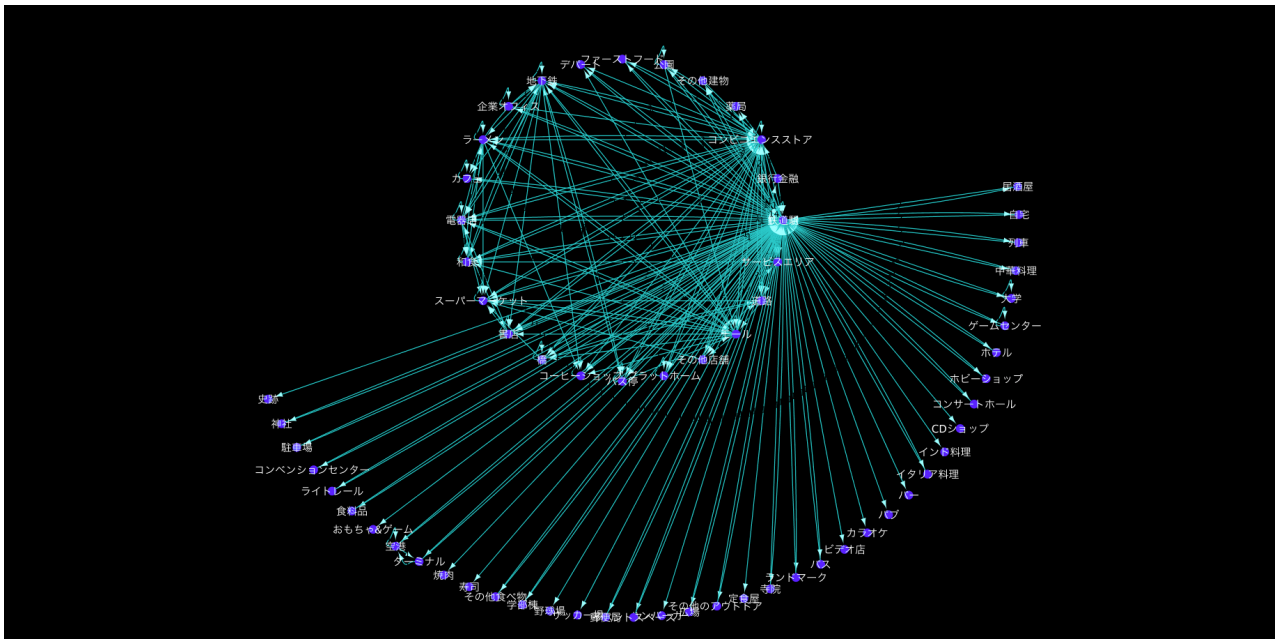

図 4.10　ネットワーク図

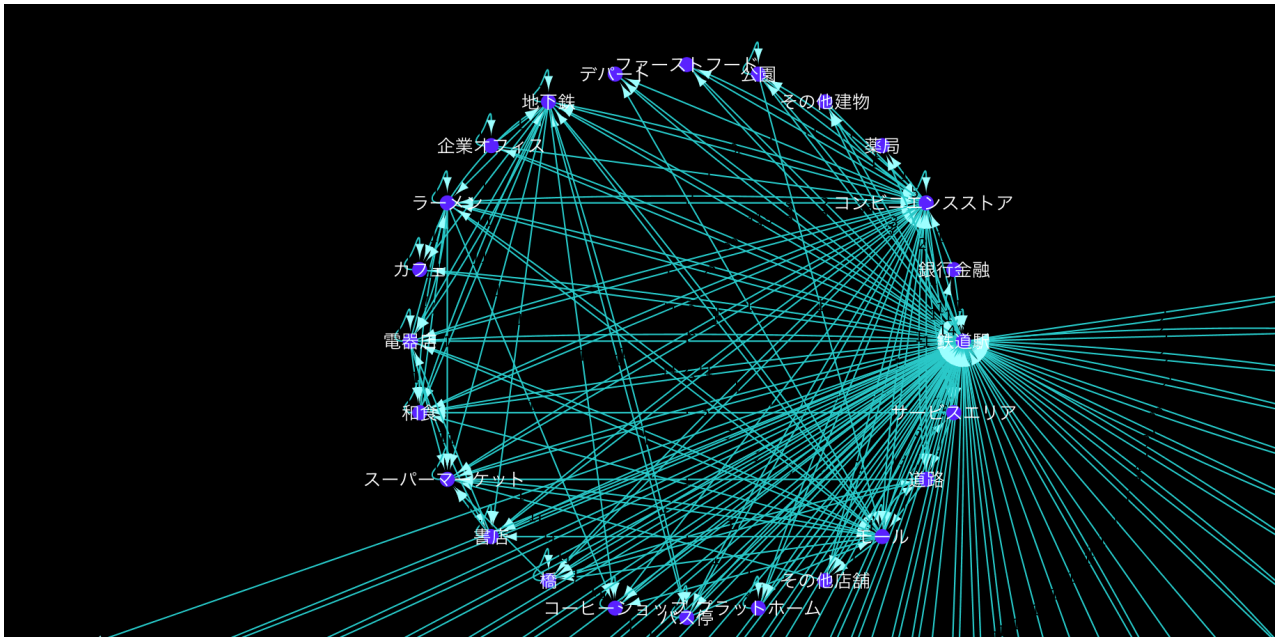

図 4.11　ネットワーク図 (拡大)

第 5 章

考察

5.1　チェックインデータに関する考察

Foursquare のチェックインデータを分析をしたところ，図 4.2 より，ユーザのチェックイン回数はべき乗分布に近いことがわかった．その中で複数回のチェックインをしたユーザは半数程度であったことから，今回抽出されたスポットパスは Foursquare を複数回利用しているユーザ層の特性を表していることになる．また，チェックインされたスポットのカテゴリを確認したところ，表 4.3 より鉄道駅が全体の約 25% を占めていた． この要因として，Foursquare が同じ場所でチェックインを重ねることで特典が得られるというサービス設計がされている点，鉄道駅という場所が通勤や通学など普段の生活の中で多く訪れるスポットであるという点，鉄道駅での電車の待ち時間はスマートフォンを操作してチェックインをするのに適している点などが考えられる．その他，ラーメンやコンビニエンスストア，和食，地下鉄，ショッピングモールといったカテゴリでのチェックインが多いのは，これらのスポットを生活の中で利用する機会が多いためと考えられる．チェックインの時間帯に着目すると，図 4.3 より全体の傾向として 9 時・13 時・19 時でのチェックインが多いことがわかった．また，カテゴリによってピークの分布が異なっていることも確認できた．

チェックインデータからシーケンスを作成した結果，表 4.4 よりシーケンスの平均長は 2.29 であることがわかった．また，シーケンスに含まれるスポットのカテゴリに着目してシーケンス長を調べると，表 4.5 より，鉄道駅や地下鉄がシーケンスに存在する場合はシーケンスが長く，ラーメンや和食，カフェが存在する場合はシーケンスが短いということがわかった．鉄道駅や地下鉄でチェックインする場合は，乗車した次の駅でチェックインをしたり，その先の目的地でチェックインすることが多いためだと思われる．一方，ラーメンや和食、カフェなどの飲食店では，ユーザの目的は飲食であり，実際にその前後でのスポットの移動が少ないためと考えられる．

5.2　スポットパスに関する考察

スポットパスを抽出した結果，340 パターンのスポットパスを抽出した．表 4.4 より，スポットパスの長さは 70% が 2 であった．スポットパスの最初のスポットに着目すると，表 4.6 よりスポットパスの半数近くが鉄道駅であった．その他にはコンビニエンスストア，地下鉄，ショッピングモール，ラーメンから始まるスポットパスが見られたが，鉄道と比べると行き先のカテゴリのバリエーションは少ないことがわかった．

最後に，抽出したスポットパスからネットワーク図を作成した．図 4.10，4.11 より，鉄道駅が移動軌跡の主要なハブとなっていること，コンビニ，ショッピングモール，スーパーマーケットなど生活に密着したカテゴリ郡も入出次数が多く，ハブとなっていることが確認された．その他，鉄道としか移動軌跡を形成していないカテゴリが存在していることが判明した．

第 6 章

まとめ

本研究では，LBSNS の代表的なサービスである Foursquare に蓄積されたチェックインデータを対象に，系列マイニングを適用することでユーザが訪問したスポットのシーケンスであるスポットパスの抽出した．

まず，Foursquare のチェックインデータの概要を分析したところ，カテゴリによってチェックイン数や，チェックインの時間帯などが異なる特徴を示すことがわかった．特に鉄道駅のチェックインが頻出しており，その他，ラーメンや和食などの飲食店，コンビニエンスストアやショッピングモールなど日々の生活で利用するカテゴリも多く見られた．ユーザ毎に 1 日 24 時間を単位として，チェックインのシーケンスを作成した結果，半数近くが長さ 2 以上のシーケンスであった．また，シーケンスに含まれるカテゴリによって，シーケンスの平均長も異なっており，交通機関に関係するカテゴリが含まれるとシーケンス長が長く，飲食に関係するカテゴリが含まれるとシーケンス長が短い傾向が見られた．

作成したシーケンスに対し PrefixSpan を適用することで，340 種類のスポットパスが抽出された．スポットパスの長さは多くが 2 および 3 であり，鉄道駅を含むスポットパスが多く抽出された．次に，抽出した 340 種類のスポットパスを長さ 2 毎のパスに分割し，スポットをノード，次のスポットへの経路をエッジとして有向グラフを生成した．その結果，鉄道カテゴリを主要なハブ，その他コンビニやショッピングモールなど生活に近いカテゴリもハブとなるようなネットワークが確認できた．

今後の課題として，Foursquare はユーザによってその利用具合が異なるため，ユーザ層によって特徴的なスポットパスを抽出することが挙げられる．そのためには，他の期間のチェックインデータでの分析や，チェックイン数やカテゴリを考慮したユーザのクラスタリングの処理が挙げられる．

謝辞

本研究を進めるにあたり，同期の大塚淳史さん，香川雄一さんには研究内外において非常に世話になりました．共に佐藤研究室で学べたこと感謝しています．また，佐藤研究室の後輩である山本修平さん，山口裕太郎さん，佐野駿さん，中岡義貴さん，堀内雅人さん，山村悟さん，関研究室の宮嶋清人さん，大山鉄郎さん，堂前友貴さん，酒井紗希さん，庄司茜さんには研究の相談から研究環境のサポートまで多くの点で協力いただきましたこと，ありがとうございます．

関洋平助教授には，ゼミや研究セミナー等で本質的なアドバイスをいただき，研究の方向性を決めていく上で非常にお世話になりました．手塚太郎准教授には，副指導教員として研究について熱心なご指導いただきました．

佐藤哲司教授には 3 年間もの間，熱心にご指導頂き，研究のみならず多くのことを学ばせて頂きました．心から感謝申し上げます．

皆様のご協力を頂き，本研究が完成に至ったことに感無量です．本当にありがとうございました．

参考文献

[1] 篠田裕之, 竹内亨, 寺西裕一, 春本要, 下條真司: 行動履歴に基づく協調フィルタリングによる行動ナビゲーション手法. 情報処理学会研究報告, 2007-DPS-132, pp. 87-92, 2007.

[2] Z. Yin, L. Cao, J. Han, J. Luo, and T. S. Huang: Diversified Trajectory Pattern Ranking in Geo-tagged Social Media. In SDM, pp. 980-991, 2011.

[3] M.D. Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel and C. Yu: Constructing Travel Itineraries from Tagged Geo-Temporal Breadcrumbs. in Proc. Int. Conf. on World Wide Web (WWW), pp. 1083-1084, 2010.

[4] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura: Travel route recommendation using geotags in photo sharing sites. In CIKM, pp. 579-588, 2010.

[5] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011.

[6] 旭直人, 山本岳洋, 中村聡史, 田中克己: 行動連鎖を用いた情報検索支援と Web からの行動連鎖の抽出. 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM2009), A7-2, 2009.

[7] 郡宏志, 服部峻, 手塚太郎, 田島敬史, 田中克己: ブログからのビジターの代表的な行動経路とそのコンテキストの抽出. 信学技報, DE2006-55(2006-7), pp.29-34, 2006.

[8] 筏島郁子, 嶋田和孝, 遠藤勉: 系列パターンを利用した評価表現の分類. 言語処理学会第 11 回年次大会, 2009.

[9] 山田 吾郎, 吉田 則裕, 井上 克郎: シーケンシャルパターンマイニングに基づくオブジェクト指向プログラムのための欠陥検出手法. 情報処理学会研究報告, Vol.2009-SE-164/Vol.2009-EMB-13/Vol.2009-CSEC-45, No.15, pp.1-8, 2009.

[10] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu: PrefixSpan mining sequential patterns Efficiently by prefix projectedpattern growth. In ICDE 2001, pp. 215-226, Heidelberg, Germany, 2001.