F-048

系列ルールマイニングにおいて時間差を考慮する確信度の提案
A Proposal of New Confidence Considered Time Lag for Sequential Pattern Mining

今村　理† Satoru Imamura　西村　聖† Satoru Nishimura　橋本 和夫† Kazuo Hashimoto

1　はじめに

近年の携帯電話技術の発達により，利用者の位置情報を用いた LBS(Location Based Service) が注目されている [1, 2]．GPS 等で取得した利用者の現在地情報を用いて地域情報の配信は既に行われているが，LBS をさらに便利なサービスにするためには，利用者の移動予測が必要である．移動予測は，利用者の訪問先の情報を事前配信することを可能とするため，高度サービスの開発に資すると考えられている．

移動予測技術には様々なものがあるが，近年のコンピュータのストレージ容量や処理速度の向上により，大量の履歴から利用者の移動パターンを抽出する系列ルールマイニングの技術が注目されている [1, 2, 3]．

移動手段ごとに道路等の物理的な制約が生じ，とりうる移動経路の自由度が減ると考えられる．したがって，系列ルールマイニングを用いて高精度の予測を行うためには，利用者の移動手段の区別が必要となる．時間情報から利用者の移動手段が推定可能と思われるが，一般的な系列ルールマイニングでは時間情報から移動手段を判別する仕組みがないため，予測精度に問題がある．

時間情報を用いた系列ルールマイニングの手法として，Chen ら [4] は時間情報をファジィ空間にマッピングするファジィ時間間隔付き系列ルールマイニングを提案している．Chen らはメンバシップ関数を適切に設定することにより，系列の時間情報に隠れたコンテキスト情報を区別した系列ルールが抽出できると主張している．移動予測の場合は，主なコンテキスト情報は移動手段となる．本論文では Chen らの手法を用いて，利用者の移動手段を考慮した移動予測を行う．

高精度の移動予測を行うためには，利用者の移動手段と予測に用いるルールの移動手段を一致させる必要がある．しかし，その手法については充分な議論がなされていない．そこで，本論文では移動手段の一致度を考慮した新たなルール選択指標を提案する．

†東北大学大学院　情報科学研究科，Graduate School of Information Sciencies, Tohoku University

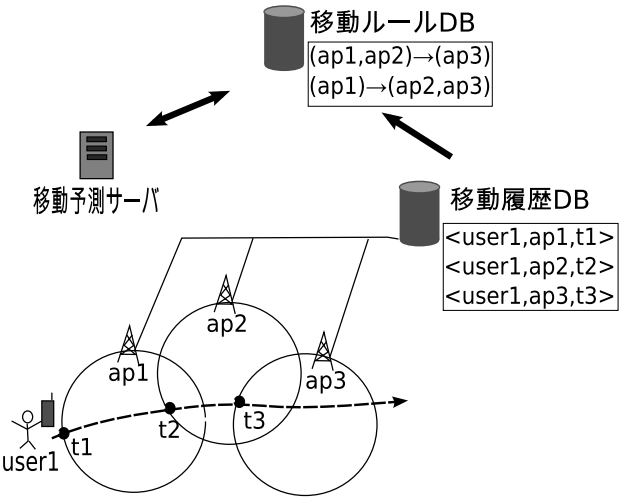

図 1: 移動履歴の取得

本論文は以下の構成になっている．第 2 章では，従来の系列ルールを用いた移動予測とその問題点をあげ，第 3 章で Chen らの提案したファジィ時間間隔付き系列ルールマイニングについて説明する．そして，第 4 章で Chen らの手法を用いた移動予測と，新たなルール選択指標の提案を行う．最後に，第 5 章で本論文のまとめを行う．

2　想定する移動予測システムとその問題

本論文で仮定する移動予測システムは，携帯通信事業者が移動予測を行い，アプリケーションに提供するものである．図 1 で示すように，携帯通信事業者の移動履歴データベースに利用者が利用したアクセスポイント ap_i とそのサービスエリアに進入した時刻 t_i が記録

されている．

利用者の典型的な移動パターンを，移動履歴データベースから系列ルールマイニング手法を用いて抽出し，移動予測を行う．移動履歴データベースに対して一定時間ごとに系列ルールマイニングが行われ，利用者の移動パターンが抽出される．移動履歴データベースから各利用者の利用したアクセスポイントの時間順系列が取り出され，系列ルールが抽出される．また，この時，各ルールに対して確信度と呼ばれるルール選択指標が計算され，抽出されたルールとともに移動ルールデータベースに格納される．

利用者 $user_i$ の移動予測は以下のように行われる．(1)$user_i$ の直前の移動系列 s を移動履歴データベースから取得する．(2)s に条件部が一致するルールを移動ルールデータベースから取り出す．(3) その中から確信度の最も大きなルールを選択し，これを用いた予測を行う．

一般に，利用者の移動手段に応じて移動先が異なることが想定されるため，精度の高い移動予測を行うためには利用者の移動手段の区別が必要となる．しかし，一般的な系列ルールマイニングは移動手段を区別する仕組みがないため，全ての移動手段を混同したルールが履歴から抽出され，予測の精度が低下する．

Chen ら [4] は，移動予測に限らず，一般的に時間間隔からメンバシップ関数を用いて情報を抽出する系列ルールマイニング手法を提案している．次章では Chen らの手法について説明する．

3 ファジィ時間間隔付き系列ルールマイニング

Chen ら [4] は，ファジィ時間間隔付き系列ルールマイニングを提案している．本章では，Chen らの手法について説明する．まず，Chen らが想定している系列を示す．

[**系列**]:アイテムの全集合を $I = \{i_1, i_2, \cdots, i_m\}$，系列 s を $\{(a_1, t_1), (a_2, t_2), \cdots, (a_n, t_n)\}$ とする．$a_j \in I$，t_j は a_j の生じた時刻である．$1 \leq j \leq n$ であり，$2 \leq j \leq n$ に対して，$t_{j-1} \leq t_j$ である．□

この系列に対して，時間間隔は $\tau_j = |t_{j+1} - t_j|$ で与えられる．時間間隔の言語表現 (linguistic term) を $LT = \{LT_j | j = 1, 2, \cdots, l\}$ とし，これに対応するメンバシップ関数を $\mu_{LT_j}(\tau)$ とする．

[**時間間隔付き系列**]:時間間隔付き系列を $\alpha = (b_1, LG_1, b_2, LG_2, \cdots, b_{r-1}, LG_{r-1}, b_r)$ と定義する．$b_i \in I$ であり，$LG_i \in LT$ である．□

これらの定義のもと，次のように包含度数を計算し，支持度を求めることを提案している．

[**包含関係**]:系列 s と時間間隔付き系列 α の包含関係を定義する．ここで，$1 \leq r \leq n$ である．言語表現 LG_i と対応するメンバシップ関数 $\mu_{LG_i}(t)$，系列 s の添字を $1 \leq w_{k,1} < w_{k,2} < \cdots < w_{k,r} \leq n (k = 1, \cdots, K)$ とする．$b_1 = a_{w_{k,1}}, b_2 = a_{w_{k,2}}, \cdots, b_r = a_{w_{k,r}}$ であるとき，系列 s は α を包含度数 $\gamma(\alpha, s)$ で包含すると定義する．$\gamma(\alpha, s)$ の定義は以下の通りである．なお，時間間隔 $\tau_{W_{k,i}} = |t_{W_{k,i+1}} - t_{W_{k,i}}|$ である．

$$\gamma(\alpha, s) = \begin{cases} 1 & (r = 1) \\ \max\limits_{1 \leq k \leq K} \min\limits_{1 \leq i \leq r-1} \{\mu_{LG_i}(\tau_{w_{k,i}})\} & (r > 1) \end{cases} \tag{1}$$

□

[**支持度**]:系列集合 S における，時間間隔付き系列 α の支持度を式 (2) で定義する．

$$supp_S(\alpha) = \frac{\sum\limits_{s \in S} \gamma(\alpha, s)}{|S|} \tag{2}$$

□

Chen らの功績によって，一般的な系列データからファジィ時間間隔付き系列ルールを求めることができるようになった．本論文では，Chen らの提案したアルゴリズムを用いて移動履歴から移動パターンを求め，移動パターンから我々の定義する系列ルールを求める．

4 ファジィ時間間隔付き系列ルールによる移動予測

本章では，Chen らが示したアルゴリズムを用いて移動履歴データベースから系列ルールを求め，移動予測

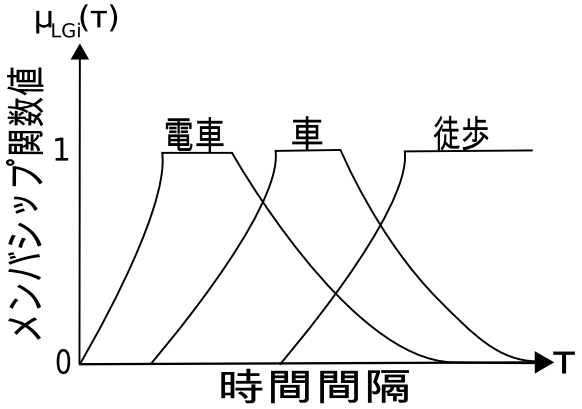

図 2: メンバシップ関数の例

を行うことを検討する．本論文では，図 2 のようにメンバシップ関数により時間間隔を移動手段の言語表現 (linguistic term) に $[0,1]$ の値で対応付けることとする．

4.1 移動予測に適した包含度数

Chen らは，一般的な系列データから系列ルールを求める手法を提案した．Chen らの提案手法を移動履歴に適用する場合には，移動履歴の特徴を考慮しなければならない．移動履歴における時間間隔は，信号待ち等で突然長くなったり移動経路の関係で突然短くなったりすることがある．そのため，最小値を用いる Chen らの定義は適さない．ここでは，平均値を用いる定義 [5] を採用する．平均値を用いると，式 (1) における $r > 1$ の時の定義は，式 (3) のように置き換えられる．

$$\gamma(\alpha, s) = \max_{1 \le k \le K} \frac{\sum_{i=1}^{r-1} \mu_{LG_i}(\tau_{w_{k,i}})}{|r-1|} \tag{3}$$

4.2 ファジィ時間間隔付き系列ルールの定義とルール選択指標の提案

系列ルールの定義

Chen らは系列ルールを予測に用いることは想定しておらず，系列ルールの定義には触れていない．我々は，移動予測に用いるための系列ルールと確信度を次のように定義する．

[系列ルール]:時間間隔付き系列を $\alpha = (b_1, LG_1, b_2, LG_2, \cdots, b_{r-1}, LG_{r-1}, b_r)$ とする．α が条件部と結論部に分割された $\beta = (b_1, LG_1, \cdots, LG_{n-1}, b_n) \Rightarrow (LG_n, b_{n+1}, LG_{n+1}, \cdots, LG_{r-1}, b_r)$ を系列ルールと定義する．ここで，$1 \le n < r$ である．□

[確信度]:系列ルール β の確信度 $conf(\beta)$ を式 (4) で定義する．

$$conf(\beta) = \frac{supp_S((b_1, LG_1, \cdots, LG_{r-1}, b_r))}{supp_S((b_1, LG_1, \cdots, LG_{n-1}, b_n))} \tag{4}$$

□

[一致ルール]:予測の際は，その時点の直前の履歴系列を用いて，予測に利用する候補となるルールを検索する．予測に用いる直前の系列の最大時間間隔を τ_{max} と定め，直前の系列 s' を $\{(a_1, t_1), (a_2, t_2), \cdots, (a_n, t_n)\}$ とする．a_n は予測時の直前に生起したアイテムであり，$\tau_{max} \ge |t_n - t_1|$ である．そして，s' が最小包含度数 γ_{min} 以上で包含する系列を α' とし，s' の一致ルールをルールの条件部が α' であるルールとする．□

ルール選択指標

移動予測を行う上で残る課題は，利用者の移動手段とルールの移動手段との適合性を移動予測に反映させることである．包含度数が系列の時間間隔と移動手段との関連性を示す指標であるので，本論文では，長期間の移動予測を行うために結論部の長さを確信度の補正に用いた Morzy ら [3] にならい，次のように確信度を補正した指標を提案する，この指標が最も大きくなるルールを移動予測のために選択する．

[ルール選択のための指標]:予測時の直前の系列を $s' = \{(a_1, t_1), (a_2, t_2), \cdots, (a_n, t_n)\}$，$s'$ の一致ルールを $\beta = (b_1, LG_1, \cdots, LG_{n-1}, b_n) \Rightarrow (LG_n, b_{n+1}, LG_{n+1}, \cdots, LG_{r-1}, b_r)$，$\beta$ の確信度を $conf(\beta)$ とする．また，ルール β の条件部 $(b_1, LG_1, \cdots, LG_{n-1}, b_n)$ を β_{head} とする．この時，直前系列の時間間隔と系列ルールの時間間隔の一致度を考慮した指標 $\theta(s', \beta)$ を式 (5) で定義する．

$$\theta(\beta, s') = \gamma(\beta_{head}, s') \cdot conf(\beta) \tag{5}$$

□

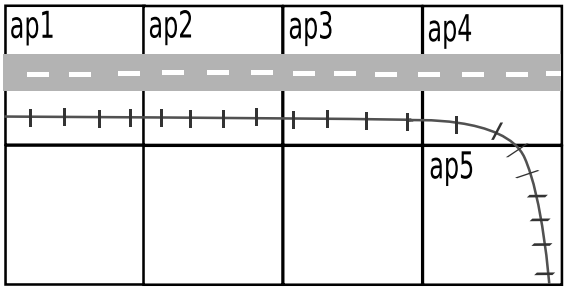

図 3: 系列の例

4.3 提案手法の効果

図 3 のように線路と道路がある時，通常の系列ルールマイニングでは，ルール $\beta_1 = (ap_1, ap_2, ap_3) \Rightarrow (ap_4)$ と $\beta_2 = (ap_1, ap_2, ap_3) \Rightarrow (ap_5)$ のように条件部が等しく結論部の異なるルールが抽出される．条件部が等しいので，移動予測の際には確信度の大きいルールが常に用いられる．

本論文で定義するファジィ時間間隔付き系列ルールを用いると，$\beta'_1 = (ap_1, LG_{車}, ap_2, LG_{車}, ap_3) \Rightarrow (LG_{車}, ap_4)$ と $\beta'_2 = (ap_1, LG_{電車}, ap_2, LG_{電車}, ap_3) \Rightarrow (LG_{電車}, ap_5)$ のようなルールが抽出でき，条件部の時間間隔から区別が可能になる．

直前系列 $s' = (ap_1, \tau_1, ap_2, \tau_2, ap_3)$ から移動予測に用いるルールを選択する時，本論文で提案するルール選択指標を用いると，$\mu_{車}(\tau_i) > \mu_{電車}(\tau_i)$ であれば β'_1 の方が選ばれやすくなる．移動手段の推定が移動予測に反映され，予測精度が上がることが期待できる．

5 まとめと今後の課題

本論文では，Chen ら [4] のファジィ時間間隔付き系列ルールマイニングを用いて利用者の移動予測を行った．従来の系列ルールでは，移動手段による系列の区別が行われていないという問題があった．Chen らは，メンバシップ関数によって時間情報から隠れたコンテキスト情報を抽出するファジィ時間間隔付き系列ルールマイニングを提案した．しかし，Chen らは系列ルールを用いた予測は考慮しておらず，利用者の移動手段と予測に用いるルールの移動手段の一致度が考慮されていなかった．そのため，本論文では移動手段の一致度を考慮した新たなルール選択指標を提案した．本論文の貢献は以下の通りである．

1. ファジィ時間間隔付き系列ルールの定義.
2. ファジィ時間間隔付き系列ルールの移動予測への適用.
3. 予測時の直前系列の移動手段を考慮した，予測に用いる系列ルールの選択指標の提案.

本論文では移動予測について取り上げたが，系列ルールの定義とルール選択の指標は一般的に言えることであり，他の目的にも応用できると考えられる．今後の課題として，提案手法の有効性を実験により確かめることが挙げられる．

参考文献

[1] G. Yavas, D. Katsaros, O. Ulusoy, and Y. Manolopoulos, “A data mining approach for location prediction in mobile environments,” *Data & Knowledge Engineering*, vol. 54, no. 2, pp. 121 – 146, 2005.

[2] J. W. Lee, O. H. Paek, and K. H. Ryu, “Temporal moving pattern mining for location-based service,” *Journal of Systems and Software*, vol. 73, no. 3, pp. 481 – 490, 2004.

[3] M. Morzy, “Prediction of moving object location based on frequent trajectories,” in *Computer and Information Sciences - ISCIS 2006, 21th International Symposium*, (Istanbul, Turkey), Nov. 1-3 2006.

[4] Y.-L. Chen and T.-K. Huang, “Discovering fuzzy time-interval sequential patterns in sequence databases,” *IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 35, pp. 959–972, Oct. 2005.

[5] A. Gyenesei, “A fuzzy approach for mining quantitative association rules,” tech. rep., Turku Centre for Computer Science, 2000.