潜在特徴関係モデルを用いた時系列金融ネットワークの解析と予測

Time dependent analysis and prediction of financial networks using supervised latent feature relational models

伊藤翔太郎 [1*]　　江口浩二 [1]
Shotaro Ito[1]　　Koji Eguchi[1]

[1] 神戸大学大学院システム情報学研究科
[1] Graduate School of System Informatics, Kobe University

Abstract: In recent years, many researchers have taken keen interest in analyzing various kinds of relational data, such as social networks and financial networks. These data can be expressed as a graph or network where each vertex or node is an entity and each edge or link is a relation between a pair of entities. Moreover, each link is often associated with continuous and/or discrete relational attributes, such as in financial networks, the interest rate for a transaction and whether the transaction is international or intranational. In this paper we focus on max-margin latent feature relational models (called Med-LFRM) that are based on Indian buffet process (IBP) and maximum entropy discrimination (MED). For the estimation of model parameters, the Bayesian estimation is deemed equivalent to minimizing an objective function, which involves misclassification errors. We focus on link prediction problem for the networks with continuous and discrete relational attributes. We also focused on the time dependent analysis for the networks, and therefore, we estimated the model parameters considering the observations in the previous time interval. We demonstrate, through experiments with inter-bank financial networks, the effectiveness of the above model in terms of the link prediction performance.

1　はじめに

近年，社会的ネットワークや金融ネットワークなどの関係データの可用性が増加しており，それらのデータを統計解析に用いることで有用な知見を得ることが課題となっている．このようなデータは，エンティティをノードで，それらの間に存在する関係をリンクで表すようなグラフ構造として表現することができる．既観測のリンクから未観測のリンクを予測するリンク予測は，このようなデータの分析における基本的な問題の一つである [1]．この問題を考えるとき，各エンティティまたはリンクの持つ属性などの情報を利用し予測を行うこともある [2][3].

確率モデルに基づいた様々な手法を用いたリンク予測に関する研究は発展を続けている．本稿において着目するのは，リンク構造の確率分布を定義するために各ノードが潜在特徴を持つと仮定し，それと共にシグモイド関数などのリンク尤度関数を利用するようなモデルである [3][4].

しかし，潜在特徴の未知の次元数を決定するために，ほとんどの手法においては交差検定などによってモデルを選択する必要がある．この時，多くの異なる訓練データによる結果を比較する必要があり，それにより大きなコストがかかってしまう．そこで，Miller らはインド料理過程（Indian Buffet Process: IBP）[5] に基づいたノンパラメトリックベイズ法を用いて，データから自動的に未知の潜在特徴の次元数を推定することを提案した [3]．これが潜在特徴関係モデル（Latent Feature Relational Model: LFRM）である．そして LFRM を発展させ，最大エントロピー識別（Maximum Entropy Discrimination: MED）[6][7] の枠組みに基づき，リンク予測の精度を測るヒンジ損失などの目的関数を直接最小化することによりリンク予測を行うモデルが Zhu によって提案された [8]．これがマージン最大化潜在特徴関係モデル（Max-Margin Latent Feature Relational Model: MedLFRM）である．このモデルは，それぞれ独立に研究されてきたノンパラメトリックベイズ法とマージン最大化法を統合したモデルである．このモデ

*連絡先：神戸大学大学院システム情報学研究科
〒 0657-0013 兵庫県神戸市灘区六甲台町 1-1
E-mail: shotaro@cs25.scitec.kobe-u.ac.jp

ルにおいては，ベイズ推定の計算が目的関数を最小化することと等価になり，それによって教師ラベルを用いることが可能となる．また，ソフトマージンにより誤分類を許容し，より柔軟なモデルを実現することができる．ソフトマージンの最大化に関する部分問題は，既存の高性能な求解法によって解くことができる．

MedLFRM が，LFRM よりもリンク予測の精度が優れていることが知られている [8]．しかしながら，それは連続値のみで表現された関係属性について結果が示されたものであり，連続値と離散値の関係属性が混在する場合については検討されていない．また，ネットワークの時系列解析についても検討されていない．本稿ではこの二点に着目して評価を行う．後者に関しては，前時区間のデータから学習したパラメータを当時区間のパラメータの初期状態に設定して学習することによって時間依存性を考慮し，リンク予測の精度の向上を図る．

本稿の実験は，LFRM や MedLFRM では連続値表現が仮定されていた関係属性を連続値と離散値が混合したものであると仮定する．これは実際の金融ネットワークに対して柔軟に対応するためである．これを踏まえて，2009 年から 2012 年の欧州の銀行間での取引をまとめたデータセットを用いてリンク予測問題に対して実験を行った．

本稿の構成は以下のようになる．第 2 章では，LFRM, MED，MedLFRM，などの既存の手法を紹介する．第 3 章では，MedLFRM による連続値・離散関係属性付きネットワークの時系列解析について述べ，その条件下におけるパラメータの推定方法について述べる．連続値・離散関係属性を考慮したリンク予測の実験結果を第 4 章で示し，第 5 章で結論および今後の課題について述べる．

2 関連研究

2.1 無限潜在特徴関係モデル

潜在特徴関係モデル（LFRM）は Miller らによって提案されたモデルである [3]．このモデルは，各ノードが二値の値をとる潜在特徴ベクトルを持つと仮定し，それらの未知の次元を自動的に推定すると共に，ネットワーク間のリンクが生成される尤度を推定するモデルである．尤度は各ノードの持つ潜在特徴とノード間のリンクに付与される関係属性，そしてそれらの重みを用いて算出される．

ネットワーク内のノード数を N とし，$N \times N$ の二値隣接行列を Y とする．この時，ノード i とノード j の間にリンクが存在する場合は $Y_{ij} = +1$ とし，リンクが存在しない場合は $Y_{ij} = -1$ とする．Y は完全には観測されておらず，既観測のリンクから未観測のリンクの有無を予測できるモデルを学習することが目的となっている．また，ノード i とノード j の間のリンクに作用する関係属性 $X_{ij} \in \mathbb{R}^D$ が観測されている場合もある．

各ノードの持つ潜在特徴の数を K とすると，各ノードは二値潜在特徴ベクトル $\mu_i \in \mathbb{R}^K$ の集合とみなすことができる．ここで Z を $N \times K$ の二値潜在特徴行列とすると，$\mathbf{Z} = [\mu_{\mathbf{1}}^\top; \cdots; \mu_{\mathbf{N}}^\top]$ となる．Z_i はノード i の二値潜在特徴ベクトルを表し，ノード i が潜在特徴 k を持つとき，$Z_{ik} = 1$ となり，そうでない場合は $Z_{ik} = 0$ となる．また，W を $K \times K$ の実数値重み行列とし，$W_{kk'}$ は，ノード i が潜在特徴 k を持ち，ノード j が潜在特徴 k' を持つとき，その二つのノード間のリンクの生成に影響を与える重みであるとする．以上より，リンク尤度は以下のように定義される．

$$p(Y_{ij} = 1 | X_{ij}, Z_i, Z_j) = \Phi(\mu + \eta^\top X_{ij} + Z_i^\top W Z_j) \quad (1)$$

ここで，Φ はシグモイド関数である．そして，μ は尤度に影響を与える大域的バイアス値であり，η は関係属性の実数値重みベクトルである．最適な事前分布を得るために，インド料理過程（IBP）[5] を Z の事前分布として用いる．これによって，Z を推定すると同時に，潜在特徴数 K も推定することができる．W は，各成分において独立して事前分布 $\mathcal{N}(0, \sigma_w^2)$ を仮定する．

IBP を確率が 1 となる無限二値行列の事前分布と仮定する．これにより生成される行列は，潜在特徴をいくつ持っていたとしても各成分は必ず正の値をとる．行列の成分のサンプリングは以下のように行われる．1 番目のノードに対応する行のうち，Poisson(β) の数だけの成分を 1 とする．ここで，β はハイパーパラメータである．次に，i 番目のノードに対応する行に属する成分のうち，既に他の行で 1 となっている成分は，その 1 となっている成分の数に比例した確立で 1 となる．また，Poisson(β/i) の数だけの成分を新しく 1 にする．これを有限個のノードの数だけ繰り返すことで潜在特徴行列の事前分布を得る．この過程は交換可能であるため，選択される行の順番には影響されない．

2.2 最大エントロピー識別

最大エントロピー識別（MED）[6][7] は，事前分布を用いて目的関数である正則化項付き相対エントロピー最小化問題を解くことにより事後分布を学習する手法である．

応答変数 Y が $\{+1, -1\}$ を取るような二値分類問題を考える．X を入力特徴ベクトルとし，$F(X; \eta) = \eta^\top X_n$ を η によってパラメータ化された識別関数とする．また，ℓ を正の損失パラメータとし，ヒンジ損失関数を

$h_\ell(x) = \max(0, \ell - x)$ と定義する．η の事前分布を $p_0(\eta)$，事後分布を $p(\eta)$ とすると，単一の η を推定する通常の SVM とは異なり，MED は $p_0(\eta)$ を用いて以下の正則化項付き相対エントロピー最小化問題を解くことにより，$p(\eta)$ を学習するものである．

$$\min_{p(\eta)} \ \mathrm{KL}(p(\eta)||p_0(\eta)) + C\mathcal{R}(p(\eta)) \tag{2}$$

ここで，C は正の定数である．$\mathrm{KL}(p(\eta)||p_0(\eta))$ は KL ダイバージェンス，すなわち相対エントロピーであり，$\mathcal{R}(p(\eta)) = \sum_n h_\ell(Y_n \mathbb{E}_{p(\eta)}[F(X_n;\eta)])$ はヒンジ損失である．

応答変数 Y の予測値は以下のようになる．

$$\hat{Y} = \mathrm{sign}\mathbb{E}_{p(\eta)}[F(X;\eta)] \tag{3}$$

2.3 マージン最大化潜在特徴関係モデル

マージン最大化潜在特徴関係モデル（MedLFRM）は Zhu によって提案されたモデルである [8]．このモデルは，MED における識別関数 F を LFRM におけるリンク尤度として定義することにより，より効果的にリンク尤度を推定できるモデルである．

LFRM と同様に，二値潜在特徴行列 Z と，実数値重み行列 W，関係属性 X_{ij} が与えられると，識別関数は以下のようになる．

$$\begin{aligned} f(Z_i, Z_j; X_{ij}, W, \eta) &= Z_i W Z_j^\top + \eta^\top X_{ij} \\ &= \mathrm{Tr}(W Z_j^\top Z_i) + \eta^\top X_{ij} \end{aligned} \tag{4}$$

η は関係属性に対する実数値重みベクトルである．ここで，$\Theta = \{W, \eta\}$ をすべてのパラメータとし，Θ と事前分布 $p_0(\Theta)$ は確率変数とする．予測を行うためには，潜在変数の不確実性を取り除く必要があるため，より効果的な識別関数を $p(Z, \Theta)$ に関する期待値として以下のように定義する．

$$f(X_{ij}) = \mathbb{E}_{p(Z,\Theta)}[f(Z_i, Z_j; X_{ij}, \Theta)] \tag{5}$$

したがって，応答変数 Y の予測値は $\hat{Y}_{ij} = \mathrm{sign} f(X_{ij})$ となる．$\mathcal{I}$ を観測されたリンクの組の集合とし，ヒンジ損失関数を 2.2 節と同様に定義すると，ヒンジ損失は以下のようになる．

$$\mathcal{R}_\ell(p(Z,\Theta)) = \sum_{(i,j)\in\mathcal{I}} h_\ell(Y_{ij} f(X_{ij})) \tag{6}$$

ここで，$p_0(Z)$ を潜在特徴行列の事前分布とすると，以上より MedLFRM を以下の問題を解くことと定義できる．

$$\min_{p(Z,\Theta)\in P} \mathrm{KL}(p(Z,\Theta)||p_0(Z,\Theta)) + C\mathcal{R}_\ell(p(Z,\Theta)) \tag{7}$$

一般的に，補助変数を導入することによりマージンの依存性を条件付き独立に変換し，推定を簡単化できるということが知られており，これを変分近似と呼ぶ．これを行うために，Teh らによって提案された IBP の棒折り過程（Stick Breaking Prior: SBP）[9] を用いる．$\pi_k \in (0,1)$ を行列 Z の列 k と対応するパラメータとし，このパラメータ π は棒折り過程によって生成される．ここで，$\pi_1 = \nu_1, \pi_k = \nu_k \pi_{k-1} = \prod_{i=1}^k \nu_i$ であり，ν_i は α をハイパーパラメータとするベータ分布 $\mathrm{Beta}(\alpha, 1)$ からサンプリングされるとする．ある π_k について，列 k 内の各 Z_{nk} は，独立にベルヌーイ分布 $\mathrm{Bernoulli}(\pi_k)$ からサンプリングされるものとする．この過程を経て，連続確率 π_k は減少し，データセット上の特徴 k を観測する確率は指数関数的に減少する．

MedLFRM におけるリンクの生成過程を以下に示す．

1. 潜在特徴行列の 1 行目である Z_1 に対して，
   (a) ポアソン分布 $\mathrm{Poisson}(\beta)$ から成分数 M を選択
   (b) ベータ分布 $\mathrm{Beta}(\alpha, 1)$ から，各成分 $i \in \{1, \ldots, M\}$ に対してパラメータ ν_i を選択
   (c) 棒折り過程 $\mathrm{SBP}(\nu)$ から，各成分 $k \in \{1, \ldots, M\}$ に対してパラメータ π_k を生成
   (d) ベルヌーイ分布 $\mathrm{Bernoulli}(\pi_k)$ から，各成分 $Z_{1k} \in \{Z_{11}, \ldots, Z_{1M}\}$ を選択
2. 潜在特徴行列の n 行目である Z_n に対して，
   (a) これまで選択されていない成分に対してはポアソン分布 $\mathrm{Poisson}(\beta/n)$ から，既に選択されたことのある成分に対してはその時用いられた確率から成分数 M を選択
   (b) ベータ分布 $\mathrm{Beta}(\alpha, 1)$ から，各成分 $i \in \{1, \ldots, M\}$ に対してパラメータ ν_i を選択
   (c) 棒折り過程 $\mathrm{SBP}(\nu)$ から，各成分 $k \in \{1, \ldots, M\}$ に対してパラメータ π_k を生成
   (d) ベルヌーイ分布 $\mathrm{Bernoulli}(\pi_k)$ から，各成分 $Z_{1k} \in \{Z_{11}, \ldots, Z_{1M}\}$ を選択
3. 正規分布 $\mathcal{N}(0, \sigma_w^2)$ から，重み行列 W の各成分を選択
4. 正規分布 $\mathcal{N}(0, \sigma_w'^2)$ から，重みベクトル η の各成分を選択
5. リンク評価関数 $\Phi(Z_i W Z_j^\top + \eta^\top X_{ij})$ を用いて，各ノード対 $(i,j) \in \mathcal{I}$ に対して応答変数 Y_{ij} を生成

補助変数を導入することによって拡張された問題を以下に示す．

$$\min_{p(\nu,Z,\Theta)} \mathrm{KL}(p(\nu, Z, \Theta)||p_0(\nu, Z, \Theta)) + C\mathcal{R}_\ell(p(Z,\Theta)) \tag{8}$$

ここで，$p_0(\nu, Z, \Theta) = p_0(\nu) p(Z|\nu) p_0(\Theta)$

3 提案手法

前章ではネットワークデータに対しての MedLFRM によるリンク予測問題について述べた．この章では，連続値で表現された関係属性に加え，離散関係属性がリンクに付与されている場合を想定したリンク予測問題について述べる．またそのようなモデルを用いたネットワークの時系列解析についても述べる．

3.1 MedLFRM による連続値・離散関係属性付きネットワークの解析

前章で述べていたモデル [8] では関係属性は連続値のみで表現されていたが，実データセットではしばしば関係属性の中に離散表現されているものが存在することもある．したがって，そのようなデータセットに対するリンク予測問題を扱うとき，関係属性を連続値表現とも離散表現ともする必要がある．したがって，X_{ij} はこれまで $X_{ij} \in \mathbb{R}^D$ と定義されていたが，X^c_{ij} を離散表現された関係属性，X^r_{ij} を連続値表現された関係属性とすると，$X^c_{ij} \in \mathbb{I}^D, X^r_{ij} \in \mathbb{R}^{D'}$ といったように再定義し議論を行う．この時，識別関数内の関係属性の重みベクトル η は連続値関係属性と離散関係属性それぞれに定義する．よって，識別関数は以下のようになる．

$$\begin{aligned} & f(Z_i, Z_j; X^c_{ij}, X^r_{ij}, W, \eta^c, \eta^r) \\ = \;& Z_i W Z_j^\top + \eta^{c\top} X^c_{ij} + \eta^{r\top} X^r_{ij} \\ = \;& \mathrm{Tr}(W Z_j^\top Z_i) + \eta^{c\top} X^c_{ij} + \eta^{r\top} X^r_{ij} \end{aligned} \quad (9)$$

ここで，$\Theta = \{W, \eta^c, \eta^r\}$ をすべてのパラメータとし，Θ と事前分布 $p_0(\Theta)$ は確率変数とする．2.3 節と同様に，ここからより効果的な識別関数を $p(Z, \Theta)$ に関する期待値として以下のように定義する．

$$f(X^c_{ij}, X^r_{ij}) = \mathbb{E}_{p(Z,\Theta)}[f(Z_i, Z_j; X^c_{ij}, X^r_{ij}, \Theta)] \quad (10)$$

したがって，応答変数 Y の予測値は $\hat{Y}_{ij} = \mathrm{sign} f(X^c_{ij}, X^r_{ij})$ となる．$\mathcal{I}$ を観測されたリンクの組の集合とし，ヒンジ損失関数を 2.2 節と同様に定義すると，ヒンジ損失は以下のようになる．

$$\mathcal{R}_\ell(p(Z, \Theta)) = \sum_{(i,j)\in\mathcal{I}} h_\ell(Y_{ij} f(X^c_{ij}, X^r_{ij})) \quad (11)$$

3.2 MedLFRM によるパラメータの推定

次に MedLFRM によるパラメータの推定方法について述べる．MedLFRM を提案した Zhu の手法 [8] に従った推定方法を以下に述べる．

切断平均場近似（truncated mean field approximation）[10] によって，$p(\nu, Z, \Theta)$ を次のように表す．

$$p(\nu, Z, \Theta) = p(\Theta) \prod_{k=1}^{K} p(\nu_k|\gamma_k) (\prod_{i=1}^{N} p(Z_{ik}|\psi_{ik})) \quad (12)$$

ここで，$p(\nu_k|\gamma_k)$ はベータ分布 $\mathrm{Beta}(\gamma_{k1}, \gamma_{k2})$ からサンプリングされたもの，$p(Z_{ik}|\psi_{ik})$ は $\mathrm{Bernoulli}(\psi_{ik})$ からサンプリングされたものである．K は切断レベルである．これらを踏まえて，MedLFRM の問題は以下の手順を反復することで解くことができる．

1. $p(\Theta)$ の推定
   $p(\nu, Z)$ が与えられたとき，部分問題を以下の制約の形で書くことができる．

$$\begin{aligned} & \min_{p(\Theta),\xi} \mathrm{KL}(p(\Theta)||p_0(\Theta)) + C \sum_{(i,j)\in\mathcal{I}} \xi_{ij} \\ & \forall (i,j) \in \mathcal{I},\ \text{s.t.} : Y_{ij}(\mathrm{Tr}(\mathbb{E}[W]\bar{\mathbf{Z}}_{ij}) + \mathbb{E}[\eta^c]^\top X^c_{ij} + \mathbb{E}[\eta^r]^\top X^r_{ij}) \\ & \qquad \geq \ell - \xi_{ij} \end{aligned} \quad (13)$$

   ここで，$\bar{\mathbf{Z}}_{ij} = \mathbb{E}_p[Z_j^\top Z_i]$ は潜在特徴ベクトルの内積の期待値であり，$\xi = \{\xi_{ij}\}$ はソフトマージンを実現するためのスラック変数である．ラグランジュの双対理論を用いることで，$p(\Theta)$ の最適解を得ることができる．$p(\Theta)$ は以下のように表すことができる．

$$\begin{aligned} p(\Theta) \propto p_0(\Theta) \mathrm{exp} \Bigg\{ & \sum_{(i,j)\in\mathcal{I}} \omega_{ij} Y_{ij} (\mathrm{Tr}(W \bar{\mathbf{Z}}_{ij})) \\ & + \eta^{c\top} X^c_{ij} + \eta^{r\top} X^r_{ij} \Bigg\} \end{aligned}$$

   $\omega = \{\omega_{ij}\}$ はラグランジュ乗数である．
   ここで，η を η^c と η^r を連接して得たベクトルであるとする．一般に使用される標準正規事前分布 $p_0(\Theta)$ により，$p(\Theta)$ の最適解を得ると以下のように表すことができる．

$$p(\Theta) = p(W)p(\eta) = \left(\prod_{kk'} \mathcal{N}(\Lambda_{kk'}, 1) \right) \left(\prod_{d} \mathcal{N}(\kappa_d, 1) \right)$$

   ここで，$\mathcal{N}(\Lambda_{kk'}, 1), \mathcal{N}(\kappa_d, 1)$ のそれぞれの期待値は，$\Lambda_{kk'} = \sum_{(i,j)\in\mathcal{I}} \omega_{ij} Y_{ij} \mathbb{E}[Z_{ik} Z_{jk'}]$，$\kappa_d = \sum_{(i,j)\in\mathcal{I}} \omega_{ij} Y_{ij} (X^c_{ijd} + X^r_{ijd})$ とする．双対問題は以下のようになる．

$$\begin{aligned} & \max_{\omega} \sum_{(i,j)} \ell \omega_{ij} - \frac{1}{2}(||\Lambda||_2^2 + ||\kappa||_2^2) \\ & \text{s.t.} : 0 \leq \omega_{ij} \leq C, \forall (i,j) \in \mathcal{I} \end{aligned}$$

   この時，部分問題は等価的に以下のように書き換えることができ，これを解くことでパラメータ Λ

と κ を直接求めることができる.

$$\min_{\Lambda,\kappa,\xi} \frac{1}{2}(||\Lambda||_2^2 + ||\kappa||_2^2) + C \sum_{(i,j)\in\mathcal{I}} \xi_{ij}$$
$$\forall (i,j) \in \mathcal{I}, \text{s.t. } : Y_{ij}(\mathrm{Tr}(\Lambda \bar{\mathbf{Z}}_{ij}) + \kappa^\top (X_{ij}^c + X_{ij}^r)) \geq \ell - \xi_{ij} \tag{14}$$

これは, SVM (Support Vector Machine) の二値分類問題の形式と一致しているため, SVMLight や LIBSVM などの既存の高性能なソルバーによって解くことができる.

2. $p(\nu, Z)$ の推定

$p(\Theta)$ が与えられると, 部分問題は以下のようになる.

$$\min_{p(\nu,Z)} \mathrm{KL}(p(\nu,Z)||p_0(\nu,Z)) + C\mathcal{R}_\ell(p(Z,\Theta))$$

切断平均場近似より, 以下の式が得られる.

$$\mathrm{Tr}(\Lambda \bar{\mathbf{Z}}_{ij}) = \begin{cases} \psi_i \Lambda \psi_j^\top & \text{if } i \neq j \\ \psi_i \Lambda \psi_i^\top + \sum_k \Lambda_{kk} \psi_{ik}(1-\psi_{ik}) & \text{if } i = j \end{cases}$$

マージンの制約は ν に依存しないので, $p(\nu)$ は Doshi-Velez ら [11] と同様にして解を得ることができる.
また, 劣勾配法を用いることで $p(Z)$ を解くことができる. ここで, 観測されたリンクの組の集合 $\mathcal{I}$ を以下のように定義する.

$$\mathcal{I}_i = \{j : j \neq i,\ (i,j) \in \mathcal{I} \ and\ Y_{ij} f(X_{ij}^c, X_{ij}^r) \leq \ell\}$$
$$\mathcal{I}_i^{'} = \{j : j \neq i,\ (j,i) \in \mathcal{I} \ and\ Y_{ji} f(X_{ji}^c, X_{ji}^r) \leq \ell\}$$

また, $g(x) \leq \ell$ であるとき, $\partial_x h_\ell(g(x))$ は $-\partial_x g(x)$ に等しく, そうでなければ 0 となる. これより, $\mathcal{R}_\ell$ の劣勾配は以下のようになる.

$$\partial_{\psi_{ik}} \mathcal{R}_\ell = -\sum_{j\in\mathcal{I}_i} Y_{ij} \Lambda_{k\cdot} \psi_j^\top - \sum_{j\in\mathcal{I}_i^{'}} Y_{ji} \psi_j \Lambda_{\cdot k}$$
$$-\mathbb{I}(Y_{ii} f(X_{ii}^c, X_{ii}^r) \leq \ell) Y_{ii} (\Lambda_{k\cdot} \psi_i^\top + \psi_i \Lambda_{\cdot k} + \Lambda_{kk}(2 - \psi_{ik}))$$

ここで, $\Lambda_{k\cdot}$ は Λ の k 番目の行を表し, $\Lambda_{\cdot k}$ は Λ の k 番目の列を表す. また, $\mathbb{I}(\cdot)$ は指示関数であり, 括弧内に示された命題が真であるとき 1, そうでないとき 0 を返す. 部分問題の劣勾配を 0 とすると, 以下の ψ_{ik} の更新式を得る.

$$\psi_{ik} = \Phi\left(\sum_{j=1}^{k} \mathbb{E}_p[\log \nu_j] - \mathcal{L}_k^\nu - C\partial_{\psi_{ik}} \mathcal{R}_\ell\right) \tag{15}$$

ここで, $\mathcal{L}_k^\nu$ は $\mathbb{E}_p[\log(1 - \prod_{j=1}^k \nu_j)]$ の下限である.

3.3 MedLFRM による時系列ネットワークの解析

ここまでは一つのネットワークにおけるリンク予測問題について着目してきた. この節では複数のネットワークを時系列的に捉えたときの予測について述べる.

これまで述べられてきていたモデルにおいては, 潜在特徴行列 Z は ψ をハイパーパラメータとする Bernoulli(ψ) からサンプリングされたものであり, この ψ をモデルの中でランダムに初期化, 学習し予測を行っていた. しかし, 複数のネットワークを時系列的に捉えたとき, 前時区間のデータの ψ の学習結果を当時区間においての初期状態として据えることで時間依存性を考慮する.

また, 識別関数 F 内の実数値重み行列 W, 実数値重みベクトル η もモデル内でランダムに初期化, 学習されるものであるが, この二つについても先ほど述べたように前時区間のデータの学習結果を当時区間においての初期状態に据えることで時間依存性を考慮することができ, 予測精度の向上につながると考える.

4 実験

この章では実データセットを用いて, 時間 $t-1$ までの学習結果を用いて時間 t のデータを学習し, 時間 $t+1$ におけるリンクを予測する実験を行い, その結果について考察する.

4.1 データセット

実験には, 欧州債務危機が起こった 2009 年 7 月から 2012 年 12 月における欧州銀行間での取引を記録したデータセットを用いる. 通年のデータを月ごとに集計し, 12 個のデータセットとして扱う. データセットには 153 の銀行と 14 の国が含まれる. データセットに含まれる各銀行をノード, 当月にある銀行と別の銀行との間に現金のやり取りがあった場合, そこにリンクがあると見なす. この時, 現金のやり取りというのは銀行 A から銀行 B へと x ユーロの現金が送られたというような状態を指す. したがって, このネットワークは有向グラフとなり, 隣接行列は非対称行列となる. リンクに付与される関係属性は, 取引された現金の量, その取引が行われた際の金利, そして取引が同じ国内の銀行間で行われたのであれば 1, 異国の銀行間で行われたのであれば-1 をとるような二値変数の 3 つとする.

4.2 実験設定

次にパラメータの設定について述べる. 切断レベル K は 40 よりも大きくすれば十分に良い結果を得られ

表 1: 提案手法と従来手法による翌月の取引有無に関する予測結果

	MAP	average AUC
time-dependent model	0.387637	0.919695
time-independent model	0.346412	0.905515

ることが知られている [8] ことから，本実験においても $K = 50$ と設定する．損失パラメータ ℓ, ハイパーパラメータ α は全ての実験において $\ell = 1, \alpha = 0.1$ と設定した．また，このデータセットは不均衡である．(つまり，正例よりも負例の方が多く存在する．) したがって，正のデータに対しては C^+，負のデータに対しては C^- という異なった正則化定数を用いることにし，$C^- = 0.1$ とした上で $C^+ = 10C^-$ とした．実験を行う時，先に述べたように前月のパラメータの学習結果を当月の初期状態として設定する．ただし，1 回目の実験を行う時，つまり 1 月のデータセットに対して実験を行う際は，W は [0,0.1] の区間で一様に初期化，ψ は 0.5 に [0,0.001] の区間で一様に分布したランダムノイズを加えたものとなるように初期化，η は平均が 0 となるように初期化する．

4.3 評価方法と結果

先に述べたパラメータを用いて実験を行う．また，それとは別に前月のパラメータの学習結果を用いず，ランダムに初期化した上で学習していく従来の手法でも実験を行う．この二つの結果の MAP（mean average precision）を比較する．MAP は予測精度評価法の 1 つであり，[0,1] の範囲の値を取りうるもので，1 に近いほど精度がよいということになる．また同様にして得られた AUC（Area Under the Curve）の平均（12ヶ月分）も評価指標として用いる．AUC も予測精度評価法の一つであり，ROC（Receiver Operating Characteristic）曲線の積分値を表すものである．完全に理想的な予測をした場合に 1 を，完全にランダムな予測をした場合に 0.5 をとる．このときの評価結果を表 1 に示す．

また，4.1 節で述べた 3 つの関係属性のうち，どれか 1 つを考慮せずにリンク予測を行うという実験を全ての関係属性に対して行い，その結果を比較する．この実験を行う際も前月のパラメータの学習結果を用いる場合と用いない場合に分けて実験を行う．このときの評価結果を表 2〜表 4 に示す．

4.4 考察

表 1 から，前月のパラメータの学習結果を当月の初期状態に据えることで，ランダムに初期化を行って予測する場合よりも精度が改善されていることが確認できる．次に，AUC よりも MAP の方が比較的向上している点に関して考察する．まず AUC はリンクが有るノード対にリンクが有ると予測できた場合とリンクが無いノード対にリンクが無いと予測できた程度が全ノード対を占める割合を表す指標であり，MAP に関してはリンクが有るところを有ると予測できた割合だけに注目した指標である．そして 4.2 節で述べたようにこのデータセットは不均衡なものであるため，リンクが有るということを予測できたことにより価値があると考えられる．よって，今回の実験においてより評価指標として意味を持つのは MAP の方であるといえる．その MAP により大きな改善が見られたことから，前月のパラメータの学習結果を当月のパラメータの初期状態として用いることで効果的に時間依存性を反映させることができたと言える．さらに，各関係属性を考慮せずに行った実験結果について考察する．金利を考慮せずに行った場合に比較的大きく予測精度が損なわれていることから，取引の有無を予測するという観点においては，金利がどの程度であったかという情報がより重要なのだと予想することができる．逆に，取引量を考慮せずに行った場合には比較的予測精度に差は見られないため，過去の取引量は将来の取引の有無を予測するにあたってはさほど重要ではないと解釈できる．

表 2: 国際取引，国内取引の別を考慮しない場合の予測結果

	MAP	average AUC
time-dependent model	0.375433	0.915590
time-independent model	0.347404	0.906652

表 3: 取引量を考慮しない場合の予測結果

	MAP	average AUC
time-dependent model	0.380175	0.918378
time-independent model	0.346403	0.905512

表 4: 金利を考慮しない場合の予測結果

	MAP	average AUC
time-dependent model	0.362422	0.911382
time-independent model	0.347434	0.905525

5 おわりに

本稿では，実問題において想定される連続値と離散値で表現される関係属性が混在する場合のネットワークに対するリンク予測問題について，マージン最大化潜在特徴関係モデル（MedLFRM）を拡張し，予測精度の評価を行った．そしてそのようなネットワークの時系列解析についても検討した．

銀行間の取引データを用いて，前時区間において学習したパラメータを当時区間のパラメータの初期状態として据えながらリンク予測問題の実験を行い，各月の AUC の平均と，MAP を用いて評価を行った．比較対象として，従来の手法と同じく時系列を考慮せず，パラメータをランダムに初期化するリンク予測問題の実験も行った．その結果，前時区間のパラメータの学習結果を用いた方が予測精度が良くなることが確認された．これより，ネットワークを時系列的に捉えて未知パラメータや潜在変数を学習することで，リンク予測の性能を改善できることを示した．

今後の展望として，損失パラメータ ℓ や正則化定数 C^- の適切な値を交差検定などの手法によって決定し，リンク予測問題において精度の更なる向上を試みることが挙げられる．また，本稿ではリンク予測問題について議論を行ったが，リンクに付与された関係属性の予測を行うといったことも考えられる．この場合，関係属性が連続値で表現されているのか離散値で表現されているのかによって，問題設定を回帰または分類に区別して考える必要がある．関係属性の予測を行うことによりノード間にあるリンクがどのようなものであるのかを予測でき，さらに本稿において用いた時系列的な考えを導入することで，より詳細で広範囲なネットワーク解析が可能になると考えられる．

謝辞

本研究を行うにあたり，助言と協力を頂いた神戸大学大学院経済学研究科の羽森茂之教授と金京拓司教授，同大学大学院システム情報学研究科谷口隆晴准教授に感謝する．本研究の一部は科学研究費補助金基盤研究 (B)(15H02703) の援助による．

参考文献

[1] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. Journal of the American society for information science and technology, Vol. 58, No. 7, pp. 1019.1031, 2007.

[2] Lars Backstrom and Jure Leskovec. Supervised random walks: predicting and recommending links in social networks. In Proceedings of the fourth ACM international conference on Web search and data mining, pp. 635.644. ACM, 2011.

[3] Kurt T Miller, Thomas L Griffiths, and Michael I Jordan. Nonparametric latent feature models for link prediction. In Advances in Neural Information Processing Systems, Vol. 22, pp. 1276.1284, 2009.

[4] Peter Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In Advances in Neural Information Processing Systems, Vol. 20, pp.737.744, 2007.

[5] Thomas L Griffiths and Zoubin Ghahramani. Infinite latent feature models and the indian buffet process. In Advances in Neural Information Processing Systems, Vol. 18, pp. 475.482, 2005.

[6] Tommi Jaakkola, Marina Meila, and Tony Jebara. Maximum entropy discrimination.In Advances in Neural Information Processing Systems, Vol. 12, pp.470.476, 1999.

[7] Tony Jebara. Machine learning: discriminative and generative. Springer, 2004.

[8] Jun Zhu. Max-margin nonparametric latent feature models for link prediction.In Proceedings of the 29th International Conference on Machine Learning, 2012.

[9] Yee W Teh, Dilan G¨or¨ur, and Zoubin Ghahramani. Stick-breaking construction for the indian buffet process. In Proceedings of the 11th International Conference on Arti

cial Intelligence and Statistics, pp. 556.563, 2007.

[10] Christopher M Bishop, et al. Pattern recognition and machine learning. springer New York, 2006.

[11] Finale Doshi, Kurt Tadayuki Miller, Jurgen Van Gael, and Yee Whye Teh. Variational inference for the indian buffet process. In Proceedings of the 12th International Conference on Arti

cial Intelligence and Statistics, pp. 137.144,2009.