3D4-6

リンク予測のための ノード情報とリンク構造を組み合わせたペアワイズカーネル

A combination pairwise kernel of node information and link structure for link prediction

長谷川 聡[*1]　佐久間 淳[*1]
Satoshi Hasegawa　Jun Sakuma

[*1]筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻
Dept. of Computer Science, Graduate school of SIE, Univ. of Tsukuba

Link prediction problem can be solved as supervised pairwise classification problem. Definition of pairwise kernel is important , because pairwise classification problem is only solved by using a pairwise kernel and labels. Existing pairwise kernels only use node information. We propose pairwise kernel using node information and link structure. Experiment result shows our pairwise kernel in predictive accuracy is better than existing pairiwise kernel.

1. はじめに

世界中には, ネットワーク構造で表現できるものが溢れており, リンクはノード間の関係性を表すのに重要な役割を持っている. 例えばSNSにおいてリンクは, ユーザ間の友達関係を表しており, またオンラインショッピングにおいてリンクは, ユーザがある商品を買うという購買関係を表している. 仮にSNSにおいてリンクの有無を予測できるとするならば, 新たな友人を予測することに相当し, それにより交友関係を広げることができる. それゆえリンクを予測することは新たな知識発見につながることから, 研究が盛んに行われている.

リンク予測は, ネットワークの一部のリンクが与えられており, それらを用いてノードペア間にリンクがあるかどうかを予測する問題である. 既知のリンクが教師となりリンクの有無を予測することから, 教師付き 2 値分類問題と定式化できる. 与えられるリンク構造を基に予測を行うことから, リンク構造は重要な情報である.

リンク構造以外にノード特徴情報が与えられている場合, そのノード特徴情報を用いることで, ノードペア同士の関係性を表すペアワイズカーネルおよびリンクの有無を表すラベルにより, リンク予測問題を教師付きペアワイズ 2 値分類問題と定式化できる. しかしペアワイズカーネルは, ノード同士の関係性を表すカーネルと比べて記憶容量が 2 乗必要になり, また計算時間もそれに伴うことから, 大規模ネットワークにおけるリンク予測では工夫が必要である.

既存のペアワイズカーネルの計算量的性質を, 表 1 にまとめる. ノードペアの特徴ベクトル同士の近さを考慮したペアワイズカーネルとして距離学習ペアワイズカーネル [7] が提案されている. このペアワイズカーネルは 2 部グラフの予測のような, V_1 と V_2 が異なるケースの場合用いることができない. 他にクロネッカー演算に基づくペアワイズカーネルとして, クロネッカー積カーネル [1][5], カルテシアンカーネル [3] が提案されている. これら既存のペアワイズカーネルは, 共通してノード情報のみを用いたペアワイズカーネルであり, リンク予測の際に重要であるリンク構造は明示的に用いていない.

本研究では, リンク構造を組み合わせることで, リンク予測精度の向上を期待する新たなペアワイズカーネルの提案を行う. 既存のノード情報に基づくペアワイズカーネルではリンク構造を用いていないため, 対象となるノードペアの情報しか用いることができなかったが, リンク構造を用いることで周りのノードペアの情報も用いるペアワイズカーネルとなり, 予測精度の向上が期待できる.

提案するペアワイズカーネルは, ペアワイズカーネル同士の行列積が伴うことから時間計算量が $O(N^3M^3)$ となり, 表 1 に示す既存のペアワイズカーネルと比べ劣る. そこで対角行列を 2 つの行列のクロネッカー積に分解により低ランク近似を行い, クロネッカー積カーネルと同等の時間計算量である $O(N^2M^2)$ に抑える方法を提案する.

提案したノード情報とリンク構造を組み合わせペアワイズカーネルが, リンク構造を加えたことにより予測精度が上がるかどうかを調べるため, 3 つのベンチマークデータを用意し, 既存のペアワイズカーネルと予測精度の比較を行った. その結果提案ペアワイズカーネルの予測精度が良い結果を示した.

2. リンク予測

この章では, リンク予測の問題定義を行い, ペアワイズカーネルを用いたリンク予測法について記す.

2.1 問題定義

リンク予測問題をペアワイズ分類問題として定式化を行う. まずノード集合 $V_1 = \{v_1^{(1)}, ..., v_1^{(N)}\}$, ノード集合 $V_2 = \{v_2^{(1)}, ..., v_2^{(M)}\}$ を考える. そしてリンクの集合を $\mathbf{E}$ とし, ノードペア $(v_1^{(i)}, v_2^{(j)})$ 間のリンクを $e_{ij} \in E$ と表す. グラフ G を $G = (V_1, V_2, E)$ と表す. またノードペア $v_1^{(i)}, v_2^{(j)}$ 間にリンクがあることを示すラベルを $y_{ij} = +1$, リンクがないことを $y_{ij} = -1$ と表す. 我々は予測関数 $f : V_1 \times V_2 \rightarrow \{-1, 1\}$ を作ることを目標とする.

ここで, $V_1 = V_2$ の際は 1 ネットワークのリンク予測であり, $V_1 \cap V_2 = 0$ の際は 2 部グラフのリンク予測である.

2.2 ペアワイズカーネル法

本論文で取り扱うリンク予測は, 補助情報として, ノード集合 V_1, V_2 の特徴情報が与えられているとする. そしてそのノード情報から非負のカーネル $k_1 : V_1 \times V_1 \rightarrow \mathbb{R}$, $k_2 : V_2 \times V_2 \rightarrow \mathbb{R}$ を考える. このカーネルに対応するグラム行列を $\mathbf{K}_1$, $\mathbf{K}_2$ とする. ノードの特徴情報に基づくカーネルであることからノード情報カーネルと呼ぶ. また, グラム行列の要素 $[\mathbf{K}_1]_{ij}$, $[\mathbf{K}_2]_{kl}$

連絡先: 長谷川 聡 筑波大学大学院システム情報工学研究科 茨城県つくば市天王台 1-1-1 hasegawa(at)mdl.cs.tsukuba.ac.jp

表 1: Compare to pairwise kernel

pairwise kernel	spase complexity	time complexity	node information	link structure
metric learning pairwise kernel[7]	$O(N^4)$	$O(N^4)$	∘	×
kronecker product kernel[1][5]	$O(N^2M^2)$	$O(N^2M^2)$	∘	×
cartesian kernel[3]	$O(N^2M+NM^2)$	$O(N^2M+NM^2)$	∘	×
tensor diffusion kernel(proposal)	$O(N^2M^2)$	$O(N^2M^2)$	×	∘
tensor propagation kernel(proposal)	$O(N^2M^2)$	$O(N^2M^2)$	∘	∘

は, ノード $v_1^{(i)}, v_1^{(j)}$ 間の類似度, ノード $v_2^{(k)}, v_2^{(l)}$ 間の類似度をそれぞれ表す.

このノードのカーネル k_1, k_2 を用いることにより, ノードペア間のカーネルであるペアワイズカーネル $k^{\mathrm{pair}} : (V_1 \times V_1) \times (V_2 \times V_2) \to \mathbb{R}$ を生成する. モデルパラメータ $\boldsymbol{\alpha} = (\alpha_{1,1}, ..., \alpha_{N,M})$, ペアワイズカーネル $k^{\mathrm{pair}}((a,b),(c,d))$ により構成される予測関数 f によってリンクの有無を予測する.

$$f(a,b) := \sum_{c,d} \alpha_{c,d} k^{\mathrm{pair}}((a,b),(c,d)) \tag{1}$$

次の節ではペアワイズカーネルの具体例として, グラム行列同士のクロネッカー積演算を用いたペアワイズカーネルについて示す.

2.3 クロネッカー積演算を用いたペアワイズカーネル

対象となるノードペア $(v_1^{(i)}, v_2^{(k)})$ とノードペア $(v_1^{(j)}, v_2^{(l)})$ の類似度を考えた場合, ノードペア $(v_1^{(i)}, v_1^{(j)})$ が似ており, かつノードペア $(v_2^{(k)}, v_2^{(l)})$ が似ていれば, ペア同士が似ているとするのが自然であるといえる. これはグラム行列 $\mathbf{K}_1$ と $\mathbf{K}_2$ のクロネッカー積演算で表現できることから, クロネッカー積カーネルと呼ばれている.

$$\mathbf{K}_{\mathrm{kp}}^{\mathrm{pair}} := \mathbf{K}_2 \otimes \mathbf{K}_1. \tag{2}$$

このクロネッカー積カーネルの要素は以下の式で計算できる.

$$[\mathbf{K}_{\mathrm{tp}}^{\mathrm{pair}}]_{ijkl} := [\mathbf{K}_2]_{kl}[\mathbf{K}_1]_{ij} \tag{3}$$

3. ノード情報とリンク構造を含むペアワイズカーネル

既存のペアワイズカーネルはノード特徴情報のみを用いた設計になっており, リンク予測の際に重要となるリンク構造を用いていない. このため, ペアワイズカーネル行列の値 $[\mathbf{K}^{\mathrm{pair}}]_{ijkl}$ を決める際に, 周りのノードペアとの構造的な関係を考慮していないことから, $(v_1^{(i)}, v_1^{(j)})$ と $(v_2^{(k)}, v_2^{(l)})$ のノード特徴情報しか用いることができなかった. しかしながらノードペア同士の類似度を与える際, 対象となるノードペア同士だけでなくリンク構造上関連の高いノードペアの情報も有用であると言える. 例えば, 共著ネットワークの共著関係の予測を考えた際, 対象となる著者の情報だけでなく, 他に共著した人の情報も類似度に影響すると想像できる.

そこでノード特徴情報にリンク構造を加えることで, 対象となるノードペアだけでなくリンク構造上関連の高いノードペアの情報も用いたペアワイズカーネルを提案する. 提案するペアワイズカーネルは, ノード情報のペアワイズカーネルであるクロネッカー積カーネル $\mathbf{K}_{\mathrm{kp}}^{\mathrm{pair}}$ に対し, リンクがランダムウォークに従って遷移することを表すペアワイズカーネル $\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}}$ を挟んで行列積を行うことにより表現する.

$$\mathbf{K}_{\mathrm{tp}}^{\mathrm{pair}} := \mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} \mathbf{K}_{\mathrm{kp}}^{\mathrm{pair}} \mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} \tag{4}$$

式 4 はクロネッカー積演算について $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A}^{\mathrm{T}})\mathrm{vec}(\mathbf{B})$ が成り立つことに注意すると

$$\mathrm{vec}(\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} \mathbf{K}_{\mathrm{kp}}^{\mathrm{pair}} \mathbf{K}_{\mathrm{td}}^{\mathrm{pair}}) = (\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} \otimes \mathbf{K}_{\mathrm{td}}^{\mathrm{pair}})\mathrm{vec}(\mathbf{K}_{\mathrm{kp}}^{\mathrm{pair}})$$

と書き表せる. これはリンクの遷移同士の関係を表現するクアッドワイズカーネルとノード特徴情報のペアワイズカーネルのベクトル表現との積であることがわかる. クアッドワイズカーネルの各行がリンクの遷移を表しているため, それとノード特徴情報のペアワイズカーネルとの積を行うことにより, ノード特徴情報がリンクの遷移に従い伝播すると言える. すなわち式 (4) は, リンクの遷移に従いノード特徴情報が伝搬し, 対象のノードペアのノード特徴情報だけでなく, リンク構造上関連の高いノード特徴情報も考慮したペアワイズカーネルである. クロネッカー積演算 (テンソル積演算) を伴い, またノード特徴情報が伝搬することから, テンソル伝搬ペアワイズカーネルと呼ぶ.

本稿ではテンソル伝搬ペアワイズカーネルの計算のため, 2 つの提案を行う. それぞれ 3.1 節, 3.2 節に示す.

テンソル伝搬ペアワイズカーネルを計算するため, まずリンクがランダムウォークに従って遷移することを表すペアワイズカーネル $\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}}$ の提案を行う (3.1 節).

テンソル伝搬ペアワイズカーネルの計算は, ペアワイズカーネル ($NM \times NM$ 行列) 同士の行列積を伴うことから, ナイーブに計算すると時間計算量が $O(N^3M^3)$ となり, 現実的に計算が困難である. そこで $\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} = \mathbf{T}_2 \otimes \mathbf{T}_1$ とし, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ を用いることで, 行列のクロネッカー積にする.

$$\begin{aligned} \mathbf{K}_{\mathrm{tp}}^{\mathrm{pair}} &= (\mathbf{T}_2 \otimes \mathbf{T}_1)(\mathbf{K}_2 \otimes \mathbf{K}_1)(\mathbf{T}_2 \otimes \mathbf{T}_1) \\ &= (\mathbf{T}_2\mathbf{K}_2\mathbf{T}_2) \otimes (\mathbf{T}_1\mathbf{K}_1\mathbf{T}_1) \end{aligned} \tag{5}$$

これにより, $N \times N$ の行列と $M \times M$ の行列とのクロネッカー積となることから, 時間計算量を $O(N^2M^2)$ となる. しかしながら $\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} = \mathbf{T}_2 \otimes \mathbf{T}_1$ となる $\mathbf{T}_2$, $\mathbf{T}_1$ は実際には求めることができないことから, 行列の低ランク近似を用いることで, 近似的に $\mathbf{T}_2$ と $\mathbf{T}_1$ を求める方法を示す (3.2 節).

3.1 テンソル積拡散カーネル

この節ではリンクがランダムウォークに従って遷移することを表すペアワイズカーネルを提案するため, ノードがランダムウォークに従って遷移することを表す拡散カーネルを拡張する.

3.1.1 拡散カーネル

拡散カーネル $\mathbf{K}_{\mathrm{d}}$[4] は, ネットワークの構造を表す隣接行列 $\mathbf{A}$ を用いたグラフラプラシアン $\mathbf{L} = \mathbf{D} - \mathbf{A}(\mathbf{D}_{ii} = \sum_j [\mathbf{A}]_{ij})$

を用いて以下のように定義される．

$$\mathbf{K}_{\mathrm{d}} = \exp(-\beta\mathbf{L}) \tag{6}$$

拡散カーネルの値 $[\mathbf{K}_{\mathrm{d}}]_{ij}$ はノード i がノード j にランダムウォークで遷移する確率を表している．

拡散カーネルは行列の指数関数演算を伴うが, グラフラプラシアンの固有値分解 $\mathbf{L} = \mathbf{Q\Lambda Q}^{\mathrm{T}}$ を用いることにより,

$$\mathbf{K}_{D} = \mathbf{Q}\exp(-\beta\mathbf{\Lambda})\mathbf{Q}^{\mathrm{T}} \tag{7}$$

と計算できる．$\exp(-\beta\mathbf{\Lambda})$ は, $-\beta\mathbf{\Lambda}$ が対角要素のみ値を持つことから, 対角要素ごとに exp 計算をすれば良い．この拡散カーネル拡張し, リンクの遷移を表現する方法を試みる．

3.1.2 拡散ペアワイズカーネル

まず既存のネットワーク構造を表現する $N \times M$ の行列 $\mathbf{F}$ を以下のように定義する．

$$[\mathbf{F}]_{ij} := \begin{cases} 1 & y_{ij} = +1, \\ 0 & y_{ij} = -1. \end{cases} \tag{8}$$

行列 $\mathbf{F}$ は 2 部グラフを表していることから, V_1, V_2 のそれぞれのノードの間にはリンクは存在しない．しかしながら, もし V_1 の 2 ノードが V_2 の 1 つのノードに対しリンクを有する場合, V_1 の 2 ノードは似ていると考えることができる．この関係を co-citation という．V_1 の co-citaion を表現する行列を $\mathbf{G}_1 = \mathbf{FF}^{\mathrm{T}}$, V_2 の co-citation を表現する行列を $\mathbf{G}_2 = \mathbf{F}^{\mathrm{T}}\mathbf{F}$ とする．この $\mathbf{G}_1$, $\mathbf{G}_2$ を隣接行列とするグラフを考える．

リンク同士の関係は, 隣接行列 $\mathbf{G}_1$ と $\mathbf{G}_2$ のクロネッカー積からテンソル積グラフで表現できる．このテンソル積グラフは, V_1 の $(v_1^{(i)}, v_1^{(j)})$ にリンクがあり, V_2 の $(v_2^{(k)}, v_2^{(l)})$ にリンクがある場合, ノードペア $(v_1^{(i)}, v_2^{(k)}), (v_1^{(j)}, v_2^{(l)})$ 同士に関係があるとする．すなわちリンクがあるノードペア同士に関係があるとするグラフである．

このリンク同士の関係を表すテンソル積グラフとノードの遷移を表現する拡散カーネルを組み合わせることで, リンクの遷移を表現するテンソル積拡散カーネルを提案する．このテンソル積拡散カーネルは, 隣接行列のクロネッカー積の正規化ラプラシアン及び, 式 (6) を用いることで表現する．クロネッカー積の正規化ラプラシアン [6] は以下の式で得られることが知られている．

$$\mathbf{L} \quad = \quad \mathbf{I}_2 \otimes \mathbf{I}_1 - (\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1) \tag{9}$$

ここで, $\tilde{\mathbf{G}}_2 = \mathbf{D}_2^{\frac{1}{2}}\mathbf{G}_2\mathbf{D}_2^{\frac{1}{2}}$, $\tilde{\mathbf{G}}_1 = \mathbf{D}_1^{\frac{1}{2}}\mathbf{G}_1\mathbf{D}_1^{\frac{1}{2}}$ である．このクロネッカー積正規化ラプラシアンを用いて, テンソル積拡散カーネルを式 (10) で定義する．

$$\mathbf{K}_{\mathrm{td}}^{\mathrm{pair}} := \exp\{-\beta(\mathbf{I}_2 \otimes \mathbf{I}_1 - (\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1))\}. \tag{10}$$

テンソル積拡散カーネルを式 (7) に基づいて計算を行うとした場合, $-\beta(\mathbf{I}_2 \otimes \mathbf{I}_1 - (\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1))$ の固有値分解の時間計算量が $O(N^3M^3)$, 空間計算量が $O(N^2M^2)$ であることから, スケーラビリティの問題が発生する．またこのままの状態では, 行列同士のクロネッカー積で表すことができない．そこで行列の低ランク近似手法を用いることにより, 時間計算量を抑え, 行列のクロネッカー積で表現する手法を提案する．

3.2 行列の低ランク近似を用いたテンソル積拡散カーネルの計算

この節では, テンソル積伝搬ペアワイズカーネルを高速に計算するため, 以下となる $\mathbf{T}_2$ および $\mathbf{T}_1$ を求めることを目標とする．

$$\exp\{-\beta(\mathbf{I}_2 \otimes \mathbf{I}_1 - (\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1))\} \approx \mathbf{T}_2 \otimes \mathbf{T}_1. \tag{11}$$

まず exp の中にある $\mathbf{I}_2 \otimes \mathbf{I}_1 - (\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1)$ を分け, $\exp(\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1)$ を固有値分解を行うことにより計算する (拡散カーネルを計算する式 (7) と同様)．$\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1$ の固有値分解は, $\tilde{\mathbf{G}}_2$, $\tilde{\mathbf{G}}_1$ の固有ベクトル並べた行列 $\mathbf{Q}_2$, $\mathbf{Q}_1$ 及び, 固有値を対角要素に持つ対角行列 $\mathbf{\Lambda}_2$, $\mathbf{\Lambda}_1$ を用いることにより,

$$\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1 = (\mathbf{Q}_2 \otimes \mathbf{Q}_1)(\mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)(\mathbf{Q}_2 \otimes \mathbf{Q}_1)^{\mathrm{T}} \tag{12}$$

となると知られている．それを用いることで

$$\begin{aligned} & \exp\{-\beta(\mathbf{I}_2 \otimes \mathbf{I}_1 - (\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1))\} \\ = \ & \exp\{-\beta(\mathbf{I}_2 \otimes \mathbf{I}_1)\}\exp\{\beta(\tilde{\mathbf{G}}_2 \otimes \tilde{\mathbf{G}}_1)\} \\ = \ & (\exp(-\tfrac{\beta}{2}\mathbf{I}_2)\otimes\exp(-\tfrac{\beta}{2}\mathbf{I}_1))(\mathbf{Q}_2\otimes\mathbf{Q}_1)(\exp\{\beta(\Lambda_2\otimes\Lambda_1)\})(\mathbf{Q}_2\otimes\mathbf{Q}_1)^{\mathrm{T}} \end{aligned}$$

となる．$\exp\{\beta(\Lambda_2 \otimes \Lambda_1)\}$ は対角行列であることから, $\exp\{\beta(\Lambda_2 \otimes \Lambda_1)\} = \mathbf{H}_2 \otimes \mathbf{H}_1$ なる対角行列 $\mathbf{H}_2$, $\mathbf{H}_1$ を求めることが出来れば,

$$\begin{aligned} & \exp(-\tfrac{\beta}{2}\mathbf{I}_2)\otimes\exp(-\tfrac{\beta}{2}\mathbf{I}_1)(\mathbf{Q}_2\otimes\mathbf{Q}_1)(\exp\{\beta(\Lambda_2\otimes\Lambda_1)\})(\mathbf{Q}_2\otimes\mathbf{Q}_1)^{\mathrm{T}} \\ = \ & \exp(-\tfrac{\beta}{2}\mathbf{I}_2)\otimes\exp(-\tfrac{\beta}{2}\mathbf{I}_1)(\mathbf{Q}_2\otimes\mathbf{Q}_1)(\mathbf{H}_2\otimes\mathbf{H}_1)(\mathbf{Q}_2\otimes\mathbf{Q}_1)^{\mathrm{T}} \\ = \ & (\exp(-\frac{\beta}{2}\mathbf{I}_2)\mathbf{Q}_2\mathbf{H}_2\mathbf{Q}_2^{\mathrm{T}}) \otimes (\exp(-\frac{\beta}{2}\mathbf{I}_1)\mathbf{Q}_1\mathbf{H}_1\mathbf{Q}_1^{\mathrm{T}}) \\ = \ & \mathbf{T}_2 \otimes \mathbf{T}_1 \end{aligned} \tag{13}$$

となり, 行列のクロネッカー積で表現することができる．

しかしながらそのような対角行列は一般的には求めることができない．というのも, $\exp(ab) = \exp(a)\exp(b)$ となる a, b を求めることと等価であり, 成り立つのは特殊なケースのみであるためである．そこで何らかの近似によるアプローチを用いて求めることを考える．

$\exp\{\beta(\Lambda_2 \otimes \Lambda_1)\}$ は対角行列であることから, $NM \times NM$ の行列を, 対角要素のみを表現する $N \times M$ の行列 $\mathbf{E}$ とする．

$$[\mathbf{E}]_{ij} = \exp\{\beta([\Lambda_2]_{ii}[\Lambda_1]_{jj})\}. \tag{14}$$

$\mathbf{E} \approx \mathbf{uv}^{\mathrm{T}}$ なる $\mathbf{u}$, $\mathbf{v}$ を求めることで, $\exp\{\beta(\Lambda_2 \otimes \Lambda_1)\} \approx \mathrm{diag}(\mathbf{u}) \otimes \mathrm{diag}(\mathbf{v})$ となることから, 求めたい $\mathbf{H}_2 = \mathrm{diag}(\mathbf{u})$, $\mathbf{H}_1 = \mathrm{diag}(\mathbf{v})$ となる．

$\mathbf{E}$ は各要素は非負であることから, 具体的な行列の近似手法として, 非負値行列分解を用いる．非負値行列分解は, 非負の行列 $\mathbf{E} \approx \mathbf{UV}^{\mathrm{T}}$ となる非負の $N \times r$ 行列 $\mathbf{U}$ と非負の $r \times M$ 行列に分解するアルゴリズムである．この非負値行列分解の rank-1 近似を行うことにより求めた $\mathbf{u}$, $\mathbf{v}$ が, $\mathbf{H}_2 = \mathrm{diag}(\mathbf{u})$, $\mathbf{H}_1 = \mathrm{diag}(\mathbf{v})$ である．

4. 実験

提案するテンソル伝搬ペアワイズカーネルおよびテンソル積拡散カーネルと既存のペアワイズカーネルとの予測精度の比較を行う．

ペアワイズカーネルは記憶容量が $O(N^2M^2)$ 必要であることから, 2 値分類で一般的に用いられているサポートベクター

表 2: Experiment result (AUC)

	metabolic	drug-target	nips
kronekcer product kernel	0.629(0.027)	**0.922(0.021)**	0.654(0.021)
cartesian kernel	0.633(0.022)	0.875(0.032)	0.610(0.023)
tensor diffusion kernel	0.595(0.026)	0.653(0.021)	0.592(0.012)
tensor propagation kernel	**0.653(0.021)**	**0.922(0.021)**	**0.667(0.021)**

マシンは使えない．そこでオンラインのマージン最大化学習である OnlinePassiveAggressiveAlgorithm[2](PA-I) を用いた．

またテンソル積拡散カーネルの $\mathbf{T}_2$ は, $[\mathbf{D}_2]_{ii} = \sum_j [\mathbf{T}_2]_{ij}$ を用いて, $\mathbf{D}_2^{\frac{1}{2}}\mathbf{T}_2\mathbf{D}_2^{\frac{1}{2}}$ のようにして正規化をし実験を行った ($\mathbf{T}_1$ も同様).

実験には 3 つのベンチマークデータを用いた．

代謝ネットワーク (metabolic) [*1] 668 のノード (タンパク質) と 2782 のリンクを持つ．リンクはタンパク質同士が反応を触媒することを示している．カーネル行列は遺伝子の発現データを用いる．

創薬ネットワーク (drug-target)[*2] 445 のノード (薬) と 664 のノード (タンパク質) と, 2,926 のリンクを持つ．薬とタンパク質の反応関係を表すネットワークである．

共著ネットワーク (nips) [*3] 2,865 のノード (著者), 4,733 のリンクを持つ．リンクは著者同士が共著することを表す．カーネル行列は著者の特徴ベクトル同士の内積により生成する．

すべてのデータセットについて, リンクの 10%を訓練データの正例とし, 同数のリンクのないノードペアを負例とする．そして残りの 90%のリンクを予測する実験を行った．このデータセットをランダムに 20 回生成し, 予測精度の評価には AUC の平均および標準偏差を用いた．

実験結果を表 2 に示す．表はそれぞれ AUC の平均値及び, 括弧内の値が標準偏差を表している．

まずノード情報を用いないテンソル積拡散カーネルにおける実験結果を見る．metabolic, drug-target, nips において, 既存のペアワイズカーネルより低い値となっている．

テンソル積伝播カーネルは, 3 つのネットワークにおいて, すべて既存のペアワイズカーネル以上の AUC であることがわかる．metabolic において, テンソル積伝播カーネルは, カルテシアンカーネルと比べ 0.020 高い結果となった．nips ではクロネッカー積カーネルと比べ 0.013 高い結果となった．これらにおいて, ノード情報を伝播させることにより予測精度が向上したと考えられる．drug-target では, クロネッカー積カーネルと同等となった．この結果になった原因として, テンソル積伝播カーネルの拡散パラメータを調べたところ $\beta = 0.0001$ の際に AUC が最も高くなっており, このデータセットにおいてはリンク構造の効果がなかったと考えられる．

5. 結論と今後の課題

ノード情報とリンク構造を組み合わせた新たなペアワイズカーネルの提案を行った．提案ペアワイズカーネルは, リンク構造を追加しつつも時間計算量を抑えており, かつ実験結果より既存のものと比べ予測精度が良いケースがあるとわかった．提案法は空間計算量がカルテシアンカーネルと比べて多く必要であったことから, 今後の課題として, 空間計算量が同等もしくはそれ以下となるようなペアワイズカーネルの提案, また別の学習方法で実験を行うことが考えられる．

*1 http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/ismb05
*2 http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/
*3 http://ai.stanford.edu/ gal/data.html

謝辞

本研究は，科学研究費若手研究 (A)12913388「大規模非構造型秘密情報のためのアウトソース型プライバシ保護データマイニング基盤」の助成を受けました．

参考文献

[1] Asa Ben-Hur and William Stafford Noble. Kernel method for prediction protein-protein interactions. *Bioinformatics*, Vol. 21, No. suppl 1, pp. i38–i46, 2005.

[2] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, Vol. 7, pp. 551–585, 2006.

[3] Hisashi Kashima, Satoshi Oyama, Yoshihiro Yamanishi, and Koji Tsuda. Cartesian kernel: An efficient alternative to the pairwise kernel. *IEICE Transactions*, Vol. 93-D, No. 10, pp. 2672–2679, 2010.

[4] Risi Imre Kondor and John D. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 315–322, 2002.

[5] Satoshi Oyama and Christopher D. Manning. Using feature conjunctions across examples for learning pairwise classifiers. In *Proceeding of the fifteenth European Conference on Machine Learning*, pp. 322–333, 2004.

[6] Rudy Raymond and Hisashi Kashima. Fast and scalable algorithms for semi-supervised link prediction on static and dynamic graphs. In *Proceeding of the 2010 European conference on Machine learning and knowledge discovery in databases*, pp. 131–147, 2010.

[7] Jean-Philippe Vert, Jian Qiu, and William Noble. A new pairwise kernel for biological network inference with support vector machines. *BMC Bioinformatics*, Vol. 8, No. Suppl 10, p. S8, 2007.