

ラプラシアンカーネルを用いた One-Class SVM とそのマーケティングへの応用

矢島 安敏, 矢田 佳久

近年, Kondor 等[7]の diffusion カーネルを始めとしたグラフラプラシアン行列から構成されるさまざまなカーネル行列を用いたパターン認識法が, 機械学習の分野で盛んに研究されるようになってきた. 本稿では, このようなパターン認識法のひとつである One-Class SVM (1-SVM) をマーケティング分野へと応用し, 顧客の好みに合致した商品やサービスの推薦, あるいは購買の見込みの高い顧客の抽出を試みる. ラプラシアン行列のスパースな構造を利用することで, データ数が数百万を超える場合でも, 高速に処理できるスケーラビリティの高い手法の構築が可能である.

キーワード: ラプラシアン行列, カーネル法, One-Class SVM, 協調フィルタリング, 推薦システム

1. はじめに

多変量解析やデータマイニングの分野で従来から用いられている多くの手法では, データはいくつかの属性によって特徴付けられ, 数理的には属性数の次元を持った実数ベクトルとして扱われている. 例えば, 商品の推薦を行う協調フィルタリングの代表的手法である相関係数法では, 顧客は商品毎の購買数や購買の有無を要素とした商品数の次元を持ったベクトルとして扱われる[9]. あるいは, 文書の分類や情報検索の場合であれば, 各文書は単語を属性としたベクトルで表現される[4]. しかし, このような属性ベクトルによるデータの表現方法が, 目的としている分析に必ずしもふさわしいとは言えない.

近年注目をされている, サポートベクターマシン(SVM)に代表されるカーネル関数を用いた手法は, 「データ間にある種の類似度(あるいは距離)を定め分析を行う方法」と考えることができる. 例えば, この分野でもっとも代表的なカーネル関数のひとつである rbf 関数は, 2つのデータ x, x' をベクトルとし, また σ を実数のパラメータとして, 次のように定義される関数である.

$$\mathcal{K}(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right) \quad (1)$$

この式よりあきらかなように, データ間のユークリッド距離(の二乗)が大きくなればなるほどカーネル関数の値は小さなものとなる. したがってこの関数は, データ x と x' のある種の類似性を表す指標と考えることができる. ただし, 類似性の指標がカーネル関数となるためには, データ i とデータ j の間の類似性を i 行 j 列要素とする対称行列 K が, 半正定値行列となる必要がある. このような行列 K はカーネル行列と呼ばれる. もちろん, 上で示した rbf 関数は行列 K が常に正定値行列となることが示されている[11].

行列 K が半正定値行列となるよう類似性を定めることができれば, SVM をはじめとするさまざまなカーネル法の適用が可能となる. 例えば, 商品の推薦を行う問題の場合では, 顧客 i と j の両方が購買した商品数を $\#(i, j)$ と記せば,

$$\mathcal{K}(i, j) = 2^{\#(i, j)} \quad (2)$$

はあきらかに類似性を示す指標であり, また, 行列 K は常に半正定値行列となることが知られている[11].

類似性の指標を定める場合, 全てのデータ間について適切な値を定めることが難しい場合が多い. 例えば, 上の(2)式を用いた場合, 共通に購買した商品が無い顧客 i と j に対しては, 類似性の値がすべて一律に $\mathcal{K}(i, j) = 1$ となってしまう. このように, 特に類似性の低いデータ間に関して適切な値の類似性を定めることが困難な場合が多い.

やじま やすとし
東京工業大学 大学院社会理工学研究科
〒152-8552 目黒区大岡山 2-12-1
やだ よしひさ
㈱ブレインパッド
〒141-0022 品川区東五反田 5-2-5

そこで、特に類似性の高いデータ間の局所的な類似性の関係から全データ間に対し類似度の値を算出し、かつ半正定値なカーネル行列を構築する方法がいくつか知られている。局所的な関係とは、例えば、Web ページの間に張られたハイパーリンクなどである。リンクの張られたページ同士は局所的な意味で類似性の高いページと考えられる。一方、直接リンクの張られていないページ同士に対しては、リンクを辿って到達するまでの経路の構造をうまく数値化することで、カーネル行列 K が半正定値となるよう類似性を定めることが可能である。しかし、一般に、これらのカーネル行列を算出するためには、逆行列の計算あるいは行列の固有ベクトルの計算が必要とされ、特にデータ数の多い大規模な問題に適用するためには、数値計算上の工夫が不可欠となる。

本稿では、まず、局所的なデータの類似性を使い定められるラプラシアンカーネル行列[12, 19]と呼ばれるカーネル行列について解説を行う。その後、One-Class SVM (1-SVM) を用いた商品の購買予測法について説明する。この際、クレジットカードなどの利用履歴データから顧客や商品の局所的な類似性を用いてラプラシアンカーネル行列を構成し、1-SVM のカーネル行列として利用する。カーネル行列の特殊な構造を用いることで、極めて大規模な問題に対しても効率的に実行可能なことを示す。最後に、実際のデータを用いて手法の検証を行い、その結果を報告する。

2. グラフラプラシアンとカーネル行列

データの局所的な類似関係は、ノード集合 V および枝集合 E からなる重み付き無向グラフ $G(V, E)$ を用いて表現することができる。グラフ $G(V, E)$ の各ノードは各データに対応し、局所的な類似性があるデータ間 $i, j \in V$ には枝 $(i, j) \in E$ を与えグラフを作成する。また、データ i, j 間の類似性の程度を表す正の実数 $b_{ij} > 0$ を枝 $(i, j) \in E$ の重さと考え、以降ではグラフ $G(V, E)$ は連結であると仮定する。さらに、 M をノード数とし、枝 $(i, j) \in E$ の重み b_{ij} を要素とする M 次正方行列を B とする。なお、 B の対角要素は $b_{ij} = 0$ とし、さらに、ノード i, j 間に枝が張られていない場合には便宜的に $b_{ij} = 0$ とする。このように作成した $G(V, E)$ を base similarity グラフ、また行列 B を base similarity 行列と呼ぶことにする。

一般に、行列 B は半正定値行列とはならず、この

ままではカーネル行列として用いることはできない。これに対して、近年、行列 B を元に半正定値なカーネル行列 K を構成するさまざまな手法が提案されている[13, 12, 3, 6]。その中の代表的なものを次に示す。

2.1 グラフ構造を用いたカーネル

まず簡単のため、base similarity 行列 B を要素が 0 または 1 の行列と仮定する。すなわち、 B はグラフ $G(V, E)$ の隣接行列と仮定する。このとき、行列 B の 2 乗、すなわち $B^2 = B \times B$ の第 i 行 j 列要素を $b_{ij}^{(2)}$ と記せば、行列の積の定義より、

$$b_{ij}^{(2)} = \sum_{k=1}^M b_{ik} b_{kj}, \quad i, j = 1, 2, \dots, M$$

であるから、 $b_{ij}^{(2)}$ は、ノード i から j へ至る枝数が 2 本のパスの本数を表していると考えられる。

そこで、適当な正の重み $\mu > 0$ を定めて次のような和：

$$B + \mu B^2$$

を考えれば、これも新たな類似性を表す行列となる。さらに、 B^3, B^4, \dots もある種の類似性を定めていると考えることができるので、次のように重みを付けて加えた行列の和

$$I + \mu B + \frac{1}{2!} \mu^2 B^2 + \frac{1}{3!} \mu^3 B^3 + \dots \quad (3)$$

を考える。ただし、 I は単位行列である。ここで、 μ は枝数の多いパスの影響をコントロールするためのパラメータと考えることができる。

式(3)はよく知られているように、

$$\exp(\mu B) = I + \sum_{k=1}^{\infty} \frac{1}{k!} \mu^k B^k \quad (4)$$

と記述された。また、この行列は常に正定値行列であり、ゆえに、カーネル行列として用いることが可能である。なお、(4)は exponential diffusion カーネルと呼ばれている。

また、和を求める際の係数を変えれば、

$$(I - \mu B)^{-1} = I + \mu B + \mu^2 B^2 + \mu^3 B^3 + \dots \quad (5)$$

と表される。 μ を十分小さな実数で

$$\mu < \|B\|_2^{-1}$$

とすれば(5)は正定値行列となる。このとき、(5)は von Neuman diffusion カーネルと呼ばれる。これらのカーネル行列は、その構成のされ方から分かるように、グラフ全体の接続関係を反映した類似性を表すカーネル行列となっていることが分かる。

さて、上で導入したカーネル行列を別の角度から眺めてみる。まず、行列 B は実対称行列であることが

ら実数の固有値

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$$

とそれぞれに対応する正規直交な固有ベクトル

$$\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M \in \mathbf{R}^M$$

を持つ。これらの固有値と固有ベクトルを用いれば、行列 B は

$$B = \sum_{j=1}^M \lambda_j (\mathbf{v}_j \mathbf{v}_j^T)$$

とランク 1 行列の和に分解された。これにしたがえば、exponential diffusion カーネル行列(4)は

$$\exp(\mu B) = \sum_{j=1}^M \exp(\mu \lambda_j) (\mathbf{v}_j \mathbf{v}_j^T)$$

と表現することが可能である。

指数関数の性質から、 $\lambda_j \ll 0$ ならば、 $\exp(\mu \lambda_j) \approx 0$ となることから、行列 $\exp(\mu B)$ は、 B の小さな固有値に相当した固有ベクトルの成分を、ほとんど無視することで構成されたカーネル行列と見ることもできる。

一般に、行列の次元が大きな場合、その固有ベクトルを全て求めることは実用的には困難である。したがって、exponential diffusion カーネル(4)を計算することも容易ではない。そこで、 $\exp(\mu B)$ の近似として、大きな固有値を k 個選択し、それに対応する固有ベクトルのみを用いて、

$$\sum_{j=1}^k \exp(\mu \lambda_j) (\mathbf{v}_j \mathbf{v}_j^T) \quad (6)$$

とした行列を $\exp(\mu B)$ の近似として用いる方法も試みられている[2]。もちろん式(6)のように近似しても半正定値行列であり、したがってカーネル行列として用いることが可能である。

2.2 ラプラシアン行列によるカーネル

まず、 D を M 次の対角行列で、その対角要素 d_{ii} は行列 B の第 i 列の和、すなわち

$$d_{ii} = \sum_{j=1}^M b_{ij}, \quad i=1, 2, \dots, M$$

と定義し、次のように行列 L を定める。

$$L = D - B \quad (7)$$

L はラプラシアン行列[5]と呼ばれ、 L を用いたさまざまなカーネル行列が提案されている。

まず、 L は次に示すように半正定値行列である。すなわち、任意の M 次ベクトル $\mathbf{r} = (r_1, r_2, \dots, r_M)^T$ に対して、二次形式を考えると

$$\mathbf{r}^T L \mathbf{r} = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M b_{ij} (r_i - r_j)^2$$

と式変形できる。 b_{ij} は定義より非負であることから、 L の半正定値性はあきらかである。さらに、 \mathbf{e} を要素がすべて 1 のベクトルとすれば、

$$L \mathbf{e} = \mathbf{0}$$

となることから、 L は固有値として 0 を持ち、ゆえに正則な行列ではないことが分かる。さらに、グラフ $G(V, E)$ が連結なグラフであれば、 L のランクは $M - 1$ となることが知られている。行列 L のさまざまな性質については、[5]を参照されたい。

このとき、 L の一般化逆行列 L^+ を考える。すなわち、 L の固有値を

$$0 = \mu_1 < \mu_2 \leq \dots \leq \mu_M,$$

とし、またそれぞれに対応する固有ベクトルを

$$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M \in \mathbf{R}^M$$

とすれば、一般化逆行列は、固有ベクトルからなるランク 1 行列の線形和として、

$$L^+ = \sum_{j=2}^M \frac{1}{\mu_j} (\mathbf{u}_j \mathbf{u}_j^T) \quad (8)$$

と定めることができた。

さらに、仮にグラフ $G(V, E)$ が正則グラフ、すなわち、

$$d_{ii} = d \quad (\text{一定}), \quad i=1, 2, \dots, M$$

であれば、ラプラシアン行列 L の定義(7)より、行列 L の固有ベクトルは行列 B のそれと等しく、また固有値も次の関係

$$\mu_i = d - \lambda_i, \quad i=1, 2, \dots, M$$

を満たす。したがって、式(8)より、 L^+ は base similarity 行列 B の 0 以外の小さな固有値に対応したランク 1 行列ほど大きな係数で加えることで構成された行列と考えることができる。よって、 L^+ も前節で導入した exponential diffusion カーネル(4)や von Neuman diffusion カーネル(5)と同様に、データ間のある種の類似性を要素とする半正定値行列と考えられる。なお、ベクトル \mathbf{e} が L の 0 の固有値に対応する固有ベクトルであることより、 L の一般化逆行列は

$$L^+ = \left(L - \frac{\mathbf{e} \mathbf{e}^T}{M} \right)^{-1} + \frac{\mathbf{e} \mathbf{e}^T}{M} \quad (9)$$

と計算できることが知られている。 L^+ は commute time カーネル行列と呼ばれ、文書分類や画像認識などで高い性能を発揮できることが報告されている[1, 18, 16, 15]。

さらに、式(8)のバリエーションとして、正の実数 t をパラメータに持つ次の行列

$$(I + tL)^{-1} = \sum_{j=1}^M \frac{1}{1 + t\mu_j} (\mathbf{u}_j \mathbf{u}_j^T) \quad (10)$$

が用いられる。本稿ではこれをラプラシアンカーネル行列と呼ぶことにする。また、パラメータ $0 \leq \gamma \leq 1$ によりラプラシアン行列を $L_\gamma = \gamma D - B$ と一般化して

用いた次の行列[6]

$$(I+tL)^{-1} \quad (11)$$

なども提案されている。

さらに、ラプラシアン行列に対して、次のようなある種の基準化を施した normalized ラプラシアン行列:

$$\tilde{L}=D^{-\frac{1}{2}}LD^{-\frac{1}{2}} \quad (12)$$

を用いる方法[12]も提案されている。行列 \tilde{L} も半正定値行列であり、全ての固有値 $\tilde{\mu}_j$ は

$$0 \leq \tilde{\mu}_j \leq 2, j=1, 2, \dots, M$$

と基準化されていることが知られている[5,12]。その上で、今までとはほぼ同様な考え方でカーネル行列を構成することが可能である。例えば、normalized ラプラシアンカーネル行列:

$$(I+t\tilde{L})^{-1} \quad (13)$$

や、normalized diffusion カーネル行列:

$$\exp(-\sigma^2\tilde{L}) \quad (14)$$

などである。このように基準化をしたラプラシアン行列を用いることで、データによってはより性能の高い識別が行えることが報告されている[17]。

3. 1-SVM を用いた購買予測

3.1 カーネル法

前節で導入したカーネル行列 K の各要素 K_{ij} をデータ i と j の内積と考え、さまざまな分析を行う方法を総称してカーネル法と呼んでいる。すなわち、 i 番目のデータを適当な空間内のベクトルとして ϕ_i と記せば、

$$K_{ij}=\phi_i^T\phi_j, i, j=1, 2, \dots, l \quad (15)$$

を満たすベクトルに対して分析を行う方法である。なかでも2クラスの判別を行う SVM[14]は、カーネル法の代表的な手法である。他にも回帰を行う SVR やカーネル主成分分析など、多変量解析の分野で用いられているほとんどの手法は、カーネル法へと自然に拡張が可能である。詳しくは[11]などを参照されたい。

上で述べたように、行列 K は半正定値行列であることから、適当な正規直交行列 P と対角行列 Λ を用いれば、

$$K=PA P^T=(P\Lambda^{1/2})(P\Lambda^{1/2})^T \quad (16)$$

と分解が可能である。そこで、行列 $(P\Lambda^{1/2})^T$ の列ベクトルを

$$(P\Lambda^{1/2})^T=[\phi_1\phi_2\cdots\phi_m]$$

と考えれば、カーネル行列から式(15)を満たすベクトル ϕ_i を算出することが可能である。カーネル法では、

特に、ベクトル ϕ_i を特徴ベクトル、またその空間を特徴空間と呼び、以降では ϕ と記す。

一般に、式(16)のように行列 K を分解し特徴ベクトル ϕ_i を得るためには、 K の固有値や固有ベクトルの算出が必要となる。しかし、カーネル法では最適化理論などを用いることで、具体的に特徴ベクトルを算出することなく行列 K のみにより分析を可能としている。このような方法はカーネルトリックとも呼ばれている。

3.2 購買スコアの計算

SVM のバリエーションのひとつに 1-SVM[10] と呼ばれるものがある。この節では、1-SVM を用いた購買スコアの算出を試みる。

前節で導入したカーネル行列 K によって特徴空間 ϕ 上のデータが、

$$\phi_1, \phi_2, \dots, \phi_l \quad (17)$$

と与えられていると仮定する。1-SVM ではこれを学習データと呼ぶ。1-SVM は、学習データ(17)に共通なパターンを学習し、新たなデータに対して、それが(17)と同一のパターンであるか否かを判定するための機械学習法のひとつである。

1-SVM では、学習データ(17)をできるだけ原点と分離するような超平面を求める。すなわち、 $w \in \phi$ を法線ベクトル、また $\rho \in \mathbf{R}$ を実数とし求める超平面を

$$w^T\phi(x)=\rho \quad (18)$$

とした上で、変数 $\xi=(\xi_1, \xi_2, \dots, \xi_l)^T$, w および ρ に関する次の二次計画問題:

$$\begin{cases} \text{最小化} & \frac{1}{2}w^T w + \frac{1}{\nu l} \sum_{j=1}^l \xi_j - \rho \\ \text{制約} & w^T\phi_j + \xi_j \geq \rho, j=1, 2, \dots, l, \\ & \xi_j \geq 0, j=1, 2, \dots, l \end{cases} \quad (19)$$

を最適化する。ただし、 $\nu \in (0, 1]$ はあらかじめ定められた正のパラメータである。ここで、 (w^*, ρ^*) を問題(19)の最適解とすれば、データ ϕ が $w^{*T}\phi - \rho^* > 0$ であれば、 ϕ は学習データ(17)と同一のパターンであると判定される。

さて、本稿の目的は、ある特定の商品(以降ターゲット商品と呼ぶ)を購入する顧客を予測することである。そこで、まず過去において実際にターゲット商品を購入したことがある顧客を学習データ(17)とし、1-SVM により超平面(18)を算出する。その上で、まだターゲット商品を購入したことが無い顧客 j のデータ ϕ_j に対して $w^{*T}\phi_j - \rho^*$ を計算し、この値が大きいほど学習データとして与えた顧客と類似したターゲット商

品を購入する傾向が高いと考える。すなわち、定数部分を省略し、

$$S_j = \mathbf{w}^{*T} \phi_j \quad (20)$$

を顧客 j がターゲット商品を購入する程度を示すスコアと考えることとした。

3.3 ラプラシアンカーネルを用いた 1-SVM

この節では、カーネル行列として、ラプラシアンカーネル行列(10)を 1-SVM に用いた場合の効率的な解法について考える。通常、SVM ではカーネル関数は問題(19)の双対問題の中で扱われるが、ラプラシアンカーネル行列を用いる場合には、下に示すように主問題(19)を扱うことで、データ数が多い大規模な問題でも高速に扱うことが可能となる。

まず簡単のため、得られている M 個のデータの中で最初の l 個 $\phi_1, \phi_2, \dots, \phi_l$ を学習データと仮定し、残りの $M-l$ 個のデータ $\phi_{l+1}, \phi_{l+2}, \dots, \phi_M$ に対して、スコアを求めることを考える。なお、学習データを表すために、 M 次の 0-1 ベクトル $\mathbf{y} = (y_1, y_2, \dots, y_M)^T$ を

$$y_j = \begin{cases} 1 & \text{if } j=1, 2, \dots, l, \\ 0 & \text{if } j=l+1, l+2, \dots, M \end{cases} \quad (21)$$

と定める。また、 M 個の変数 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_M)^T$ を導入し、求めたい超平面(18)の法線ベクトル \mathbf{w} がデータ ϕ_j の線形結合で

$$\mathbf{w} = \sum_{k=1}^M \alpha_k \phi_k \quad (22)$$

と表現されると考える。このとき、内積 $\phi_i^T \phi_j$ がカーネル行列 K の第 i 行 j 列成分 K_{ij} によって与えられていることから、簡単な計算で $\mathbf{w}^T \mathbf{w} = \boldsymbol{\alpha}^T K \boldsymbol{\alpha}$ であり、また

$$\mathbf{w}^T \phi_j = \sum_{k=1}^M \alpha_k K_{kj} = (K\boldsymbol{\alpha})_j$$

である。これらを問題(19)へ代入すれば次の定式化：

$$\begin{cases} \text{最小化} & \frac{1}{2} \boldsymbol{\alpha}^T K \boldsymbol{\alpha} + \frac{1}{\nu l} \mathbf{y}^T \boldsymbol{\xi} - \rho \\ \text{制約} & (K\boldsymbol{\alpha})_j + \xi_j \geq \rho, \quad j=1, 2, \dots, l, \\ & \xi_j \geq 0, \quad j=1, 2, \dots, l \end{cases} \quad (23)$$

を得る。ただし $(K\boldsymbol{\alpha})_j$ はベクトル $K\boldsymbol{\alpha}$ の j 番目の要素を表すものとする。

このとき、式(10)より K の逆行列が

$$K^{-1} = (I + tL)$$

と得られていることから、新たに変数ベクトル $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_M)^T \in \mathbf{R}^M$ を導入し、 $\boldsymbol{\beta} \equiv K\boldsymbol{\alpha}$ と変数変換を行えば、

$$\boldsymbol{\alpha}^T K \boldsymbol{\alpha} = \boldsymbol{\beta} K^{-1} \boldsymbol{\beta} = \boldsymbol{\beta} (I + tL) \boldsymbol{\beta}$$

となる。また任意の $j=1, 2, \dots, M$ で $\beta_j = (K\boldsymbol{\alpha})_j$ とな

ることから、問題(23)はベクトル $\boldsymbol{\beta}$ を変数として、次のように変形が可能である。

$$\begin{cases} \text{最小化} & \frac{1}{2} \boldsymbol{\beta}^T (I + tL) \boldsymbol{\beta} - \rho + \frac{1}{\nu l} \mathbf{y}^T \boldsymbol{\xi} \\ \text{制約} & \beta_j + \xi_j \geq \rho, \quad j=1, 2, \dots, l, \\ & \xi_j \geq 0, \quad j=1, 2, \dots, l. \end{cases} \quad (24)$$

ここで、ラプラシアン行列 L の特殊な構造を用いると、次の定理を得ることができる[15]。

定理 3.1 $(\boldsymbol{\beta}^*, \boldsymbol{\xi}^*, \rho^*)$ を、問題(24)の最適解とする。このとき、次の等式

$$\beta_j^* + \xi_j^* = \rho^*, \quad \forall j=1, 2, \dots, l \quad (25)$$

が成立する。

そこで、等式(25)を用いて変数 ξ_j を消去することで、問題(24)は次の定式化へと帰着される。

$$\begin{cases} \text{最小化} & \frac{1}{2} \boldsymbol{\beta}^T (I + tL) \boldsymbol{\beta} + \frac{1-\nu}{\nu} \rho - \frac{1}{\nu l} \mathbf{y}^T \boldsymbol{\beta} \\ \text{制約} & \beta_j \leq \rho, \quad j=1, 2, \dots, l. \end{cases} \quad (26)$$

なお、この問題は制約の無い凸計画問題として扱うことが可能であり、大規模な問題でも高速に最適化が可能である。詳しくは[8]などを参照されたい。

さらに、パラメータを $\nu=1.0$ とした場合には、目的関数で変数 ρ の係数が 0 になることから、問題(24)は単なる二次関数の最小化：

$$\text{最小化 } W(\boldsymbol{\beta}) = \frac{1}{2} \boldsymbol{\beta}^T (I - tL) \boldsymbol{\beta} - \frac{1}{l} \mathbf{y}^T \boldsymbol{\beta} \quad (27)$$

に帰着される。この問題は解析的に解くことが可能である。すなわち、目的関数の勾配ベクトルを $\mathbf{0}$ とすればよく、最適解 $\boldsymbol{\beta}^*$ は

$$\nabla W(\boldsymbol{\beta}) = (I - tL) \boldsymbol{\beta}^* - \frac{1}{l} \mathbf{y} = \mathbf{0} \quad (28)$$

より求めることが可能である。これは単純な連立一次方程式であり、行列 $(I - tL)$ がスパースな行列であれば、CG 法などを用いることで、変数の個数が極めて多い大規模な場合でも効率良く解を算出できる。さらに、式(20)より、

$$S_j = \mathbf{w}^{*T} \phi_j = \beta_j^*$$

であったので、顧客 j のスコアは β_j^* に他ならない。なお、次節ではこの方法を用い手法の検証を行っている。

最後に、本稿ではページ数の都合上、ラプラシアンカーネル行列(10)に関してのみ 1-SVM の解法を説明したが、2 節で導入した commute time カーネル行列(9)や、normalized ラプラシアンカーネル行列(13)など、逆行列を用いて定義されるタイプのカーネル行列であれば、同様な式変形が可能である。

4. 実データを用いた手法の検証

この節では、1-SVM にラプラシアンカーネルを用いた手法を、実際のクレジットカードデータに適用した結果を述べる。ある保険商品をターゲット商品として、この保険に加入する顧客の予測を行った。

まず、過去1年間の利用履歴を顧客毎に集計し、店舗毎に利用の有無によって二部グラフを作成した。すなわち、顧客に対応したノード集合 N_c と、店舗あるいは保険などのサービスに対応したノード集合 N_s を定め、顧客 $i \in N_c$ が店舗 $j \in N_s$ を利用したことがあれば、顧客 i は店舗 j と類似性が高いと考え、対応するノード i, j 間に枝を張り二部グラフを構築した。また、枝の重みは単純に全て $b_{ij}=1$ を用いた。したがって base similarity 行列 B はグラフの隣接行列である。今回用いたデータでは、2人以上の利用があった店舗は約43万店であり、また、その店舗を1回以上利用したことのある顧客約160万人の履歴データを用いた。すなわち、行列 B は約200万次元の行列となる。ここで注目すべき点は、行列 B の非ゼロ要素の割合が約 4.2×10^{-4} % と非常にスパースな行列となることである。この程度の行列であれば、以下に述べるようなパソコンの環境で十分計算可能なサイズである。

ターゲット商品を購入したことのある顧客を1-SVMの学習データとし、式(21)のようにベクトル \mathbf{y} を定める。方程式(28)を解くことで、各顧客 i 毎にターゲット商品に対するスコア $S_i = \beta_i^*$ を算出した。この検証では、ラプラシアンカーネル行列(10)のパラメータを $t=1.0$ とした。なお、計算は Xeon (3.6 GHz) を搭載した主記憶2GBのWindowsパソコンで、MATLAB 7の関数 `cgs()` を用いCG法により連立一次方程式を解いた。本検証での問題サイズの場合、CG法は約15回程度の繰り返しで残差の相対ノルムが 10^{-6} 以下となり、これに要した計算時間も10秒程度と非常に高速に解くことが可能であった。

検証は、今回ターゲット商品となった保険契約のある顧客の50%を学習データとして用い、残り50%をテストデータとして予測の精度を検討した。図1は、得られたスコアを用いてテストデータに対して作成したゲインチャートである。比較のため、各顧客の利用店舗数をスコアとして用いた場合の結果も示した。図を見ると、特にスコア上位の部分では提案手法の方が、単純な店舗数を用いたものに比べて2倍以上の精度で

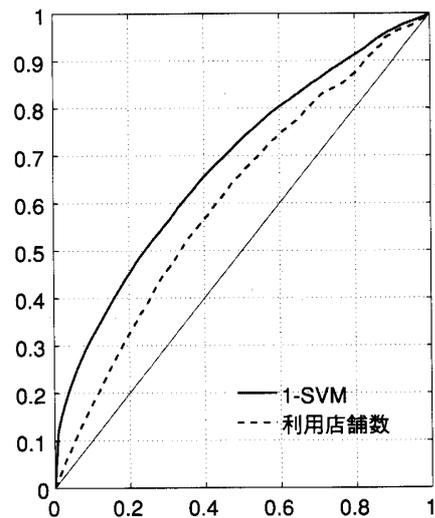


図1 テストデータに対するゲインチャート

予測していることが分かる。

5. おわりに

本稿では、1-SVMにより顧客の購買履歴データから、ある特定の商品購入を予測するスコアの算出を試みた。これには、すでにその商品を購入している顧客を学習データとして用い1-SVMを解き、学習データとの類似性を用いてスコアとした。

一般にSVMでは、どのようなカーネル行列を用いるのかが非常に重要である。本稿では、近年注目されているラプラシアン行列から構成されるカーネル行列に注目し、スケーラビリティのある手法を構築した。通常、最も頻繁に用いられるrbf関数の場合では、カーネル行列が完全に稠密な行列になってしまうため、データ数の増加により計算機上での扱いが非常に困難となる。SMOアルゴリズムなどのSVMに特化したアルゴリズムを用いても、問題の構造が悪くサポートベクターの個数が増加する場合には、大規模な問題を解くことは不可能である。本稿で扱ったマーケティングなどの分野では、通常は最低でも数百万顧客のデータを扱わなくてはならず、従来からのカーネル関数を用いたSVMの適用には限界があると思われる。

一方、本稿で取り上げたラプラシアンカーネル行列をはじめとしたデータの局所的な構造により構成されるカーネル行列は、一般に非常にスパースな構造を持つ行列となる。したがって、適切にデータ構造を用いたプログラムで処理すれば、数百万データの問題であっても容易に扱うことが可能である。ここで紹介したカーネル行列は、1-SVMに限らず、通常の2クラスSVMなどの他のさまざまなカーネル法を大規模な問

題へと適用する際に、もっとも有力なカーネル行列のひとつと考えられる。

参考文献

- [1] M. Belkin, I. Matveeva, and P. Niyogi. Regularization and semi-supervised learning on large graphs. In *COLT*, 2004.
- [2] M. Belkin and P. Niyogi. Using manifold structure for partially labelled classification. In *Advances in Neural Information Processing Systems*, volume 15, 2003.
- [3] M. Belkin and P. Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56: 209-239, 2004.
- [4] M. W. Berry, Z. Drmac, and E. R. Jessup. Matrices, vector spaces, and information retrieval. *SIAM Review*, 41: 335-362, 1999.
- [5] F. R. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [6] T. Ito, M. Shimbo, T. Kudo, and Y. Matsumoto. Application of kernels to link analysis. In *KDD'05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 586-592, New York, NY, USA, 2005. ACM Press.
- [7] R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *ICML*, 2002.
- [8] O. L. Mangasarian and M. V. Solodov. Nonlinear complementarity as unconstrained and constrained minimization. *Math. Programming*, 62 (2, Ser. B): 277-297, 1993.
- [9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, pages 175-186, Chapel Hill, North Carolina, 1994. ACM.
- [10] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13: 1443-1471, 2001.
- [11] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, 2004.
- [12] A. Smola and I. Kondor. Kernels and regularization on graphs. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the Annual Conference on Computational Learning Theory*, Lecture Notes in Computer Science. Springer, 2003.
- [13] M. Szummer and T. Jaakkola. Partially labeled classification with Markov random walks. In *Advances in Neural Information Processing Systems*, volume 14, pages 945-952, 2002.
- [14] V. N. Vapnik. *The nature of statistical learning theory*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 2000.
- [15] Y. Yajima. One-class support vector machines for recommendation tasks. In *PAKDD*, 2006.
- [16] Y. Yajima and T. Hoshiba. Optimization approaches for semi-supervised learning. In *ICMLA*, 2005.
- [17] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16: 321-328, 2004.
- [18] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *ICML*, pages 912-919, 2003.
- [19] X. Zhu, J. Lafferty, and Z. Ghahramani. Semi-supervised learning: From Gaussian fields to Gaussian processes. Technical Report CMU-CS-03-175, Carnegie Mellon University, 2003.