数理解析研究所講究録 1409 巻 2005 年 205-218

Electric Network Kernel for Support Vector Machines

京都大学 数理解析研究所　平井 広志 (Hirai Hiroshi)
Research Institute for Mathematical Sciences, Kyoto University

東京大学大学院 数理情報学専攻　室田 一雄 (Kazuo Murota)
Graduate School of Information Science and Technology, University of Tokyo

ジャストシステム 技術戦略室　力徳 正輝 (Masaki Rikitoku)
Justsystem Corporation R&D Strategy Department

Abstract

This paper investigates support vector machine (SVM) with a discrete kernel, named electric network kernel, defined on the vertex set of an undirected graph. Emphasis is laid on mathematical analysis of its theoretical properties with the aid of electric network theory. SVM with this kernel admits physical interpretations in terms of resistive electric networks; in particular, the SVM decision function corresponds to an electric potential. Preliminary computational results indicate reasonable promise of the proposed kernel in comparison with the Hamming and diffusion kernels.

1 Introduction

Support vector machine (SVM) has come to be very popular in machine learning and data mining communities. SVM is a binary classifier using an optimal hyperplane learned from given training data. Through *kernel functions*, which are a kind of similarity functions defined on the data space, the data can be implicitly embedded into a high (possibly infinite) dimensional Hilbert space. With this *kernel trick*, SVM achieves a nonlinear classification with low computational cost.

Input data from real world problems, such as text data, DNA sequences and hyperlinks in World Wide Web, is often endowed with discrete structures. Theory and application of "kernels on discrete structures" are pioneered by D. Haussler [5], C. Watkins [14] and R. I. Kondor and J. Lafferty [8]. Haussler and Watkins independently introduced the concept of *convolution kernels*. Kondor and Lafferty utilized spectral graph theory to introduce *diffusion kernels*, which are discrete kernels defined on vertices of graphs.

In this paper we propose a novel class of discrete kernels on vertices of an undirected graph. Our approach is closely related to that of Kondor and Lafferty, but is based on electric network theory rather than on spectral graph theory. Accordingly we will name the proposed kernels *electric network kernels*. SVM using an electric network kernel admits natural physical interpretations. The vertices with positive label and negative label correspond, respectively, to terminals with $+1$ electric potential and -1 electric potential. The resulting decision function corresponds to an electric potential, and the separating hyperplane to points with potential equal to zero.

Emphasis is laid on mathematical analysis of the electric network kernel with the aid of electric network theory. Another interesting special case is where the underlying graph is a tensor product of complete graphs. By exploiting symmetry of this graph, we provide an explicit formula for the electric network kernel, which makes it possible to apply the electric network kernel to large-scale practical problems. In our preliminary computational experiment the electric network kernel shows fairly good performance for some data sets, as compared with the Hamming and diffusion kernels.

This paper is organized as follows. In Section 2, we review SVM and its formulation as optimization problems. In Section 3, we propose our kernel and investigate its properties. Physical interpretations to SVM with our kernel are also explained. In Section 4, we deal with the case of a tensor product of complete graphs, and show some computational results for some real world problems.

This paper is based on [6] but with new results described in Section 4.

2 Support Vector Machines

In this section, we review SVM and its formulation as optimization problems; see [11], [13] for details. Let $\mathcal{X}$ be an input data space, e.g. $\mathbf{R}^n$, $\{0,1\}^n$, text data and DNA sequence, etc. A symmetric function $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ is said to be a *kernel* on $\mathcal{X}$ if it satisfies the *Mercer condition*:

$$\begin{array}{l}\text{For any finite subset } Y \text{ of } \mathcal{X} \\ \text{matrix } (K(x,y) \mid x,y \in Y) \text{ is positive semidefinite.}\end{array} \tag{2.1}$$

For a kernel K, it is well known that there exists some Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ and a map $\phi : \mathcal{X} \to \mathcal{H}$ such that

$$K(x,y) = \langle \phi(x), \phi(y) \rangle \quad (x, y \in \mathcal{X}).$$

Given a labeled training set $(x_1, \eta_1), (x_2, \eta_2), \ldots, (x_m, \eta_m) \in \mathcal{X} \times \{\pm 1\}$, SVM classifier is obtained by solving the optimization problem

$$\begin{array}{ll}\min\limits_{\alpha \in \mathbf{R}^m} & \dfrac{1}{2} \sum\limits_{1 \leq i,j \leq m} \alpha_i \alpha_j \eta_i \eta_j K(x_i, x_j) - \sum\limits_{1 \leq i \leq m} \alpha_i \\ \text{s.t.} & \sum\limits_{1 \leq i \leq m} \eta_i \alpha_i = 0, \ 0 \leq \alpha_i \leq C \quad (i = 1, \ldots, m),\end{array}$$

where C is a penalty parameter that is a positive real number or $+\infty$. If $C = +\infty$, it is called the *hard margin SVM* formulation. If $C < +\infty$, it is called the *1-norm soft margin SVM* formulation.

For our purpose, it is convenient to consider the equivalent problem

$$
\begin{aligned}
[\mathrm{SVM}]:\ \min_{u\in\mathbf{R}^m}\quad & \frac{1}{2}\sum_{1\le i,j\le m} u_iu_jK(x_i,x_j) - \sum_{1\le i\le m}\eta_i u_i \\
\text{s.t.}\quad & \sum_{1\le i\le m} u_i = 0, && (2.2)\\
& 0\le \eta_i u_i \le C \quad (i=1,\dots,m), && (2.3)
\end{aligned}
$$

where $u_i = \eta_i\alpha_i$ for $i = 1,\dots,m$.

Let $u^* \in \mathbf{R}^m$ be an optimal solution of the problem [SVM] and $b^* \in \mathbf{R}$ be the Lagrange multiplier of constraint (2.2) at u^*, where the Lagrange function of [SVM] is supposed to be defined as

$$
\begin{aligned}
L(u,\lambda,\mu,b) &= \frac{1}{2}\sum_{1\le i,j\le m} u_iu_jK(x_i,x_j) - \sum_{1\le i\le m}\eta_i u_i \\
&\quad - \sum_{1\le i\le m}\lambda_i\eta_i u_i - \sum_{1\le i\le m}\mu_i(\eta_i u_i - C) + b\sum_{1\le i\le m} u_i,
\end{aligned}
$$

where $u \in \mathbf{R}^m$, $\lambda,\mu \in \mathbf{R}^m_{\ge 0}$, and $b \in \mathbf{R}$. Then the decision function $f : \mathcal{X} \to \mathbf{R}$ is given as

$$f(x) = \sum_{i=1}^{m} u_i^* K(x_i,x) + b^* \quad (x \in \mathcal{X}). \tag{2.4}$$

That is, we classify a given data x according to the sign of $f(x)$. A data x_i with $\eta_i u_i^* > 0$ is called a *support vector*. In the case of the 1-norm soft margin SVM, a support vector x_i is called a *normal support vector* if $0 < \eta_i u_i^* < C$ and a *bounded support vector* if $\eta_i u_i^* = C$.

3 Proposed Kernel and Its Properties

Let (V, E, r) be a resistive electric network with vertex set V, edge set E, and resistors on edges with the resistances represented by $r : E \to \mathbf{R}_{>0}$. We assume that the graph (V, E) is connected. Let $D : V \times V \to \mathbf{R}$ be a distance function on V defined as

$$D(x,y) = \text{resistance between } x \text{ and } y \quad (x,y \in V). \tag{3.1}$$

Fix some vertex $x_0 \in V$ as a root, and define a symmetric function $K : V \times V \to \mathbf{R}$ on V as

$$K(x,y) = \{D(x,x_0) + D(y,x_0) - D(x,y)\}/2 \quad (x,y \in V). \tag{3.2}$$

Seeing that $K(x_0, y) = 0$ for all $y \in V$, we define a symmetric matrix $\hat{K}$ by

$$\hat{K} = (K(x,y) \mid x,y \in V \setminus \{x_0\}). \tag{3.3}$$

Remark 3.1. Given a distance function D, the function K defined by (3.2) is called the *Gromov product*.

Let L be the *node admittance matrix* defined as

$$L(x,y) = \begin{cases} \sum\{(r(e))^{-1} \mid x \text{ is an endpoint of } e \in E\} & \text{if } x = y \\ -(r(xy))^{-1} & \text{if } x \neq y \end{cases} \qquad (x, y \in V). \quad (3.4)$$

If all resistances are equal to 1, then L coincides with the *Laplacian matrix* of graph (V, E). Let $\hat{L}$ be a symmetric matrix defined as

$$\hat{L} = (L(x,y) \mid x, y \in V \setminus \{x_0\}).$$

Note that $\hat{L}$ satisfies

$$\hat{L}(x,y) \leq 0 \quad (x \neq y), \qquad (3.5)$$

$$\sum_{z \in V\setminus\{x_0\}} \hat{L}(x,z) \geq 0 \quad (x \in V \setminus \{x_0\}). \qquad (3.6)$$

Hence $\hat{L}$ is a nonsingular diagonally dominant symmetric *M-matrix.* In particular, $\hat{L}$ is positive definite. A matrix whose inverse is an M-matrix is called an *inverse M-matrix.* The following relationship between K and L is well known in electric network theory; see [4] for example.

Proposition 3.2. *We have $\hat{K}^{-1} = \hat{L}$. In particular $\hat{K}$ is an inverse M-matrix.*

Hence, K in (3.2) satisfies the Mercer condition. We shall call such K an *electric network kernel.*

Remark 3.3. An electric network kernel K of (V, E, r) with root x_0 coincides with *discrete Green's function* of (V, E, r) taking $\{x_0\}$ as a boundary condition [2].

We consider the SVM on electric network (V, E, r) with the kernel K of (3.2). Let $\{(x_i, \eta_i)\}_{i=1,\ldots,m} \subseteq V \times \{\pm 1\}$ be a training data set, where we assume that x_i ($i = 1, \ldots, m$) are all distinct. Just as the SVM with a diffusion kernel, we assume that $\{x_1, \ldots, x_m\}$ is a subset of the vertex set V; accordingly we put $V = \{x_1, \ldots, x_n\}$ with $n \geq m$.

Lemma 3.4. *The optimization problem* [SVM] *is determined independently of the choice of a root $x_0 \in V$.*

Proof. The objective function of [SVM] is in fact independent of x_0, since its quadratic term can be rewritten as

$$\begin{aligned} \textstyle\sum_{i,j} u_i u_j K(x_i, x_j) &= \textstyle\sum_{i,j} u_i u_j (D(x_i, x_0) + D(x_j, x_0) - D(x_i, x_j))/2 \\ &= \textstyle\sum_j u_j \sum_i u_i D(x_i, x_0) - (1/2) \sum_{i,j} u_i u_j D(x_i, x_j) \\ &= -(1/2) \textstyle\sum_{i,j} u_i u_j D(x_i, x_j), \end{aligned}$$

where the last equality follows from the constraint (2.2). □

Next we give physical interpretations to the problem [SVM] with the aid of nonlinear network theory (see [7, Chapter IV]). Suppose that we are given an electric network (V, E, r) and labeled training data set $\{(x_i, \eta_i)\}_{i=1,\ldots,m} \subseteq V \times \{\pm 1\}$, where $x_1, \ldots, x_m$ are all distinct. We connect voltage sources to (V, E, r) as follows:

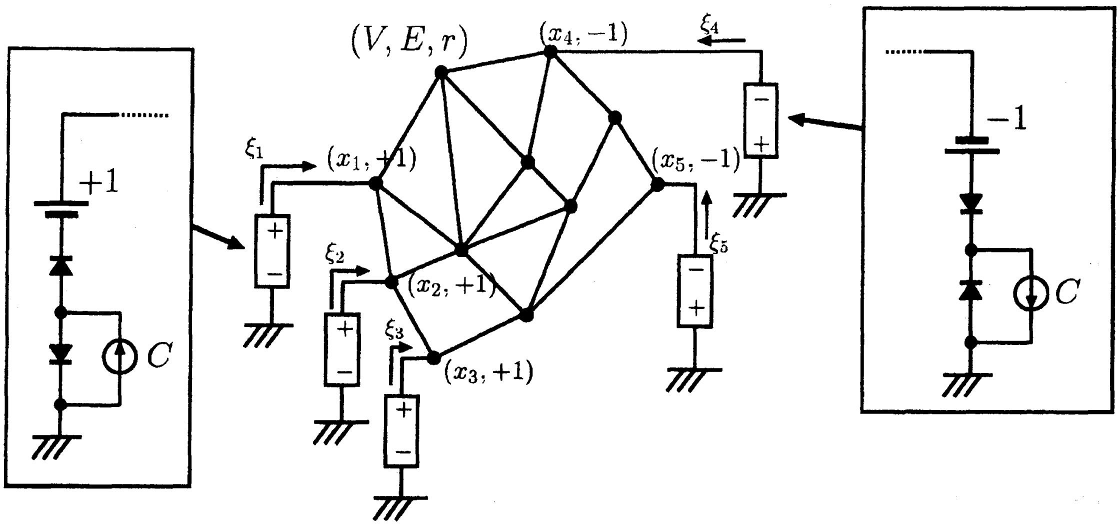

Figure 1: Physical interpretation

For each x_i with $1 \leq i \leq m$, connect to the earth a voltage source whose electric potential is η_i and the current flowing into x_i is restricted to $[0, C]$ if $\eta_i = 1$ and $[-C, 0]$ if $\eta_i = -1$.

By using voltage sources, current sources and diodes, this network can be realized as in Figure 1.

Let $A = (A(x,e) \mid x \in V, e \in E)$ be the incidence matrix of (V, E) with some fixed orientation of edges and let $R = \operatorname{diag}(r(e) \mid e \in E)$ be the diagonal matrix whose diagonals are the resistances of edges.

The electric current in this network is given as an optimal solution of the problem:

$$
\begin{aligned}
[\text{FLOW}] : \min_{(\zeta,\xi)} \quad & \frac{1}{2}\zeta^\top R\zeta - \sum_{i=1}^{m} \eta_i \xi_i \\
\text{s.t.} \quad & A\zeta = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \\
& \sum_{1 \leq i \leq m} \xi_i = 0, \quad 0 \leq \eta_i \xi_i \leq C \ (i = 1, \ldots, m),
\end{aligned}
$$

where ζ represents the currents in edges and ξ_i represents the current flowing into x_i for $i = 1, \ldots, m$. The first and second terms of the objective function of [FLOW] represents current potential of edges E and of the voltage sources respectively. The electric potential of this network is given as an optimal solution of the problem:

$$
[\text{POT}] : \min_{p \in \mathbf{R}^n} \quad \frac{1}{2} p^\top A R^{-1} A^\top p + C \sum_{i=1}^{m} \max\{0, 1 - \eta_i p_i\},
$$

where p_i represents the potential on vertex x_i for $i = 1, \ldots, n$. The first and second terms of the objective function of [POT] represent voltage potentials of edges E and of the voltage sources respectively.

Proposition 3.5. *The electric current (ζ^*, ξ^*) in this network is uniquely determined. If there exists $i \in \{1, \dots, m\}$ with $0 < \eta_i \xi_i^* < C$, then the electric potential is also uniquely determined.*

Proof. The first assertion follows from the uniqueness theorem [7, Theorem 16.2]. Note that [FLOW] and [POT] are a dual pair. Hence if such ξ_i^* exists, from complementarity condition, any optimal solution p^* of [POT] must satisfy $p_i^* = \eta_i$. Consequently, the potentials of other vertices are also uniquely determined by Ohm's law $p(x) - p(y) = R(xy)\zeta(xy)$ for $xy \in E$, $x, y \in V$. □

The following theorem indicates the relationship between SVM problem and this electric network.

Theorem 3.6. *Let u^* be the optimal solution of* [SVM]. *Then u_i^* coincides with the electric current flowing into x_i for $i = 1, \dots, m$. Moreover, the decision function f of* (2.4) *for* [SVM] *is an electric potential.*

Proof. The problem [FLOW] is equivalent to

$$
\begin{aligned}
[\text{FLOW}'] : \min_{\xi} \quad & W(\xi) - \sum_{i=1}^{m} \eta_i \xi_i \\
\text{s.t.} \quad & \sum_{1 \le i \le m} \xi_i = 0, \quad 0 \le \eta_i \xi_i \le C \ (i = 1, \dots, m),
\end{aligned}
$$

where $W : \mathbf{R}^m \to \mathbf{R}$ is defined as

$$
W(\xi) = \min_{\zeta} \left\{ \frac{1}{2} \zeta^\top R \zeta \;\middle|\; A\zeta = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \right\}.
$$

By the Lagrange multiplier method, we can easily show that

$$
W(\xi) = \frac{1}{2} \sum_{1 \le i,j \le m} \xi_i \xi_j K(x_i, x_j).
$$

This implies that the problem [FLOW′] coincides with [SVM]. Next we show the latter part. From the fact that [FLOW] and [POT] are a dual pair, it can be shown that the decision function $f : V \to \mathbf{R}$ defined by (2.4) satisfies the optimality condition of [POT]. □

From Proposition 3.5 and Theorem 3.6, we see that the electric potential coincides with the decision function of [SVM], provided that the optimal solution of [SVM] has a normal support vector. Furthermore, the Lagrange multiplier b^* corresponds to the electric potential of the root vertex x_0, if the potential is normalized in such a way that the earth has zero electric potential.

Next we consider the case of the hard-margin SVM. The following proposition indicates that solving [SVM] with $C = +\infty$ reduces to solving linear equations.

Proposition 3.7. *For the electric network kernel K, an optimal solution u^* of the unconstrained optimization problem*

$$\begin{aligned} [\text{SVM}'] : \min_{u \in \mathbf{R}^m} \quad & \frac{1}{2} \sum_{1 \le i,j \le m} u_i u_j K(x_i, x_j) - \sum_{1 \le i \le m} \eta_i u_i \\ \text{s.t.} \quad & \sum_{1 \le i \le m} u_i = 0 \end{aligned}$$

is also optimal to [SVM] *with* $C = +\infty$.

Proof. Suppose that $\eta_i = +1$ for $1 \le i \le k$ and $\eta_i = -1$ for $k+1 \le i \le m$. By a variant of Lemma 3.4, we may take x_m as the root. Then problem [SVM'] is equivalent to

$$\min_{u \in \mathbf{R}^{m-1}} \frac{1}{2} \sum_{1 \le i,j \le m-1} u_i u_j K(x_i, x_j) - \sum_{1 \le i \le k} 2u_i,$$

where we substitute $u_m = -\sum_{1 \le i \le m-1} u_i$ in [SVM']. Let $\overline{K} = (K(x_i, x_j) \mid 1 \le i, j \le m-1)$. Hence the optimal solution $u^* \in \mathbf{R}^m$ is given by

$$\begin{aligned} u_i^* &= 2 \sum_{1 \le j \le k} (\overline{K}^{-1})_{ij} \quad (1 \le i \le m-1), \\ u_m^* &= -2 \sum_{1 \le j \le k} \sum_{1 \le h \le m-1} (\overline{K}^{-1})_{hj}. \end{aligned}$$

Since $\overline{K}$ is an inverse M-matrix by Proposition 3.2, we have

$$u_i^* \ge 0 \quad (1 \le i \le k), \quad u_i^* \le 0 \quad (k+1 \le i \le m).$$

Hence u^* satisfies the inequality constraint of [SVM] and is optimal. □

Hence, in the case of the hard-margin SVM, the following correspondence holds.

SVM	electric network
positive label data	$+1$ voltage sources
negative label data	-1 voltage sources
optimal solution of [SVM]	electric current from voltage sources
decision function	electric potential

The following corollaries immediately follow from these physical interpretation, where we assume the hard-margin SVM.

Corollary 3.8. *Let $u^* \in \mathbf{R}^m$ be the optimal solution of* [SVM]. *Then, for $i \in \{1, \ldots, m\}$, x_i is a support vector, i.e., $u_i^* > 0$ if and only if there exists a path from x_i to some x_j with $\eta_i \ne \eta_j$ such that it contains no other labeled training vertex (data).*

Suppose that there exists some training data x such that the deletion of x from (V, E) makes two or more connected components, i.e., x is an articulation point of (V, E). Let $U_1, \ldots, U_k$ be the vertex sets of the connected components after the deletion of x. Let $(U_1 \cup \{x\}, E_1), \ldots, (U_k \cup \{x\}, E_k)$ be subgraphs of (V, E). Restricting training data set to each subgraph, we obtain SVM problems [SVM$_1$], …, [SVM$_k$].

Corollary 3.9. *Under the above assumption, the optimal solution of* [SVM] *can be represented as the sum of optimal solutions of* $[\mathrm{SVM_1}], \ldots, [\mathrm{SVM_k}]$. *Consequently, for each* $i \in \{1, \ldots, k\}$, *the restriction to* $U_i \cup \{x\}$ *of the decision function of the hard-margin* [SVM] *coincides with the decision function of* $[\mathrm{SVM}_i]$.

Remark 3.10. SVM with an electric network kernel falls in the scope of *discrete convex analysis* [9], which is a theory of convex functions with additional combinatorial structures. Specifically, the objective function of [SVM] with an electric network kernel is an *M-convex function* in continuous variables, and the optimization problem [SVM] is an M-convex function minimization problem.

Remark 3.11. Smola and Kondor [12] consider various kernels constructed from the Laplacian matrix L of an undirected graph (V, E). In particular, they introduced the kernel

$$K = (I + \sigma L)^{-1},$$

where σ is a positive parameter. In our view, this kernel corresponds to the electric network kernel of a modified graph $(V \cup \{x_0\}, E \cup \{yx_0 \mid y \in V\})$ with a newly introduced root vertex x_0.

The computation of elements of D or K through numerical inversion of $\hat{L}$ is highly expensive because the size of $\hat{L}$ is usually very large. In Section 4, we consider an N-tensor product of k-complete graphs that admits efficient computation of the elements of K.

4 SVM on Tensor Product of Complete Graphs

4.1 Explicit formula for the resistance

In this section, we consider the case where (V, E) is an N-tensor product of k-complete graphs defined as

$$\begin{aligned} V &= \{0, 1, 2, \ldots, k-1\}^N, \\ E &= \{xy \mid x, y \in V,\ d_{\mathrm{H}}(x, y) = 1\}, \end{aligned}$$

where $d_{\mathrm{H}} : V \times V \to \mathbf{R}$ is the Hamming distance defined as

$$d_{\mathrm{H}}(x, y) = \{i \in \{1, \ldots, N\} \mid x^i \neq y^i\},$$

where x^i denotes the ith component of $x \in V$.

In the case of $k = 2$, this graph coincides with an N-dimensional hypercube. We regard (V, E) as an electric network where all resistances of edges are equal to 1. Hence the node admittance matrix of (V, E) coincides with the Laplacian matrix.

By symmetry of this graph, the resistance D between two vertex pair is given as a function in the Hamming distance of the pair as follows. The proof is presented in Subsection 4.3.

Theorem 4.1. *The resistance $D : V \times V \to \mathbf{R}$ of an N-tensor product of k-complete graphs (V, E) is given by*

$$D(x,y) = \begin{cases} \dfrac{1}{2^{N-2}} \displaystyle\sum_{s=1,3,5,\ldots}^{d} \sum_{u=0}^{N-d} \binom{d}{s} \binom{N-d}{u} \frac{1}{2(s+u)} & \text{if } k = 2, \\ \displaystyle\sum_{s=1,3,5,\ldots}^{d} \sum_{t=0}^{d-s} \sum_{u=0}^{N-d} \binom{d}{s} \binom{d-s}{t} \binom{N-d}{u} & \\ \qquad \times \dfrac{2^{2-s} k^{-N+s+t+u}}{k(s+t+u)} \left(\dfrac{1}{2} - \dfrac{1}{k}\right)^t \left(1 - \dfrac{1}{k}\right)^u & \text{if } k \geq 3, \end{cases} \tag{4.1}$$

where $d = d_{\mathrm{H}}(x, y)$.

The theorem implies, in particular, that each element of kernel K can be computed with $O(N^4)$ arithmetic operations. This makes it possible to apply the electric network kernel to large-scale practical problems on this class of graphs.

Remark 4.2. The computational efficiency of the formula (4.1) relies essentially on the fact that the number of distinct eigenvalues of the Laplacian matrix of (V, E) is bounded by $O(N)$; see Subsection 4.3. In a more general case where the graph is an N-tensor product of complete graphs of different sizes, the number of distinct eigenvalues may possibly be exponential in N. Our approach in Subsection 4.3 hints at a difficulty of obtaining a computationally efficient formula for this class of graphs.

4.2 Experimental results

Here, we describe preliminary experiments with our electric network kernels on tensor products of complete graphs. In order to estimate the performance, we compare the electric network kernel with the Hamming kernel and the diffusion kernel [8] using benchmark data having binary attributes. By the Hamming kernel we mean the kernel defined as

$$K(x,y) = N - d_{\mathrm{H}}(x,y) \quad (x, y \in \{0,1\}^N).$$

The diffusion kernel and the electric network kernel are implemented to LIBSVM package [1], which is one of the common SVM package programs. For benchmark data sets, we use `Hepatitis`, `Votes`, `LED2-3`, and `Breast Cancer` taken from UCI Machine Learning Repository [10] (Table 4.1). For the first three data sets, we regard these input spaces as hypercubes, i.e., $k = 2$ in (4.1). In `Hepatitis` data set, we use 12 binary attributes of all 20 attributes. `LED2-3` data set is made through the data generating tool in [10] by adding 10% noise. For `Breast Cancer` data set, we regard its input space as 9-tensor product of 10-complete graphs, i.e., $N = 9$ and $k = 10$ in (4.1).

Table 4.2 shows the experimental results with Hamming kernel (HK), diffusion kernel (DK), and electric network kernel (ENK) for these data sets, where Acc means the ratio of correct answers averaged over 40 random 5-fold cross validations and SVs is the number of support vectors for whole data set. Results are reported for the setting of

Table 4.1: Data sets

Data set	Size	Positive	Negative	Attribute
Hepatitis	155	32	123	12
Votes	435	168	267	16
LED2-3	1914	937	977	7
Breast Cancer	699	251	458	9

Table 4.2: Experimental results

	HK		DK		ENK	
Data set	SVs	Acc (C)	SVs	Acc (C, β)	SVs	Acc (C)
Hepatitis	60	79.125 (64)	60	79.775 (256, 3.5)	106	77.725 (256)
Votes	36	95.975 (4.0)	53	96.025 (512, 3.0)	274	84.450 (128)
LED2-3	386	89.550 (2.0)	392	89.700 (0.2, 3.0)	388	89.800 (70)
Breast Cancer	152	97.271 (0.02)	242	97.000 (2, 0.3)	463	81.713 (200)

HK = Hamming kernel, DK = diffusion kernel, ENK = electric network kernel.

the soft margin parameter C and the diffusion coefficient β achieving the best cross validated error rate.

For Hepatitis and LED2-3 data sets, three kernels show almost equivalent performance. For Votes and Breast Cancer data set, However, our ENK shows somewhat poor performance than others. In Hepatitis, Votes, and Breast Cancer, ENK has larger SVs than other kernels. This phenomenon can be explained by Corollary 3.8 as follows. Since these three data sets are well separated than LED2-3, the soft margin SVM with ENK is close to the hard margin SVM. Hence it is expected from Corollary 3.8 that these SVM with ENK have many SVs.

The above results indicate that our electric network kernel works well as an SVM kernel. It is fair to say, however, that more extensive experiments against various kinds of data sets are required before its performance can be confirmed with more precision and confidence. Comprehensive computational study is left as a future research topic.

4.3 Proof of Theorem 4.1

First, we consider the spectra of a k-complete graph. Let L be the Laplacian matrix of a k-complete graph with vertex set $[k] = \{0, 1, \ldots, k-1\}$, i.e.,

$$L(x, y) = \begin{cases} k-1 & \text{if } x = y \\ -1 & \text{otherwise} \end{cases} \qquad (x, y \in [k]).$$

The spectra of L is given as follows (see [3]).

Lemma 4.3. *L has the eigenvalues 0 with multiplicity 1 and k with multiplicity $k-1$. The eigenvector for 0 is given by*

$$p_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{4.2}$$

and the eigenvectors for $k-1$ are given by

$$p_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ \\ \end{pmatrix}, \; p_2 = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \\ \vdots \\ \end{pmatrix}, \; p_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \\ 0 \\ \vdots \end{pmatrix}, \ldots, p_{k-1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ -k+1 \end{pmatrix}. \tag{4.3}$$

Let Λ and g be diagonal matrices with diagonal elements given by

$$\Lambda(x) = \begin{cases} 0 & \text{if } x = 0 \\ k & \text{otherwise} \end{cases} \quad (x \in [k]), \tag{4.4}$$

$$g(x) = \begin{cases} 1/k & \text{if } x = 0 \\ 1/x(x+1) & \text{otherwise} \end{cases} \quad (x \in [k]), \tag{4.5}$$

and P and Q be matrices defined as

$$P(x,y) = p_y^x, \quad Q(x,y) = g(x)P(y,x) \quad (x, y \in [k]), \tag{4.6}$$

where p_y^x means the x-component of the eigenvector p_y. By the orthogonality of p_y's, we have

$$\sum_{z\in[k]} P(x,z)Q(z,y) = \delta(x,y) \quad (x,y \in [k]), \tag{4.7}$$

where δ is the Kronecker's delta. Then L can be diagonalized as

$$L(x,y) = \sum_{z\in[k]} \Lambda(z)P(x,z)Q(z,y) \quad (x,y \in [k]). \tag{4.8}$$

Second, we derive the diagonalization of the Laplacian matrix L_N of an N-tensor product of k-complete graphs. Note that L_N can be represented as

$$L_N(x,y) = \sum_{i=1}^{N} L(x^i, y^i) \prod_{j=1,\ j\neq i}^{N} \delta(x^j, y^j) \quad (x, y \in [k]^N).$$

From (4.7) and (4.8), L_N can be diagonalized as

$$\begin{aligned} L_N(x,y) &= \sum_{i=1}^{N}\sum_{z^i\in[k]} \Lambda(z^i)P(x^i,z^i)Q(z^i,y^i) \prod_{j=1,j\neq i}^{N} \sum_{z^j\in[k]} P(x^j,z^j)Q(z^j,y^j) \\ &= \sum_{z\in[k]^N} \left(\sum_{i=1}^{N}\Lambda(z^i)\right) \prod_{j=1}^{N} P(x^j,z^j)Q(z^j,y^j) \\ &= \sum_{z\in[k]^N} \Lambda_N(z)P_N(x,z)Q_N(z,y) \quad (x,y\in[k]^N), \end{aligned}$$

where Λ_N, P_N and Q_N are defined as

$$\Lambda_N(x) = \sum_{i=1}^{N}\Lambda(x^i),\ P_N(x,y) = \prod_{j=1}^{N} P(x^j,y^j),\ Q_N(x,y) = \prod_{j=1}^{N} Q(x^j,y^j) \quad (x,y\in[k]^N). \tag{4.9}$$

Note that Q_N is the inverse of P_N.

Finally, we drive the formula for the resistance. We take $\mathbf{0} = (0,0,\dots,0)$ as the root. Let $\hat{L}_N$, $\hat{P}_N$ and $\hat{Q}_N$ be the restrictions of L_N, P_N and Q_N to $[k]^N\setminus\{\mathbf{0}\}$ respectively. Then we have

$$\begin{aligned} \hat{L}_N(x,y) &= \sum_{z\in[k]^N\setminus\{\mathbf{0}\}} \Lambda_N(z)\hat{P}_N(x,z)\hat{Q}_N(z,y) \quad (x,y\in[k]^N\setminus\{\mathbf{0}\}), \\ (\hat{L}_N)^{-1}(x,y) &= \sum_{z\in[k]^N\setminus\{\mathbf{0}\}} (1/\Lambda_N(z))(\hat{Q}_N)^{-1}(x,z)(\hat{P}_N)^{-1}(z,y) \quad (x,y\in[k]^N\setminus\{\mathbf{0}\}). \end{aligned}$$

From the fact that P_N is the inverse of Q_N and definitions of P_N, Q_N, P and Q, it is easy to verify that

$$\begin{aligned} (\hat{P}_N)^{-1}(x,y) &= Q_N(x,y) - Q_N(x,\mathbf{0})Q_N(\mathbf{0},y)/Q_N(\mathbf{0},\mathbf{0}) \\ &= g_N(x)(P_N(y,x)-1), \\ (\hat{Q}_N)^{-1}(x,y) &= P_N(x,y) - P_N(x,\mathbf{0})P_N(\mathbf{0},y)/P_N(\mathbf{0},\mathbf{0}) \\ &= P_N(x,y)-1, \end{aligned}$$

where $g_N(y) = \prod_{i=1}^{N} g(y^i)$. Hence we have

$$D(x,\mathbf{0}) = K(x,x) = \hat{L}_N^{-1}(x,x) = \sum_{z\in[k]^N\setminus\{\mathbf{0}\}} \frac{g_N(z)}{\Lambda_N(z)}(P_N(x,z)-1)^2. \tag{4.10}$$

By symmetry of the graph, if $d_H(y,z) = d$ for $y,z\in[k]^N$, then we have $D(y,z) = D(x,\mathbf{0})$ with $x = (\underbrace{1,\cdots,1}_{d},\underbrace{0,\cdots,0}_{N-d})$. Hence it suffices to consider $D(x,\mathbf{0})$ for such x.

For $A,B\subseteq\{1,\dots,d\}$ with $A\cap B=\emptyset$ and $C\subseteq\{d+1,\dots,N\}$, a subset $S_{A,B,C}\subseteq[k]^N$ is defined as

$$S_{A,B,C} = \{z\in[k]^N \mid z^i = 1\ (i\in A),\ z^i\geq 2\ (i\in B),\ z^i\geq 1\ (i\in C)\ z^i = 0\ (i\notin A\cup B\cup C)\}.$$

Let $s = \#A$, $t = \#B$ and $u = \#C$ with $s + t + u \neq 0$. Then we have

$$\begin{aligned}
&\sum_{z \in S_{A,B,C}} \frac{g_N(z)}{\Lambda_N(z)} (P_N(x,z) - 1)^2 \\
&= \sum_{z \in S_{A,B,C}} \frac{g_N(z)}{k(s+t+u)} \left(\prod_{i=1}^{d} P(1, z^i) \prod_{i=d+1}^{N} P(0, z^i) - 1 \right)^2 \\
&= \frac{\{(-1)^s - 1\}^2}{k(s+t+u)} \sum_{z \in S_{A,B,C}} \prod_{i=1}^{N} g(z^i) \\
&= \frac{\{(-1)^s - 1\}^2}{k(s+t+u)} \sum_{z \in S_{A,B,C}} g(1)^s g(0)^{N-s-t-u} \prod_{i \in B} g(z^i) \prod_{j \in C} g(z^j) \\
&= \frac{\{(-1)^s - 1\}^2}{k(s+t+u)} 2^{-s} k^{-N+s+t+u} \\
&\qquad \times \prod_{i \in B} \left(\frac{1}{6} + \cdots + \frac{1}{k(k+1)} \right) \prod_{j \in C} \left(\frac{1}{2} + \frac{1}{6} + \cdots + \frac{1}{k(k+1)} \right) \\
&= \frac{\{(-1)^s - 1\}^2}{k(s+t+u)} 2^{-s} k^{-N+s+t+u} \left(\frac{1}{2} - \frac{1}{k} \right)^t \left(1 - \frac{1}{k} \right)^u .
\end{aligned}$$

In the case of $k \geq 3$, we have

$$\begin{aligned}
&D(x, \mathbf{0}) \\
&= \sum_{\substack{A,B \subseteq \{1,\ldots,d\}, A \cap B = \emptyset \\ C \subseteq \{d+1,\ldots,N\}}} \sum_{z \in S_{A,B,C}} \frac{g_N(z)}{\Lambda_N(z)} (P_N(x,z) - 1)^2 \\
&= \sum_{s=1,3,\ldots}^{d} \sum_{t=0}^{d-s} \sum_{u=0}^{N-d} \binom{d}{s} \binom{d-s}{t} \binom{N-d}{u} \frac{2^{2-s} k^{-N+s+t+u}}{k(s+t+u)} \left(\frac{1}{2} - \frac{1}{k} \right)^t \left(1 - \frac{1}{k} \right)^u .
\end{aligned}$$

In the case of $k = 2$, B must be empty, and hence we have

$$\begin{aligned}
D(x, \mathbf{0}) &= \sum_{\substack{A \subseteq \{1,\ldots,d\}, \\ C \subseteq \{d+1,\ldots,N\}}} \sum_{z \in S_{A,\emptyset,C}} \frac{g_N(z)}{\Lambda_N(z)} (P_N(x,z) - 1)^2 \\
&= \sum_{s=1,3,\ldots}^{d} \sum_{u=0}^{N-d} \binom{d}{s} \binom{N-d}{u} \frac{2^{2-s} 2^{-N+s+u}}{2(s+u)} \left(1 - \frac{1}{2} \right)^u \\
&= \frac{1}{2^{N-2}} \sum_{s=1,3,\ldots}^{d} \sum_{u=0}^{N-d} \frac{1}{2(s+u)} \binom{d}{s} \binom{N-d}{u} .
\end{aligned}$$

Acknowledgements

The authors thank Satoru Iwata and Shiro Matuura for helpful suggestions. This work is an outcome of a joint project between University of Tokyo and Justsystem Co. This work is also supported by the 21st Century COE Program on Information Science and Technology Strategic Core, and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] C. C. Chang and C. J. Lin: LIBSVM a library for support vector machines, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

[2] F. Chung and S.-T. Yau: Discrete Green's functions, *Journal of Combinatorical Theory Series A* **91** (2000), no. 1-2, 191–214.

[3] D. M. Cvetković, M. Doob and H. Sachs: *Spectra of Graphs*, 3rd. ed., Johann Ambrosius Barth Verlag, Heidelberg–Leipzig, 1995.

[4] M. Fiedler: Some characterizations of symmetric inverse M-matrices, *Linear Algebra and Its Applications* **275/276** (1998), 179–187.

[5] D. Haussler: Convolution kernels on discrete structures, Technical report UCSC-CRL-99-10, Department of Computer Science, University of California at Santa Cruz, 1999.

[6] H. Hirai, K. Murota and M. Rikitoku: SVM Kernel by Electric Network, METR 2004-41, University of Tokyo, August, 2004.

[7] M. Iri: *Network Flow, Transportation and Scheduling*, Academic Press, New York, 1969.

[8] R. I. Kondor and J. Lafferty: Diffusion kernels on graphs and other discrete structures, in: C. Sammut and A. Hoffmann, eds., *Proceedings of the 19th International Conference on Machine Learning,* Morgan Kaufmann, 2002, 315–322.

[9] K. Murota: *Discrete Convex Analysis*, SIAM, Philadelphia, PA, 2003.

[10] P. M. Murphy and D. W. Aha: *UCI Repository of machine learning databases*, `http://www.ics.uci.edu/~mlearn/MLRepository.html`, Irvine, CA: University of California, Department of Information and Computer Science, 1994.

[11] B. Schölkopf and A. J. Smola: *Learning with Kernels,* MIT Press, 2002.

[12] A. J. Smola and R. Kondor: Kernels and regularization on graphs, in: B. Schölkopf and M. Warmuth, eds., *Proceedings of the Annual Conference on Computational Learning Theory and Kernel Workshop,* Lecture Notes in Computer Science 2777, Springer, 2003.

[13] V. Vapnik: *Statistical Leaning Theory,* Wiley, 1998.

[14] C. Watkins: Dynamic alignment kernels, in: A. J. Smola, B. Schöklopf, P. Barlett, and D. Schuurmans, eds., *Advances in Large-Margin Classifiers,* MIT Press 2000, 39–50.