Diffusion Kernels on Graphs

Bruno Jedynak

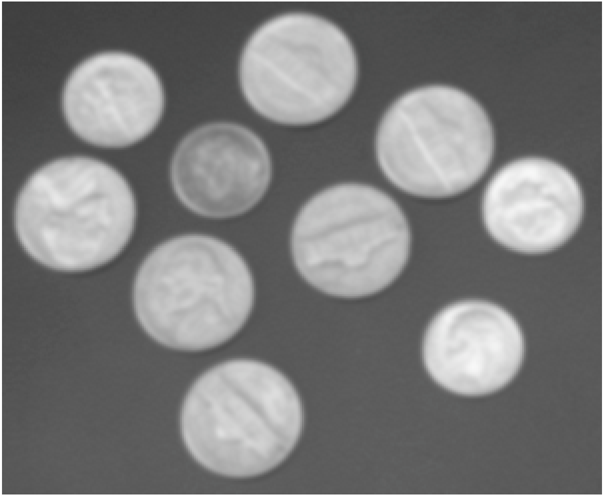
Johns Hopkins University

February 9, 2007

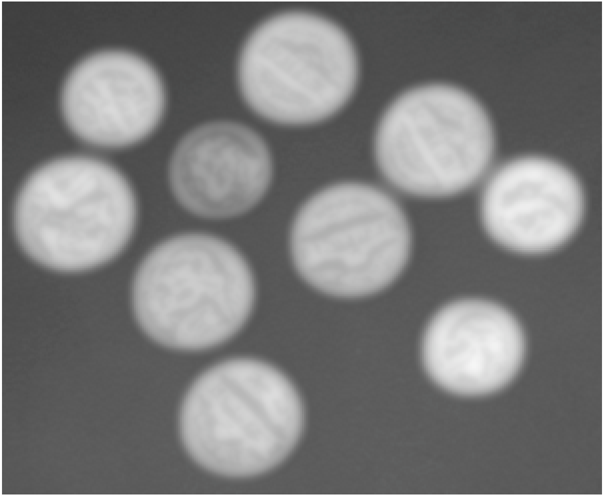
Outline

- Examples of Diffusion
- Motivations for a heat equation when the space variable Belongs at the vertices of a graph
- heat equation when the space variable belongs $\boldsymbol{R}^2$
- Heat equation over a graph
- Application: Models for natural language

Example of diffusion



Example of diffusion



Motivation for a Heat equation over a graph. Image processing

Image smoothing

Image: function $g : S \mapsto \boldsymbol{R}$, S is a discrete set of pixels.

Noisy image: $g(s) + \epsilon(s)$

Recover g by diffusing.

Idea: Build a graph. The vertices are S. Define the edges ... Start at $g(s) + \epsilon(s)$ and diffuse and stop ...

Motivation for a Heat equation over a graph. Language processing

Closed vocabulary V, $\#V = K \approx 10^5$
Training set of words $x_1, \ldots, x_m$, counts $n(w_1), \ldots, n(w_K)$
Want to build a probability mass function π over the words of V
i.e $\pi(w_k) \geq 0$ and $\sum_{k=1}^{K} \pi(w_k) = 1$
What is the probability assigned to unseen words ? ...
Idea: Build a graph. Vertices = V, Define the edges ... start at
$\pi_0(w_k) = m^{-1} n(w_k)$ and diffuse and stop ...

Heat equation in $\boldsymbol{R}^2$

$x = (x_1, x_2) \in \boldsymbol{R}^2, y = (y_1, y_2) \in \boldsymbol{R}^2, t \geq 0, \alpha > 0$
$K_t(x, y)$ is the temperature at time t at x when starting at time $t = 0$ with all the heat concentrated at y. It is called a diffusion kernel.

$$\text{for all } x, \text{ for all } t > 0, \frac{\partial}{\partial t} K_t(x, y) = \alpha \triangle K_t(x, y)$$

$\triangle$ stands for Laplacian.

$$\triangle K_t(x, y) = \frac{\partial^2}{\partial^2 x_1} K_t((x_1, x_2), y) + \frac{\partial^2}{\partial^2 x_2} K_t((x_1, x_2), y)$$

Heat equation in R^2

Without restricting the domain, the solution is given by

$$K_t(x,y) = \frac{1}{4\pi\alpha t} \exp\left(-\frac{1}{4\alpha t}\left((x_1-y_1)^2+(x_2-y_2)^2\right)\right)$$

$K_t(x,y)$ is the density of a

$$N(y, 2\alpha t\ Id)$$

If now the temperature at time 0 is given by $g(x)$ then the solution of the heat equation is the convolution

$$\int\int K_t(x,y)g(y)dy$$

Discretization of the Laplacian

$x = (x_1, x_2) \in \boldsymbol{R}^2,\ f : \boldsymbol{R}^2 \mapsto \boldsymbol{R}$

$$
\begin{aligned}
\frac{\partial}{\partial x_1} f(x_1, x_2) &\approx \frac{1}{h}\left(f(x_1 + \frac{h}{2}, x_2) - f(x_1 - \frac{h}{2}, x_2) \right) \\
\frac{\partial^2}{\partial^2 x_1} f(x_1, x_2) &\approx \frac{1}{h}\left(\frac{\partial}{\partial x_1} f(x_1 + \frac{h}{2}, x_2) - \frac{\partial}{\partial x_1} f(x_1 - \frac{h}{2}, x_2) \right) \\
&\approx \frac{1}{h}\left(\frac{1}{h}\left(f(x_1 + h, x_2) - f(x_1, x_2)\right) - \right. \\
&\qquad \left. \frac{1}{h}\left(f(x_1, x_2) - f(x_1 - h, x_2)\right) \right) \\
&\approx \frac{1}{h^2}\left(f(x_1 + h, x_2) + f(x_1 - h, x_2) - 2f((x_1, x_2)\right) \\
\frac{\partial^2}{\partial^2 x_2} f(x_1, x_2) &\approx \frac{1}{h^2}\left(f(x_1, x_2 + h) + f(x_1, x_2 - h) - 2f((x_1, x_2)\right)
\end{aligned}
$$

Discretization of the Laplacian

$$
\begin{aligned}
\triangle f(x_1, x_2) &= \frac{\partial^2}{\partial^2 x_1} f(x_1, x_2) + \frac{\partial^2}{\partial^2 x_2} f(x_1, x_2) \\
&= \frac{1}{h^2} \left(f(x_1 + h, x_2) + f(x_1 - h, x_2) - 2f((x_1, x_2)) + \right. \\
&\quad \frac{1}{h^2} \left(f(x_1, x_2 + h) + f(x_1, x_2 - h) - 2f((x_1, x_2)) \right)
\end{aligned}
$$

Define $\mathcal{V}(x) = \{(x_1 + h, x_2), (x_1 - h, x_2), (x_1, x_2 - h), (x_1, x_2 + h)\}$ and $d(x) = \#\mathcal{V}(x)$ then

$$
\triangle f(x) = \frac{1}{h^2} \left(\left(\sum_{y \in \mathcal{V}(x)} f(y) \right) - d(x) f(x) \right)
$$

Heat equation over a graph

$G(V, E)$ a non oriented graph.
$V = \{x_1, \ldots, x_n\}$ is the finite set of vertices.
$E \subset V \times V$ is the set of edges. If $(x, y) \in E$, we denote $x \sim y$. Assume no edge from a vertex to itself. Assume G is connected.
The degree of $x \in V$ is $d(x) = \sum_{y \in V} \delta(x \sim y)$
$f : V \mapsto \boldsymbol{R}$ can be seen as a function or as a vector $(f(x_1), \ldots, f(x_n))^T$
$H : V \times V \mapsto \boldsymbol{R}$ can be seen as a function or as a $n \times n$ matrix.

Define the Laplacian (choose $h = 1$)

$$\begin{aligned}
\triangle f(x) &= \left(\sum_{y \in \mathcal{V}(x)} f(y) \right) - d(x) f(x) \\
&= \left(\sum_{y : y \sim x} f(y) \right) - d(x) f(x) \\
&= \sum_{y \in V} \left(f(y) \delta(y \sim x) - d(y) f(y) \delta(x = y) \right) \\
&= \sum_{y \in V} \left(\delta(y \sim x) - d(y) \delta(x = y) \right) f(y) \\
&= \sum_{y \in V} H(x, y) f(y) \\
&= H f(x)
\end{aligned}$$

Laplacian

$$H(x, y) = \delta(y \sim x) - d(y)\delta(x = y)$$

$$H = A - D$$

$A(x, y) = \delta(x \sim y)$ is the adjacency matrix of G
$D(x, y) = d(x)\delta(x = y)$ is the degree matrix. D is diagonal.

Heat Equation

$x, y \in V$, $t \geq 0$.
$K_t(x, y)$ is the temperature at x at time t when starting with a unit temperature at y at time 0.
$K_0(x, y) = \delta(x = y)$ which in matrix notation is $K_0 = Id$
We define the heat equation for a fixed $y \in V$ as:

$$\text{for each } x \in V, \text{ for each } t > 0, \frac{\partial}{\partial t} K_t(x, y) = HK_t(x, y)$$

Notate $u_t(x) = K_t(x, y)$

$$\begin{aligned} \frac{\partial}{\partial t} u_t(x) &= \sum_{z \in V} H(x, z) u_t(z) \\ &= \left(\sum_{z : z \sim x} u_t(z) \right) - d(x) u_t(x) \\ &= d(x) \left(\left(\frac{1}{d(x)} \sum_{z : z \sim x} u_t(z) \right) - u_t(x) \right) \end{aligned}$$

Claims:

The heat equation admits a unique solution $K_t = e^{tH}$

$$e^{tH} = Id + tH + \frac{t^2}{2!}H^2 + \frac{t^3}{3!}H^3 + \dots$$

$$e^{tH} = \lim_{k \to +\infty} (Id + \frac{t}{k}H)^k$$

Starting with a temperature $\pi(x)$, $x \in V$, the solution to the heat equation is $K_t \pi$

If for all x, $\pi(x) \geq 0$ and $\sum_{x \in V} \pi(x) = 1$ then for all $x \in V$ and all $t > 0$, $K_t \pi(x) > 0$ and $\sum_{x \in V} K_t \pi(x) = 1$

Markov Chain interpretation ...

Examples

- Complete graph with n vertices. $x \sim y \iff x \neq y$

$$\begin{aligned} K_t(x,y) &= \frac{1}{n}\left(1+(n-1)e^{-nt}\right) \text{ if } x \neq y \\ &= \frac{1}{n}\left(1-e^{-nt}\right) \text{ if } x \neq y \end{aligned}$$

- Vertices are binary strings of length n.
  $x \sim y \iff Hamming(x,y) = 1$

$$K_t(x,y) = \frac{1}{2^n}(1+e^{-2t})^n(\tanh(t))^{H(x,y)}$$

- small graphs. Diagonalize H

Application to language modeling

Closed vocabulary V, $\#V = K \approx 10^5$
Training set of words $x_1, \dots, x_m$, counts $n(w_1), \dots, n(w_K)$
Want to build a probability mass function π over the words of V
i.e $\pi(w_k) \geq 0$ and $\sum_{k=1}^{K} \pi(w_k) = 1$ From observed counts
$\pi_0(w_k) = m^{-1} n(w_k)$Vertices $= V$,

- Choose the complete graph over V. $x \sim y \iff x \neq y$. Start at π_0 Then ... for all $w \in V$, $K_t \pi(w) = (1 - \lambda(t))\pi_0(w) + \lambda(t)\frac{1}{K}$
- Choose $v \sim w \iff |n(v) - n(w)| \leq 1$, compute $(Id + \frac{t}{3}H)^3 \pi_0$ with $t = \frac{1}{K}$ yields fast and competitive results.

Thank you