Laplace-Matrix

Autor und Herausgeber

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Zusammenfassung

Die Laplace-Matrix eines gewichteten Graphen mit n Knoten ist die $(n \times n)$-Matrix $\mathbf{L}^{(\mathcal{G})} = \mathbf{D}^{(\mathcal{G})} - \mathbf{A}^{(\mathcal{G})}$, die Gradmatrix minus die Adjazenzmatrix der Kantengewichte. Ihre quadratische Form $\mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x}$ ist die Summe über die Kanten aus dem Kantengewicht mal der quadrierten Differenz der beiden Einträge von $\mathbf{x}$ an den Enden der Kante. Die Laplace-Matrix ist daher eine symmetrische Matrix und positiv semidefinit, der konstante Vektor ist ein Eigenvektor zum Eigenwert 0, und die Vielfachheit des Eigenwerts 0 ist gleich der Anzahl der zusammenhängenden Komponenten des Graphen. Zwei normierte Varianten teilen die Kantengewichte durch die Knotengrade; sie haben dieselben Eigenwerte, welche zwischen 0 und 2 liegen. Im maschinellen Lernen ist die quadratische Form der Strafterm der Laplace-Regularisierung, die dafür sorgt, dass die an zusammenhängenden Knoten gelernten Werte einander ähneln, und den Wert an einem unmarkierten Knoten als das gewichtete Mittel seiner Nachbarn vorhersagt. Die Eigenvektoren ihrer kleinsten Eigenwerte sind die Koordinaten, die das Spektrales Clustering an k-Means übergibt.

Definition

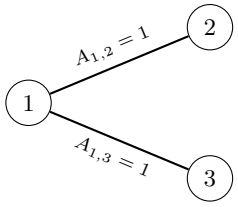
Zwölf Wetterstationen rund um Krems an der Donau, von Melk und Zwettl im Westen bis Retz und Langenlebarn im Osten, haben jeweils eine Jahresmitteltemperatur für das Jahr 2024 aufgezeichnet: 12.6 Grad Celsius in Krems, 9.6 auf dem Berg Jauerling. Zwei höchstens 35 km voneinander entfernte Stationen sind durch eine Kante verbunden, deren Gewicht mit ihrem Abstand abnimmt, sodass die Stationen einen gewichteten Graphen $\mathcal{G}$ mit 27 Kanten bilden (Abb. 2(a)). Wie stark schwankt die Temperatur entlang der Kanten von $\mathcal{G}$? Eine Antwort ist die Summe über die Kanten aus dem Kantengewicht mal der quadrierten Differenz der beiden Temperaturen an ihren Enden. Die Laplace-Matrix von $\mathcal{G}$ ist die Matrix, die diese Summe als quadratische Form im Vektor der zwölf Temperaturen berechnet.

Betrachtet wird ein ungerichteter Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ mit Knoten $\mathcal{V} = \{1, \ldots, n\}$ und nichtnegativen Kantengewichte $A_{i,i'} = A_{i',i}$, wobei $A_{i,i'} = 0$ gilt, wenn $\mathcal{G}$ keine Kante $\{i, i'\}$ besitzt (von Luxburg, 2007, Abschn. 2.1). Die Kantengewichte sind die Einträge der Adjazenzmatrix $\mathbf{A}^{(\mathcal{G})} \in \mathbb{R}^{n \times n}$, und der Knotengrad $d^{(i)} := \sum_{i'=1}^{n} A_{i,i'}$ des Knotens i ist die Summe der Gewichte

seiner Kanten. Die Laplace-Matrix von $\mathcal{G}$ ist die Differenz

$$\mathbf{L}^{(\mathcal{G})} := \mathbf{D}^{(\mathcal{G})} - \mathbf{A}^{(\mathcal{G})}, \text{ mit der Gradmatrix } \mathbf{D}^{(\mathcal{G})} := \operatorname{diag}\left\{d^{(1)}, \ldots, d^{(n)}\right\}. \quad (1)$$

Der Diagonaleintrag $L_{i,i}^{(\mathcal{G})}$ der Laplace-Matrix (1) ist der Knotengrad $d^{(i)}$, und der Außerdiagonaleintrag $L_{i,i'}^{(\mathcal{G})}$ ist das negierte Kantengewicht $-A_{i,i'}$, sodass die Summe jeder Zeile null ergibt (Hastie et al., 2009, Abschn. 14.5.3; von Luxburg, 2007, Abschn. 3.1). Abb. 1 führt die Konstruktion für einen Graphen mit drei Knoten und zwei Kanten durch. Beim Stationsgraphen hat die Kante zwischen Krems und dem 8 km entfernten Langenlois das Gewicht 0.92, die längsten Kanten haben das Gewicht 0.23, und Krems hat durch seine sechs Nachbarn den Grad 3.18.



$$\mathbf{L}^{(\mathcal{G})} = \mathbf{D}^{(\mathcal{G})} - \mathbf{A}^{(\mathcal{G})} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Abbildung 1: Ein ungerichteter Graph mit drei Knoten und zwei Kanten, beide mit dem Kantengewicht 1, und seine Laplace-Matrix (1). Knoten 1 hat den Grad 2 und die Knoten 2 und 3 haben den Grad 1, was den Diagonaleinträgen entspricht; jede Kante trägt -1 zu den beiden Außerdiagonaleinträgen bei, die sie verbindet, und die übrigen Einträge sind 0

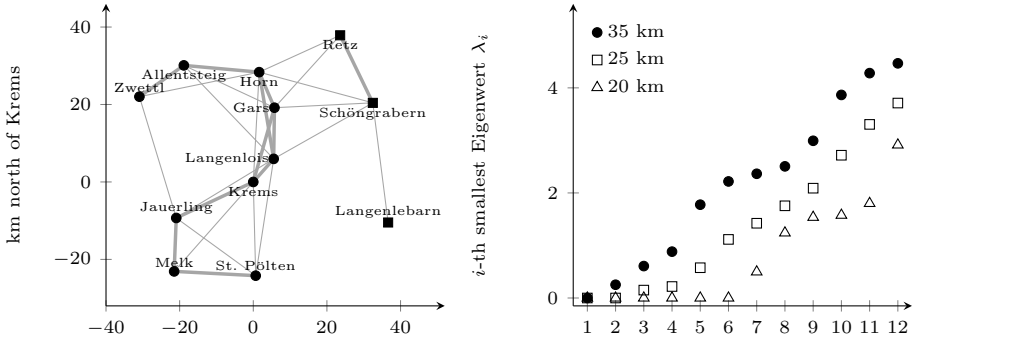
Die Laplace-Matrix ist durch ihre quadratische Form charakterisiert. Für jeden Vektor $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, mit einem Eintrag pro Knoten, gilt

$$\mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x} = \sum_{\{i,i'\} \in \mathcal{E}} A_{i,i'} \left(x_i - x_{i'}\right)^2. \quad (2)$$

Die Identität folgt durch Ausmultiplizieren von $\mathbf{x}^\top \mathbf{D}^{(\mathcal{G})} \mathbf{x} - \mathbf{x}^\top \mathbf{A}^{(\mathcal{G})} \mathbf{x} = \sum_{i=1}^n d^{(i)} x_i^2 - \sum_{i=1}^n \sum_{i'=1}^n A_{i,i'} x_i x_{i'}$ und Zusammenfassen der beiden Terme jeder Kante (Hastie et al., 2009, Abschn. 14.5.3; von Luxburg, 2007, Prop. 1). Für den Stationsgraphen und den Vektor der Mitteltemperaturen beträgt die quadratische Form (2) 26.9 Quadratgrad; die Kante zwischen Jauerling und Melk, mit einem Gewicht von 0.79 und einer Temperaturdifferenz von 2.8 Grad, trägt 6.0 dazu bei.

Aus der quadratischen Form (2) folgen drei Eigenschaften der Laplace-Matrix (von Luxburg, 2007, Prop. 1, Prop. 2; Mohar, 1991). Erstens ist $\mathbf{L}^{(\mathcal{G})}$ eine symmetrische Matrix und positiv semidefinit, da jeder Term der Summe nichtnegativ ist, sodass ihre Eigenwerte reell sind und als $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ geordnet werden können. Zweitens lässt der konstante Vektor $\mathbf{1} = (1, \ldots, 1)^\top$ jeden Term verschwinden, und es gilt $\mathbf{L}^{(\mathcal{G})} \mathbf{1} = \mathbf{0}$, da jede Zeile in der Summe null

ergibt, sodass $\lambda_1 = 0$ mit dem Eigenvektor $\mathbf{1}$ gilt. Drittens ist die Vielfachheit des Eigenwerts 0 gleich der Anzahl der zusammenhängenden Komponenten von $\mathcal{G}$. Ein Vektor $\mathbf{x}$ mit $\mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x} = 0$ nimmt an den beiden Enden jeder Kante denselben Wert an, ist also auf jeder Komponente konstant, und die IndikatorVektoren der Komponenten spannen den Nullraum von $\mathbf{L}^{(\mathcal{G})}$ auf. Der StationsGraph ist zusammenhängend, sodass $\lambda_1 = 0$ einfach und $\lambda_2 = 0.25$ positiv ist. Verbindet man nur Stationen, die höchstens 25 km voneinander entfernt sind, bleibt Langenlebarn ohne Kante und der Graph hat zwei Komponenten, und die Grenze von 20 km liefert sechs Komponenten; der Eigenwert 0 hat dann die Vielfachheit zwei bzw. sechs (Abb. 2(b)). Der zweitkleinste Eigenwert λ_2 ist die algebraische Konnektivität von $\mathcal{G}$, und der Eigenvektor zu λ_2 ist der Fiedler-Vektor (Fiedler, 1973). Die Vorzeichen seiner Einträge teilen die Knoten in zwei Cluster mit wenigen Kanten zwischen ihnen auf (Fiedler, 1975). Beim Stationsgraphen haben die drei mehr als 20 km östlich von Krems gelegenen Stationen, Langenlebarn, Schöngrabern und Retz, negative Einträge und die übrigen neun positive Einträge.



(a) the station Graph (b) Eigenwerte of the Laplacian matrix of three Graphen

Abbildung 2: (a) Zwölf Wetterstationen rund um Krems an der Donau, eingezeichnet an ihren Positionen in km östlich und nördlich von Krems und verbunden durch eine Kante, wenn sie höchstens 35 km voneinander entfernt sind; eine dicke Linie kennzeichnet ein Kantengewicht über 1/2, eine dünne Linie eines darunter. Ausgefüllte Kreise markieren die Stationen mit einem nichtnegativen Eintrag des Fiedler-Vektors, ausgefüllte Quadrate jene mit einem negativen Eintrag. (b) Die Eigenwerte der Laplace-Matrix in aufsteigender Reihenfolge für den Graphen in (a) (ausgefüllte Kreise) sowie für die Graphen, die Stationen in höchstens 25 km (Quadrate) und 20 km (Dreiecke) Entfernung verbinden, wobei die Legende die größte Kantenlänge angibt. Der Eigenwert 0 tritt einmal, zweimal bzw. sechsmal auf, was der Anzahl der zusammenhängenden Komponenten des jeweiligen Graphen entspricht. Daten erzeugt mit `pythondemos/LapMat.py`

Zwei normierte Varianten der Laplace-Matrix teilen die Kantengewichte

durch die Knotengrade (von Luxburg, 2007, Abschn. 3.2),

$$\mathbf{L}^{(\mathrm{sym})} := \left(\mathbf{D}^{(\mathcal{G})}\right)^{-1/2} \mathbf{L}^{(\mathcal{G})} \left(\mathbf{D}^{(\mathcal{G})}\right)^{-1/2} \text{ und } \mathbf{L}^{(\mathrm{rw})} := \left(\mathbf{D}^{(\mathcal{G})}\right)^{-1} \mathbf{L}^{(\mathcal{G})},$$

für einen Graphen ohne isolierte Knoten. Die erste ist eine symmetrische Matrix, deren quadratische Form die Summe (2) ist, wobei jedes x_i durch $x_i/\sqrt{d^{(i)}}$ ersetzt wird. Die zweite ist $\mathbf{I} - \left(\mathbf{D}^{(\mathcal{G})}\right)^{-1} \mathbf{A}^{(\mathcal{G})}$, die Einheitsmatrix minus die Übergangsmatrix der Irrfahrt auf $\mathcal{G}$, die mit einer zum Kantengewicht proportionalen Wahrscheinlichkeit zu einem Nachbarn wechselt. Die beiden Varianten haben dieselben Eigenwerte, welche zwischen 0 und 2 liegen (Chung, 1997, Lemma 1.7); der Eigenwert 0 hat den Eigenvektor $\mathbf{1}$ für $\mathbf{L}^{(\mathrm{rw})}$ und $\left(\mathbf{D}^{(\mathcal{G})}\right)^{1/2}\mathbf{1}$ für $\mathbf{L}^{(\mathrm{sym})}$, und seine Vielfachheit zählt wiederum die zusammenhängenden Komponenten (von Luxburg, 2007, Prop. 3, Prop. 4). Beim Stationsgraphen beträgt der größte Eigenwert von $\mathbf{L}^{(\mathrm{sym})}$ 1.57. Welche der drei Matrizen ein Verfahren verwendet, ist eine Entwurfsentscheidung (von Luxburg, 2007, Abschn. 8.5): Das Spektrales Clustering wurde mit $\mathbf{L}^{(\mathrm{rw})}$ (Shi and Malik, 2000), mit $\mathbf{L}^{(\mathrm{sym})}$ (Ng et al., 2001) und mit $\mathbf{L}^{(\mathcal{G})}$ (von Luxburg, 2007, Abschn. 4) formuliert.
Im maschinellen Lernen tritt die Laplace-Matrix über ihre quadratische Form (2) als Strafterm in Erscheinung. Betrachtet seien Datenpunkte, die die Knoten von $\mathcal{G}$ bilden, bei dem es sich dann um einen Ähnlichkeitsgraph handelt, wobei ein Label y_i nur für die Knoten in einer Teilmenge $\mathcal{A} \subset \mathcal{V}$ bekannt ist. Beispiele hierfür sind Messstationen mit vorliegenden Aufzeichnungen oder die Nutzer eines sozialen Netzwerks, die ein bestimmtes Profilfeld ausgefüllt haben. Eine Hypothese ist ein Vektor $\mathbf{x} \in \mathbb{R}^n$, der jedem Knoten einen Wert zuordnet, und der Hypothesenraum ist $\mathbb{R}^n$. Die Laplace-Regularisierung addiert die quadratische Form zum durchschnittlichen quadratischen Fehler auf den markierten Knoten,

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{|\mathcal{A}|} \sum_{i\in\mathcal{A}} \left(x_i - y_i\right)^2 + \alpha\, \mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x}, \tag{3}$$

mit einem Hyperparameter $\alpha > 0$ (Belkin et al., 2004; Zhou and Schölkopf, 2004). Das Hinzufügen eines Strafterms ist einer der drei Wege, die die Regularisierung unterscheidet, und dieser setzt die Annahme um, dass Knoten, die durch eine stark gewichtete Kante verbunden sind, ähnliche Werte aufweisen. Der Trainingsfehler allein sagt nichts über einen Knoten außerhalb von $\mathcal{A}$ aus; der Strafterm ist der einzige Term, der dessen Wert einschränkt, sodass die Generalisierung von den markierten auf die unmarkierten Knoten aus den Kanten resultiert. Die Zielfunktion (3) ist konvex, und ihr Minimierer löst ein lineares Gleichungssystem, da der Gradient der quadratischen Form $2\mathbf{L}^{(\mathcal{G})}\mathbf{x}$ lautet. Für $\alpha \to 0$ hält der Minimierer $x_i = y_i$ auf den markierten Knoten fest und minimiert die quadratische Form über die verbleibenden Einträge. Der Eintrag von $\mathbf{L}^{(\mathcal{G})}\mathbf{x}$ an einem unmarkierten Knoten ist dann null. Ein solcher Knoten nimmt das gewichtete Mittel seiner Nachbarn an, $x_i = (1/d^{(i)}) \sum_{i'=1}^{n} A_{i,i'} x_{i'}$, was der harmonischen Lösung von Zhu et al. (2003) entspricht.

Lässt man nacheinander jede Station aus Abb. 2(a) weg und sagt ihre Durchschnittstemperatur durch dieses Mittel ihrer Nachbarn voraus, ergibt sich ein

mittlerer absoluter Fehler von 0.86 Grad. Sagt man stattdessen den Mittelwert der anderen elf Stationen voraus, ergibt sich ein Wert von 1.12 Grad. Der größte Fehler tritt am Jauerling auf, 955 m über dem Meeresspiegel, mit einer Vorhersage von 12.3 Grad gegenüber gemessenen 9.6 Grad. Die Kantengewichte kodieren die Abstand, aber nicht die Höhenlage, sodass der Strafterm von einer Bergstation verlangt, den umliegenden Talstationen zu ähneln. Dieselbe quadratische Form, mit den Modellparameter lokaler Modelle anstelle der Skalare x_i, ist die verallgemeinerte totale Variation, die die verallgemeinerte Totalvariationsminimierung beim föderierten Lernen bestraft. Die Eigenvektoren zu den kleinsten Eigenwerte einer Laplace-Matrix sind die Koordinaten, die das Spektrales Clustering an k-Means übergibt (von Luxburg, 2007, Abschn. 4).

Siehe auch: Graph, gewichteter Graph, Kantengewicht, Adjazenzmatrix, Knotengrad, Matrix, positiv semidefinit, Eigenwert, Eigenvektor, Eigenwertzerlegung, zusammenhängend, algebraische Konnektivität, Fiedler-Vektor, Spektrales Clustering, Graph-Clustering, Ähnlichkeitsgraph, Regularisierung, Strafterm, GTV, GTVMin.

Querverweise

Literatur

1. Chung (1997). Spectral Graph Theory. American Mathematical Society, Providence, RI, USA.

2. Fiedler (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), pp. 298–305.

3. Fiedler (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. Czechoslovak Mathematical Journal, 25(4), pp. 619–633.

4. von Luxburg (2007). A tutorial on spectral clustering. Statist. Comput., 17(4), pp. 395–416.

5. Mohar (1991). The Laplacian Spectrum of Graphs. In: Graph Theory, Combinatorics, and Applications, pp. 871–898. Wiley.

6. Ng et al. (2001). On Spectral Clustering: Analysis and an Algorithm. In: Adv. Neural Inf. Process. Syst., pp. 849–856.

7. Shi and Malik (2000). Normalized Cuts and Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 22(8), pp. 888–905.

8. Belkin et al. (2004). Regularization and semi-supervised learning on large graphs. In: COLT, pp. 624–638.

9. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.

10. Zhou and Schölkopf (2004). A regularization framework for learning from graph data. In: ICML workshop on statistical relational learning and Its connections to other fields, pp. 67–68.

11. Zhu et al. (2003). Semi-supervised learning using gaussian fields and harmonic functions. In: ICML, pp. 912–919.