Spektrales Clustering

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

Spektrales Clustering ist ein Verfahren des Graph-Clusterings, das die Cluster eines Datensatzes aus den Eigenvektoren der Laplace-Matrix eines Graphen berechnet, dessen Knoten die Datenpunkte sind und dessen Kantengewichte ihre Ähnlichkeit ausdrücken. Die Laplace-Matrix ist positiv semidefinit, und die Vielfachheit ihres Eigenwerts 0 ist gleich der Anzahl der zusammenhängenden Komponenten des Graphen, mit deren Indikator-Vektoren als Eigenvektoren. Für k Cluster bilden die Einträge der k Eigenvektoren zu den kleinsten Eigenwerten für jeden Knoten einen neuen Merkmalsvektor, und k-Means auf diesen Merkmalsvektoren liefert die Cluster. Für zwei Cluster wendet das Verfahren einen Schwellenwert auf den Fiedler-Vektor an. Spektrales Clustering ist eine Relaxierung des Ratio-Cuts, der das Gesamtgewicht der Kanten zwischen den Clustern durch deren Größen teilt, und der zweitkleinste Eigenwert ist eine untere Schranke für den Ratio-Cut jeder Bipartition. Im Gegensatz zu k-Means auf den ursprünglichen Merkmalsvektoren findet spektrales Clustering Cluster, die im Graphen zusammenhängend, aber im Merkmalsraum nicht konvex sind. Die Anzahl der Cluster und die Konstruktion des Graphen sind seine Hyperparameter.

Definition

Jeder Tag des Jahres 2024 an der Wetterstation Krems an der Donau (Österreich) ist ein Datenpunkt mit einem Merkmalsvektor $\mathbf{x} \in \mathbb{R}^2$, der den Messwert von 06:00 Uhr und die Temperatur um 15:00 Uhr (UTC) des Tages enthält. Verbindet man jeden Tag mit seinen zehn nächsten Tagen, gemessen an der euklidischer Abstand zwischen ihren Merkmalsvektoren, werden die 365 Tage zu den Knoten $i = 1, \ldots, n$ eines k-Nächste-Nachbarn-Graphen $\mathcal{G}$ mit $n = 365$ Knoten und 2214 Kanten (Abb. 2(a)). Ein Cluster dieses Graphen ist eine Menge von Tagen, die untereinander durch viele Kanten und mit den übrigen Tagen durch wenige Kanten verbunden sind. Spektrales Clustering ist ein Verfahren des Graph-Clusterings, das solche Cluster aus den Eigenvektoren der Laplace-Matrix des Graphen bestimmt (Hastie et al., 2009, Abschn. 14.5.3; von Luxburg, 2007; Ng et al., 2001; Shi and Malik, 2000).
Betrachtet wird ein Ähnlichkeitsgraph des Datensatzes: ein gewichteter Graph $\mathcal{G}$ mit Knoten $i = 1, \ldots, n$, einem Knoten pro Datenpunkt, und Kantengewichte

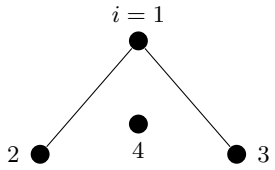
$A_{i,i'} = A_{i',i} \geq 0$, welche die Ähnlichkeit zwischen den Datenpunkten i und i' ausdrücken; eine fehlende Kante hat das Gewicht 0. Die Kantengewichte bilden die Adjazenzmatrix $\mathbf{A}^{(\mathcal{G})} \in \mathbb{R}^{n \times n}$, und der Knotengrad $d^{(i)} = \sum_{i'=1}^{n} A_{i,i'}$ des Knotens i ist die Summe seiner Kantengewichte. Die Laplace-Matrix von $\mathcal{G}$ ist

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

Für jeden Vektor $\mathbf{u} \in \mathbb{R}^n$ mit Einträgen u_i erfüllt die Laplace-Matrix (1)

$$\mathbf{u}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{u} = \frac{1}{2} \sum_{i=1}^{n} \sum_{i'=1}^{n} A_{i,i'} \left(u_i - u_{i'}\right)^2, \tag{2}$$

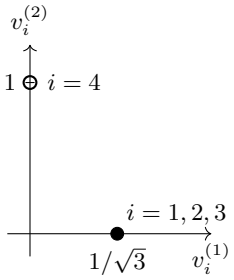
eine gewichtete Summe quadrierter Differenzen über die Kanten hinweg (von Luxburg, 2007, Prop. 1). Die quadratische Form (2) ist niemals negativ, sodass $\mathbf{L}^{(\mathcal{G})}$ eine symmetrische Matrix ist, die positiv semidefinit ist. Sie besitzt eine Eigenwertzerlegung $\mathbf{L}^{(\mathcal{G})} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ mit einer orthogonalen Matrix $\mathbf{V} = \left(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)}\right)$ von Eigenvektoren und mit Eigenwerten $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Der konstante Vektor ist ein Eigenvektor zum Eigenwert 0, da sich jede Zeile von $\mathbf{L}^{(\mathcal{G})}$ zu null summiert. Allgemeiner ist die Vielfachheit des Eigenwerts 0 gleich der Anzahl der zusammenhängenden Komponenten von $\mathcal{G}$, und die Indikator-Vektoren der Komponenten spannen seine Eigenvektoren auf (von Luxburg, 2007, Prop. 2). Der Grund dafür ist die quadratische Form (2): $\mathbf{L}^{(\mathcal{G})}\mathbf{u} = \mathbf{0}$ erzwingt $u_i = u_{i'}$ über jede Kante hinweg, sodass $\mathbf{u}$ auf jeder Komponente konstant ist. Abb. 1 zeigt einen ungerichteter Graphen mit zwei Komponenten; seine beiden Eigenvektoren zum Eigenwert 0 platzieren die Knoten jeder Komponente auf einem einzigen Punkt eines Streudiagramms.



(a)

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

(b)



(c)

$$\mathbf{v}^{(1)} = \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

(d)

Abbildung 1: (a) Ein ungerichteter Graph $\mathcal{G}$ mit vier Knoten $i = 1, 2, 3, 4$ und Einheits-Kantengewichte, wobei jeder Knoten ein Datenpunkt ist; Knoten 4 hat keine Kante, sodass der Graph zwei zusammenhängende Komponenten hat. (b) Seine Laplace-Matrix $\mathbf{L}^{(\mathcal{G})} \in \mathbb{R}^{4\times 4}$, deren Eigenwert 0 die Vielfachheit zwei hat. (c) Streudiagramm der Knoten unter Verwendung der neuen Merkmalsvektoren $\widetilde{\mathbf{x}}^{(i)} = \left(v_i^{(1)}, v_i^{(2)}\right)^\top$: Die Knoten 1, 2, 3 der ersten Komponente teilen sich einen Punkt (ausgefüllter Kreis) und Knoten 4 liegt an einem anderen (offener Kreis). (d) Die beiden Eigenvektoren $\mathbf{v}^{(1)}, \mathbf{v}^{(2)} \in \mathbb{R}^4$ zum Eigenwert 0, die normierten Indikator-Vektoren der beiden Komponenten

Ein Graph, dessen Cluster durch wenige Kanten mit kleinem Gewicht verbunden sind, ist eine Störung eines Graphen mit mehreren Komponenten. Die Eigenvektoren seiner kleinsten Eigenwerte sind dann eine Störung der Indikator-Vektoren (von Luxburg, 2007, Abschn. 7). Spektrales Clustering mit k Clustern verwendet daher diese Eigenvektoren als neue Merkmale: Knoten i wird durch den Merkmalsvektor

$$\widetilde{\mathbf{x}}^{(i)} := \left(v_i^{(1)}, \ldots, v_i^{(k)}\right)^\top \in \mathbb{R}^k, \tag{3}$$

dargestellt, der die i-ten Einträge der k Eigenvektoren zu den kleinsten Eigenwerten enthält. Anschließend wird k-Means auf die Merkmalsvektoren $\widetilde{\mathbf{x}}^{(1)}, \ldots, \widetilde{\mathbf{x}}^{(n)}$ angewendet (von Luxburg, 2007, Abschn. 4). Algorithmus 1 gibt das Verfahren vollständig an.

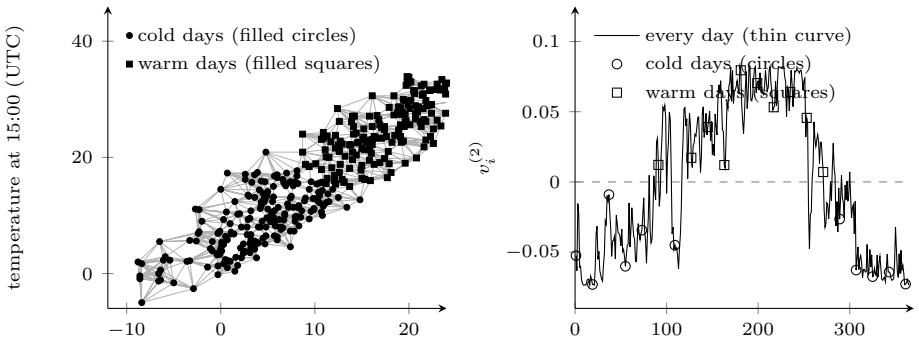
Algorithm 1 Spektrales Clustering (von Luxburg, 2007, Sect. 4)

Input: gewichteter Graph $\mathcal{G}$ with nodes $i = 1, \ldots, n$ and Kantengewichte $A_{i,i'}$, Cluster count k

1: compute the Laplace-Matrix (1) of $\mathcal{G}$
2: compute Eigenvektoren $\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(k)}$ of $\mathbf{L}^{(\mathcal{G})}$ for its k smallest Eigenwerte
3: form the Merkmalsvektoren (3) for $i = 1, \ldots, n$
4: run k-Means with k Cluster on $\widetilde{\mathbf{x}}^{(1)}, \ldots, \widetilde{\mathbf{x}}^{(n)}$

Output: the Cluster of each node i, i.e., of each Datenpunkt

Für $k = 2$ ist der erste Eigenvektor konstant, sodass das neue Merkmal von Knoten i sein Eintrag $v_i^{(2)}$ im Fiedler-Vektor ist, dem Eigenvektor zur algebraische Konnektivität λ_2 (Fiedler, 1973). Der Fiedler-Vektor ist orthogonal zum konstanten Vektor, besitzt also Einträge beider Vorzeichen, und k-Means auf seinen Einträgen läuft auf einen Schwellenwert hinaus. Bei den Tagen in Krems ist der Graph zusammenhängend mit $\lambda_2 = 0.044$ und $\lambda_3 = 0.179$, und der Fiedler-Vektor hat 190 negative und 175 positive Einträge (Abb. 2(b)). K-Means auf diesen Einträgen liefert 194 kalte Tage und 171 warme Tage. Jeder Tag im Januar, Februar, November und Dezember ist ein kalter Tag, und jeder Tag von Mai bis August ist ein warmer Tag. Jedes der beiden Cluster ist im Graphen zusammenhängend, und die beiden sind durch 39 der 2214 Kanten verbunden. K-Means auf den ursprünglichen Merkmalsvektoren $\mathbf{x}^{(i)}$, dessen Cluster stets konvexe Regionen des Merkmalsraums sind, liefert fast dasselbe hartes Clustering: Die beiden Verfahren weichen nur an zwei Tagen im Oktober voneinander ab. Die Verfahren unterscheiden sich, sobald die Cluster nicht konvex sind. Datenpunkte auf zwei konzentrischen Ringen werden durch die Kanten eines k-Nächste-Nachbarn-Graphen entlang jedes Rings verbunden; spektrales Clustering gibt die Ringe zurück, während k-Means auf den ursprünglichen Merkmalsvektoren sie durchschneidet (Hastie et al., 2009, Abschn. 14.5.3).



(a) the 10-nearest-neighbor graph of the 365 days

(b) entries of the Fiedler-Vektor $\mathbf{v}^{(2)}$

Abbildung 2: (a) Die 365 Tage des Jahres 2024 in Krems, jeder als ein an seinem Merkmalsvektor gezeichneter Knoten, und die 2214 Kanten (dünne graue Linien) des 10-Nächste-Nachbarn-Graphen; alle Tage sind eingezeichnet, da die Teilabbildung den Zusammenhang des Graphen zeigt. Die beiden durch spektrales Clustering mit $k = 2$ gefundenen Cluster sind die kalten Tage (ausgefüllte Kreise) und die warmen Tage (ausgefüllte Quadrate). (b) Der Eintrag $v_i^{(2)}$ des Fiedler-Vektors für jeden Tag (dünne Kurve) aufgetragen gegen den Tag des Jahres, wobei jeder 18. Tag, insgesamt 21 Tage, als kalter Tag (Kreise) oder warmer Tag (Quadrate) markiert ist; die gestrichelte Linie ist die Nulllinie. Daten erzeugt durch `pythondemos/specclustering.py`

Spektrales Clustering löst näherungsweise ein Partitionierungsproblem für Graphen (von Luxburg, 2007, Abschn. 5). Für eine Teilmenge $\mathcal{C} \subseteq \mathcal{V}$ der Knoten ist der Schnitt $\operatorname{cut}(\mathcal{C}) := \sum_{i \in \mathcal{C}} \sum_{i' \notin \mathcal{C}} A_{i,i'}$ das Gesamtgewicht der Kanten, die $\mathcal{C}$ verlassen. Die Teilmenge mit dem kleinsten Schnitt ist oft ein einzelner Knoten, weshalb der Ratio-Cut den Schnitt durch die Größen beider Teile dividiert (Hagen and Kahng, 1992; von Luxburg, 2007, Abschn. 5),

$$\operatorname{RatioCut}(\mathcal{C}) := \operatorname{cut}(\mathcal{C}) \left(\frac{1}{|\mathcal{C}|} + \frac{1}{|\mathcal{V} \setminus \mathcal{C}|} \right). \tag{4}$$

Das Minimieren des Ratio-Cuts (4) über alle Teilmengen $\mathcal{C}$ ist NP-schwer (von Luxburg, 2007, Abschn. 5.1). Der Vektor $\mathbf{u}$ mit den Einträgen $u_i = \sqrt{|\mathcal{V} \setminus \mathcal{C}|/|\mathcal{C}|}$ für $i \in \mathcal{C}$ und $u_i = -\sqrt{|\mathcal{C}|/|\mathcal{V} \setminus \mathcal{C}|}$ andernfalls erfüllt $\mathbf{u}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{u} = n \operatorname{RatioCut}(\mathcal{C})$ aufgrund der quadratischen Form (2), sowie $\mathbf{u}^\top \mathbf{1} = 0$ und $\|\mathbf{u}\|_2^2 = n$ (von Luxburg, 2007, Abschn. 5.1). Lässt man die Bedingung fallen, dass $\mathbf{u}$ nur diese beiden Werte annimmt, erhält man das relaxierte Problem

$$\min_{\mathbf{u} \in \mathbb{R}^n} \mathbf{u}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{u} \quad \text{unter den Nebenbedingungen } \mathbf{u}^\top \mathbf{1} = 0 \text{ und } \|\mathbf{u}\|_2^2 = n, \tag{5}$$

und das relaxierte Problem (5) hat nach dem Satz von Rayleigh–Ritz das Minimum $n\lambda_2$, das bei $\sqrt{n}\,\mathbf{v}^{(2)}$ angenommen wird (von Luxburg, 2007, Abschn. 5.1).

Die Anwendung eines Schwellenwerts auf den Fiedler-Vektor überführt diesen Minimierer wieder in eine Teilmenge, und jede Teilmenge hat einen Ratio-Cut von mindestens λ_2. Bei den Tagen in Krems hat die spektrale Bipartition einen Ratio-Cut von 0.429, die Bipartition von k-Means auf den ursprünglichen Merkmalsvektoren einen von 0.485 (sie schneidet 44 Kanten) und eine zufällige Bipartition einen von 12.1, gegenüber $\lambda_2 = 0.043$. Für allgemeines k führt dieselbe Relaxierung des Ratio-Cuts zu den ersten k Eigenvektoren und dem Merkmalsvektor (3) (von Luxburg, 2007, Abschn. 5.2). Ersetzt man $|\mathcal{C}|$ durch die Summe der Knotengrade in $\mathcal{C}$, ergibt sich der normierte Schnitt (Shi and Malik, 2000), dessen Relaxierung auf die Eigenvektoren der normierten Laplace-Matrix $\mathbf{D}^{-1}\mathbf{L}^{(\mathcal{G})}$ führt (von Luxburg, 2007, Abschn. 5.3). Eine Variante verwendet die Eigenvektoren von $\mathbf{D}^{-1/2}\mathbf{L}^{(\mathcal{G})}\mathbf{D}^{-1/2}$ und skaliert jeden Merkmalsvektor (3) auf die Norm eins (Ng et al., 2001).
Das spektrale Clustering ist keine Fixpunktiteration: Die Eigenwertzerlegung wird einmal berechnet, und Iterationen treten nur im k-Means-Schritt auf. Dieser Schritt ist eine empirische Risikominimierung (ERM), bei der der Clusteringfehler in den Merkmalsvektoren (3) statt in den ursprünglichen gemessen wird. Die gelernte Hypothese weist nur jedem Knoten des gegebenen Graphen ein Cluster zu; ein neuer Datenpunkt erfordert einen neuen Knoten, eine neue Laplace-Matrix und eine neue Eigenwertzerlegung. Die Hyperparameter sind die Anzahl k der Cluster und die Konstruktion des Graphen: wie viele Nachbarn der k-Nächste-Nachbarn-Graph verbindet, oder die Bandbreite eines Gaußschen Kernels, der Abstände in Kantengewichte umwandelt (von Luxburg, 2007, Sect. 8.1). Eine Lücke im Spektrum, wie zwischen $\lambda_2 = 0.044$ und $\lambda_3 = 0.179$ in Krems, ist der übliche informelle Richtwert für k (von Luxburg, 2007, Sect. 8.3); mit $k = 3$ teilen sich die Tage in Krems in 102 kalte, 171 milde und 92 warme Tage auf, wobei jedes Cluster im Graphen zusammenhängend ist.

Das spektrale Clustering wird für die Bildsegmentierung verwendet, bei der die Knoten Pixel oder Bildausschnitte sind und die Kantengewichte mit dem Unterschied in Farbe und Position abnehmen (Shi and Malik, 2000), sowie für die Erkennung von Gemeinschaften in sozialen Netzwerken, wo die Kanten gegeben sind und kein Merkmalsvektor benötigt wird (Fortunato, 2010).

Siehe auch: Clustering, Graph-Clustering, Cluster, k-Means, Clusteringfehler, hartes Clustering, Laplace-Matrix, Adjazenzmatrix, Knotengrad, Eigenwertzerlegung, Eigenwert, Eigenvektor, Fiedler-Vektor, algebraische Konnektivität, zusammenhängend, positiv semidefinit, k-Nächste-Nachbarn-Graph, Ähnlichkeitsgraph, gewichteter Graph, Kantengewicht, Bildsegmentierung, Hyperparameter.

Querverweise

Literatur

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

2. Fortunato (2010). Community detection in graphs. Physics Reports, 486(3–5), pp. 75–174.

3. Hagen and Kahng (1992). New Spectral Methods for Ratio Cut Partitioning and Clustering. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 11(9), pp. 1074–1085.

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

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

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

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