

Master-Thesis

Convolutional Neural Networks auf Graphrepräsentationen von Bildern

Matthias Fey

26. Februar 2017

Gutachter:

Prof. Dr. Heinrich Müller

M.Sc. Jan Eric Lenssen

Lehrstuhl Informatik VII
Graphische Systeme
TU Dortmund

Inhaltsverzeichnis

1. Gedachter Inhalt	1
2. Einleitung	2
2.1. Motivation	2
2.2. Aufbau der Arbeit	2
3. Grundlagen	3
3.1. Notationen	3
3.2. Graphentheorie	3
4. Graphrepräsentationen von Bildern	5
4.1. Kantengewichte	5
5. Räumliche Graphentheorie	6
5.1. Patchy-SAN	6
6. Spektrale Graphentheorie	7
6.1. Einführung	7
6.1.1. Eigenwerte, Eigenvektoren und Eigenfunktionen	7
6.1.2. Der Laplacian und seine Eigenwerte	7
6.1.3. Visuelle Interpretation des Laplacian	8
6.1.4. Eigenschaften	9
6.2. Graph-Fourier-Transformation	9
6.3. Spectral Graph Domain	10
6.4. Diskrete Fourier Transformation	11
6.5. Faltung	11
6.5.1. Faltung in CNNs	11
6.5.2. Faltung auf Graphen	12
6.5.3. Offene Fragen	12
6.6. Chebyshev Polynome	13
6.7. Probleme	13
6.8. Pfadlänge	13

6.9. Polynomielle Approximation	13
6.9.1. Tschebyschow-Polynome	13
6.10. Graph Convolutional Networks	14
6.11. Weisfeiler Lehman Analogie	16
6.12. Erweiterung für mehrere Kantenattribute	16
6.12.1. Übertragung auf räumlich eingebettete Graphen	16
6.13. Pooling-Ebene	18
6.13.1. Clustering von Graphen	18
6.13.2. Pooling-Operation	19
6.14. Beispiel	20
A. Weitere Informationen	22
Symbolverzeichnis	23
Abbildungsverzeichnis	25
Algorithmenverzeichnis	27
Literaturverzeichnis	29

1. Gedachter Inhalt

Einleitung: Motivation Aufbau der Arbeit

Grundlagen: Graphen, insbesondere planare Graphen Mathematische Notationen: Vektor, Matrix, Tensor Neuronale Netze (Was ist ein CNN, wie ist der Convolution Operator definiert, nicht lineare Aktivierungsfunktion)

Graphrepräsentationen von Bildern Grid Superpixel Superpixelalgorithmen Merkmalsextraktion (Momente) Merkmalselektion (Cov, PCA)

Lernen auf Graphen: Stand der Forschung: Spatial vs Spectral

Spatial: Patchy Zentralität Canonical Labeling Übertragung auf planare Graphen <- EIGENER ANTEIL (z.B. Grid Spiral) Komplexität Vorteile (einfache Architektur)/Nachteile (keine direkte Nachbarschaftsberücksichtigung möglich, keine Graph Coarsening möglich, Vorverarbeitung ist recht teuer und muss Preprocessed werden weil man das nicht über Matrixoperationen ausdrücken kann)

Spectral: Laplacian, Fourier Transformation GCN und kGCN (weisfeiler Lehman) Übertragung auf planare Graphen (Adjazenzpartitionierung) <- EIGENER ANTEIL Pooling/Coarsening Komplexität Vorteile (z.B. Nachbarschaftsberücksichtigung/keine Ordnung nötig)/Nachteile (rotationsinvariant)

Deep Learning auf variabler Input-Menge (SPP)

Augmentierung von Graphen (ist das überhaupt möglich)

Realisierung (Experimente) und Evaluation Adam-Optimizer Sparse Tensors Vorstellung Datensätze (MNIST, PascalVOC, CIFAR-10, ImageNet) Tensorflow Dropout L2-Regularisierung

Zusammenfassung und Ausblick

2. Einleitung

2.1. Motivation

2.2. Aufbau der Arbeit

3. Grundlagen

3.1. Notationen

$\text{diag}: \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ einen Vektor \mathbf{d} in Diagonalform \mathbf{D} bringt mit $\mathbf{D}_{ii} = \mathbf{d}_i$ und $\mathbf{D}_{ij} = 0$ für $i \neq j$. Die Inverse $\text{diag}^{-1}: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^n$ liefert zu einer beliebigen Diagonalmatrix \mathbf{D} dessen Diagonalvektor \mathbf{d} . brauch ich glaub ich nicht mehr

Wir erlauben, dass wir eine eindimensionale Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ auch elementweise auf Vektoren, Matrizen bzw. Tensoren anwenden dürfen.

Identitätsmatrix \mathbf{I}

awda

3.2. Graphentheorie

Graph $G = (\mathcal{V}, \mathcal{E}, w)$

$\mathcal{V} = \{v_i\}_{i=1}^n$

$|\mathcal{V}| = n < \infty$

Umschreibung in Tensor/Dense Matrix

$\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$

Falls $(u, v) \in \mathcal{E}$, dann sind u und v adjazent und wir schreiben $u \sim v$

Gewichtsfunktion $w: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_+$

ungewichtet: $w: \mathcal{V} \times \mathcal{V} \rightarrow \{0, 1\}$

Falls $(u, v) \notin \mathcal{E}$, dann $w(u, v) = 0$

Im ungewichteten Fall ist Gewichtsfunktion implizit durch \mathcal{E} gegeben

ungerichtet: $u \sim v$ genau dann, wenn $v \sim u$ und

$$w(u, v) = w(v, u) \quad (3.1)$$

Fordern wir für den Verlauf dieser Arbeit (also keine gerichteten Graphen)

Als *Schleife* wird eine Kante bezeichnet, die einen Knoten mit sich selbst verbindet, d.h. $w(v, v) > 0$. Ein Graph ohne Schleifen wird *schleifenloser Graph* genannt. Für den weiteren Verlauf dieser Arbeit fordern wir schleifenlose Graphen.

Adjazenzmatrix $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ eines Graphen G mit $\mathbf{A}_{ij} = w(v_i, v_j)$
 Wir sagen ein Knoten v_i hat Position i in \mathbf{A} . Umschreibung in Sparse Matrix/Tensor

$G = (\mathcal{V}, \mathcal{E}, w)$ ist eindeutig definiert durch \mathbf{A} .

Der *Grad* eines Knotens v ist die Anzahl der Knoten, die adjazent zu ihm sind, d.h.

$$\deg(v) = |\{u: (v, u) \in \mathcal{E}\}| \quad (3.2)$$

Im Falle von gewichteten Graphen wird der Grad eines Knotens von v auch oft über

$$d(v_i) = \sum_{j=1}^n \mathbf{A}_{ij} \quad (3.3)$$

definiert. Die unterschiedliche Notation macht deutlich, wann wir welchen Grad eines Knotens meinen.

Die Gradmatrix $\mathbf{D} \in \mathbb{R}_+^{n \times n}$ eines Graphen G ist definiert als Diagonalmatrix

$$\mathbf{D} = \text{diag}\left([d(v_1), \dots, d(v_n)]^\top\right) \quad (3.4)$$

Umschreibung in Sparse Matrix/Tensor

Ein Graph heißt *k-regulär* falls $\deg(v_i) = k$ für alle $1, \dots, n$.

Ein *ebener Graph* ist eine konkrete Darstellung eines Graphen auf der zweidimensionalen Ebene \mathbb{R}^2 . Jedem Knoten v ist eine Positionsfunktion $p: \mathcal{V} \rightarrow \mathbb{R}^2$ zugeordnet, die die Position eines Knotens auf der Ebene eindeutig definiert.

Ein *Weg* ist eine Folge von Knoten $(v_{x(1)}, v_{x(2)}, \dots, v_{x(k)})$, sodass $v_{x(i)} \sim v_{x(i+1)}$ für alle $1 \leq i < k$ mit Länge k , wobei $x: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ eine Permutation auf der Anzahl der Knoten.

Ein Graph ist *verbunden*, falls er nur eine Komponente hat. Ein Graph ist *verbunden*, falls es von jedem Knoten u einen Weg zu jedem Knoten v gibt. Für den weiteren Verlauf dieser Arbeit fordern wir, dass G verbunden ist.

Ein *Pfad* ist ein Weg, sodass $v_{x(i)} \neq v_{x(i+1)}$. Im Kontext von schleifenlosen Graphen sind die Begriffe Weg und Pfad äquivalent. Wir schreiben $s(u, v)$ einer Funktion $s: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{N}$ für die Länge des kürzesten Pfades von u nach v .

In Graphen mit *Mehrfachkanten*, auch *Multigraphen* genannt, können zwei Knoten durch mehrere Kanten verbunden sein. Multigraphen lassen sich als Tensor über einen Vektor von Adjazenzmatrizen $[\mathbf{A}_1, \dots, \mathbf{A}_m] \in \mathbb{R}_+^{m \times n \times n}$ schreiben. Graphen mit Mehrfachkanten können ebenso als eine Menge von Graphen mit gleicher Knotenmenge betrachtet werden.

nimmt noch
 ganz, ge-
 a auch
 ere Kno-
 und graph
 onenten,
 eren
 te knoten

4. Graphrepräsentationen von Bildern

planarer Graph (MUSS NICHT UNBEDINGT SEIN), gegenbeispiel, ist aber auch egal

4.1. Kantengewichte

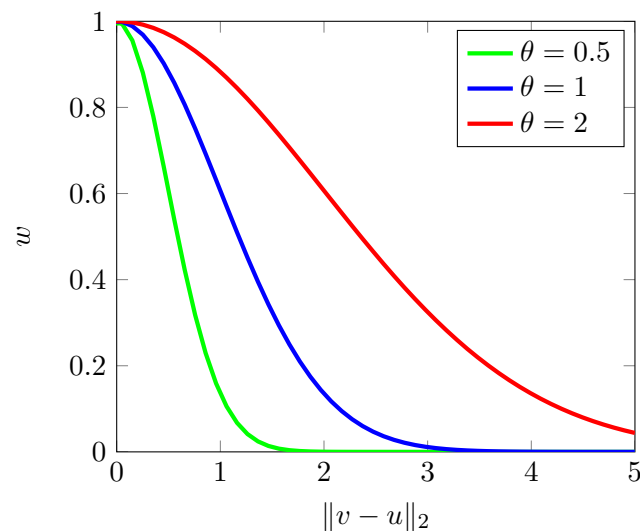
Kantengewichte werden ermittelt aus der euklidischen Distanz der Zentren zweier adjazenter Regionen $\|u, v\|_2$. Distanz entspricht aber nicht der üblichen Bedeutung von Kantengewichten auf Graphen. Je höher das Gewicht, desto ähnlicher bzw. näher sind zwei Knoten.

Ein üblicher Weg die Distanz zweier Knoten zueinander als Kantengewicht darzustellen ist über einen gewichteten *Gaussian-Kernel* [1]

$$w(v, u) = \exp\left(-\frac{\|v - u\|_2}{2\theta^2}\right) \quad (4.1)$$

falls $v \sim u$ und einem Parameter $\theta \in \mathbb{R}$.

Die Wahl von θ ist dabei abhängig von der Ausdehnung der Distanzen eines Graphen.



5. Räumliche Graphentheorie

Isomorphismus, Automorphismus, Canonical Labeling

Labeling / Node Partitions

5.1. Patchy-SAN

6. Spektrale Graphentheorie

6.1. Einführung

- *Spektrum* eines Graphen zur Untersuchung seiner Eigenschaften
- *algebraische* oder *spektrale Graphentheorie* genannt

Algebraische Methoden sind sehr effektiv bei Graphen, die regulär und symmetrisch sind.

6.1.1. Eigenwerte, Eigenvektoren und Eigenfunktionen

$$\mathbf{M}\mathbf{u} = \lambda\mathbf{u}$$

Zu einem Eigenwert λ gibt es unendlich viele (skalierte) Eigenvektoren \mathbf{u} . Wir definieren einen Eigenvektor \mathbf{u} dann eindeutig über $\|\mathbf{u}\|_2 = 1$. Wenn \mathbf{M} symmetrisch ist und \mathbf{u}_1 und \mathbf{u}_2 zwei unterschiedliche Eigenvektoren, dann gilt $\mathbf{u}_1 \perp \mathbf{u}_2$. Jede symmetrische Matrix $\in \mathbb{R}^{n \times n}$ hat genau n Eigenwerte mit $\lambda_1 \leq \dots \leq \lambda_n$

Wir definieren $\mathbf{\Lambda} = \text{diag}([\lambda_1, \dots, \lambda_n])$. Wir definieren die orthogonale Matrix $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$. Dann gilt $\mathbf{M}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$. Daraus folgt sofort

$$\mathbf{M} = \mathbf{M}\mathbf{U}\mathbf{U}^\top = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \quad (6.1)$$

mit $\mathbf{U}\mathbf{U}^\top = \mathbf{I}$.

6.1.2. Der Laplacian und seine Eigenwerte

- diskrete Analogie des ∇^2 Operators
- man nimmt eine Funktion und approximiert sie mit Hilfe eines Graphen, so dass Knoten, die dichter beieinander liegen eine größere zweite Ableitung besitzen.

Der nicht-normalisierte Laplacian \mathbf{L} eines Graphen G ist definiert als $\mathbf{L} = \mathbf{D} - \mathbf{A}$ [3]. Er wird auch oft kombinatorischer Laplacian genannt. Der normalisierte Laplacian $\tilde{\mathbf{L}}$ ist definiert als $\tilde{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$ [3]. Es gilt die Konvention, dass $\left(\mathbf{D}^{-\frac{1}{2}}\right)_{ii} = 0$ falls $\mathbf{D}_{ii} = 0$ (im Falle von isolierten Knoten) Für verbundene Graphen gilt weiterhin $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ [3]. Jeder Eintrag der Diagonalen von $\tilde{\mathbf{L}}$ ist damit 1. $\tilde{\mathbf{L}}$ ist weiterhin symmetrisch, das wäre bei einer Normierung der Form $\mathbf{D}^{-1}\mathbf{L}$ nicht der Fall.

Eigenfunkt
brauch ich
haupt?

wie nennt m
ihn auch?

\mathbf{L} und $\tilde{\mathbf{L}}$ sind keine ähnlichen Matrizen. Insbesondere sind ihre Eigenvektoren unterschiedlich. Die Nutzung von \mathbf{L} oder $\tilde{\mathbf{L}}$ ist damit abhängig von dem Problem, welches man betrachtet. [2].

Wir schreiben \mathcal{L} wenn die Wahl des Laplacian \mathbf{L} oder $\tilde{\mathbf{L}}$ irrelevant ist.

6.1.3. Visuelle Interpretation des Laplacian

$$\nabla^2 f = \nabla \cdot \nabla f$$

Die Divergenz eines Vektorfeldes ist ein Skalarfeld, das an jedem Punkt angibt, wie sehr die Vektoren in einer kleinen Umgebung des Punktes auseinanderstreben.

The Laplace operator measures how much a function differs at a point from the average of the values of the function over small spheres centered at that point. As it turns out, the Laplacian of a graph does something completely analogous: namely, it measures how much a function on a graph differs at a vertex from the average of the values of the function over the neighbors of the vertex.

Im n -dimensionalen euklidischen Raum

$$\nabla^2 f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \quad (6.2)$$

in einer Dimension reduziert sich der Laplace-Operator auf die zweite Ableitung $\nabla^2 f = f''$.

Der *diskrete Laplace-Operator* ist eine Analogie zum diskreten Laplace-Operator, der finite Differenzen $x \pm h$ zur Approximation von $\nabla^2 f$ nutzt. Approximation des Laplace-Operators für finite Elemente

Sei $f: \mathcal{V} \rightarrow \mathbb{R}$ eine Funktion auf den Knoten eines Graphen. f kann ebenso als Vektor $\mathbf{f} \in \mathbb{R}^n$ betrachtet werden mit der Ordnung der Knoten, die die Adjazenzmatrix vorgibt.

Dann gilt für \mathcal{L} , dass

$$(\mathcal{L}\mathbf{f})_i = \sum_{\substack{j=0 \\ j \neq i}}^n -\mathcal{L}_{ij}(\mathbf{f}_i - \mathbf{f}_j) \quad (6.3)$$

Für einen Graphen, der ein reguläres Gitter aufspannt mit gleichen Kantengewichten $\frac{1}{h^2} \in \mathbb{R}$ gilt für einen Knoten an Position (x, y) : Abusing the index notation

$$(\mathbf{L}\mathbf{f})_{x,y} = \frac{4\mathbf{f}_{x,y} - \mathbf{f}_{x+1,y} - \mathbf{f}_{x-1,y} - \mathbf{f}_{x,y+1} - \mathbf{f}_{x,y-1}}{h^2} \quad (6.4)$$

beschreibt die *5-Punkte-Stern* Approximation $-\nabla^2 f$. mit $\nabla^2 f$ definiert auf den fünf Punkten $\{(x, y), (x + h, y), (x - h, y), (x, y + h), (x, y - h)\}$.

$$\mathcal{L}f \approx -\nabla^2 f \quad (6.5)$$

Damit kann der Graph Laplacian als eine Generalisierung des diskreten Laplacian auf einem Gitter verstanden werden.

Eigenwerte und Eigenvektoren werden benutzt, um zu verstehen was passiert, wenn wir einen Operator (hier \mathcal{L}) mehrfach auf einen Vektor \mathbf{x} anwenden (hier Merkmal auf den Knoten).

Wir können \mathbf{x} als Linearkombination der *Eigenbasis* schreiben mit

$$\mathbf{x} = \sum_i c_i \mathbf{u}_i \quad (6.6)$$

und berechnen dann

$$\mathcal{L}^k \mathbf{x} = \sum_i c_i \mathcal{L}^k \mathbf{u}_i = \sum_i c_i \lambda_i \mathcal{L}^{k-1} \mathbf{u}_i = \sum_i c_i \lambda_i^k \mathbf{u}_i \quad (6.7)$$

Wenn wir einen Operator haben, der einen Graphen beschreibt, dann können Eigenschaften dieses Operators und damit des Graphen selber durch dessen Eigenwerte und Eigenvektoren beschrieben werden.

6.1.4. Eigenschaften

Jede Reihen- und Spaltensumme von \mathcal{L} ist 0, d.h. $\sum_i \mathcal{L}_{ij} = 0$ und $\sum_i \mathcal{L}_{ji} = 0$ für alle $i \in \{1, \dots, n\}$.

$\mathcal{L} \in \mathbb{R}^{n \times n}$ hat genau n Eigenwerte $\{\lambda_i\}_{i=1}^n$, wobei die Eigenwerte für gewöhnlich aufsteigend sortiert werden, d.h. $\lambda_i \leq \lambda_{i+1}$.

\mathcal{L} ist eine symmetrische reelle Matrix, d.h. insbesondere liegen ihre Eigenwerte λ_i in \mathbb{R}_+ . Damit ist \mathcal{L} positiv semidefinit, d.h. $\mathbf{x}^\top \mathcal{L} \mathbf{x} \geq 0$ für alle $\mathbf{x} \in \mathbb{R}^n$.

Anzahl der Eigenvektoren gleich Null ist die Anzahl an Komponenten, die ein Graph besitzt. Insbesondere gilt $\lambda_1 = 0$, da $[1, \dots, 1]^\top \in \mathbb{R}^n$ Eigenvektor von \mathcal{L} .
 $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$ wenn Graph verbunden.

Für einen Graphen G definieren wir $\lambda_G := \lambda_2$ und $\lambda_{\max} := \lambda_n$. Für $\tilde{\mathbf{L}}$ gilt $\lambda_{\max} \leq 2$.

Für \mathcal{L}^k mit $k \in \mathbb{N}$ gilt $(\mathcal{L}^k)_{ij} = 0$ genau dann, wenn $s(v_i, v_j) > k$ [2]. Damit beschreibt \mathcal{L}^k bildlich gesprochen die Menge an Knoten, die maximal k Kanten entfernt liegen.

Eine Verschrumpfung eines Graphen G kann beschrieben werden über zwei verschiedene Knoten u und v zu einem neuen Knoten v^* mit

$$w(x, v^*) = w(x, u) + w(x, v) \quad (6.8)$$

$$w(v^*, v^*) = w(u, u) + w(v, v) + 2w(u, v) \quad (6.9)$$

Für einen Graphen G gilt für einen Graphen H , der aus G verkleinert wurde,

$$\lambda_G \leq \lambda_H \quad (6.10)$$

6.2. Graph-Fourier-Transformation

Directly extending this construction to arbitrary weighted graphs is problematic, as it is unclear how to define scaling and translation on an irregular graph. We approach this

quelle

quelle, warum
gilt das

quelle

quelle

problem by working in the spectral graph domain, i.e. the space of eigenfunctions of the graph Laplacian \mathcal{L} . This tool from spectral graph theory, provides an analogue of the Fourier transform for functions on weighted graphs.

Eigenwerte werden als Frequenz aufgefasst, die das Spektrum des Graphen beschreiben. Die Eigenvektoren beschreiben die Signale zu den gegebenen Frequenzen.

Fourier-Transformation beschreibt die gleiche Funktion f , aber in einer völlig anderen Domäne. Nicht in der Vertex-Domäne, sondern in der Spectrum-Domäne, d.h. auf Basis der Eigenwerte.

Ein Signal in der Fourier-Transformierten wird daher beschrieben durch den „Anteil“ oder die Amplituden der Eigenwerte des Graphen.

Fourier Transformation:

$$\hat{\mathbf{f}}_i = \hat{f}(\lambda_i) = \sum_{j=1}^n f(v_j)(\mathbf{u}_i)_j = \sum_{j=1}^n \mathbf{f}_j(\mathbf{u}_i)_j \quad (6.11)$$

oder in Matrixschreibweise

$$\hat{\mathbf{f}} = \mathbf{U}^\top \mathbf{f} \quad (6.12)$$

Inverse Fourier Transformation

$$\mathbf{f}_i = f(v_i) = \sum_{j=1}^n \hat{f}(\lambda_j)(\mathbf{u}_j)_i = \sum_{j=1}^n \hat{\mathbf{f}}_j(\mathbf{u}_j)_i \quad (6.13)$$

oder in Matrisschreibweise

$$\mathbf{f} = \mathbf{U} \hat{\mathbf{f}} \quad (6.14)$$

Fourier-Transformation hat gute Eigenschaften (Faltung ist reine Multiplikation) Mittels der Fouriertransformierten kann man die Faltung zweier Funktionen als Produkt ihrer Fouriertransformierten ausdrücken.

6.3. Spectral Graph Domain

- *Spectral Graph Domain*: Der Raum der Eigenfunktionen von \mathcal{L}
- Analogon (Nachbildung) einer *Fourier-Transformation* von Funktionen auf gewichteten Graphen

Eine beliebige Funktion $f : V \rightarrow \mathbb{R}$ kann als ein Vektor in \mathbb{R}^n gesehen werden. Dies impliziert eine Ordnung auf den Knoten. Wir schreiben $f \in \mathbb{R}^n$ für Funktionen auf den Knoten eines Graphen und $f(m)$ für den Wert des m ten Knoten.

Dann gilt für eine beliebige Funktion $f \in \mathbb{R}^n$

$$\mathcal{L}f(x) = \sum_{x,y} w(x,y) \cdot (f(x) - f(y)) \quad (6.15)$$

wobei die Summe über x, y die Summierung über alle Knoten y beschreibt, die adjazent zu x sind.

Angenommen G ist als ein reguläres Gitter definiert der Breite und Höhe M . Dann hat ein Knoten $v_{x,y}$ genau 4 Nachbarn mit Kantengewicht $\frac{1}{(\delta w)^2}$, bei dem δw die euklidische Distanz zwischen zwei Gitterpunkten beschreibt.

Für eine Funktion $f : M \times M \rightarrow \mathbb{R}$ gilt dann:

$$\mathcal{L}f(x, y) = \frac{4f(x, y) - f(x+1, y) - f(x-1, y) - f(x, y+1) - f(x, y-1)}{(\delta w)^2} \quad (6.16)$$

Damit kann ein Signal f mit der Multiplikation mit \mathcal{L} als eine Weiterpropagation von f unter der Berücksichtigung der lokalen Nachbarn verstanden werden (*5-point Stencil*, d.h. $\mathcal{L}f \approx -\nabla^2 f$).

6.4. Diskrete Fourier Transformation

\mathcal{L} besitzt genau n orthogonal zueinander stehende Eigenvektoren $\{u_l\}_{l=1}^n \in \mathbb{R}^n$. Eigenvektoren u_i sind auf 1 normiert, d.h. $\|u_i\|_2 = 1$. Diese werden auch *Graph Fourier Modes* genannt. Diesen sind Eigenwerte $\{\lambda_l\}_{l=1}^n \in \mathbb{R}$ zugeordnet, die die „Frequenzen“ bzw. das Spektrum des Graphen beschreiben oder visuell betrachtet die Ausdehnung des Raumes, den die Eigenvektoren aufspannen. Bemerke dass $\lambda_0 = 0$, da für den Eigenvektor $\vec{u}_0 = (1, 1, \dots, 1)^T$ gilt, dass $\mathcal{L}\vec{u}_0 = 0$. \mathcal{L} ist diagonalisierbar über $\mathcal{L} = U\Lambda U^T$, wobei $U = [u_1, \dots, u_n] \in \mathbb{R}^{n \times n}$ die *Fourier Basis* und $\Lambda = \text{diag}([\lambda_0, \dots, \lambda_n]) \in \mathbb{R}^{n \times n}$. Die *Fourier Transformation* eines Signals $x \in \mathbb{R}^n$ ist dann definiert als $\hat{x} = U^T x$ und die Inverse als $x = U\hat{x}$.

6.5. Faltung

Wir suchen einen Operator $x *_G g$, der eine Faltung zweier Eingangssignale x, g zu einem Ausgangssignal umleitet. x beschreibt dabei die Knotenattribute und g die Gewichte.

6.5.1. Faltung in CNNs

In der Funktionalanalysis beschreibt die *Faltung* einen mathematischen Operator, der für zwei Funktion f und g eine dritte Funktion $f * g$ liefert. Die Faltung kann als ein Produkt von Funktionen verstanden werden.

Anschaulich ist $(f * g)(x)$ der *gewichtete Mittelwert* von f , wobei die Gewichtung durch g gegeben ist.

Angenommen wir wollen über einer Matrix mit einem *Filter* falten. Sei unsere Eingangsmatrix 3×4 und unsere Filtergröße 2×2 .

Dann gilt zum Beispiel für den Faltungsoperator $*$ in einem Convolutional Neural Network:

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 1 \\ 7 & 8 & 9 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 12 & 16 & 11 \\ 24 & 28 & 17 \end{pmatrix} \quad (6.17)$$

$f : 3 \times 4 \rightarrow \mathbb{R}$ und $g : 2 \text{ times } 2 \rightarrow \mathbb{R}$, dann ist $*$ definiert als

$$(f * g)(x, y) = \sum_{x_i \in [x, x+1] y_i \in [y, y+1]} f(x_i, y_i) g(x - x_i, y - y_i) \quad (6.18)$$

6.5.2. Faltung auf Graphen

Da wir keinen Translationsoperator auf der Domäne der Knoten x beschreiben können, müssen wir unseren Faltungsoperator in der Fourier-Domäne beschreiben. Dafür wandeln wir unsere Knotenmenge x zuerst in \hat{x} um.

Wir definieren $*_G$ in der Fourier-Domäne als

$$x *_G g = U \cdot (U^T \cdot x \odot \hat{g}) \quad (6.19)$$

wobei $\odot(A, B) = (a_{ij} \cdot b_{ij})$ die elementweise Multiplikation bzw. das *Hadamard-Produkt*.

Das Hadamard-Produkt löst sich auf, wenn \hat{g} als eine Diagonalmatrix repräsentiert wird. Dann gilt

$$x *_G g = U \begin{pmatrix} \hat{g}(\lambda_0) & \cdots & 0 \\ 0 & \cdots & \hat{g}(\lambda_n) \end{pmatrix} U^T x = U \hat{g}(\Lambda) U^T x \quad (6.20)$$

Dann beschreibt $\hat{g}(\Lambda) = \text{diag}(\theta)$ eine Gewichtsfunktion mit n Variablen, $\theta \in \mathbb{R}^n$. Damit ist die Faltung bzw. die Gewichtung abhängig von der Input-Größe n , was extrem schlecht ist.

6.5.3. Offene Fragen

- Wie erklärt sich noch einmal der normalisierte Laplacian?
- Warum wird \hat{g} als Diagonalmatrix repräsentiert?
- Wie kommt die Convolution zustande mit dem $*$ Operator?
- Was passiert bei gerichteten Graphen???? Wir haben keinen symmetrischen und insbesondere keinen positiv definiten

6.6. Chebyshev Polynome

6.7. Probleme

Rotationsinvariant

6.8. Pfadlänge

wenn $d_G(m, n) > k$, dann $(L^k)_{m,n} = 0$ (normalisiert sowie unnormalisiert (siehe Wavelet Lemma 5.4))

6.9. Polynomielle Approximation

- bisheriger Ansatz skaliert nicht gut für große Graphen
- schneller Algorithmus zur Approximation des Filters notwendig \Rightarrow Polynome niedriger Ordnung
- Größe des Filters soll unabhängig zu den Daten sein
- approximiere $g(\mathcal{L})$ durch Polynom, dass rekursiv durch \mathcal{L} berechnet werden kann

Beweisidee: $\tilde{\mathbf{L}}^k = (\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top)^k = \mathbf{U}\mathbf{\Lambda}^k\mathbf{U}^\top$ [6].

Das sieht man leicht: $\mathbf{U}(\sum_i \mathbf{\Lambda}^k)\mathbf{U}^\top = \sum_i \mathbf{U}\mathbf{\Lambda}^k\mathbf{U}^\top = \sum_i \mathbf{L}^k$

Diese Polynome formen eine Orthogonalbasis Polynome formen eine Orthogonalbasis für
 $L^2\left([-1, 1], \frac{dx}{\sqrt{1-x^2}}\right)$, auch *Hilbertraum* genannt
Spektrale Filter, die repräsentiert werden durch ein Polynom vom Grad k sind *k-lokalisiert*.

stable recursion property

die Grundidee der Polynomapproximation m hierhin + Übergang zu Chebyshev

6.9.1. Tschebyschow-Polynome

- bisher Filterung eines Signals \mathbf{x} zu $\mathbf{y} = \mathbf{U}g_\theta(\mathbf{\Lambda})\mathbf{U}^\top \mathbf{x} \approx \mathbf{U}g'_{\theta'}(\mathbf{\Lambda})\mathbf{U}^\top \mathbf{x} = g'_{\theta'}(\mathbf{L})\mathbf{x}$
- $g_\theta(\mathbf{\Lambda})$ kann über ein Polynom k ten Grades $g'_{\theta'}(\mathbf{\Lambda})$ approximiert werden, $\theta' \in \mathbb{R}^k$
- Aber: Filterung ist sehr teuer aufgrund der Multiplikation der dichten Matrix \mathbf{U} , d.h. $\mathcal{O}(n^2)$
- Lösung: Parametrisiere $g_\theta(\mathbf{L})$ als eine polynomielle Funktion, die rekursiv aus \mathbf{L} berechnet werden kann.
- Warum sollte das effizienter sein? \mathbf{L} ist nicht dicht besetzt, und hat nur $|\mathcal{E}| + n \ll n^2$ Einträge mit $n \leq |\mathcal{E}|$

Tschebyschow-Polynome (engl. *Chebyshev*) bezeichnen eine Menge von Polynomen $T_n(x): \mathbb{R} \rightarrow \mathbb{R}$ mit dem rekursiven Zusammenhang

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x) \quad (6.21)$$

mit $T_0(x) = 1$ und $T_1(x) = x$. Ein Tschebyschow-Polynom T_n ist ein Polynom n -ten Grads.

Für $x \in [-1, 1]$ gilt $T_k(x) \in [-1, 1]$

Rescale \mathbf{A} zu $\tilde{\mathbf{A}} = \frac{2}{\lambda_{\max}} \mathbf{A} - \mathbf{I} \in [-1, 1]^{n \times n}$. λ_{\max} ist der Wert des größten Eigenvektors von \mathbf{L} .

Dann ist $T_k(\tilde{\mathbf{A}}) \in [-1, 1]^{n \times n}$

$$g_\theta(\mathbf{A}) \approx g'_{\theta'}(\mathbf{A}) = \sum_{i=0}^{k-1} \theta'_i T_i(\tilde{\mathbf{A}}) \quad (6.22)$$

sidee nach-

Es zeigt sich, dass

$$\mathbf{U} g'_{\theta'}(\mathbf{A}) \mathbf{U}^\top = g'_{\theta'}(\mathbf{L}) \quad (6.23)$$

wobei $g'_{\theta'}(\mathbf{L}) = \sum_{i=0}^{k-1} \theta'_i T_i(\tilde{\mathbf{L}})$ mit $\tilde{\mathbf{L}} = \mathbf{U} \tilde{\mathbf{A}} \mathbf{U}^\top = \frac{2}{\lambda_{\max}} \mathbf{L} - \mathbf{I}$

Jetzt lässt sich $y = g'_{\theta'}(\mathbf{L}) \mathbf{x} = \sum_{i=0}^{k-1} \theta'_i T_i(\tilde{\mathbf{L}}) \mathbf{x}$ sehr schnell berechnen:

1. berechne $\bar{\mathbf{x}}_i := T_i(\tilde{\mathbf{L}}) \mathbf{x}$ für alle $i \in \{0, 1, \dots, k-1\}$ mit Hilfe von Rekursion:

a) $\bar{\mathbf{x}}_0 = \mathbf{x}$

b) $\bar{\mathbf{x}}_1 = \tilde{\mathbf{L}} \mathbf{x}$

c) $\bar{\mathbf{x}}_i = 2\tilde{\mathbf{L}}\bar{\mathbf{x}}_{i-1} - \bar{\mathbf{x}}_{i-2}$

2. berechne $\mathbf{y} = [\bar{\mathbf{x}}_0, \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_{k-1}] \theta'$

Laufzeit

- anstatt \mathbf{L}^k zu berechnen mit Komplexität $\mathcal{O}(n^2)$ haben wir nur noch k Multiplikationen mit der Matrix $\tilde{\mathbf{L}}$
- da $\tilde{\mathbf{L}}$ für große Graphen sehr dünnbesetzt ist, d.h. $|\mathcal{E}| \ll n^2$, haben wir bei Verwendung von *dünnbesetzten Matrizen* nur noch eine Laufzeit von $\mathcal{O}(k|\mathcal{E}|)$ [2, 5]
- für den zweiten Schritt gilt $\mathcal{O}(kn)$, mit $n \leq |\mathcal{E}|$ bedingt damit nur Schritt Eins die Laufzeit

6.10. Graph Convolutional Networks

- aus [6]
- Idee: setze $k = 2$ für alle Faltungsebenen

- Begründung: Faltung über $i < k = 2$ berücksichtigt nur alle Knoten, die zum jeweiligen Faltungsknoten adjazent sind und den Knoten selber (Begründung weiter oben)
- für CNNs hat sich gezeigt, dass kleine Filtergrößen wie 3×3 keine negativen Auswirkungen haben
- viele kleine Faltungsebenen ohne Pooling propagieren die Informationen weit entfernter Knoten weiter
- kann das Problem des Overfitting reduzieren für Graphen mit hohem Grad

Annahme: $\lambda_{\max} \approx 2$ Es gilt $0 \leq \lambda_0 \leq \dots \leq \lambda_{n-1} \leq 2$ [3] (aber nur für bipartite Graphen?) Wenn das gilt, dann wird $\lambda_{\max} := 2$ einfach auf die obere Schranke gesetzt Begründung: Netzparameter werden die Veränderung der Skalierung von $\tilde{\mathbf{L}}$ annehmen/ausgleichen. Dann ist $\tilde{\mathbf{L}} = \mathbf{L} - \mathbf{I}$.

Dann gilt für die Filterung eines Signals \mathbf{x} über g_θ

$$g_\theta \star \mathbf{x} = \mathbf{U} g_\theta(\mathbf{\Lambda}) \mathbf{U}^\top \mathbf{x} \approx g'_{\theta'}(\mathbf{L}) \mathbf{x} = \sum_{i=0}^{k-1} \theta'_i T_i(\mathbf{L} - \mathbf{I}) \mathbf{x} = \theta'_0 \mathbf{x} + \theta'_1 (\mathbf{L} - \mathbf{I}) \mathbf{x} \quad (6.24)$$

Wenn wir den normalisierten Laplacian benutzen, gegeben durch $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$, dann lässt sich weiter vereinfachen mit

$$g_\theta \star \mathbf{x} \approx \theta'_0 \mathbf{x} - \theta'_1 \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{x} \quad (6.25)$$

Um die Gefahr des Overfittings und die Anzahl an Berechnungen weiter zu reduzieren, können die Anzahl der Parameter weiter reduziert werden. Mit $\theta = \theta'_0 = -\theta'_1$ gilt dann

$$g_\theta \star \mathbf{x} \approx \theta \left(\mathbf{I} + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \mathbf{x} \quad (6.26)$$

$\mathbf{I} + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ hat nun Eigenwerte im Bereich $[0, 2]$. Wiederholte Anwendungen dieses Operators können daher zu numerischen Instabilitäten und dann zu explodierenden oder verschwindenden Gradienten führen. Um dies zu verhindern, wird folgende Renormalisierung vorgenommen: $\mathbf{I} + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \rightarrow \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$ mit $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ und $\tilde{\mathbf{D}}$ ist nun die Gradmatrix der renormalisierten Adjazenzmatrix $\tilde{\mathbf{A}}$.

$$g_\theta \star \mathbf{x} \approx \theta \left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \right) \mathbf{x} \quad (6.27)$$

$$H^{(l+1)} = f(H^{(l)}, A) \quad (6.28)$$

$$f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)}) \quad (6.29)$$

VGG Paper
oder he et a
deep residu
learning

ist das begr
det, woher
kommt dies
Zahl

warum?

warum

$$D_{ii} = \sum_j A_{ij} \quad (6.30)$$

Für die Potenz $x \in \mathbb{R}$ einer Diagonalmatrix $D \in \mathbb{R}^{N \times N}$ gilt:

$$D^x = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix}^x = \begin{pmatrix} d_{11}^x & 0 & \cdots & 0 \\ 0 & d_{22}^x & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}^x \end{pmatrix} \quad (6.31)$$

6.11. Weisfeiler Lehman Analogie

6.12. Erweiterung für mehrere Kantenattribute

Graph Convolutional Networks berücksichtigen nur eine Adjazenzmatrix. Das bedeutet insbesondere, dass ein Graph nur über ein Kantenattribut verfügen kann. Das ist für ungewichtete Graphen die Markierung einer Kante ($a_{ij} \in \{0, 1\}$) oder für gewichtete Graphen das Gewicht einer Kante ($a_{ij} \in \mathbb{R}_+$). Eine Menge von Kantenattributen kann über mehrere Adjazenzmatrizen definiert werden. Damit ist es ebenfalls möglich unterschiedliche Kanten für unterschiedliche Attribute zu definieren.

Eine Menge von Adjazenzmatrizen $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$ mit $A_i \in \mathbb{R}^{n \times n}$ beschreibt damit eine Menge von m Graphen über der gleichen Knotenmenge \mathcal{V} mit Kardinalität n .

$\mathcal{A} \in \mathbb{R}^{m \times n \times n}$ kann zu einer zweidimensionalen Matrix $A \in \mathbb{R}^{m \cdot n \times n}$ geglättet werden.

Dann ist $A \cdot H^{(l)} \in \mathbb{R}^{m \cdot n \times d}$. Reshape zu $\mathbb{R}^{n \times m \cdot d}$ und Gewichtsmatrix $G \in \mathbb{R}^{m \cdot d \times x}$.

$$H^{(l+1)} = f(H^{(l)}, \tilde{\mathcal{A}}) = \sigma \left(\frac{1}{|\tilde{\mathcal{A}}|} \sum_{\tilde{A}_i \in \tilde{\mathcal{A}}} \tilde{D}_i^{-\frac{1}{2}} \tilde{A}_i \tilde{D}_i^{-\frac{1}{2}} H^{(l)} W_i^{(l)} \right) \quad (6.32)$$

$\sigma(\cdot)$ kennzeichnet eine Aktivierungsfunktion wie zum Beispiel $\text{ReLU}(\cdot) = \max(0, \cdot)$.

6.12.1. Übertragung auf räumlich eingebettete Graphen

Graphknoten haben im Allgemeinen keine Position oder Lage im Raum. Knoten, die Regionen in einer vorhandenen Segmentierung darstellen, haben jedoch offensichtlich eine gewisse Lage im Raum, die zum Beispiel über das Zentrum der Region definiert werden kann. Diese Information ist vorhanden und wichtig und sollte demnach auch nicht verloren gehen. Anstatt diese lokal im Knoten zu speichern, bietet es sich eher an diese Information in den Kanten zu speichern um eine bessere Faltung zu garantieren. Die euklidische Distanz zwischen zwei benachbarten Regionszentren wahrt zwar die Information der Distanz zweier Knoten zueinander, verliert aber die Information der Position zweier Knoten zueinander. Es bietet sich daher an, die horizontalen und vertikalen Abstände in einer Koordinate

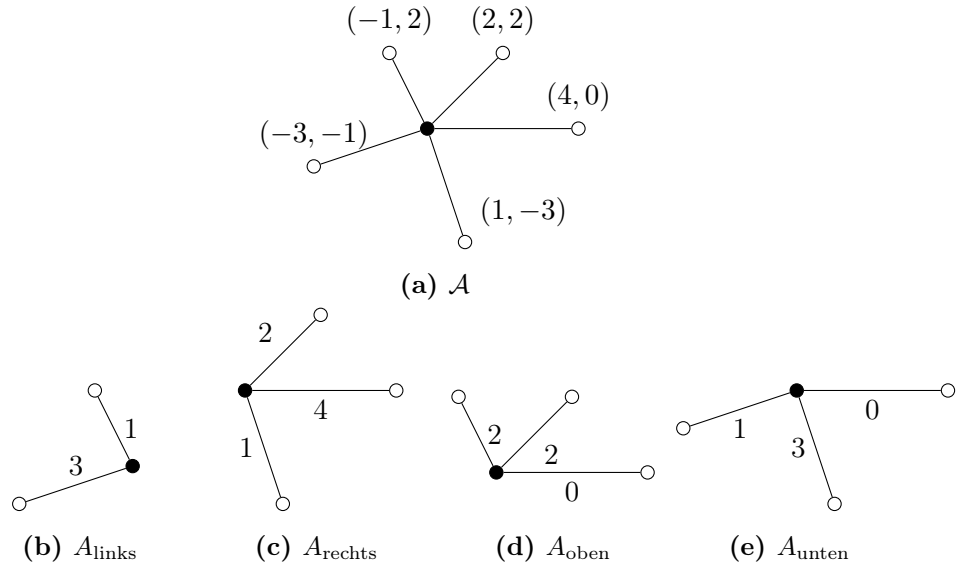


Abbildung 6.1.: Aufteilung einer Adjazenzmatrix in vier räumlich eingebettete Bereiche.

an den Kanten zu speichern. Es ist zu beachten, dass wir dadurch zu einem gerichteten Graphen übergehen, bei dem jede Kante von v nach w auch eine Kante von w nach v besitzt.

Wir haben damit zwei Adjazenzmatrizen. Da Graph Convolutional Networks nicht mit negativen Gewichten funktionieren, müssen wir negative Koordinaten in eine weitere Adjazenzmatrix schreiben. Wir gelangen damit zu vier Adjazenzmatrizen, die die Verbindungen von einem Knoten beschreibt, die links, rechts, oben oder unten zu ihm liegen. Wir definieren diese Adjazenzmatrizen respektive als A_{links} , A_{rechts} , A_{oben} und A_{unten} (vgl. Abbildung 6.1). Falls eine Kante horizontal bzw. vertikal liegt, so definieren wir $a_{ij} = 1$ respektive für beide „gegenüberliegenden“ Adjazenzmatrizen.

Kantenattribute bzw. Positionen von Knoten sollten skalierungsinvariant gespeichert werden. Dafür werden die Abstände auf den Einheitskreis gemappt, wobei der Knoten mit der längsten Distanz zum Wurzelknoten genau auf dem Einheitskreis liegt (vgl. Abbildung 6.2).

Für die Anwendung auf das Graph Convolutional Network müssen die Gewichte aller Adjazenzmatrizen $a_{xij} \in [0, 1]$ invertiert werden, damit nähere Knoten einen größeren Einfluss haben. Ebenso müssen *Self Loops* für alle Knoten hinzugefügt werden. Wir definieren unsere Adjazenzmatrix $\tilde{A} \in \mathbb{R}^{N \times N}$ aus einer Adjazenzmatrix $A \in \mathbb{R}^{N \times N}$ dann über

$$\tilde{A}_{ij} = \begin{cases} 1, & \text{falls } i = j, \\ (a_{ij} + 1)^{-1}, & \text{falls } a_{ij} \neq 0, \\ 0, & \text{sonst.} \end{cases} \quad (6.33)$$

Dann ist $\tilde{a}_{ij} \in [1, 0.5]$

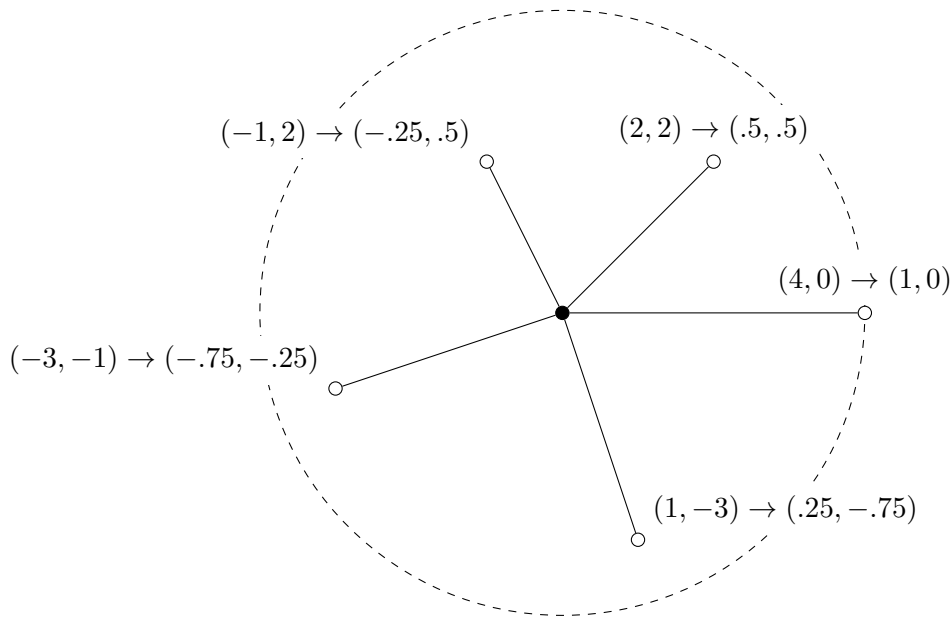


Abbildung 6.2.: Abbildung der lokalen Nachbarschaftsknoten auf den Einheitskreis.

Diagonalmatrix ist schwierig. Man will ja die Normalisierung damit $H^{(l)}$ nicht überskalieren. Ich würde auch die gewichtete Matrix normalisieren. Denke das macht Sinn. Dann fallen die Werte ab, wenn viele Knoten weit entfernt sind.

6.13. Pooling-Ebene

6.13.1. Clustering von Graphen

Pooling-Ebenen des Netzes sollen über das Clustering bzw. die logische Zusammenfassung von Knoten realisiert werden.

Anforderungen:

- mehrstufiges Clustering von Graphen für mehrere Pooling-Ebenen
- Reduzierung der Knotenanzahl soll den Blick auf einen Graphen bei unterschiedlichen Auflösungen zeigen
- Cluster-Algorithmen, die die Größe eines Graphen um den Faktor zwei für jede Anwendung reduzieren erlauben eine feine Kontrolle über die zu benutzenden Pooling-Größen.
- effiziente Approximation, da Graph-Clustering NP-schwer (vgl. 5)

Es existieren einige Cluster-Techniken auf Graphen wie das populäre *spektrale Clustering* [?].

h ich evtl
cht

Dieser erfüllt aber nicht die Voraussetzungen (warum nicht?). Stimmt doch garnicht!!

Defferrard et al. [5] benutzen für die Pooling-Ebene eines Netzes auf Graphen die *Vergrößerungsphase* des mehrstufigen Cluster-Algorithmus *Graculus* [4]. Dabei wird der initiale Graph G_0 sukzessive in kleinere Graphen G_1, G_2, \dots, G_m mit $|\mathcal{V}_0| > |\mathcal{V}_1| > \dots > |\mathcal{V}_m|$ transformiert. Für die Transformation von einem Graphen G_i zu einem Graphen G_{i+1} mit kleinerer Knotenanzahl $|\mathcal{V}_{i+1}| < |\mathcal{V}_i|$ werden aus disjunkten Knotenuntermengen von \mathcal{V}_i *Superknoten* für \mathcal{V}_{i+1} gebildet.

Die Auswahl der Untermengen erfolgt gierig. Die Knoten des Graphen werden als unmarkiert initialisiert und zufällig durchlaufen. Für jeden Knoten $v \in \mathcal{V}_i$, der noch unmarkiert ist, wird ein lokaler, ebenfalls noch unmarkierter, Nachbarschaftsknoten $u \in \mathcal{N}(v)$ nach einer zuvor definierten Strategie bestimmt und v sowie w zu einem Superknoten $v^* := \{v, w\} \in \mathcal{V}_{i+1}$ verschmelzt. Anschließend werden v und w markiert. Falls v keinen unmarkierten Nachbarn besitzt, wird v allein als *Singleton-Superknoten* $v^* := \{v\} \in \mathcal{V}_{i+1}$ deklariert und markiert [4].

Strategien für die Nachbarschaftsauswahl basieren üblicherweise auf der Maximierung von w_{uv} oder $w_{uv} \left(\frac{1}{d_u} + \frac{1}{d_v} \right)$ (*Normalized Cut*) .

erklären

Graculus reduziert die Knotenanzahl eines beliebigen Graphen näherungsweise um die Hälfte, d.h. $2 \cdot |\mathcal{V}_{i+1}| \approx |\mathcal{V}_i|$. Ausnahmen sind zum Beispiel Graphen $G = (\mathcal{V}, \mathcal{E})$ mit $\mathcal{E} = \emptyset$. In der Praxis zeigt sich jedoch, dass Graculus nur sehr wenige Singleton-Knoten generiert [5].

Nach der spektralen Graphentheorie [3] gilt für Kanten eines Graphen $G = (\mathcal{V}, \mathcal{E})$ nach Verschmelzung von u und v zu v^*

$$w_{xv^*} = w_{xu} + w_{xv} \quad (6.34)$$

$$w_{v^*v^*} = w_{uu} + w_{vv} + 2w_{uv} \quad (6.35)$$

für einen Knoten $x \in \mathcal{V}$, $x \neq v^*$. Insbesondere gilt für einen Graphen H , der auf diese Weise konstruiert wurde, $\lambda_G \leq \lambda_H$, wobei λ_G , λ_H jeweils die ersten Eigenvektoren λ_1 von G respektive H [3].

Übertragung auf planare Graphen

- Mittelwert der Positionen wird gebildet (auch gewichtet über d_i ?)
- $\mathcal{E}_{v^*} = \mathcal{E}_u \cup \mathcal{E}_v$

nur die Kar
Gewichte w
neu berech

6.13.2. Pooling-Operation

Anhand eines kleineren, vergrößerten Graphen G_{i+1} und der eindeutigen Zuweisung von Knoten $u, v \in \mathcal{V}_i$ zu $v^* \in \mathcal{V}_{i+1}$ können nun die Pooling-Operation der Knotenattribute von \mathcal{V}_i zu \mathcal{V}_{i+1} definiert werden:

- **Max-Pooling:** $v^* := \max(u, v)$
- **L2-Pooling:** $v^* := \|u, v\|_2$

6.14. Beispiel

Wir betrachten eine einfache 3×3 Adjazenzmatrix, d.h. $|\mathcal{V}| = n = 3$.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (6.36)$$

mit Diagonalmatrix $D = \text{diag}(1, 2, 1)$.

Der Laplacian $\mathcal{L} = D - A$ ist dann

$$\mathcal{L} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \quad (6.37)$$

Nun müssen die Eigenvektoren der Matrix und dessen Eigenwerte bestimmt werden, d.h. wir müssen das folgende Eigenwertproblem lösen

$$\mathcal{L} \cdot \vec{u} = \lambda \cdot \vec{u} \quad (6.38)$$

Wir erhalten 3 Eigenvektoren und Eigenwerte mit

$$\lambda_0 = 0, \vec{u}_0 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.58 \\ 0.58 \\ 0.58 \end{pmatrix}, \lambda_1 = 1, \vec{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.71 \\ 0 \\ 0.71 \end{pmatrix}, \lambda_2 = 3, \vec{u}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.41 \\ -0.82 \\ 0.41 \end{pmatrix} \quad (6.39)$$

Dann sind U , Λ und U^T definiert als

$$U \approx \begin{pmatrix} 0.58 & -0.71 & 0.41 \\ 0.58 & 0 & -0.82 \\ 0.58 & 0.71 & 0.41 \end{pmatrix}, \Lambda = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}, U^T \approx \begin{pmatrix} 0.58 & 0.58 & 0.58 \\ -0.71 & 0 & 0.71 \\ 0.41 & -0.82 & 0.41 \end{pmatrix} \quad (6.40)$$

Angenommen wir haben ein Signal $x = (100, 10, 1)^T$, dann ist der Wert dieses Signals transformiert in die Fourier Domäne definiert als $\hat{x} \approx (64.09, -70.00, 33.07)^T$. Führen wir \hat{x} auf x mittels $U \cdot \hat{x}$ zurück, erhalten wir korrekterweise $x = (100, 10, 1)^T$.

Es gilt $\lambda_{\max} = 3$ Jetzt ist $\tilde{\Lambda}$ definiert als

$$\tilde{\Lambda} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6.41)$$

Wir überprüfen die Approximation durch die Polynome mit $k = 2$:

$$g_{\theta}(\mathbf{\Lambda}) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, g_{\theta prime} = (wd) \quad (6.42)$$

A. Weitere Informationen

Symbolverzeichnis

\deg Gradfunktion der Knoten eines Graphen G mit $\deg: \mathcal{V} \rightarrow \mathbb{N}$. 4, 23

λ Eigenwert. 7, 9–12

\mathbb{N} Menge der natürlichen Zahlen. 4, 9, 23

\mathbb{R}_+ Menge der positiven reellen Zahlen inklusive Null. 3, 4, 9, 16, 23

\mathbb{R} Menge der reellen Zahlen. 4, 5, 7–9, 14, 16, 23

\mathcal{E} Kantenmenge eines Graphen G mit $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. 3, 4, 23

\mathcal{V} Knotenmenge $\{v_i\}_{i=1}^n$ eines Graphen G . 3, 4, 8, 23

\perp Orthogonalität. 7

\sim Adjazenzrelation zweier Knoten eines Graphen G mit $u \sim v$ genau dann, wenn u und v adjazent. 3–5, 23

diag Diagonalfunktion. 4, 7

d gewichtete Gradfunktion der Knoten eines Graphen G mit $d: \mathcal{V} \rightarrow \mathbb{R}_+$. 4, 23

p Positionsfunktion auf den Knoten \mathcal{V} mit $p: \mathcal{V} \rightarrow \mathbb{R}^2$. 4, 23

s kürzeste Distanzfunktion mit $s: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{N}$. 4, 9, 23

w Gewichtsfunktion der Kanten eines Graphen G mit $w: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_+$. 3–5, 9, 23

A Adjazenzmatrix eines Graphen G . 4, 7

D gewichtete Gradmatrix. 4, 7

I Identitätsmatrix. 3, 7

L Laplacian, unnormalisiert. 7, 8

U Eigenvektormatrix. 7, 10

Λ Diagonalmatrix der Eigenwerte des Laplacian. 7

\mathcal{L} Laplacian, normalisiert oder unnormalisiert. 8–10

$\tilde{\mathbf{L}}$ Laplacian, normalisiert. 7–9

G Graph. 3, 4, 7, 9, 23

\mathbf{u} Eigenvektor mit $\|\mathbf{u}\|_2 = 1$. 7, 9, 10, 24

Abbildungsverzeichnis

6.1. Aufteilung einer Adjazenzmatrix in vier räumlich eingebettete Bereiche. . .	17
6.2. Abbildung der lokalen Nachbarschaftsknoten auf den Einheitskreis.	18

List of Algorithms

Literaturverzeichnis

- [1] D. I. SHUMAN AND S. K. NARANG AND P. FROSSARD AND A. ORTEGA AND P. VANDERGHEYNST: *Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Data Domains*. CoRR, 2012.
- [2] D. K. HAMMOND AND P. VANDERGHEYNST AND R. GRIBONVAL: *Wavelets on graphs via spectral graph theory*. Applied and Computational Harmonic Analysis, Seiten 129–150, 2011.
- [3] F. R. K. CHUNG: *Spectral Graph Theory*. American Mathematical Society, 1997.
- [4] I. S. DHILLON AND Y. GUAN AND B. KULIS: *Weighted Graph Cuts Without Eigenvectors: A Multilevel Approach*. IEEE, Seiten 1944–1957, 2007.
- [5] M. DEFFERRARD AND X. BRESSON AND P. VANDERGHEYNST: *Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering*. CoRR, 2016.
- [6] T. N. KIPF AND M. WELLING: *Semi-Supervised Classification with Graph Convolutional Networks*. CoRR, 2016.

Eidesstattliche Versicherung

Name, Vorname

Matr.-Nr.

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit/Masterarbeit* mit dem Titel

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ort, Datum

Unterschrift

*Nichtzutreffendes bitte streichen

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -)

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird gfls. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Ort, Datum

Unterschrift