

Master-Thesis

Convolutional Neural Networks auf Graphrepräsentationen von Bildern

Matthias Fey

17. Februar 2017

Gutachter:

Prof. Dr. Heinrich Müller

M.Sc. Jan Eric Lenssen

Lehrstuhl Informatik VII
Graphische Systeme
TU Dortmund

Inhaltsverzeichnis

1. Einleitung	1
2. Spektrale Graphentheorie	3
3. Graph Convolutional Networks	5
3.1. Erweiterung für mehrere Kantenattribute	5
3.1.1. Übertragung auf räumlich eingebettete Graphen	6
A. Weitere Informationen	9
Symbolverzeichnis	11
Abbildungsverzeichnis	13
Algorithmenverzeichnis	15
Literaturverzeichnis	17

1. Einleitung

\mathbb{R} und \mathbb{N} sind mathematische Symbole [1].

2. Spektrale Graphentheorie

- *Spektrum* eines Graphen zur Untersuchung seiner Eigenschaften
- *algebraische* oder *spektrale Graphentheorie* genannt
- als Spektrum eines Graphen bezeichnet man die (nach Größe geordnete) Folge der Eigenwerte λ seiner Adjazenzmatrix, d.h. $A \cdot x = \lambda x$ (x Eigenvektoren)

Algebraische Methoden sind sehr effektiv bei Graphen, die regulär und symmetrisch sind. Als *Schleife* wird in der Graphentheorie eine Kante bezeichnet, die einen Knoten mit sich selbst verbindet. Ein Graph ohne Schleifen wird *schleifenloser* Graph genannt.

Sei d_v der Grad eines Knotens v eines Graphen G . Der *Laplacian* \mathcal{L} eines Graphen ohne Schleifen und Mehrfachkanten ist definiert als

$$\mathcal{L}(u, v) = \begin{cases} d_v, & \text{wenn } u = v, \\ -1, & \text{wenn } u \text{ und } v \text{ adjazent,} \\ 0, & \text{sonst.} \end{cases} \quad (2.1)$$

Damit ist $\mathcal{L} = I - A$. \mathcal{L} kann normalisiert werden über $\mathcal{L}_{\text{norm}} = T^{-\frac{1}{2}} \mathcal{L} T^{-\frac{1}{2}}$, wobei T die Diagonalmatrix beschreibt mit $T(v, v) = d_v$. Für einen *isolierten* Knoten v , d.h. $d_v = 0$, gilt die Konvention $T^{-1}(v, v) = 0$. Ebenso lässt sich $\mathcal{L}_{\text{norm}}$ definieren als

$$\mathcal{L}_{\text{norm}}(u, v) = \begin{cases} 1, & \text{wenn } u = v \text{ und } d_v \neq 0, \\ -\frac{1}{\sqrt{d_u d_v}}, & \text{wenn } u \text{ und } v \text{ adjazent,} \\ 0, & \text{sonst.} \end{cases} \quad (2.2)$$

Wenn G k -regulär ist, d.h. $T = \text{diag}(k)$, dann gilt $\mathcal{L}_{\text{norm}} = I - \frac{1}{k} A$.

Da \mathcal{L} symmetrisch ist, sind seine Eigenwerte alle reell und positiv.

Einem gewichtetem ungerichtetem Graph G kann eine Gewichtsfunktion $w : V \times V \rightarrow \mathbb{R}$ zugeschrieben werden, sodass $w(u, v) = w(v, u)$ und $w(u, v) \geq 0$. Falls $\{u, v\} \notin \mathcal{E}$, dann $w(u, v) = 0$. Damit sind ungewichtete Graphen nur ein Spezialfall bei dem alle Gewichte 0 oder 1 sind. Der Grad d_v eines Knoten v ist dann definiert als

$$d_v = \sum_u w(u, v). \quad (2.3)$$

2. Spektrale Graphentheorie

Dann gilt

$$\mathcal{L} = \begin{cases} 1 - \frac{w(v,v)}{d_v}, & \text{wenn } u = v \text{ und } d_v \neq 0, \\ -\frac{w(u,v)}{\sqrt{d_u d_v}}, & \text{wenn } u \text{ und } v \text{ adjazent,} \\ 0, & \text{sonst.} \end{cases} \quad (2.4)$$

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

$$w(x, v^*) = w(x, u) + w(x, v) \quad (2.5)$$

und

$$w(v^*, v^*) = w(u, u) + w(v, v) + 2w(u, v) \quad (2.6)$$

Mit $\lambda_G := \lambda_1$ für einen Graphen G , gilt für einen Graphen H der aus G verkleinert wurde

$$\lambda_G \leq \lambda_H \quad (2.7)$$

3. Graph Convolutional Networks

$$H^{(l+1)} = f(H^{(l)}, A) \quad (3.1)$$

$$f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)}) \quad (3.2)$$

$$D_{ii} = \sum_j A_{ij} \quad (3.3)$$

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

$$D^x = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix}^x = \begin{pmatrix} d_{11}^x & 0 & \cdots & 0 \\ 0 & d_{22}^x & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}^x \end{pmatrix} \quad (3.4)$$

3.1. 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 $G \in \mathbb{R}^{m \cdot d \times x}$.

$$H^{(l+1)} = f(H^{(l)}, \tilde{\mathcal{A}}) = \sigma \left(\frac{1}{|\tilde{\mathcal{A}}|} \sum_{\tilde{A}_i \in \tilde{\mathcal{A}}} \tilde{D}_i^{-\frac{1}{2}} \tilde{A}_i \tilde{D}_i^{-\frac{1}{2}} H^{(l)} W_i^{(l)} \right) \quad (3.5)$$

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

3. Graph Convolutional Networks

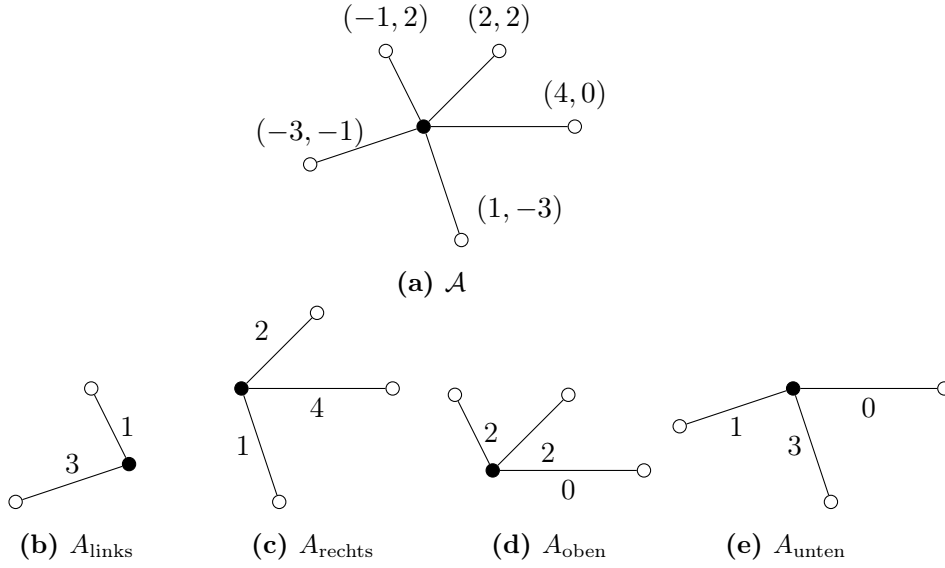


Abbildung 3.1.: Aufteilung einer Adjazenzmatrix in vier räumlich eingebettete Bereiche.

3.1.1. 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 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 3.1). 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

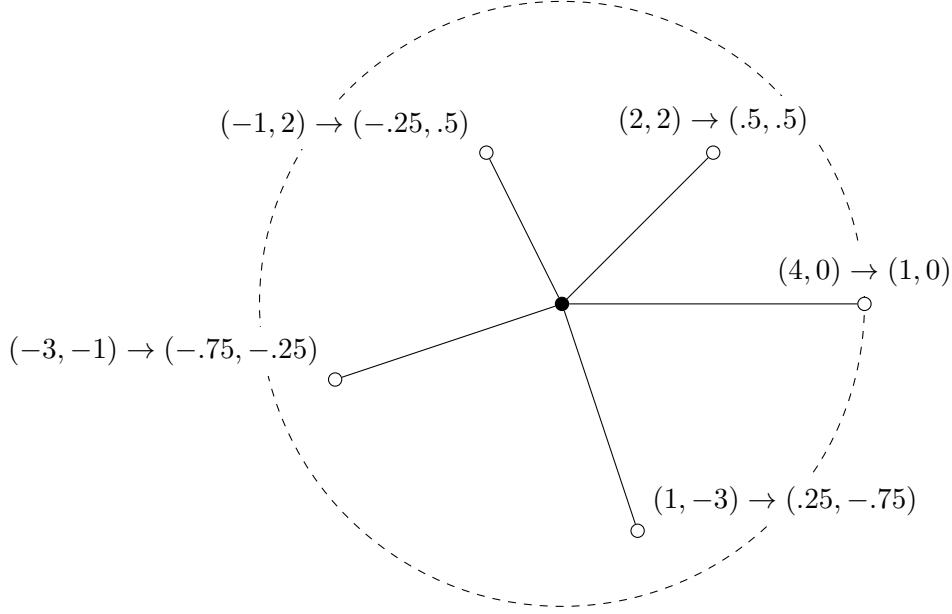


Abbildung 3.2.: Abbildung der lokalen Nachbarschaftsknoten auf den Einheitskreis.

der längsten Distanz zum Wurzelknoten genau auf dem Einheitskreis liegt (vgl. Abbildung 3.2).

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

$$\tilde{A}_{ij} = \begin{cases} 1, & \text{falls } i = j, \\ (a_{ij} + 1)^{-1}, & \text{falls } a_{ij} \neq 0, \\ 0, & \text{sonst.} \end{cases} \quad (3.6)$$

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

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

A. Weitere Informationen

Symbolverzeichnis

\mathbb{N} Menge der natürlichen Zahlen. 1

\mathbb{R}^+ Menge der positiven reellen Zahlen inklusive Null. 5

\mathbb{R} Menge der reellen Zahlen. 1, 5

Abbildungsverzeichnis

3.1. Aufteilung einer Adjazenzmatrix in vier räumlich eingebettete Bereiche. . .	6
3.2. Abbildung der lokalen Nachbarschaftsknoten auf den Einheitskreis.	7

Algorithmenverzeichnis

Literaturverzeichnis

- [1] NIELSEN, M. A.: *Neural Networks and Deep Learning*. Determination Press, 2015.

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