Vergleich von Distanzen und Kernel für Klassifikatoren zur Optimierung der Annotation von Bildern

Alexander Askinadze

Heinrich-Heine-Universität Düsseldorf
Universitätsstrasse 1
D-40225 Düsseldorf
alexander.askinadze@uni-duesseldorf.de

Abstract: Die stetig steigende Anzahl von Bildern erfordert Verfahren zur maschinellen Annotation. Um automatisch semantische Informationen aus den Bildern zu extrahieren, repräsentieren wir die Bilder durch numerische Vektoren, sogenannte BoW-Histogramme und klassifizieren diese auf vorgegebene Klassen. Als Klassifikatoren werden Nearest-Centroid (NC) und Support Vector Machine (SVM) eingesetzt. Auf der Caltech 101 Bilder-Datenbank liefert der SVM-Klassifikator mit dem empfohlenen RBF-Kernel bessere Ergebnisse als der NC-Klassifikator mit der Euklidischen Distanz. Wir vergleichen verschiedene Distanzfunktionen wie z.B. die Bhattacharyya- und Hellinger-Distanz und zeigen, wie sich die Mahalanobis-Distanz für eine Modifikation des NC-Klassifikators nutzen lässt. Nach einer Evaluation folgern wir, dass der NC-Klassifikator mit anderen Distanzfunktionen die SVM-Ergebnisse erreichen kann und eine Normierung der BoW-Histogramme sich ebenfalls positiv auswirkt. Außerdem zeigen wir, dass sich die Ergebnisse des SVM-Klassifikators signifikant durch den Einsatz des Chi-Quadrat- und Histogrammschnitt-Kernels verbessern können.

1 Einleitung

Die stetig wachsende Anzahl von Bildern erfordert einen erheblichen Arbeitsaufwand durch manuelle Annotation. Die Annotation ist notwendig, um Bilder innerhalb der großen Bestände wiederfinden zu können. Maschinell ausgeführte Annotationsverfahren werden dadurch immer wichtiger. Diese Verfahren müssen in der Lage sein, den semantischen Inhalt der Bilder zu erkennen und dadurch passende Metainformationen zu extrahieren.

Ein typischer Ansatz für diese Aufgabe ist die Umwandlung eines Bildes in einen numerischen Vektor und die anschließende Klassifikation dieses Vektors auf n vorgegebene Klassen, die sich als Annotationen des Bildes verwenden lassen. Die Qualität der Annotation eines Bildes hängt also von der Qualität der numerischen Repräsentation des Bildes und des Klassifikators ab.

Nach der Wahl einer festen numerischen Repräsentation des Bildes soll untersucht werden, welche Faktoren und Parametereinstellungen beliebter Klassifikatoren sich positiv auf die Klassifikationsergebnisse auswirken können. Hierzu werden die Klassifikatoren mit üblichen oder empfohlenen Einstellungen getestet und anschließend wird gezeigt, wie

sich die Wahl anderer Einstellungen auswirkt.

2 Grundlagen

Ein beliebter Ansatz zur Transformation eines Bildes in einen numerischen Vektor ist das Prinzip “bag of keypoints” [CDF+04], dass auf das “Bag of Words (BoW)”-Modell zurückzuführen ist. Die Bilder werden hierbei in eine einheitliche Darstellung von gleichdimensionalen Histogrammen (BoW-Histogrammen) gebracht.

Ablauf der Tansformation Mit einem Detektor werden markante Punkte im Bild gefunden und durch Deskriptoren beschrieben. Zur Detektion und Extraktion der Bildmerkmale wird das SIFT (Scalable Invariant Feature Transform)-Verfahren [Low04] verwendet, dass nach [vdSGS08] vergleichsweise gute deskriptive Eigenschaften aufweist. Das SIFT-Verfahren beschreibt die gefundenen markanten Punkte durch 128-dimensionale Deskriptoren. Die Anzahl der SIFT-Deskriptoren aller Trainingsbilder wird durch Clustering (K-Means) auf K, durch das Clustering entstandene Zentroide reduziert. Basierend auf diesem Clusteringergebnis der SIFT-Deskriptoren aller Trainingsbilder können die K-dimensionalen BoW-Histogramme für die Bilder erstellt werden. Zu beachten ist, dass die Dimension K der BoW-Histogramme sich direkt aus der Vorauswahl des Parameters K im K-Means Algorithmus ergibt.

Klassifikation Sobald ein Bild durch ein BoW-Histogramm repräsentiert ist, kann dieses als Eingabe an einen Klassifikator weitergegeben werden. Als Klassifikatoren lassen sich beispielsweise die beliebten Klassifikatoren Nearest-Centroid (NC) auch als ”Rocchio classification” [MRS08] bekannt, Abwandlungen von NC: ”Center-based nearest neighbor classifier”(CNN) [GW07] oder Support Vector Machine (SVM) [CV95] einsetzen. Der Klassifikationsvorgang für ein neues Bild läuft wie in Abbildung 1 ab.

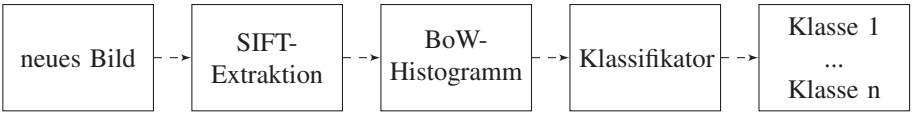

Abbildung 1: Klassifikationsablauf für ein neues Bild

Baseline Im weiteren Verlauf des Papers soll untersucht werden, welche Parametereinstellungen in einzelnen Schritten des Klassifikationsvorgangs positive Auswirkungen auf das Klassifikationsergebnis haben. Hierzu wird die Bild-Datenbank Caltech 101 [FFFP04] mit über 9000 Bildern aus 102 Klassen verwendet. Diese Bilddatenbank eignet sich zur Evaluation, da sie aus vielen semantisch unterschiedlichen Klassen (Flugzeuge, Insekten, Pflanzen, ...) besteht und sich somit nicht auf einen speziellen semantischen Bereich

spezialisiert. Als Baseline sollen die Klassifikatoren NC und SVM mit üblichen Parametereinstellungen dienen, deren Ergebnisse in Tabelle 1 abgebildet sind. Pro Klasse werden hierbei je 15 Trainings- und Testbilder verwendet. Jeder Versuch wird 3 mal wiederholt. Für den NC-Klassifikator wird wie in [MRS08] die Euklidische Distanz verwendet. Für die SVM werden die in [wHcCjL10] empfohlenen Einstellungen zur Skalierung der Vektoren und zur Wahl des SVM-Kernels (RBF-Kernel) verwendet. Für die Multiklassen-SVM-Strategie wird die in [CL11] empfohlene "one-against-one" - Strategie verwendet.

	100		500		1000		1500	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re
NC (Euklidische Dist.)	0,25	0,25	0,31	0,29	0,36	0,33	0,38	0,32
SVM (RBF-Kernel)	0,3	0,3	0,38	0,35	0,39	0,35	0,41	0,36

Tabelle 1: Precision- und Recallergebnisse mit Standardeinstellungen für die BoW-Histogramm-Größen 100,500,1000 und 1500.

2.1 Distanzfunktionen

Es folgt eine Übersicht über einige Klassen von Distanzfunktionen.

2.1.1 Durch Normen induzierte Distanzfunktionen

Eine Norm ist eine Abbildung $||.|| : V \rightarrow \mathbb{R}_+, x \rightarrow ||x||$ von einem Vektorraum über einem Körper (in unserem Fall $\mathbb{R}$) in die Menge $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$. Eine ganze Reihe von Normen, die p-Normen lassen sich durch $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ mit $x \in \mathbb{R}^n$ und $p \in \mathbb{R}, 1 \leq p \leq \infty$ angeben.

Bekannte p-Normen sind die 1-Norm (induziert die Manhattan Distanz), 2-Norm (induziert die Euklidische Distanz) und für $p \rightarrow \infty$ die Maximum Distanz. Mit Hilfe der folgenden Gleichung lassen sich aus beliebigen p-Normen Distanzen ableiten: $d(x, y) = ||x-y||$.

2.1.2 Durch Matrizen induzierte Distanzfunktionen

Eine weitere Klasse von Distanzfunktionen lässt sich durch eine Matrix A angeben:

$$d_A(x, y) = ||x - y||_A = \sqrt{(x - y)^T A(x - y)} \quad (1)$$

Ist A die Einheitsmatrix, so entspricht d_A der Euklidischen Distanz. Interessant ist also, wie sich von der Einheitsmatrix abweichende Matrizen auf das Klassifikationsergebnis auswirken und wie diese Matrizen möglichst passend zu den Daten bestimmt werden können.

Mahalanobis-Distanz Eine Möglichkeit solche Matrizen zu bestimmen, bietet sich aus der Berechnung der Kovarianzmatrix S und der Nutzung ihrer inversen Matrix S^{-1}. Die Distanzfunktion $d_{S^{-1}}$ heißt Mahalanobis-Distanz [MT09].

Es ist zunächst unklar, wie eine Kovarianzmatrix für eine Trainingsmenge mit n p-dimensionalen BoW-Histogrammen zu berechnen ist. Wir benutzen einen statistischen Ansatz zur Schätzung einer Kovarianzmatrix aus Daten [Hel09] und wenden diesen Ansatz folgendermaßen auf die vorliegenden Trainings-BoW-Histogramme an: Sei $X = (X_1, ..., X_p)$ ein multivariater Zufallsvektor der Zufallsvariablen $X_i, 1 \leq i \leq p$. Wir nehmen an, dass die BoW-Histogramme durch X realisiert wurden. Die Kovarianz von X lässt sich berechnen mit:

$$Cov(X) = \begin{pmatrix} Cov(X_1, X_2) & Cov(X_1, X_2) & ... & Cov(X_1, X_p) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & ... & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & Cov(X_n, X_2) & ... & Cov(X_p, X_p) \end{pmatrix} \tag{2}$$

Die multivariaten Verteilungsparameter der X_i liegen nicht vor und müssen deshalb geschätzt werden. Durch die vorliegenden Trainings-BoW-Histogramme liegen die $p-dimensionalen$ Realisationen von X vor, aus denen geschätzt werden kann. Sei $B = (b_{ij})$ eine $n \times p$ Matrix mit den BoW-Trainings-Histogrammen in den Zeilen. Seien $\overline{b_j}$ die Mittelwerte der Spalten. Dann gewinnen wir die zentrierte Matrix $B^* = (b^*_{ij})$ durch $b^*_{ij} = b_{ij} - \overline{b_j}$. Die Kovarianz $Cov(X_i, X_j)$ können wir dann mit der korrigierten Stichprobenkovarianz $Cov(X_j, X_k) = \frac{1}{n-1} \sum_{i=1}^{n} b^*_{ij} b^*_{ik}$ schätzen und die $p \times p$ Kovarianzmatrix $S = Cov(X) = (Cov(X_i, X_j))$ ist damit komplett angegeben.

Aus der Annahme, dass die Trainings-Histogramme aus der gleichen Klasse durch die gleiche Verteilung entstanden sind, berechnen wir pro Klasse eine solche Kovarianzmatrix. Damit ergibt sich pro Klasse eine Mahalanobis-Distanzfunktion, mit der sich der Abstand eines Bildes zu einer Klasse mit

$$d_C(x) = \sqrt{(x - \mu_C)^T S_C^{-1} (x - \mu_C)} \tag{3}$$

berechnen lässt, wobei μ_C der Mittelwert über die Trainings-BoW-Histogramme der Klasse C und S_C die Kovarianzmatrix der Trainingsvektoren der Klasse C ist.

2.1.3 Bhattacharyya-Distanz

Aus dem Bhattacharyya-Koeffizienten $BC(x, y) = \sum_{i=1}^{n} \sqrt{x_i y_i}$ (Ähnlichkeitsfunktion) für zwei BoW-Histogramme $x, y \in \mathbb{R}$ lässt sich die Bhattacharyya-Distanz mit

$$d_{B_1}(x, y) = -ln(BC(x, y)) \tag{4}$$

angeben [Cha07]. Falls $0 \leq BC(x, y) \leq 1$, dann ist $0 \leq d_{B_1} \leq \infty$.

2.1.4 Hellinger-Distanz

Eine weitere Möglichkeit aus dem Bhattacharyya-Koeffizienten $BC(x, y)$ (siehe 2.1.3) eine Distanz abzuleiten [Cha07] ist ebenfalls für $0 \leq BC(x, y) \leq 1$ gegeben mit:

$$d_{B_2}(x, y) = \sqrt{1 - BC(x, y)} \in [0, 1] \tag{5}$$

2.1.5 Histogrammschnitt-Distanz

Ähnlich wie bei der Bhattacharyya-Distanz gewinnen wir hier aus einem Ähnlichkeitsmaß eine Distanzfunktion. Aus dem Histogrammschnitt $\sum_{i=1}^{n} min(x_i, y_i) \in [0, 1]$ für zwei normierte BoW-Histogramme $x, y \in \mathbb{R}^n$ leiten wir diese Distanz ab:

$$d_H(x, y) = 1 - \sum_{i=1}^{n} min(x_i, y_i) \in [0, 1] \tag{6}$$

2.2 Ziele

Zur Verbesserung der Klassifikationsgüte wollen wir folgende Parameter untersuchen:

- Distanzfunktionen: Viele Klassifikatoren arbeiten basierend auf Distanzen, indem sie die Distanzen zwischen den BoW-Histogrammen berechnen. In der Literatur wird oft die Euklidische Distanz verwendet. Nun soll die Auswirkung weiterer Distanzfunktionen auf das Ergebnis untersucht werden.
- Kernel: Für den SVM-Klassifikator wird in der Anleitung der SVM-Bibliothek LibSVM [CL11] der RBF-Kernel empfohlen, da er allgemein gute Ergebnisse liefern soll. Nun soll die Klassifikationsgüte von SVM mit anderen Kernel untersucht werden.
- Normierung der BoW-Histogramme: Es soll der Einfluss der Normierung auf die Klassifikationsgüte der Distanz-basierten Klassifikatoren untersucht werden.

3 Hauptteil: Optimierung der Klassifikatoren

3.1 Optimierung NC-basierter Klassifikatoren

NC-basierte Klassifikatoren ordnen einer Eingabe x die Klassen C nach dem Prinzip $\arg\min_C d_C(x)$ zu, wobei $d_C(x)$ der Abstand von x zum Zentrum der Klasse C ist. Üblicherweise wird hier die Euklidische Distanz verwendet. Wir untersuchen nun, welche Auswirkungen weitere Distanzen und modifizierte NC-Ansätze auf die Ergebnisse haben.

3.1.1 NC-Klassifikator

NC-Klassifikator mit normierten BoW-Histogrammen

	100		250		500		1000		1500	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re
(1*)	0,27	0,27	0,31	0,31	0,34	0,32	0,33	0,32	0,34	0,32
(2*)	0,27	0,26	0,32	0,29	0,32	0,29	0,38	0,32	0,39	0,32
(3*)	0,27	0,26	0,31	0,3	0,31	0,3	0,35	0,34	0,36	0,33
(4*)	0,28	0,28	0,3	0,31	0,3	0,31	0,39	0,36	0,39	0,35
(5*)	0,28	0,27	0,34	0,32	0,33	0,31	0,38	0,35	0,41	0,36

Tabelle 2: Klassifikationsergebnisse des NC-Klassifikators mit normierten BoW-Histogrammen für die BoW-Größen 100,250,500,1000 und 1500 sowie den Distanzfunkionen: (1*) Manhatten (2*) Euklid (3*) HistoIntersection (4*) Bhattacharyya (5*) Hellinger

NC-Klassifikator mit nicht normierten BoW-Histogrammen

	100		250		500		1000		1500	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re
(1*)	0,24	0,21	0,26	0,2	0,25	0,15	0,19	0,08	0,13	0,07
(2*)	0,24	0,23	0,29	0,27	0,3	0,29	0,34	0,3	0,34	0,3

Tabelle 3: Klassifikationsergebnisse des NC-Klassifikators mit nicht normierten BoW-Histogrammen für die BoW-Größen 100,250,500,1000 und 1500 sowie den Distanzfunkionen: (1*) Manhatten (2*) Euklid

Diskussion zum NC-Klassifikator Mit normierten BoW-Histogrammen erreicht der NC-Klassifikator mit der BoW-Größe 1500 sowie der Hellinger-Distanz das Ergebnis (Recall: 0.36, Precision: 0.41), womit er die SVM-Ergebnisse der Baseline einholt und die NC-Ergebnisse der Baseline deutlich übertrifft. Das zeigt, dass auch einfach zu implementierende Klassifiktoren sich zur weiteren Verwendung und Optimierung eignen. Insbesondere sind NC-Klassifikatoren interessant, weil sie schnellere Trainingszeiten haben, als SVM-Klassifikatoren. Für nicht normierte Histogramme können wir nur die normierungsunabhängigen Distanzen vergleichen. Hier sehen wir, dass die Ergebnisse mit nicht normierten BoW-Histogrammen schlechter sind. Für die Manhatten-Distanz zeigt sich hier ein signifikanter Unterschied.

3.1.2 CNN

CNN (Center-based nearest neighbour) [GW07] ist ein modifizierter NC-Ansatz und berechnet nicht den Abstand eines BoW-Histogramms x zum Klassenzentrum o^c, sondern den Abstand von x zu seiner Projektion auf der Strecke $\overline{x_i^c o^c}$, wobei x_i^c ein Punkt der

Klasse c ist. Da es N_c Trainingspunkte in der Klasse c gibt, werden N_c Projektionen berechnet und anschließend minimiert, sodass die Distanz zur Klasse c sich mit

$$d^c_{CNN} = \min_{1 \leq i \leq N_c} d(x, \overline{x^c_i o^c}) \tag{7}$$

berechnen lässt. Die Ergebnisklasse wird nochmals aus allen Klassendistanzen minimiert: $c_{Result} = \arg\min_{c=1,2,...C} d^c_{CNN}$. Im Originalpaper wird die Euklidische Distanz verwendet. Wir untersuchen hier den Einsatz weiterer Distanzen.

CNN mit normierten BoW-Histogrammen

	100		250		500		1000	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re
(1*)	0,28	0,26	0,34	0,32	0,33	0,34	0,38	0,35
(2*)	0,28	0,27	0,34	0,3	0,36	0,33	0,4	0,34
(3*)	0,31	0,29	0,36	0,33	0,35	0,35	0,37	0,34

Tabelle 4: Ergebnisse des CNN-Klassifikators mit normierten BoW-Histogrammen für die BoW-Größen 100,250,500 und 1000 sowie den Distanzfunktionen: (1*) Manhatten (2*) Euklid (3*) HistoIntersection

CNN mit nicht normierten BoW-Histogrammen

	100		250		500		1000	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re
(1*)	0,28	0,26	0,31	0,25	0,29	0,19	0,26	0,1
(2*)	0,29	0,28	0,33	0,31	0,35	0,33	0,36	0,33

Tabelle 5: Ergebnisse des CNN-Klassifikators mit nicht normierten BoW-Histogrammen für die BoW-Größen 100,250,500 und 1000 sowie den Distanzfunktionen: (1*) Manhatten (2*) Euklid

Diskussion des CNN-Klassifikators Der CNN-Klassifikator liefert für normierte BoW-Histogramme mit der BoW-Größe von $k = 1000$ die besten Ergebnisse. Hierbei liefert die Manhatten Distanz einen Recall-Wert von 0.35 und kommt damit nah an den besten Recall-Wert (0.36) des SVM-Klassifikators aus der Baseline heran. Bei nicht normierten BoW-Histogrammen sind die Ergebnisse durchgehend schlechter, wobei auch hier der deutliche Unterschied bei der Manhattan Distanz zu bemerken ist.

3.1.3 Mahalanobis-Distanz-Klassifikator

Dieser Klassifikator ergibt sich aus der Mahalanobis-Distanz (3). Die beste Klasse für ein durch das BoW-Histogramm x repräsentiertes Bild lässt sich dann als eine Modifikation des NC-Klassifikators mit $\arg\min_C d_C(x)$ berechnen.

Die Ergebnisse des Mahalanobis-Distanz-Klassifikators sind in Tabelle 6 abgebildet.

	100		250		500		1000	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re
(1*)	0,2	0,22	0,27	0,27	0,31	0,3	0,38	0,32
(2*)	0,3	0,28	0,34	0,32	0,34	0,33	0,39	0,37

Tabelle 6: Klassifikationsergebnisse des Mahalanobis-Distanz-Klassifikators für die BoW-Größen 100,250,500 und 1000 sowie für folgende Einstellungen: (1*) normierte BoW-Histogramme (2*) nicht normierte BoW-Histogramme

Diskussion für den Mahalanobis-Distanz-Klassifikator Bereits ab der BoW-Histogramm-Größe von 1000 und nicht normierten Histogrammen übertrifft der NC-Klassifikator mit der Mahalanobis-Distanz die Ergebnisse des NC-Klassifikators mit der Euklidischen Distanz und das beste Recall-Ergebnis des SVM-Klassifikators mit dem RBF-Kernel. Anders als bei vorherigen Versuchen wirkt sich die Normierung deutlich negativ aus.

3.2 Optimierung des SVM-Klassifikators

3.2.1 SVM-Kernel

Folgende Kernel werden zur Klassifikation der BoW-Histogramme untersucht:

- RBF-Kernel (Radial Basis Function): $K(u,v) = exp(-\frac{1}{2\sigma^2}||u-v||^2)$. Der σ-Parameter wird mit der folgenden Heuristik [JDH99] berechnet:
  $\sigma = median(\{dist(u,v)|u \neq v\})$.
- Polynom-Kernel: $K(x,y) = (\alpha x^T y + c)^d$
- Linear-Kernel (Spezialfall des Polynom-Kernels): $K(x,y) = x^T y + c$
- Hyperbolic Tangent (Sigmoid) Kernel: $K(x,y) = tanh(\alpha x^T y + c)$
- Inverse Multiquadric Kernel [Mic86]: $K(x,y) = \frac{1}{\sqrt{||x-y||^2+1}}$
- χ^2 Kernel: $K(x,y) = 1 - \sum_{i=1}^{n} \frac{(x_i-y_i)^2}{\frac{1}{2}(x_i+y_i)}$, abgeleitet aus [Cha07] und nur für normierte BoW-Histogramme sinnvoll
- Histogrammschnitt-Kernel: $K(x,y) = \sum_{i=1}^{n} min(x_i, y_i)$

3.2.2 Evaluation verschiedener Kernel

Für die Evaluation wird die DAG Multiklassen-Strategie verwendet. Bei DAG (Directed Acyclic Graph) [PCST99] handelt es sich um einen azyklisch gerichteten Graphen. In jedem Knoten des Graphen wird eine binäre Entscheidung über die Klassenzugehörigkeit eines betrachteten BoW-Histogramms getroffen. In Tabelle 7 wird nur das Ergebnis für $C = 2$ präsentiert, wobei C der C-SVM-Parameter [CL11] ist.

	100		250		500		1000		1500	
	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re
(1*)	0,31	0,29	0,38	0,36	0,4	0,37	0,43	0,4	0,42	0,38
(2*)	0,31	0,28	0,37	0,32	0,35	0,31	0,4	0,32	0,4	0,31
(3*)	0,32	0,3	0,38	0,36	0,4	0,37	0,41	0,39	0,41	0,38
(4*)	0,25	0,2	0,32	0,24	0,29	0,21	0,33	0,22	0,32	0,2
(5*)	0,26	0,2	0,31	0,23	0,31	0,22	0,35	0,23	0,29	0,21
(6*)	0,26	0,2	0,32	0,22	0,31	0,21	0,32	0,21	0,33	0,21
(7*)	0,27	0,21	0,3	0,23	0,31	0,23	0,36	0,23	0,32	0,21

Tabelle 7: Klassifikationsergebnisse für die BoW-Größen 100,250,500,1000 und 1500 sowie folgende Kernel: (1*) χ^2 (2*) RBF mit σ-Heuristik (3*) Histogramm-Schnitt (4*) Inverse Multiquadric (5*) Linear ($c = 1$) (6*) Polynomial ($\alpha = 1, c = 1, d = 3$) (7*) Sigmoid ($c = 1$)

Diskussion der verschiedenen SVM-Kernel Die Kernel χ^2- und der Histogrammschnitt-Kernel verbessern die SVM-Baseline Ergebnisse deutlich. So wird mit dem χ^2-Kernel der Recall-Wert von 0.36 auf 0.4 und der Precision-Wert von 0.41 auf 0.43 verbessert. Zu bemerken ist, dass die RBF-Kernel σ-Heuristik für schlechtere Recall-Werte sorgt, die Precision-Werte aber nicht deutlich schlechter werden. Die polynomiellen Kernel müssen noch weiter untersucht werden, da ein 3-dimensionales Beispiel nicht zur Ablehnung ausreicht. Der Inverse Multiquadric und Sigmoid Kernel sind aufgrund vergleichsweise schlechter Recall-Werte nicht für diese Aufgabe empfehlenswert.

4 Fazit

Die distanzbasierte Optimierung hat gezeigt, dass ein vergleichsweise einfacher Klassifikator, wie der NC-Klassifikator mit der Hellinger-Distanz ähnlich gute Ergebnisse wie eine SVM mit dem oft empfohlenen RBF-Kernel liefern kann. Auch der Mahalanobis-Distanz Klassifikator konnte das Baseline Ergebnis des NC-Klassifikators deutlich verbessern. Dies zeigt, dass ein Vergleich von Distanzen sinnvoll ist, da das Ergebnis so erheblich beinflusst werden kann. Andererseits stellt sich die Frage, ob die Wahl der Distanzfunktion allein aus der Definition und nicht aus Experimenten bestimmt werden kann. So bleibt zu prüfen, ob die Hellinger Distanz auch auf normierten Histogrammen aus anderen Bereichen gute Ergebnisse liefert, um eine allgemeine Empfehlung für solche Daten auszusprechen. Da NC-Klassifikatoren einfach zu implementieren sind und im Vergleich zu SVM-Klassifikatoren schnellere Trainingszeiten aufweisen, können die vergleichsweise guten Ergebnisse als Motivation zur weiteren Untersuchung und Optimierung dienen.

Nicht zu vernachlässigen ist auch der Einfluss der Normierung auf die Klassifikationsergebnisse. In den meisten Fällen hat die Normierung positive Tendenzen gezeigt, außer bei der Mahalanobis-Distanz. Insbesondere sollte die Normierung für den Einsatz der Manhatten Distanz berücksichtigt werden, da hier die größten Differenzen aufgetreten sind.

Ein weiterer zu berücksichtigender Faktor ist die Wahl der Kernels für die SVM. Auch hier

kann die Wahl eines anderen geeigneteren Kernels deutliche Verbesserungen hervorrufen. Die Kernel χ^2- und Histogrammschnitt-Kernel können für die Klassifikation der BoW-Histogramme aufgrund der eindeutig besseren Ergebnisse empfohlen werden.

Literatur

[CDF+04] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski und Cédric Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, Seiten 1–22, 2004.

[Cha07] Sung-Hyuk Cha. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307, 2007.

[CL11] Chih-Chung Chang und Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[CV95] Corinna Cortes und Vladimir Vapnik. Support-Vector Networks. In *Machine Learning*, Seiten 273–297, 1995.

[FFFP04] L. Fei-Fei, R. Fergus und Pietro Perona. Learning Generative Visual Models From Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. 2004.

[GW07] Qing-Bin Gao und Zheng-Zhi Wang. Center-based nearest neighbor classifier. *Pattern Recognition*, 40(1):346–349, 2007.

[Hel09] Reiner Hellbrück. Multivariate Korrelationsanalyse. In *Angewandte Statistik mit R*, Seiten 165–188. Springer, 2009.

[JDH99] Tommi Jaakkola, Mark Diekhans und David Haussler. Using the Fisher Kernel Method to Detect Remote Protein Homologies. In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, Seiten 149–158. AAAI Press, 1999.

[Low04] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[Mic86] Charles A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2:11–22, 1986.

[MRS08] Christopher D. Manning, Prabhakar Raghavan und Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[MT09] P. Mather und B. Tso. *Classification Methods for Remotely Sensed Data, Second Edition*. Environmental engineering. Taylor & Francis, 2009.

[PCST99] John C Platt, Nello Cristianini und John Shawe-Taylor. Large Margin DAGs for Multiclass Classification. In *nips*, Jgg. 12, Seiten 547–553, 1999.

[vdSGS08] K. E. A. van de Sande, T. Gevers und C. G. M. Snoek. Evaluation of Color Descriptors for Object and Scene Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[wHcCjL10] Chih wei Hsu, Chih chung Chang und Chih jen Lin. A practical guide to support vector classification, 2010.