

Bewertung von Strategien des aktiven Lernens am Beispiel der Landbedeckungsklassifikation

SEBASTIAN WUTTKE^{1,2}, WOLFGANG MIDDELMANN¹ & UWE STILLA²

Zusammenfassung: Aktives Lernen ist ein überwachtes Lernverfahren. Die Beschaffung der zum Training nötigen Ground Truth verursacht Kosten (manuelle Interaktion). Daher sollten möglichst wenige, aber aussagekräftige Trainingsbeispiele verwendet werden. Aktives Lernen versucht dies durch die Anwendung einer Selektionsstrategie. Ein Problem ist deren Auswahl, da diese stark vom verwendeten Klassifikator abhängt. Dieser Beitrag bewertet Selektionsstrategien für einen Maximum-Likelihood-Klassifikator. Zwei Szenarien werden betrachtet: 1) Alle Trainingsbeispiele stehen auf einmal zur Verfügung (passives Lernen) und 2) die Selektionsstrategie wählt die Trainingsbeispiele iterativ aus (aktives Lernen). Als Anwendungsbeispiel wird Landbedeckungsklassifikation in einem RGB-NIR Luftbild durchgeführt.

1 Einleitung und Problemstellung

Landbedeckungsklassifikation ist ein aktives Forschungsthema, da es für die globale Umweltentwicklung sehr relevant ist (FOODY 2002). Im Unterschied zur Landnutzungsklassifikation findet lediglich eine Beurteilung der Landoberfläche statt und es wird nicht nach Nutzung unterschieden.

Die fortschreitende Verbesserung der Sensorik in der Fernerkundung führt zu immer höheren Auflösungen und Datenmengen. Deshalb ist Landbedeckungsklassifikation auf manuellem Wege nicht mehr wirtschaftlich. Viele Arbeiten beschäftigen sich daher mit dem Einsatz von maschinellen Lernverfahren in der Landbedeckungsklassifikation (DEMIR 2011, SCHINDLER 2012). Kommen dabei überwachte Lernverfahren zum Einsatz, ist es zwingend erforderlich Trainingsbeispiele zur Verfügung zu stellen. Die Beschaffung dieser ist oftmals mit hohen Kosten verbunden (meist in Form von menschlicher Arbeitszeit). Die Kosten stehen dabei in direktem Verhältnis zur Anzahl der benötigten Trainingsbeispiele. Eine gute Möglichkeit Trainingskosten einzusparen ist also, die Anzahl der benötigten Beispiele zu reduzieren. Da dies oftmals zu einer Verschlechterung der Klassifikationsgüte führt, ist es ratsam bestimmte Strategien zur Auswahl der zu verwendenden Beispiele zu verfolgen. Hierbei hilft aktives Lernen.

Dieser Beitrag untersucht und bewertet die Interaktion zwischen Klassifikator und Selektionsstrategie genauer. Dazu wurde der Maximum-Likelihood-Klassifikator als einfacher und gut verstandener Klassifikator gewählt. Voruntersuchungen finden auf einem synthetischen Datensatz statt und werden auf einem 4-Kanal Luftbild überprüft.

- 1) Fraunhofer IOSB, Gutleuthausstraße 1, 76275 Ettlingen,
E-Mail: Sebastian.Wuttke@IOSB.Fraunhofer.de, Wolfgang.Middelmann@IOSB.Fraunhofer.de
- 2) Technische Universität München, Boltzmannstraße 15, 85748 München,
E-Mail: Sebastian.Wuttke@tum.de, Stilla@tum.de

Es ergibt sich folgendes Szenario:

- Es stehen ungelabelte Daten in großer Menge „kostenlos“ zur Verfügung
- Der Klassifikator steht fest
- Das Erheben von Ground Truth ist möglich, verursacht jedoch hohe Kosten
- Es soll aktives Lernen eingesetzt werden, um diese Kosten gering zu halten

Die daraus folgenden Aufgabenstellungen lauten: Welche Selektionsstrategie passt zum gewählten Klassifikator? Welche Auswirkungen hat die Wahl der Selektionsstrategie auf die Klassifikationsleistung? Lassen sich die Strategien anpassen, um bessere Leistungen zu erzielen?

2 Verwandte Arbeiten

Das Thema aktives Lernen ist ein aktuelles Forschungsfeld (SETTLES 2012, DEMIR 2014). Statistiker kennen ähnliche Methoden unter dem Namen „optimal experimental design“.

2.1 Aktives Lernen

Eine der frühesten Verwendungen des Begriffs „aktives Lernen“ ist ein Beitrag von COHN in *Machine Learning* 15 (1994). In seiner Arbeit definiert er aktives Lernen als einen Lernprozess, bei dem das Lernverfahren einen gewissen Grad an Kontrolle über die Eingaben hat, mit denen es trainiert wird. Er vergleicht es mit dem Lernen in natürlichen Systemen (wie zum Beispiel dem Menschen). Er beschränkt sich auf die intuitivste Form des aktiven Lernens: Konzeptlernen durch Zugehörigkeitsanfragen („*concept learning via membership queries*“). Dabei wählt der Lernprozess eine Instanz aus der Eingabedomäne und erfährt von einem Orakel deren wahre Klasse. Das Orakel ist dabei der Mensch oder eine andere Informationsquelle von der das wahre Label – unter Einsatz von Ressourcen – erhalten werden kann. SETTLES (2009) nennt diese Art der Anfragen „*membership query synthesis*“, da hierbei der Lerner die angefragte Instanz selbst erzeugt (synthetisiert). Dieser Ansatz führt bei manchen Lernaufgaben jedoch zu Problemen, wenn sich nicht alle Instanzen der Eingabedomäne eindeutig einer Klasse zuordnen lassen (LANG & BAUM 1992). Dieses Problem tritt nicht auf, wenn der Lerner nur bereits existierende Instanzen aus der Eingabedomäne wählen darf. SETTLES unterscheidet hierbei zwei Szenarien: *Stream-Based* und *Pool-Based*. Bei ersterem treffen die unbekanntenen Trainingsbeispiele nacheinander ein und der Lerner entscheidet für jedes einzelne, ob es beim Orakel angefragt wird. Im zweiten Szenario stehen dem Lerner alle ungekennzeichneten Instanzen gleichzeitig zur Verfügung und er wählt frei daraus aus.

Eine weitere Unterscheidung ist die Anzahl der beim Orakel gleichzeitig angefragten Instanzen. Die Redundanz in den Anfragen kann am stärksten reduziert werden, indem Instanzen nur einzeln angefragt werden und nach jeder Anfrage eine erneute Auswertung stattfindet. In manchen Anwendungen (zum Beispiel zur Parallelisierung) ist es jedoch erforderlich mehrere Instanzen auf einmal auszuwählen und dem Orakel (oder den Orakeln) vorzulegen. Dies wird auch *Batch-Mode Active Learning* genannt (BRINKER 2003, XU 2007).

Dieser Beitrag beschäftigt sich mit (offline-) Landbedeckungsklassifikation. Das heißt, dass das Luftbild zum Zeitpunkt des Trainings bereits vollständig vorliegt. Daher ist das Pool-Based Szenario am geeignetsten. Eine Parallelisierung der Orakelanfragen findet nicht statt, daher werden die Instanzen sequentiell angefragt.

2.2 Abgrenzung zu anderen Lernmethoden

Die meisten Lernverfahren beschäftigen sich mit der Fragestellung wie aus einer gegebenen Menge von Trainingsbeispielen das beste Klassifikationsergebnis für neue, unbekannte Daten erreicht werden kann. Aktives Lernen hingegen versucht die Menge der benötigten Trainingsbeispiele zu reduzieren. Eine verwandte Klasse von Verfahren sind *Semi-Supervised* Lernmethoden (ZHU 2005, ZHU & GOLDBERG 2009). Diese versuchen ebenfalls möglichst viele Informationen aus der großen Menge noch ungelabelter Instanzen zu gewinnen. Die Herangehensweise ist jedoch genau gegensätzlich. Während aktives Lernen versucht durch Auswahl der informativsten Beispiele die Trainingskosten zu reduzieren, vertraut Semi-Supervised Learning dem gelernten Modell. Dieses wird dazu genutzt das bisher gelernte Wissen auf ungelabelte Instanzen zu extrapolieren, um so die Menge der verfügbaren Trainingsbeispiele zu erhöhen, ohne weitere Ground-Truth-Kosten zu verursachen.

Ein weiterer Ansatz zur Verringerung der Trainingskosten ist *Feature Selection* (BRUZZONE 2000). Die Intention ist, dass eine leichter trennbare Klassenverteilung entsteht und der Klassifikator mit weniger Trainingsbeispielen auskommt. Ebenso gibt es Verfahren, um gezielt einen Klassifikator für die konkrete Problemstellung auszuwählen.

3 Methoden

Die Verwendung eines Frameworks für aktives Lernen ermöglicht es vergleich- und wiederholbare Ergebnisse zu erzielen. Einzelne Komponenten sind durch den modularen Aufbau gezielt austausch- und erweiterbar. Da sich immer gleiche Rahmenbedingungen herstellen lassen, ist es einfach möglich verschiedene Verfahren zu vergleichen. Für diesen Beitrag wird, wie in Kapitel 1 erwähnt, ein Maximum-Likelihood-Klassifikator verwendet. Die untersuchten Selektionsstrategien sind: *Random Sampling*, *Uncertainty Sampling* und *Query by Committee*. Weitere mögliche Strategien sind: *Expected Error Reduction* (ROY 2001) und Strategien, die die Verteilung und Struktur der Daten im Featureerraum beachten (*Density Weighted Methods*).

Die Voruntersuchungen dieses Beitrags fanden mit einem synthetischen Datensatz statt. Um das Selektionsverhalten der verschiedenen Strategien einfach zu visualisieren, wurden zwei-dimensionale Daten gewählt. Der Datensatz besteht aus 9.000 Instanzen dreier normalverteilter Klassen (Abbildung 3-1).

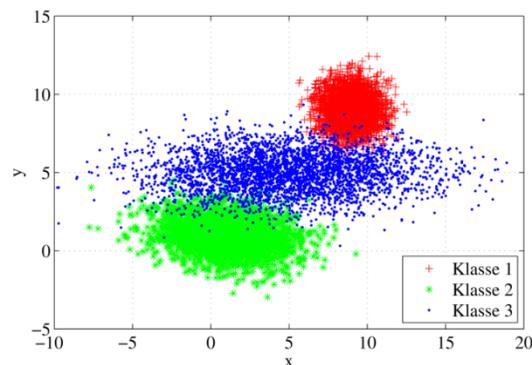


Abbildung 3-1: Synthetische Daten mit drei normalverteilten Klassen.

3.1 Maximum-Likelihood-Klassifikator

Der Maximum-Likelihood-Klassifikator ist ein sehr oft verwendeter und sehr gut verstandener Klassifikator (PRATT 1976, SAVAGE 1976), der auch auf dem Gebiet des Remote Sensing eingesetzt wird (BRUZZONE 2001, RAJAN 2008, SCHINDLER 2012). Er basiert auf der Annahme, dass die Klassen normalverteilt sind. Es wird der Mittelwert bestimmt: $\mu_c = \frac{1}{N_c} X_c e$ mit $e = (1 \dots 1)^T$ und X_c als Matrix aus Spaltenvektoren aller Trainingsbeispiele der Klasse c und N_c ihrer Anzahl. Ebenso die Kovarianzmatrix: $\Sigma_c = \frac{1}{N_c} (X_c - \mu_c e^T) (X_c - \mu_c e^T)^T$. Die sich ergebenden multivariaten Gaußverteilungen werden bei der anschließenden Klassifikation an der Stelle des angefragten Beispiels ausgewertet. Aus den berechneten Zugehörigkeitswahrscheinlichkeiten (*Likelihoods*) zu jeder Klasse wird nun die *maximale* bestimmt und das Beispiel dieser Klasse zugewiesen:

$$y(x) = \operatorname{argmax}_c \log \frac{1}{N_c} - \log |\Sigma_c| - (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c), \quad (1)$$

3.2 Selektionsstrategien aus der Literatur

Im Folgenden werden einige in der Literatur vertretene Selektionsstrategien genauer erläutert.

Random Sampling

Diese Strategie wählt die Beispiele zufällig. Sie hat im hier vorliegenden Szenario einige Vorteile: Der Maximum-Likelihood-Klassifikator modelliert eine Normalverteilung. Diese approximiert die wirkliche Verteilung dann am genauesten, wenn ihre Eingabewerte ebenfalls Normalverteilt sind. Werden die Beispiele durch Random Sampling gleichverteilt gezogen, bildet die Trainingsmenge die wahre Verteilung nach. Da dies im vorliegenden Fall eine Normalverteilung ist, ist zu erwarten das Random Sampling im Zusammenspiel mit dem Maximum-Likelihood-Klassifikator gute Ergebnisse liefert.

Uncertainty Sampling

Die Redundanz in den Trainingsbeispielen kann durch Vermeiden von Anfragen, über die sich der Lerner bereits sehr sicher ist, verringert werden. Es sind also Beispiele anzufragen, deren Klassifikationsunsicherheit besonders groß ist. Solche Anfragestrategien werden auch als *Uncertainty Sampling* bezeichnet (LEWIS 1994). Um solche Strategien zu nutzen, muss der Klassifikator ein Maß für die Klassifikationssicherheit angeben. Beim Maximum-Likelihood-Klassifikator ist dies die berechnete Likelihood.

Es gibt verschiedene Unterarten dieser Strategie. Im einfachsten Fall wird das Beispiel angefragt, dessen Klassifikationssicherheit am geringsten ist. Dies wird als *Least Confident Sampling* bezeichnet. Hier ist x^* das anzufragende Beispiel und $P_\theta(\hat{y}|x)$ die Wahrscheinlichkeit des Modells θ , dass x zur Klasse \hat{y} gehört: $x^* = \operatorname{argmax}_x 1 - P_\theta(\hat{y}|x)$.

Das Betrachten der Likelihood nur einer einzigen Klasse führt dazu, dass bevorzugt Beispiele aus Randregionen gewählt werden (siehe Abbildung 3-2). Dies beeinträchtigt den Maximum-Likelihood-Klassifikator, da die Varianz der approximierten Verteilungen groß wird und die Klassifikationsleistung stark sinkt. Ein Vergleich der beiden sichersten Klassifikationen \hat{y}_1 und \hat{y}_2 eines Beispiels reduziert dies. Die gewonnenen Informationen geben Aufschluss über die Klassengrenze. Daher wird diese Variante auch *Margin Sampling* genannt:

$$x^* = \operatorname{argmax}_x [P_\theta(\hat{y}_2|x) - P_\theta(\hat{y}_1|x)] \quad (2)$$

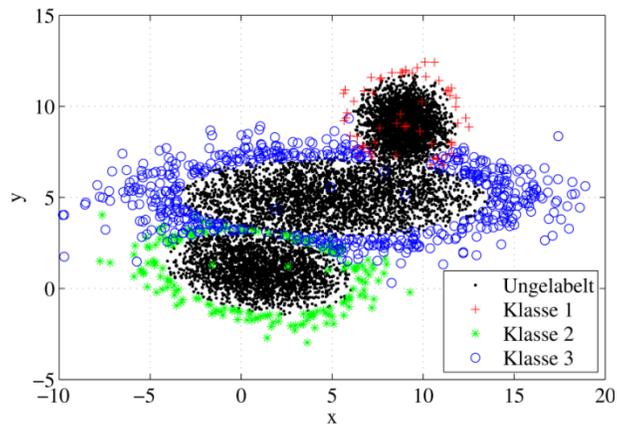


Abbildung 3-2: Selektionsverhalten von Uncertainty Sampling in der Variante Least Confident.

Query by Committee

Steht kein Maß für die Klassifikationssicherheit zur Verfügung, kann Uncertainty Sampling nicht angewendet werden. Eine Alternative stellt Query by Committee dar (SEUNG 1992). Mehrere Klassifikatoren bilden ein Komitee und die Übereinstimmung ihrer Antworten wird bewertet. Eine Möglichkeit das Komitee zu erstellen, ist einen Klassifikator mit verschiedenen Teilmengen der Trainingsbeispiele zu trainieren und so mehrere Variationen zu erhalten. In dieser Arbeit wurden die Komiteemitglieder jeweils mit 70% der verfügbaren Trainingsdaten trainiert (zufälliges Ziehen ohne Zurücklegen). Die Klassifikationsentscheidung des Komitees ergibt sich durch eine gewichtete Mehrheitsentscheidung. Die Gewichte sind hier die Klassifikationsgüten der einzelnen Mitglieder auf 100% der aktuell verfügbaren Trainingsbeispiele. Die Instanz, welche die maximale Unstimmigkeit im Komitee erzeugt, wird als nächstes angefragt (*Vote Entropy*):

$$x^* = \operatorname{argmax}_x - \sum_y \frac{\operatorname{vote}_C(y,x)}{|C|} \log \frac{\operatorname{vote}_C(y,x)}{|C|} \quad (3)$$

Hierbei steht $\operatorname{vote}_C(y,x)$ für die Anzahl der Stimmen des Komitees C , die bezüglich Beispiel x für Klasse y stimmen. $|C|$ ist die Anzahl der Komiteemitglieder. Das erhaltene Label wird zur Menge der Trainingsdaten hinzugefügt und in der nächsten Iteration wird ein neues Komitee erstellt. Das Selektionsverhalten ist in Abbildung 3-3 visualisiert. Es ist deutlich zu erkennen, dass sich das Komitee an den Klassengrenzen uneinig ist. Dies führt jedoch ebenso wie beim Uncertainty Sampling dazu, dass der Maximum-Likelihood-Klassifikator die wahre Verteilung nicht richtig approximiert.

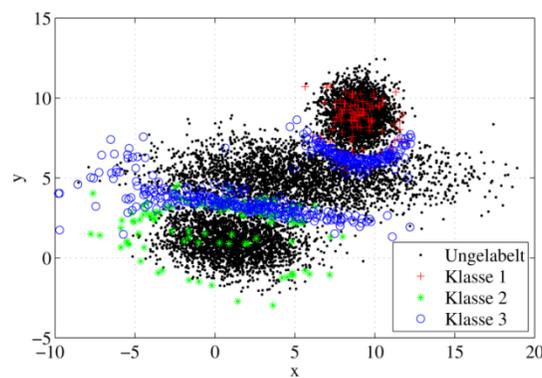


Abbildung 3-3: Selektionsverhalten von Query by Committee mit einem Komitee aus 5 Maximum-Likelihood-Klassifikatoren.

3.3 Angepasste Selektionsstrategien

In ersten Untersuchungen hat sich wie erwartet gezeigt, dass Random Sampling durch das Zusammenspiel mit dem Maximum-Likelihood-Klassifikator bessere Ergebnisse liefert als Uncertainty Sampling und Query by Committee (siehe Abbildung 3-4). Es soll nun untersucht werden, ob dies weiter verbessert werden kann.

Durch gleichverteilte Anfragen werden Beispiele aus Regionen hoher Dichte häufiger gewählt als aus weniger dichten Regionen. Dies ist erwünscht, erhöht jedoch auch die Redundanz in den Trainingsbeispielen und bietet Einsparungspotenzial. Dazu wird Random Sampling erweitert: Die neue Variante sperrt alle ungelabelten Beispiele in der Nachbarschaft des ausgewählten für zukünftige Selektionen (*Random Sampling mit Sperrung*). Dies reduziert die Redundanz in angefragten Beispielen und sorgt für eine explorativere Auswahl. Eine Schwierigkeit dabei ist die Wahl des Radius der zu sperrenden Umgebung. Hier wurde er auf einen Wert von 10% der Standardabweichung des Datensatzes gelegt. In Datensätzen, die große Skalenunterschiede zwischen einzelnen Dimensionen aufweisen, könnte dies zu Problemen führen und sollte weiter untersucht werden. Bei den hier verwendeten Datensätzen wurden keine Probleme festgestellt. Diese Variante führt einerseits zur gewünschten Verminderung der Redundanz in den Anfragen, aber andererseits stört sie die Approximation der wahren Klassenverteilung, da beim Maximum-Likelihood-Klassifikator alle Beispiele gleichmäßig zur Bestimmung der Verteilung beitragen. Eine weitere Variante gleicht dies aus, indem sie die selektierten Beispiele mit der Anzahl der Sperrungen gewichtet (*Random Sampling mit gewichteter Sperrung*). Hier ist W_c der Zeilenvektor der Gewichte für Beispiele aus Klasse c :

$$\mu_c = \frac{1}{W_c e} X_c \text{diag}(W_c) e \quad (4)$$

$$\Sigma_c = \frac{1}{W_c e} (X_c \text{diag}(W_c) - \mu_c e^T) (X_c \text{diag}(W_c) - \mu_c e^T)^T \quad (5)$$

$$y(x) = \operatorname{argmax}_c \log \frac{1}{W_c e} - \log |\Sigma_c| - (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) \quad (6)$$

So tragen Beispiele aus Regionen hoher Dichte stärker zur Approximation bei als Beispiele aus Regionen geringerer Dichte. Hier stammt das Label vom Orakel und eine Fehlklassifikation ist somit ausgeschlossen (unter der Annahme, dass das Orakel fehlerfrei ist). Dessen Zuweisung an die Nachbarn kann dennoch falsch sein. Dies tritt besonders häufig an Klassengrenzen auf. Da diese Regionen in der Regel jedoch nur eine geringe Dichte aufweisen, ist der zu erwartende Fehler begrenzt. Die Ergebnisse der 5 vorgestellten Selektionsstrategien sind in Abbildung 3-4 dargestellt.

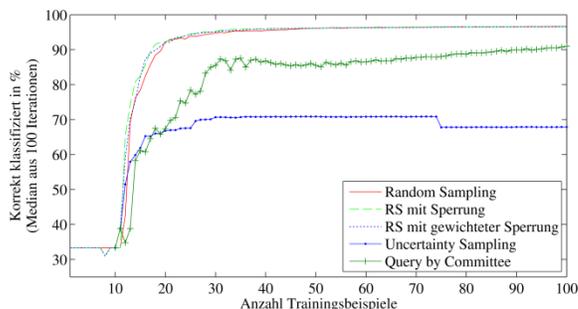


Abbildung 3-4: Lernkurven des Maximum-Likelihood-Klassifikators auf den synthetischen Daten. Dargestellt ist der Median aus 100 Iterationen.

3.4 Bewertung der Strategien

Die Bewertung der Selektionsstrategien erfolgt anhand der Lernkurven. Das heißt, der korrekt klassifizierte Anteil in Abhängigkeit der Anzahl verwendeter Trainingsbeispiele. Das bereits etablierte Verfahren der Fläche unter der Receiver Operating Characteristic Kurve (AUC-ROC) kann hier nicht angewendet werden. Die Selektionsstrategien hängen – besonders bei geringer Anzahl von Beispielen – zu stark vom Datensatz ab, so dass keine statistisch verlässlichen Aussagen gemacht werden können, wie sie für die AUC-ROC benötigt werden. Daher erfolgt hier eine visuelle Bewertung der Lernkurven. Es bleibt für den einzelnen Anwendungsfall zu diskutieren, ob eine frühzeitige Steigung (geringe Trainingskosten) oder ein möglichst hoher Maximalwert (gute Klassifikationsleistung) zu bevorzugen ist.

4 Anwendung auf Fernerkundungsdaten

Die theoretischen Untersuchungen in diesem Beitrag wurden auf einem synthetischen Datensatz durchgeführt. Die erhaltenen Ergebnisse werden nun auf dem natürlichen Datensatz angewendet.

4.1 Natürliche Daten

Dieser Datensatz stammt aus einer Befliegung des Fraunhofer IOSB am 5.9.2013 über Bremerhaven. Die Daten wurden mit dem AISA Eagle II Hyperspektralsensor der Firma Specim aufgenommen. Die Georeferenzierung wurde mit einer IOSB-eigenen Software durchgeführt. Für diesen Beitrag wurden die Daten mit einer Kanalbündelung vorverarbeitet, so dass ein 4-Kanal-Bild (RGB-NIR) entsteht. Die Bündelung orientiert sich am Pleiades Sensor und wurde mit folgenden Werten durchgeführt: Blau (455-525 nm), Grün (530-590 nm), Rot (625-695 nm), nahes Infrarot (760-890 nm). Es fand keine weitere Feature Extraction statt (zum Beispiel Berechnung des Normalized Difference Vegetation Index). Der Datensatz umfasst eine Fläche von 4,5 km² bei einer Bodenauflösung von 0,5 m pro Pixel. Die RGB-Kanäle sind in Abbildung 4-1 dargestellt.

Die Ground Truth für diesen Datensatz wurde manuell mit Hilfe eines Orthophotos und der Software ENVI erstellt. Insgesamt stehen 121.727 Beispiele aus 6 Klassen zur Verfügung: Laubwald, Asphalt, Wasser, Acker, Feld und Wiese.

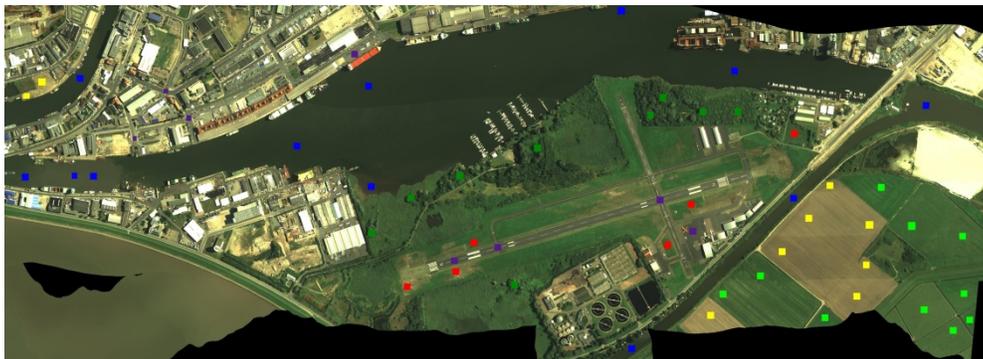


Abbildung 4-1: RGB-Kanäle des natürlichen Datensatzes. Der Ausschnitt zeigt Bremerhaven (Norden ist im Bild links). Hervorgehoben sind die manuellen Trainingsgebiete der 6 Klassen.

4.2 Experiment

Die Varianten des Random Sampling zeigen auf den natürlichen Daten ein ähnliches Verhalten wie auf den synthetischen (vergleiche Abbildung 3-4 und Abbildung 4-2). Query by Committee und Uncertainty Sampling erreichten jedoch lediglich eine Klassifikationsgüte von unter 20% und sind daher nicht für diese Kombination von Klassifikator und Datensatz geeignet.

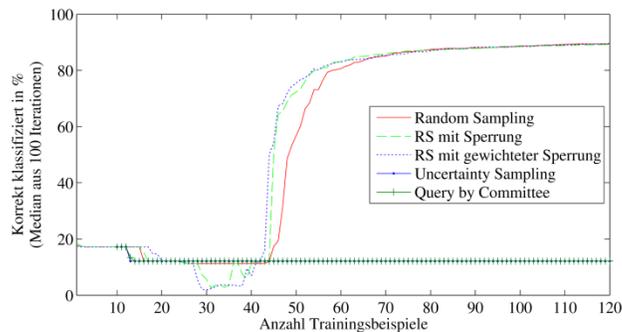


Abbildung 4-2: Lernkurven des Maximum-Likelihood-Klassifikators auf den natürlichen Daten. Dargestellt ist der Median aus 100 Iterationen.

Für die Landbedeckungsklassifikation wurde eine zusätzliche Rückweisungsklasse eingeführt. Die Rückweisung tritt insbesondere bei Hausdächern auf, da diese nicht eintrainiert wurden. Bei genauerer Betrachtung ist positiv festzustellen, dass die Markierungen auf der Landebahn nicht der Asphalt-Klasse zugewiesen wurden. Fehl-Klassifikationen treten jedoch an einem kleinen See auf, sowie an den Boots-Anlegestegen in der Mitte des Bildes auf. Die Klassen Feld und Wiese unterscheiden sich in der Stärke des Vegetationsbewuchses und konnten gut von einander getrennt werden.

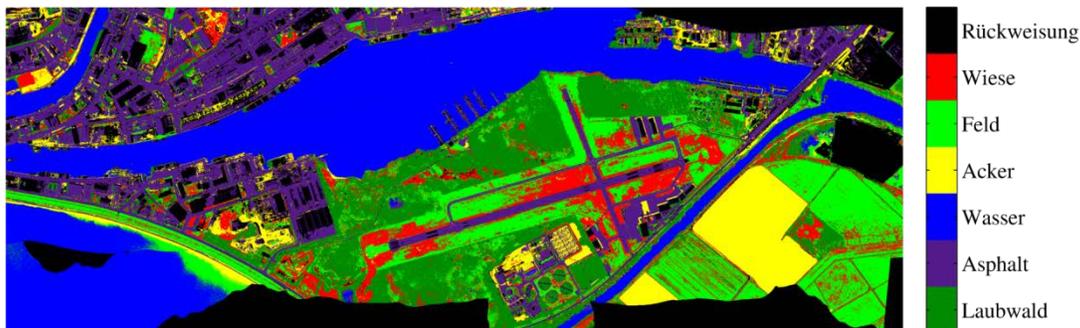


Abbildung 4-3: Ergebnis der Landbedeckungsklassifikation des Maximum-Likelihood-Klassifikators mit der Selektionsstrategie Random Sampling mit gewichteter Sperrung.

5 Diskussion

Es fand keine Feature Extraction statt, um die Rahmenbedingungen für die theoretischen Untersuchungen einfach zu halten. Auch wurde ein einfacher Klassifikator gewählt. Im Szenario „Passives Lernen“ wurde eine maximal erreichbare Klassifikationsleistung von 95,9% auf den synthetischen und 91,6% auf den natürlichen Daten bestimmt.

Im Szenario „Aktives Lernen“ wurden zur Bewertung der Selektionsstrategien die Berechnungen 100-mal wiederholt. Der Median der erreichten Klassifikationsgüte mit 50 Trainingsbeispielen beträgt für Random Sampling 56,8%. Dass die Anpassungen erfolgreich waren, hat sich auf den

echten Daten durch eine Steigerung dieser Klassifikationsgüte gezeigt: mit Sperrung 72,3% und mit gewichteter Sperrung 75,5%. Nach weiteren 50 Trainingsbeispielen liegt die Klassifikationsgüte aller drei Random Sampling Varianten bei rund 88,5%. Dies zeigt, dass eine weitere wichtige Entscheidung ist, wann das Training gestoppt werden kann. Untersuchungen dazu sind für zukünftige Arbeiten vorgesehen.

Uncertainty Sampling und Query by Committee hängen stark von der Ausgangssituation ab. Sind in den ersten Iterationen nicht alle Klassen vertreten, kommt es oftmals dazu, dass die unbekannteren Klassen nie angefragt werden. Um dies zu umgehen, wurden in der ersten Iteration 10 Trainingsbeispiele auf einmal gewählt, um das Training zu initialisieren. Anschließend wurde wie beim Random Sampling in jeder Iteration nur ein Beispiel hinzugefügt. Diese Startkonfiguration wurde für jeden Durchlauf neu zufällig ausgewählt. Dennoch hat sich die Anfälligkeit auf das schlechte Abschneiden der beiden Strategien ausgewirkt. Eine mögliche Anpassung ist die Forderung, dass in den initialisierenden Trainingsbeispielen alle vorkommenden Klassen vertreten sein müssen. Dies ist jedoch im hier untersuchten Szenario nicht vorgesehen.

6 Fazit und Ausblick

Dieser Beitrag hat untersucht, wie sich bei fester Wahl eines Klassifikators (Maximum-Likelihood) unterschiedliche Selektionsstrategien auswirken. Die Untersuchungen haben gezeigt, dass eine Reduktion der Trainingskosten unter Beibehaltung der Klassifikationsgüte nicht garantiert ist. Vielmehr muss das Zusammenspiel zwischen Selektionsstrategie und Klassifikator detailliert geprüft werden. Wird eine unpassende Strategie gewählt, kann dies zu einer Verschlechterung der Gesamtleistung führen. Im Falle eines Maximum-Likelihood-Klassifikators und normalverteilten Daten hat sich gezeigt, dass Random Sampling geeigneter ist als Uncertainty Sampling und Query by Committee. Die vorgestellten Anpassungen konnten die maximale Klassifikationsgüte nicht erhöhen, jedoch die Trainingskosten senken.

In zukünftigen Arbeiten ist zu untersuchen, wie sich die Selektionsstrategien bei Wahl von anderen Klassifikatoren (z. B. Nearest-Neighbor oder SVM) verhalten. Besonderes Augenmerk ist dabei auf den theoretischen Hintergrund zu richten und es ist zu prüfen, ob sich allgemeine Regeln und Zusammenhänge ergeben, die sich auf weitere Klassifikatoren übertragen lassen. Ebenso muss untersucht werden, welche Strategien und Klassifikatoren für bestimmte Aufgaben geeignet sind. Eine offene Fragestellung ist beispielsweise, ob eine geeignete Kombination zur Verarbeitung von Satellitenbildern mit drei Kanälen ebenso zur Verarbeitung von 7-Kanal-Bildern geeignet ist.

7 Literaturverzeichnis

BRINKER, KLAUS (2003): Incorporating Diversity in Active Learning with Support Vector Machines. In: Proceedings of the 20th ICML, S. 59–66.

BRUZZONE, LORENZO (2000): An approach to feature selection and classification of remote sensing images based on the Bayes rule for minimum cost. In: IEEE Transactions on Geoscience and Remote Sensing 38 (1), S. 429–438. DOI: 10.1109/36.823938.

- BRUZZONE, LORENZO; PRIETO, D.F. (2001): Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. In: *IEEE Transactions on Geoscience and Remote Sensing* 39 (2), S. 456–460. DOI: 10.1109/36.905255.
- COHN, DAVID; ATLAS, LES; LADNER RICHARD (1994): Improving generalization with active learning. In: *Machine Learning* (15), S. 201–221.
- DEMIR, BEGÜM; MINELLO, LUCA; BRUZZONE, LORENZO (2014): Definition of Effective Training Sets for Supervised Classification of Remote Sensing Images by a Novel Cost-Sensitive Active Learning Method. In: *IEEE Trans. Geosci. Remote Sensing* 52 (2), S. 1272–1284.
- DEMIR, BEGUM; BOVOLO, FRANCESCA; BRUZZONE, LORENZO (2011): Active-learning based cascade classification of multitemporal images for updating land-cover maps. In: *Analysis of Multi-temporal Remote Sensing Images*. 6th. International Workshop, S. 57–60.
- FOODY, GILES M. (2002): Status of land cover classification accuracy assessment. In: *Remote Sensing of Environment* 80 (1), S. 185–201. DOI: 10.1016/S0034-4257(01)00295-4.
- LEWIS, DAVID D.; CATLETT, JASON (1994): Heterogeneous Uncertainty Sampling for Supervised Learning. In: *Proceedings of the 11th ICML*, S. 148–156.
- PRATT, JOHN W. (1976): F. Y. Edgeworth and R. A. Fisher on the Efficiency of Maximum Likelihood Estimation. In: *Ann. Statist.* 4 (3), S. 501–514. DOI: 10.1214/aos/1176343457.
- ROY, NICHOLAS; MCCALLUM, ANDREW (2001): Toward Optimal Active Learning through Monte Carlo Estimation of Error Reduction. In: *Proceedings of the 18th ICML*.
- SAVAGE, LEONARD J. (1976): On Rereading R. A. Fisher. In: *Ann. Statist.* 4 (3), S. 441–500.
- SCHINDLER, KONRAD (2012): An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. In: *IEEE Trans. on Geosci. and Rem. Sensing* 50 S. 4534–4545.
- SETTLES, BURR (2009): Active Learning Literature Survey. University of Wisconsin, Madison (Computer Sciences Technical Report, 1648).
- SETTLES, BURR (2012): Active Learning. Morgan & Claypool Publishers (Synthesis Lectures on Artificial Intelligence and Machine Learning, 6).
- SEUNG, H. S.; OPPER, M.; SOMPOLINSKY, H. (1992): Query by committee. In: *Proceedings of the 5th COLT, ACM*. S. 287–294.
- XU, ZUOBING; AKELLA, RAM; ZHANG, YI (2007): Incorporating Diversity and Density in Active Learning for Relevance Feedback. In: *Proceedings of the 29th ECIR*. S. 246–257.
- ZHU, XIAOJIN (2005): Semi-Supervised Learning Literature Survey. University of Wisconsin-Madison (1530).
- ZHU, XIAOJIN; GOLDBERG, ANDREW B. (2009): Introduction to Semi-Supervised Learning: Morgan & Claypool Publishers (Synthesis Lectures on Artificial Intelligence and Machine Learning, 3).