

Entwicklung eines GDI-basierten Ansatzes zur Validierung klassifizierter Fernerkundungsdaten mithilfe von Crowdsourcing

Christoph Wiedekind, Stefan Wiemann und Pierre Karrasch
Technische Universität Dresden · stefan.wiemann@tu-dresden.de

Short paper

Zusammenfassung

Der vorliegende Beitrag beschäftigt sich mit dem Nutzungspotenzial von Crowdsourcing bei der Validierung von Fernerkundungsdaten, beziehungsweise daraus abgeleiteten Klassifikationsergebnissen, im Kontext von Geodateninfrastrukturen. Die Schwerpunkte liegen auf der Konzeption einer geeigneten Dienststruktur für einen möglichst automatisierten Validierungsprozess, der Beschreibung und Verarbeitung der verschiedenen Qualitätscharakteristika sowie einer technischen Realisierung auf Grundlage offener Standards. Als Anwendungsbeispiel wird die Validierung einer Landbedeckungs- und Landnutzungsklassifikation verwendet.

1 Einleitung

Die Nutzung fernerkundlicher Daten für ein operationelles und weitgehend automatisiertes Monitoring der Landbedeckung bzw. Landnutzung hängt unter anderem von der Güte durchgeführter Klassifikationen ab. Zur Überprüfung der Genauigkeiten können verschiedene Strategien herangezogen werden. Klassische Ansätze verfolgen dabei beispielsweise die Idee der Nutzung lokalen Wissens durch ausgewiesene Fachleute, die durch vereinzelte Feldbegehungen und mithilfe visueller Interpretation, mehr oder weniger zufällig ausgewählter Referenzpunkte, in der Lage sind, Klassifikationsgenauigkeiten abzuleiten. Andere Ansätze nutzen spektrale Bibliotheken, um auf Basis von Referenzsignaturen oder Spektralmessungen Klassifikationsergebnisse zu überprüfen. Gemein ist diesen Methoden der zum Teil enorme zeitliche und damit verbundene finanzielle Aufwand für die notwendige Qualitätskontrolle.

Auf der anderen Seite erlauben die Verbesserungen in der mobilen Positionsbestimmung und Datenerfassung, Teile der Bevölkerung in diese Arbeitsläufe zu integrieren. Unter dem Begriff der Citizen Science summieren sich verschiedene Methoden, von denen das Crowdsourcing gerade im Umweltmonitoring einen besonderen Stellenwert einnimmt. Neben der Nutzung in der jeweiligen Community und wissenschaftlichen Arbeiten, besteht auch auf der Verwaltungsebene von Schutzgebieten ein Interesse an diesen Daten (KARRASCH & WIEMANN 2014). Ein prominentes Beispiel für die Beteiligung der Bevölkerung an der

Validierung von Fernerkundungsdaten ist die *Geo-Wiki* Plattform¹, die primär auf die visuelle Interpretation von Bildern am Computer setzt (FRITZ et al. 2012). Feldbasierte Umweltbeobachtungen, insbesondere zu Flora und Fauna, können beispielsweise über die Plattformen Naturgucker², Artenfinder³ oder ornitho.de⁴ gemeldet werden. Dabei werden sowohl experten- als auch communitybasierte Verfahren eingesetzt, um Beobachtungen zu validieren. Nach einer Expertenvalidierung werden Beobachtungen aus dem Artenfinder überdies in das Landschaftsinformationssystem Rheinland-Pfalz überführt und erlangen somit administrativen Charakter.

Neben den kollaborativ gesammelten Daten im Web wird weiterhin der Aufbau behördlicher Geodateninfrastrukturen (GDI) vorangetrieben. Die INSPIRE (Infrastructure for Spatial Information in the European Community) Richtlinie (2007/2/EC) zielt dabei auf eine einheitliche und harmonisierte Bereitstellung von Umweltdaten auf europäischer Ebene und umfasst sowohl Daten zur Landbedeckung als auch zur Landnutzung. Auf technischer Ebene basiert die Umsetzung auf den Standards des Open Geospatial Consortium (OGC), die einheitliche Schnittstellen für die Visualisierung, Bereitstellung, Prozessierung und Registrierung von Geodaten beschreiben.

Der Fokus dieser Arbeit liegt auf in situ Beobachtungsdaten, die durch eine Crowd erhoben und für die Validierung einer Fernerkundungsklassifikation verwendet werden. Die Architektur auf Basis von OGC Standards ermöglicht die Anbindung an bestehende Geodateninfrastrukturen sowie die Adaption an weitere Anwendungsfälle.

2 Konzeption

Der grundlegende Informationsfluss und die Feedbackmechanismen für den angestrebten Validierungsprozess sind in Abb. 1 dargestellt. Fernerkundungsdaten können, dem Subsidiaritätsprinzip folgend, vom entsprechenden Anbieter über die OGC Web Coverage Service (WCS) Schnittstelle bereitgestellt werden. Ein geeignetes Klassifizierungsverfahren kann offline oder, sofern automatisierbar, über einen OGC Web Processing Service (WPS) erfolgen. Das Klassifikationsergebnis kann anschließend, je nach Format, über einen WCS oder Web Feature Service (WFS) bereitgestellt werden. Auf der anderen Seite können Client-Anwendungen für die in situ Beobachtungen den OGC GeoPackage Standard verwenden, um Beobachtungsdaten zu übertragen und einen interoperablen Zugriff darauf zu ermöglichen. Dieser ist speziell für mobile Anwendungen konzipiert und kapselt Geodaten auf Grundlage eines einheitlichen SQLite Datenbankschemas. Für die weitere Verwendung innerhalb der GDI können die Beobachtungen dann über eine WFS oder Sensor Observation Service (SOS) Schnittstelle zur Verfügung gestellt werden. Der zentrale Validierungsprozess kann über die jeweiligen OGC-Standards auf die Klassifikationsergebnisse und Beobachtungen zugreifen und wiederum über die WPS-Schnittstelle gekapselt werden. Auf die gleiche Weise können die Ausgangsdaten und Validierungsergebnisse in einem Fu-

¹ <http://www.geo-wiki.org>

² <http://www.naturgucker.de>

³ <http://artenfinder.de>

⁴ <http://ornitho.de>

sionsprozess genutzt werden, um Daten zu verdichten, anzureichern, zu aktualisieren oder weiter statistisch zu analysieren (WIEMANN & BERNARD 2014).

Das Feedback aus dem Validierungsprozess kann an drei verschiedenen Stellen genutzt werden. Zunächst können Beobachtungsdaten mit Informationen zum Klassifikationsergebnis an der jeweiligen Position angereichert werden. Auf die gleiche Weise kann das Klassifikationsergebnis mit existierenden Beobachtungen verknüpft werden. In regelmäßigen Abständen und abhängig von der Anzahl verfügbarer Beobachtungen kann das Validierungsergebnis auch zur Reanalyse und Anpassung des Klassifikationsverfahrens genutzt werden.

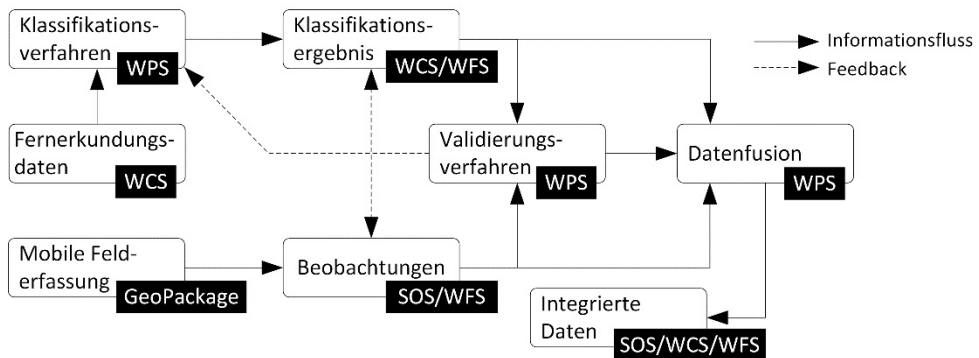


Abb. 1: Informationsfluss zur Validierung und Fusion klassifizierter Fernerkundungsdaten mit Beobachtungsdaten; Schnittstellen basierend auf OGC-Standards

Um Prozesse sinnvoll in einer GDI anbieten zu können, sollten diese möglichst automatisierbar sein, um Flexibilität und Wiederverwendbarkeit zu gewährleisten. Da automatisierte Validierungsverfahren insbesondere bei heterogen strukturierten Daten, wie sie beim Crowdsourcing in der Regel vorliegen, nur bedingt zuverlässig sind, sollte theoretisch jeder automatisierten Validierung auch eine manuelle Expertenprüfung nachgeschaltet werden.

Die zentrale Frage in dem Validierungsprozess ist die Erfassung, Bewertung und Verarbeitung der Datenqualität. Dies muss an mehreren Stellen innerhalb des Workflows erfolgen und in den Metadaten beschrieben und mitgeführt werden. Auf die interne Validierung verschiedener Klassifikationsverfahren für Fernerkundungsdaten wird im Weiteren nicht weiter eingegangen, da der Fokus auf den Beobachtungsdaten des Crowdsourcing liegt. Diese durchlaufen im Wesentlichen die folgenden drei Schritte (Abb. 2):

1. *Interne Validierung* nutzt Informationen der einzelnen Beobachtung, um grundlegende Konsistenz- und Plausibilitätsprüfungen durchzuführen. Dabei können auch Vollständigkeit und Konformität bezüglich eines vorgegebenen Beobachtungsschemas geprüft werden.
2. *Gegenseitige Validierung* vergleicht Beobachtungen untereinander und ermöglicht die Ermittlung eines Konfidenzmaßes für einzelne Beobachtungen und eine Ausreißerdetektion. Eine Gewichtung kann anhand verschiedener Faktoren, wie etwa der Anzahl übereinstimmender Beobachtungen in der näheren Umgebung, Genauigkeit der Positionsbestimmung oder Evidenz durch zeitgleich erstellte Fotos, erfolgen. Zudem kön-

nen Nutzerprofile in die Bewertung einfließen, um zusätzlich anhand der Nutzererfahrung zu gewichten.

3. *Externe Validierung* kann zunächst eine Filterung geeigneter Beobachtungen, beispielsweise anhand des Zeitstempels beinhalten, um die grundsätzliche Vergleichbarkeit der Eingangsdaten zu gewährleisten. Wie in Abbildung 2 dargestellt, kann eine externe Validierung zur qualitativen Verbesserung beider Eingangsdatensätze dienen, indem als ungenau eingestufte Daten entsprechend gekennzeichnet oder gefiltert werden.

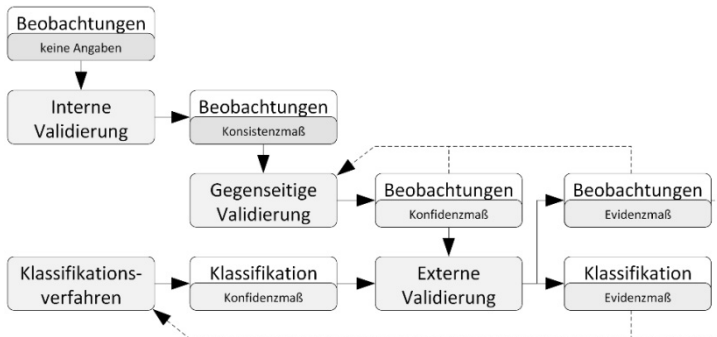


Abb. 2: Schematische Entwicklung der Qualitätsmerkmale für Beobachtungen aus Crowdsourcing und Klassifikationsergebnisse durch verschiedene Validierungsprozesse

3 Implementierung

Die Implementierung setzt das vorgestellte Konzept prototypisch um und verfolgt das Ziel, Daten zur Landbedeckung über nutzergenerierte Beobachtungen zu validieren. Zu Testzwecken werden Landsat-8-Szenen klassifiziert und über den WCS bereitgestellt. Die Landbedeckungsklassifikation erfolgt beispielhaft über eine unüberwachte Klassifikation in der Software GRASS GIS⁵. Dabei werden die Level-1-Typen der CORINE-Nomenklatur (Coordination of Information on the Environment) abgeleitet: (1) Artificial Surface, (2) Agricultural Area, (3) Forest Semi Natural Area, (4) Wetlands und (5) Waterbodies. Da derzeit noch keine Feldbeobachtungen für das Gebiet vorhanden sind, werden diese in der Testphase mithilfe von CORINE Land Cover Daten simuliert. Über ein stratified random sampling können Beobachtungen generiert, und wiederum den fünf oben genannten Klassen zur Landbedeckung zugeordnet werden. Diese Beobachtungen werden anschließend über einen WFS bereitgestellt.

Die prototypische Implementierung der Dienstarchitektur folgt offenen Standards und basiert ausschließlich auf Open Source Software. Eingesetzt wurden *GeoServer*⁶ zur Umsetzung OGC-konformer Dienste (WFS, WCS und WPS), *PostGIS*⁷ zur Speicherung von Beobachtungen sowie *Leaflet*⁸ zur clientseitigen Visualisierung. Die Applikation ist bei-

⁵ <http://grass.osgeo.org>

⁶ <http://geoserver.org>

⁷ <http://postgis.net>

⁸ <http://leaflet.org>

spielhaft in Abbildung 3 dargestellt. Über die farbliche Codierung kann festgestellt werden, ob die Klassifikation der Fernerkundungsdaten und der (simulierten) Beobachtungsdaten übereinstimmt (grün = Übereinstimmung, rot = keine Übereinstimmung). Zudem können Eigenschaften der Beobachtungen sowie die zugeordnete Landbedeckung abgefragt werden. Im Falle von dicht beieinanderliegenden Beobachtungen, werden diese visuell aggregiert und erst in einer höheren Zoomstufe einzeln angezeigt.

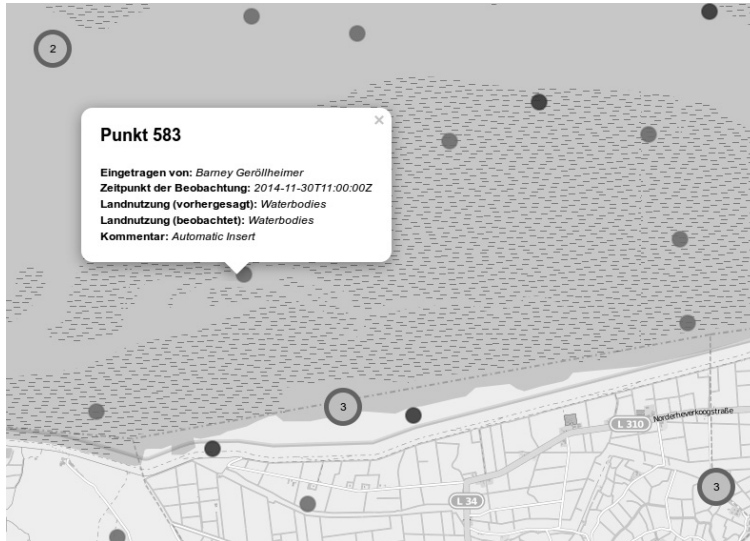


Abb. 3: Client-Applikation zum interaktiven Vergleich von Klassifikations- und Beobachtungsdaten (Hintergrundkarte: OpenStreetMap)

Die Anwendung kann über die genannten OGC-Schnittstellen auf beide Ausgangsdatensätze zugreifen und ermöglicht eine Genauigkeitsanalyse auf Grundlage einer Konfusionsmatrix (Abb. 4) sowie daraus abgeleiteten Genauigkeitsmaßen wie der *user's accuracy*, der *producer's accuracy* oder der *portmanteau accuracy* (vgl. COMBER et al. 2013). Zudem können Kappa-Werte berechnet werden, um eine Aussage zur Signifikanz des Ergebnisses zu treffen.

Klassifikation	Beobachtung						Total	User's Accuracy
	Artificial Surfaces	Agricultural Area	Forest Semi Natural Area	Wetlands	Waterbodies			
Artificial Surfaces	83	33	4	5	0	125	66.4 %	
Agricultural Area	34	210	6	1	3	254	82.68 %	
Forest Semi Natural Area	29	36	50	9	1	125	40 %	
Wetlands	9	10	1	126	104	250	50.4 %	
Waterbodies	0	0	1	30	225	256	87.89 %	
Total	155	289	62	171	333	1010		
Producer's Accuracy	53.55 %	72.66 %	80.65 %	73.68 %	67.57 %		68.71 %	

Abb. 4: Erzeugte Konfusionsmatrix auf Grundlage von Beobachtungen und Landnutzungs-klassifikation

Die vorgestellte Implementierung demonstriert die technische Machbarkeit des Konzeptes und liefert daher keine Aussage zur Genauigkeit der verwendeten Eingangsdaten. Zukünftig ist geplant, die Anwendung im Kontext des EU FP7 Projektes COBWEB⁹ (*Citizen Observatory Web*) für bestehende Klassifikationsergebnisse und echte Crowdsourcing-Daten zu verwenden.

4 Fazit

Das vorgestellte Konzept und die prototypische Implementierung stellen einen OGC-konformen technischen Unterbau für die Validierung von klassifizierten Fernerkundungsdaten mithilfe von Crowdsourcing dar. Der Vorteil gegenüber existierenden Verfahren ist die standardkonforme Umsetzung und Verwendbarkeit im Kontext von INSPIRE und existierenden GDI.

Die prototypische Umsetzung kann zur Validierung und Bewertung einer Fernerkundungsklassifikation eingesetzt werden. Diese Informationen können darüber hinaus in Form einer Reanalyse wiederum genutzt werden, um das Klassifikationsverfahren und Klassifikationsergebnis entscheidend zu verbessern.

Um regelmäßig und möglichst gleich verteilt Beobachtungsdaten mittels Crowdsourcing zu generieren und einen echten Mehrwert zu schaffen, ist es notwendig, Teilnehmer möglichst nachhaltig für ein Projekt, in diesem Fall das Monitoring der Landbedeckung bzw. Landnutzung, zu motivieren. Eine entsprechende Citizen-Science-Strategie muss auf potenzielle Teilnehmer, das Untersuchungsgebiet und den Anwendungsfall zugeschnitten werden.

Literatur

- COMBER, A., SEE, L., FRITZ, S., VAN DER VELDE, M., PERGER, C. & FOODY, G. (2013), Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation*, 23, 37-48.
- FRITZ, S., MCCALLUM, I., SCHILL, C., PERGER, C., SEE, L., SCHEPASCHENKO, D., VAN DER VELDE, M., KRAXNER, F. & OBERSTEINER, M. (2012), Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31, 110-123.
- KARRASCH, P. & WIEMANN, S. (2014), Evaluierung des Nutzungspotentials von Crowdsourcing in Schutzgebietsverwaltungen. In: STROBL, J., BLASCHKE, T., GRIESEBNER, G. & ZAGEL, B. (Hrsg.), *Angewandte Geoinformatik 2014. Beiträge zum 26. AGIT-Symposium Salzburg*. Wichmann Verlag.
- WIEMANN, S. & BERNARD, L. (2014), Linking crowdsourced observations with INSPIRE. In: HUERTA, J., SCHADE, S. & GRANELL, C. (Eds.), *Connecting a Digital Europe through Location and Place. Proceedings of the 17th AGILE International Conference on Geographic Information Science, Castellón, 2014*.

⁹ <https://cobwebproject.eu>