

CloudServiceMarket: Klassifizieren und Bewerten von Cloud Computing Services auf der Basis von Künstlichen Neuronalen Netzen

Frank Teuteberg, Benedikt Martens

Universität Osnabrück, Unternehmensrechnung und Wirtschaftsinformatik
Katharinenstr. 1, 49078 Osnabrück, frank.teuteberg@uos.de|benedikt.martens@uos.de

Abstract: Im Rahmen dieses Beitrags wird die Konzeption, prototypische Umsetzung und Evaluation einer (semi-)automatischen Klassifizierung und Bewertung von Cloud Computing Services auf der Basis Künstlicher Neuronaler Netze (KNN) für CloudServiceMarket beschrieben. CloudServiceMarket ist eine Marktplattform, die Cloud Computing Services beschreibt und das Ziel verfolgt, die Markttransparenz im Bereich Cloud Computing zu erhöhen. Die bisher nicht einheitliche Namensgebung und fehlende Standardisierungen für Cloud Computing Services erschweren jedoch derzeit noch die Identifizierung und Bewertung geeigneter Cloud Computing Services auf Nachfragerseite. Die im Rahmen dieses Beitrags vorgestellten IT-Artefakte (die Marktplattform Cloud-ServiceMarket sowie ein Künstliches Neuronales Netz zum Klassifizieren und Bewerten von Cloud Computing Services) können Abhilfe schaffen.

1 Einleitung und Motivation

Anwenderunternehmen steht eine unüberschaubare Menge an Cloud Computing Services zur Verfügung, sodass eine gezielte Suche, Bewertung und Klassifikation von Cloud Computing Services auf dem weitgehend unstrukturierten und dynamischen Cloud Computing Markt sehr aufwendig wird. Um Informationen über die verschiedenen Services zu erhalten, ist mitunter eine aufwendige Marktanalyse notwendig. Eine solche Recherche, selbst wenn sie gründlich durchgeführt wird, enthält vermutlich nur einen Ausschnitt der vorhandenen und passenden Angebote, da unter anderem eine inkonsistente Namensgebung die Inhalte der Angebote unscharf werden lässt. Außerdem mangelt es an einer neutralen und ausführlichen Evaluation der Anbieter und Services. Aufgrund der hohen Anzahl an verfügbaren Angeboten an Cloud Computing Services ist es für Unternehmen und Privatanwender daher sehr kostenintensiv, sich einen Überblick über den derzeitigen Cloud Computing Markt zu verschaffen und damit Cloud Computing Services zu identifizieren, die den jeweiligen Unternehmensanforderungen genügen. Der in diesem Beitrag vorgestellte CloudServiceMarket (<http://www.cloudservicemarket.info/>) sowie der vorgestellte und prototypisch implementierte Ansatz, relevante Cloud Computing Services auf der Basis von Künstlichen Neuronalen Netzen (KNN) zu identifizieren und zu bewerten, sollen Abhilfe schaffen. Beide IT-Artefakte können es Unternehmen ermöglichen, relevante Cloud Computing Services zu identifizieren, Bewertungen abzugeben und Bewertungen von anderen Anwendern zu den jeweils in der Datenbank von CloudServiceMarket

erfassten Services einzusehen. Die Bewertungen basieren hierbei auf den Präferenzen der Anwender, da die KNN mit diesen Präferenzprofilen durch sogenanntes überwachtes Lernen trainiert werden. Den Untersuchungen in diesem Beitrag liegt ein Such-, Bewertungs-, und Klassifikationsproblem zugrunde, das sich bereits beim Aufbau der vorgestellten Datenbank von CloudServiceMarket stellte. In dieser Datenbank werden vor allem solche Angebote aufgenommen, die aus betriebswirtschaftlicher Sicht interessant sind und einen innovativen Charakter aufweisen. Die Datenbankeinträge entstanden aufgrund manueller und aufwendiger Bewertung jedes Cloud Computing Services. Eine (semi-)automatisierte Bewertung und ein (semi-)automatisches Identifizieren von innovativen Cloud Computing Services sind daher wünschenswert. Der Beitrag ist wie folgt aufgebaut: In Abschnitt 2 werden zunächst in der gegebenen Kürze die notwendigen Grundlagen zum Cloud Computing und zu KNN vorgestellt. Abschnitt 3 geht auf die verwendete Forschungsmethodik ein. Das Design und die prototypische Implementierung von CloudServiceMarket und dessen Datenbank werden in Abschnitt 4 vorgestellt. In Abschnitt 5 wird schließlich der prototypisch implementierte Ansatz zur (semi-)automatisierten Bewertung von Cloud Computing Services auf der Basis von KNN erläutert. Erste experimentelle Ergebnisse der Anwendung des entwickelten KNN und daraus abgeleitete Handlungsempfehlungen sind Gegenstand von Abschnitt 6. Im letzten Abschnitt 7 werden die wesentlichen Ergebnisse für Wissenschaft und Praxis zusammengefasst und sowohl Limitationen als auch mögliche Ansatzpunkte für zukünftige Forschung aufgezeigt.

2 Grundlagen

2.1 Cloud Computing

Aufgrund der schnellen Entwicklung des Cloud Computing erhöht sich die Anzahl der angebotenen Cloud Computing Services fortwährend. Als mittlerweile in Wissenschaft und Praxis weit verbreitete und anerkannte Definition gilt die des National Institute of Standards and Technology (NIST), die Cloud Computing wie folgt definiert ([MG09]):

„Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management *effort or service provider interaction.*”

Es wird weiterhin zwischen drei Dienstmodellen unterschieden ([OBM10], [RWZ11], [MT11]): Software as a Service (SaaS): ermöglicht die Nutzung von Anwendungen über Thin Clients, wie z. B. mittels eines Browsers; Platform as a Service (PaaS): stellt Entwicklungs- und Laufzeitumgebungen für diverse Programmiersprachen zur Verfügung und Infrastructure as a Service (IaaS): stellt Infrastrukturressourcen, wie z. B. Rechenleistung oder Speicherplatz, über das Internet zur Verfügung. Die Art der Bereitstellung unterscheidet sich in ([MG09]): Private Clouds, die nur für eine einzelne Organisation Leistungen erbringen; Community Clouds, in denen die Ressourcen der Cloud von mehreren Organisationen mit bestimmten Gemeinsamkeiten geteilt werden; Public Clouds, wobei die Dienste für die breite Öffentlichkeit verfügbar sind und schließlich Hybrid Clouds, die einen Zusammenschluss von mindestens zwei Clouds

unterschiedlicher Bereitstellungsarten darstellen. Besonders bei Public Clouds sollte bei der Auswahl eines Anbieters großes Augenmerk auf die verwendeten Sicherheitsmaßnahmen gelegt werden ([Wei09]).

Um einen Überblick über die Anbieter einer bestimmten Serviceart zu erhalten, ist eine entsprechend intensive Recherche notwendig. So wird beispielsweise für einen Infrastrukturservice der Datenspeicherkapazität bereitstellt nicht nur der Begriff „Storage Service“, sondern auch „Data Storage as a Service“ verwendet (vgl. [YBD08] und [Wei09]). Weiterhin ist die Qualität der angebotenen Services nicht direkt erkennbar und eine ganzheitliche, alle Perspektiven und Präferenzen der Anwender umfassende Beurteilung der jeweiligen Anbieter derzeit nur auf der Basis der von den Anbietern selbst bereitgestellten Informationen möglich. Die nachfolgenden Initiativen und Standardisierungsbemühungen existieren bereits und bieten als verwandte Arbeiten zu dem hier vorgestellten CloudServiceMarket erste Unterstützung:

- www.eurocloud.org: Zusammenschluss von Cloud Computing Anbietern innerhalb Europas (Branchenverband);
- www.cloudsecurityalliance.org: Ziel ist die Erhöhung der Sicherheit von Cloud Computing Services;
- www.dmtf.org (Distributed Management Task Force): Entwicklung des Cloud-Standards Incubator, der die Standardisierung von Protokollen für die Interaktion von Cloud Computing Umgebungen fördern soll;
- www.snia.org/forums/csi/ (Storage Networking Industry Association; Cloud Storage Initiative): Non-Profit-Organisation mit dem Ziel der Verbesserung von Methoden und Technologien im Bereich Datenspeicherung und -management; die offene Schnittstelle „Cloud Data Management Interface“ wurde im Frühjahr 2010 veröffentlicht;
- www.deltacloud.org: eine Initiative von Redhat, die es ermöglicht, mit einer API mehrere Cloud Computing Services in die Unternehmens-IT einzubinden.
- www.cloudcommons.com: ein Marktplatz zum Kaufen und Verkaufen von Cloud Computing basierten Applikationen.

Die vorgestellten Initiativen erlauben jedoch keine (semi-)automatische Klassifizierung und Bewertung von Cloud Computing Services, so wie es das Forschungsziel dieses Beitrags ist.

2.2 Künstliche Neuronale Netze

Künstliche Neuronale Netze (KNNs) werden verstärkt für Klassifikationsprobleme eingesetzt, da sie im Unterschied zu anderen Klassifikationsverfahren (z. B. Logistische Regression, Clusterverfahren) in der Lage sind, aus verrauschten Daten zu lernen sowie zu generalisieren [Bi95]. Des Weiteren sind KNN in der Lage sich an Veränderungen in Datenbeständen bzw. Datenmustern anzupassen, indem sie im Rahmen einer Trainingsphase mit den neuen Datenmustern ihre bisherigen Verbindungsgewichte zwischen einzelnen Neuronen anpassen. Seit der Entwicklung des Perceptron durch Rosenblatt Ende der 50er Jahre wurden verschiedene Netzwerktopologien vorgestellt [B97]. Da es eine große Zahl von Topologien gibt, musste für das hier vorgestellte

Klassifikations- und Bewertungsproblem eine Auswahl getroffen werden. Untersucht wurden das Multilayer Perceptron (MLP), Modulare Netzwerke, sowie Generalized-Feedforward-Netzwerke. Die Wahl fiel letztendlich auf das MLP, da es in der Lage ist, nicht nur binäre (0, 1), sondern auch kontinuierliche Ein- und Ausgabewerte zu verarbeiten und zu erzeugen. Kontinuierliche Werte spielen einerseits eine Rolle, weil verschiedene Merkmale, anhand derer Cloud Computing Services klassifiziert bzw. bewertet werden, nicht binär gemessen werden können (z. B. Anzahl der Zertifikate oder Grad der Verfügbarkeit des Cloud Computing Services in Prozent (z. B. durchschnittlich 99,6 % Verfügbarkeit in 24 Stunden)). Zum andern erlauben kontinuierliche Werte beim Netzoutput eine differenzierte Interpretation der Ergebnisse. Ein einfaches Multilayer Perceptron (MLP) besteht aus drei Schichten (vgl. Abbildung 1), einer Eingabeschicht, einer verdeckten Schicht und einer Ausgabeschicht, wobei die Anzahl der verdeckten Schichten variieren kann. Bevor das MLP für eine bestimmte Klassifikationsaufgabe eingesetzt werden kann, muss es trainiert werden. Hierzu werden verschiedene Merkmalsvektoren, welche die Cloud Computing Services repräsentieren, an der Eingabeschicht angelegt. Die Eingabewerte sind technologische und strukturelle Merkmale der Cloud Computing Services (vgl. Abschnitt 4). Diese Eingabewerte werden in den verdeckten Schichten verarbeitet und an die Ausgabeschicht propagiert. Mathematisch lässt sich zeigen, dass MLPs mit zwei oder mehr verdeckten Schichten in der Lage sind, jedes Klassifikationsproblem zu lösen [Kh90, S. 70 ff.]. Die Stärke der Verbindungen zwischen diesen Neuronen kann positiv, negativ oder neutral sein. Das Wissen eines neuronalen Netzes ist entsprechend in seinen Gewichten gespeichert. Ein KNN lernt folglich, wenn Gewichte zwischen den Neuronen verändert werden.

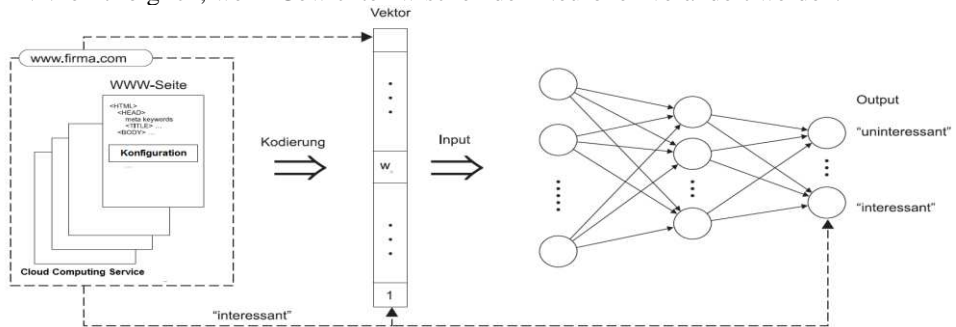


Abbildung 1: Trainieren KNN durch Präsentation von Merkmalsvektoren [KST00, KST98]

3 Forschungsmethodik

Der vorliegende Beitrag verwendet einen Design Science Ansatz [He07, HMP04, MS95] und beschreibt die Konzeption, prototypische Umsetzung als auch Evaluation einer (semi-)automatischen Klassifizierung und Bewertung von Cloud Computing Services auf der Basis von KNN. Tabelle 1 gibt einen Überblick über sieben Richtlinien zur Design Science Forschung [OO10] und zeigt auf, inwieweit diese Anforderungen durch den vorliegenden Beitrag erfüllt werden.

| Nr. | Beschreibung | Instanziierung/Berücksichtigung im Rahmen dieses Beitrags |
|-----|-------------------------------|---|
| 1 | Gestaltung als Artefakt | Die hier vorgestellten Prototypen (CloudServiceMarket sowie der Prototyp zur Bewertung von Cloud Computing Services) sind typische Design Science Artefakte [OO10]. |
| 2 | Problemrelevanz | Die Problemrelevanz wurde in Abschnitt 1 motiviert. |
| 3 | Gestaltungsevaluation | Die Evaluation des Artefakts erfolgt im Rahmen dieses Beitrags auf der Basis etablierter Fehlermaße zur Evaluierung von KNN. |
| 4 | Wissenschaftlicher Beitrag | Der Beitrag erweitert den Stand der Wissenschaft um ein Konzept zur (semi-)automatischen Klassifizierung und Bewertung von Cloud Computing Services auf Basis von KNN. Darüber hinaus werden Ansatzpunkte für zukünftige Forschung aufgezeigt. |
| 5 | Wissenschaftliche Rigorosität | Das Vorgehen in diesem Beitrag folgt etablierten Vorgehensmodellen, Richtlinien und Empfehlungen der Design Science Forschung [He07,HMP04,MS95,OO10]. |
| 6 | Gestaltung als Suchprozess | Im Rahmen dieses Beitrags wird ein Iterationszyklus des Vorgehensmodells durchlaufen. Weitere Zyklen aus Konzeption, Umsetzung und Evaluation sollen zukünftig folgen. |
| 7 | Kommunikation der Forschung | Sowohl die hier vorgestellten Forschungsergebnisse als auch die Ergebnisse zukünftiger Iterationen des Vorgehensmodells sollen auf wissenschaftlichen Konferenzen und in wissenschaftlichen Zeitschriften veröffentlicht werden. Die Forschungsergebnisse werden zudem unter www.cloudservicemarket.info veröffentlicht und zur weiteren Evaluation der interessierten Community zur Verfügung gestellt. |

Tabelle 1: Richtlinien für Design Science Forschung [OO10]

4 (Semi-)automatisches Bewerten von Cloud Computing Services

4.1 Community-Plattform CloudServiceMarket

Die von den Autoren dieses Beitrags entwickelte Community-Plattform CloudServiceMarket.info (www.cloudservicemarket.info) beinhaltet derzeit 207 Dienste von 179 Anbietern (Mai 2012), die anhand einiger Leistungsparameter, wie bspw. Dienstmodellen, Zertifizierungen, Compliance-Regelungen oder Standorte der Rechenzentren erfasst sind. Mithilfe zweier Suchmasken (Anbieter und Dienste), kann das erfasste Angebot durchsucht werden. Für jeden Anbieter und jeden Dienst steht eine eigene Seite zur Verfügung auf der Informationen eingesehen werden können. Darüber hinaus kann jeder registrierte Benutzer Bewertungen für sowohl Anbieter als auch Dienste verfassen, mit denen er bereits Erfahrungen gesammelt hat. Diese Bewertungen können entweder in einfacher Form mit einem Kommentar und einer Bewertung auf einer Skala von eins („keine Empfehlung“) bis fünf („Empfehlung“) oder ausführlich mit Kommentaren und Bewertungen zu sieben Kategorien, wie z. B. Support/Kommunikation mit Anbieter, Schnittstellen oder Kostenübersicht/-kontrolle,

vorgenommen werden. Da diese Bewertungen subjektiven Eindrücken unterliegen und nur mit hohem Aufwand auf ihren Wahrheitsgehalt überprüft werden können, wird es Benutzern in einer zukünftigen Version von CloudServiceMarket.info möglich sein die Bewertungen anderer Benutzer wiederum zu kommentieren und so Feedback zu der erwähnten Kritik zu geben. Des Weiteren können Benutzer neue Dienste und Anbieter über entsprechende Masken einsenden. Darüber hinaus werden Informationen zu Forschern, Forschungsprojekten und Studien zum Thema Cloud Computing gesammelt.

4.2 Bewertungsmerkmale

Abbildung 2 zeigt Merkmale in einer ER-Notation, die zur Bewertung der Cloud Computing Services auf der Basis von KNNs herangezogen werden.

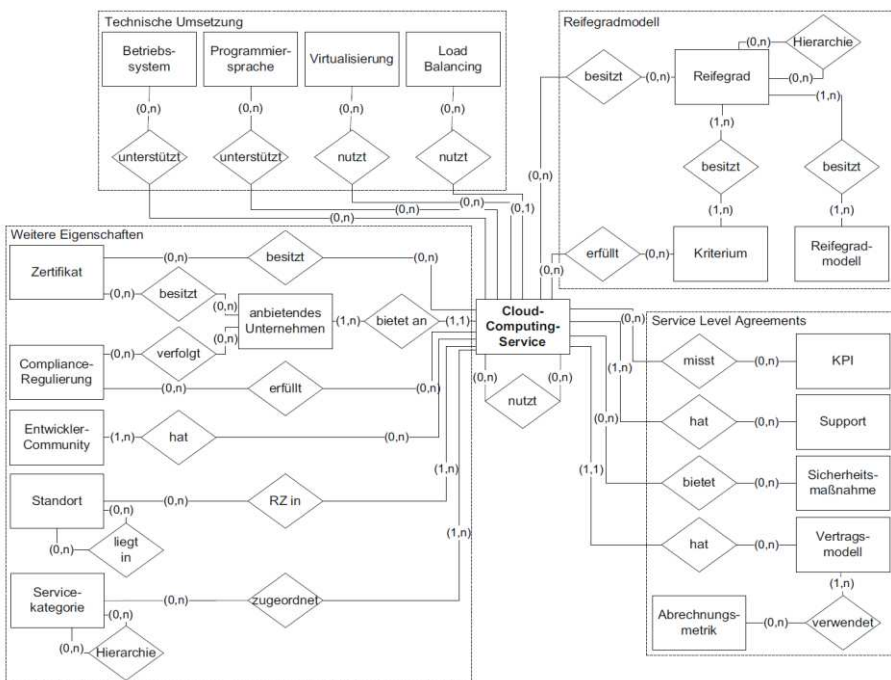


Abbildung 2: Bewertungsmerkmale (ER-Modell)

Eine Kurzbeschreibung der herangezogenen Bewertungsmerkmale aus den Bereichen 1) Standards & Rahmenwerke, 2) Compliance, 3) Schnittstellen, 4) Datensicherheit, 5) Netzwerksicherheit, 6) Technische Sicherheit, 7) Ausfallsicherung sowie 8) Weitere Merkmale ist unter der folgenden URL abrufbar: www.uwi.uni-osnabrueck.de/bewertungsmerkmale.pdf. Die Merkmale wurden im Rahmen einer explorativen Studie und auf der Basis von Experteninterviews ermittelt sowie in einer Analyse der Beschreibungen von Cloud Computing Services identifiziert. Merkmale werden binär (0/1) codiert. Wenn das entsprechende Merkmal vorhanden ist (z. B. redundante Stromversorgung) dann wird dieses Merkmal mit 1 anderenfalls mit 0

codiert. Der vorhandene Reifegrad wird durch den maximalen Reifegrad geteilt. Damit ist sichergestellt, dass auch hier die Werte zwischen 0 und 1 liegen. Bei den Bezahlverfahren werden für eine Flatrate der Wert 0,3, für leistungsbezogen der Wert 0,7 und sofern beides möglich ist der Wert 1 vergeben. Die leistungsbezogene Abrechnung beim Cloud Computing wird höher bewertet, da sie eine flexiblere, der tatsächlich in Anspruch genommenen Leistung entsprechende Bezahlung ermöglicht. Die Kombination aus beiden Bezahlverfahren erhält mit dem Wert 1 bspw. die beste Empfehlung. Die Bewertungen/Codierungen wurden im Rahmen eines Workshops von insgesamt vier Personen vorgenommen, wobei nicht ganz eindeutige Zuordnungen jeweils diskutiert wurden, sodass letztendlich jeweils ein Konsens gefunden werden konnte. Die Mehrzahl der analysierten Web-Seiten der Cloud Computing Service Anbieter sind vergleichsweise unübersichtlich, und Informationen sind auf den Seiten nur schwer zu extrahieren. Teilweise war die Leistungsbeschreibung nicht online bzw. Service Level Agreements nicht aufgeführt. Es gibt keine (gesetzlichen) Vorgaben, was genau in einer Servicebeschreibung zu stehen hat. Entsprechende Standards sind daher wünschenswert. Tabelle 2 zeigt die vergleichsweise geringe Informationsqualität von 207 untersuchten Cloud Computing Service Beschreibungen auf den Web-Seiten der Anbieter wieder.

| Informationsqualität (Level) | % (n =207) |
|--|-------------------|
| 1: Marginale Informationsbereitstellung | 20,47 % |
| 2: Informationsbereitstellung zu Zertifikaten | 46,78 % |
| 3: Beschreibung von SLAs | 23,98 % |
| 4: Umfassende Informationsbereitstellung (Zertifikate, SLAs, Referenzen, Statistiken zu Verfügbarkeiten, Einhaltung von (Sicherheits-)Standards, etc.) | 8,77 % |

Tabelle 2: Informationsqualität der Servicebeschreibungen

Informationen über den jeweils angebotenen Cloud Computing Service für detailliertere Beschreibungen in CloudServiceMarket konnten zum Teil nachträglich nur sehr zeitaufwendig über persönlichen Kontakt (per E-Mail oder Telefon) zum Anbieter eingeholt werden.

5 Prototypische Implementierung

Zur Implementierung des Künstlichen Neuronales Netzes und des Graphical User Interfaces (GUI) zum KNN wurde Eclipse Java EE IDE für Web Entwickler in der Version Helios (3.6) verwendet.

```

<?xml version="1.0"?>
- <WebApi.CCS__attributes>
  <serviceName>NEWCCS</serviceName>
  <serviceProvider>UWI</serviceProvider>
  <laas>0.0</laas>
  <paas>0.0</paas>
  <saas>1.0</saas>
  <headquarterEurope>1.0</headquarterEurope>
  <headquarterUsa>0.0</headquarterUsa>
  <headquarterOther>0.0</headquarterOther>
  <sla>0.4</sla>
  <scale>0.5</scale>
  <paying>0.3</paying>
  <sas70I>0.0</sas70I>
  <sas70II>0.0</sas70II>
  <iso9000>0.0</iso9000>
  <iso27001>1.0</iso27001>
  <pci>0.0</pci>
  <safeHarbor>0.0</safeHarbor>
  <fisma>1.0</fisma>
  <hipaa>0.0</hipaa>
  <sox>1.0</sox>
  <itil>0.0</itil>
  <rest>0.0</rest>
  <json>0.0</json>
  <soap>1.0</soap>
  <xml>0.0</xml>
  <ssl>0.0</ssl>
  <connectivity>0.9899999999999999</connectivity>
  <usvy>0.0</usvy>
  <internetProvider>0.3535533905932738</internetProvider>
  <dataRedundancy>0.816496580927726</dataRedundancy>
  <powerRedundancy>0.5773502691896257</powerRedundancy>
  <coolingRedundancy>0.816496580927726</coolingRedundancy>
  <firewall>1.0</firewall>
  <hardwareFirewall>1.0</hardwareFirewall>
  <intrusionDetection>0.0</intrusionDetection>
  <antivirus>1.0</antivirus>
  <supportchat>0.0</supportchat>
  <physicalSecurity>0.0</physicalSecurity>
  <grammlLeachBliley>1.0</grammlLeachBliley>
  <informationContent>1.0</informationContent>
  <resultKnn>0.6972369463568715</resultKnn>
</WebApi.CCS__attributes>

```

Abbildung 3: Bewertungsmerkmale und Codierung (XML-Darstellung)

Als Webserver diente lokal ein Apache Tomcat 7 Server. Dieser beinhaltet auch das notwendige Java-Archiv `servlet-api.jar` um Java Servlets ausführen zu können. Um ein Servlet auszuführen wird zusätzlich auch eine `web.xml` benötigt, welche das Servlet kurz beschreibt. Zur Darstellung des GUI wird eine Stylesheet-Datei `style.css` verwendet.

Das KNN wurde mit Hilfe von Neuroph (<http://neuroph.sourceforge.net/download.html>) in der Version 2.5.1 entwickelt. Neuroph [Ne11] ist ein Java Framework zur Entwicklung von KNN. Hierfür werden drei externe Java-Archive benötigt: `neuroph-2.5.1.jar`, `Neuroph 2.5.1` sowie `encog-core-2.5.3.jar`. Neuroph wurde gewählt, da es neben dem vollständigen Java-Quellcode, Binaries, Javadoc, den Klassendiagrammen und einem Getting Started Tutorial auch einige Beispiele zu Künstlichen Neuronalen Netzen umfasst. Da Neuroph die Programmiersprache Java verwendet ist zudem ein Zugriff über das Web möglich. XStream Version 1.3.1 diente schließlich zum Import/Export ins XML-Format und benötigt die Archive `xstream-1.3.1.jar` und `xpp3_min-1.1.4c.jar`. Abbildung 3 zeigt ein XML-File, welches in codierter Form (binär und kontinuierlich im Intervall von 0 bis 1) die Merkmale zur Bewertung eines Cloud Computing Services darstellt. Auf Basis der Merkmals-Wert-Kombinationen hat das zugrundeliegende KNN eine Bewertung mit ca. 0,69 vorgenommen (ca. 70 %). Die bestmögliche Bewertung beträgt 100 %. Das KNN konnte somit in diesem Fall nur eine eingeschränkte Empfehlung für den betreffenden Cloud Computing Service abgeben. Ein Klassendiagramm in der Notation UML ist unter folgender URL abrufbar:

www.uwi.uni-osnabrueck.de/klassendiagramm.pdf. Die Klasse CCS_attributes dient zum Auslesen der Bewertungsmerkmale sowie zur Repräsentation der Merkmalsvektoren in XML (vgl. Abbildung 3). Mit Hilfe der Klasse CloudGUI wird das GUI erzeugt (vgl. Abbildung 4). Die Klasse DBtransfer dient zum Speichern und Laden der Merkmalsvektoren. Die Klasse Codierung unterstützt die Codierung und Transformation der Merkmalswerte als binäre Werte sowie als kontinuierliche Werte im Intervall [0,1]. Die Klasse initNeuro erzeugt ein KNN zum Trainieren und Testen. Die Klasse WebAPI wird schließlich zur Initialisierung des GUI benötigt. Abbildung 4 zeigt das GUI zum KNN. Der Anwender hat über das GUI die Möglichkeiten KNN zu erstellen, zu trainieren und zu testen, Fehlerraten zu prüfen sowie die entsprechenden Merkmale eines Cloud Computing Services einzutragen, zu dem eine Empfehlung abgegeben werden soll.



Abbildung 4: GUI zum KNN

6 Experimente

6.1 Netztopologie

Im Rahmen der Experimente wurde ein MLP mit einer verdeckten Schicht verwendet. Zwischen der Eingabeschicht (Merkmalsvektor), der verdeckten Schicht und dem Output-Neuron bestehen Verbindungsgewichte (vgl. Abbildung 1). Der an der Eingabeschicht angelegte Merkmalsvektor wird in der verdeckten Schicht verarbeitet und an die Ausgabeschicht propagiert. Im Rahmen der Experimente wurde ein MLP verwendet (zur Begründung siehe Abschnitt 2.2). Es wurden 38 Merkmale codiert (vgl. Tabelle 2) und an die Eingabeschicht des MLP als Merkmalsvektor übergeben. Die Dokumentation von Neuroph empfiehlt für den Hidden-Layer des KNN eine Anzahl von $2 \cdot n + 1$ Knoten, wobei n der Anzahl der Eingangsknoten entspricht. Aus diesem Grund wurden 77 Knoten im Hidden-Layer verwendet. Die Ausgangsschicht besteht nur aus einem Knoten, welcher eine Bewertung der dem KNN präsentierten Cloud Computing Services im Intervall [0;1] ermöglicht.

6.2 Testläufe

Mehrere „Trial and Error“-Testläufe wurden nach der Trainingsphase zur Ermittlung der Fehlerquote der Testphase durchgeführt. Bei KNN wird dabei der Netzwerkoutput mit dem Sollwert (Benutzerprofil gemäß Trainingsdaten) verglichen. Es wurden zwei Fehlermetriken herangezogen, welche auch Bestandteil der Benutzeroberfläche des Administrators sind, sodass dieser zu jeder Zeit die Fehlerrate des KNN berechnen kann. Zum einen wurde die mittlere quadratische Abweichung als Fehlermaß hinzugezogen. Die Formel hierzu lautet: $1/n \sum [Istwert(i) - Sollwert(i)]^2$ für $i = 1$ bis n . Auf Basis aller Werte kann hiermit der mittlere Fehlerwert überprüft werden. Durch das Quadrieren wird ein aussagekräftigeres Ergebnis erzielt, da anderenfalls große Abweichungen zwischen Ist- und Sollwerten auftreten würden. Des Weiteren wurde die Klassifikationsrate des entwickelten KNN herangezogen. Die Klassifikationsrate beschreibt das Verhältnis von (im Sinne der Benutzerpräferenzen) korrekt bewerteten Cloud Computing Services im Verhältnis zu der Gesamtzahl der getesteten Services. Um aussagekräftige und bessere Bewertungen zu ermöglichen wurden drei Intervalle erzeugt: Intervall $[0,0;0,3]$ entspricht einer 0-30 % Empfehlung, Intervall $[0,31;0,7]$ entspricht einer 31-70 % Empfehlung und Intervall $[0,71;1,0]$ entspricht einer 71-100 % Empfehlung. Cloud Computing Services aus dem ersten Intervall sollten aufgrund der geringen Prozentwerte keine Empfehlung erhalten. Im zweiten Intervall sollten Empfehlungen nur unter Vorbehalt ausgegeben werden und im dritten Intervall kann eine Empfehlung ohne größere Bedenken für die jeweils in diesem Prozent-Intervall eingestuften Cloud Computing Services ausgegeben werden.

| Testmetrik | Wert |
|---|---|
| Middle Square Error | <ul style="list-style-type: none"> 0,004780857670412155 |
| Klassifikationsrate im Intervall $[0,0;0,3]$ | <ul style="list-style-type: none"> 111 Cloud Computing Services korrekt klassifiziert; 22 wurden falsch mit 0-30 % klassifiziert/bewertet. Durchschnittliche Abweichung: 0,21929163774791963 Quadrierte mittlere Abweichung: 0,016729529611374982 |
| Klassifikationsrate im Intervall $[0,31;0,7]$ | <ul style="list-style-type: none"> 40 Cloud Computing Services korrekt klassifiziert; 3 wurden falsch mit 31-70 % klassifiziert/bewertet. Durchschnittliche Abweichung: 0,2865705573055708 Quadrierte mittlere Abweichung: 0,026184027186666934 |
| Klassifikationsrate im Intervall $[0,71;1]$ | <ul style="list-style-type: none"> 9 Cloud Computing Services korrekt klassifiziert; 0 falsch als 71-100 % klassifiziert. Durchschnittliche Abweichung: 0,22376922224040255 Quadrierte mittlere Abweichung: 0,04790546484103173 |
| Klassifikationsrate insgesamt | <ul style="list-style-type: none"> Insgesamt 160 Cloud Computing Services wurden korrekt klassifiziert; 25 wurden falsch klassifiziert Klassifikationsrate: 86,48648648648648 % Minimaler Fehler: 8,922543130873706E-5 Maximaler Fehler: 0,3150233109120054 |

Tabelle 3: Testmetriken (Testphase)

Der in den Testläufen erzielte durchschnittliche quadrierte Fehler ist mit ca. 0,5 % vergleichsweise gering. Auch die Klassifikationsrate des Netzes pendelte sich nach mehreren Testläufen ($n > 50$) bei durchschnittlich 86 % ein, d. h. 86 % der getesteten und dem KNN zuvor unbekanntem Cloud Computing Services wurden vom KNN gemäß

seiner Benutzerpräferenzen (Profil) korrekt bewertet. Die Testläufe erfolgten auf der Basis von 185 Services, die zum Zeitpunkt der Tests (Sept. 2011) in der Datenbank von CloudServiceMarket.info (<http://www.cloudservicemarket.info/>) enthalten waren.

7 Zusammenfassung und Ausblick

7.1 Erkenntnisse

Im Rahmen dieses Beitrags wurde ein Konzept für eine (semi-)automatische Klassifizierung bzw. Bewertung von Cloud Computing Services auf der Basis von KNN konzipiert, prototypisch umgesetzt und anhand von Metriken evaluiert. Die Evaluationsergebnisse zeigen, dass der vorgestellte Ansatz vielversprechend ist. Die Entwicklung sowie das Trainieren und Testen von KNN mit Neuroph erwies sich als benutzerfreundlich. Trotz der Empfehlungen, die Unternehmen und Privatanwender aus dem hier vorgestellten CloudServiceMarket sowie den KNN erhalten können, ist eine anschließende intensive Analyse und Qualitätsprüfung der in die Unternehmens-IT eingebundenen Cloud Computing Services (z. B. mit Pilotprojekten) ratsam. Ein möglicher Lock-in-Effekt, der in diesem frühen Entwicklungsstadium durch heterogene Schnittstellen, SLAs und Softwareeinsatz mit erhöhter Wahrscheinlichkeit eintreten kann, führt zu einer Bindung, die auf hohe Wechselkosten zurückzuführen ist. Dieser Effekt kann jedoch durch die Nutzung von CloudServiceMarket und der KNN im Vorfeld der Entscheidung für bestimmte Cloud Computing Services abgeschwächt werden.

7.2 Limitationen

Der verfolgte Forschungsansatz unterliegt, wie jede Forschungsarbeit, Limitationen, welche mögliche Ansatzpunkte für zukünftige Forschungsarbeiten darstellen. Die automatische Codierung der Merkmalsvektoren bspw. durch Suchroboter, welche Merkmale automatisch aus den Beschreibungen der Cloud Computing Services generieren, ist grundsätzlich möglich. Derzeit erfolgt die Codierung jedoch noch manuell, indem auf die Beschreibungen aus den in der Datenbank in CloudServiceMarket gespeicherten Cloud Computing Services zurückgegriffen wurde und diese Beschreibungen zur Generierung der Merkmalsvektoren verwendet wurden. Dennoch ist es mit Hilfe der vorgestellten IT-Artefakte möglich, dass sich die Anwender von CloudServiceMarket ihre individuellen KNN mit ihren jeweiligen Benutzerpräferenzen trainieren und das jeweilige KNN dann neu in der Datenbank abgelegte Services automatisiert und proaktiv den Anwendern empfehlen kann. Ein zeitaufwendiges Durchsuchen der Datenbank von CloudServiceMarket ist mit Hilfe der KNN zukünftig nicht mehr erforderlich. Eine generelle Limitation liegt jedoch in den KNN an sich begründet, die „richtige“ Auswahl der Datensätze und deren Merkmale als Trainingsgrundlage sowie das „richtige“ Trainieren (Vermeiden von „Overfitting“) und Konfigurieren von KNN ist ein „Trial and Error“-Prozess sowie eine „black box“ [RW10].

7.3 Weitere Forschung

KNN erlernen die Zuordnungsregeln auf einer subsymbolischen Ebene, d. h., sie sind für den Benutzer nicht explizit ersichtlich („black box“). Entsprechende Begleitinformationen für die Benutzer zur besseren Interpretation der KNN-Bewertungen durch den Benutzer sowie die Ausgabe von Handlungsempfehlungen sind daher für zukünftige Forschung geplant. In zukünftigen Forschungsarbeiten sollen zudem weitere Netztopologien untersucht, getestet und der bestehende Ansatz im Hinblick auf Anwenderfreundlichkeit (Usability) analysiert und weiter verfeinert werden. Auch soll eine Wunschkonfiguration für Cloud Computing Services auf CloudServiceMarket implementiert werden, die als Ergebnis diejenigen Services ausgibt, die der jeweils ausgewählten Konfiguration (den Benutzerprofilen) besonders nahe kommen.

Literaturverzeichnis

- [B97] Braun, H.: Neuronale Netze: Optimierung durch Lernen und Evolution. Berlin, 1997.
- [Bi95] Bishop, C. M.: Neural Networks for Pattern Recognition. Oxford, 1995.
- [He07] Hevner, A. R.: A Three Cycle View of Design Science Research. *Inf. Sys.* 19, 2007, (2):87-92.
- [HMP04] Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design Science in Information Systems Research. *MIS Quarterly* 28, 2004, (1):75-105.
- [Kh90] Khanna, T.: Foundations of Neural Networks. New York, 1990.
- [KST00] Kurbel, K.; Szulim, D.; Teuteberg, F.: Künstliche Neuronale Netze zum Filtern und Klassifizieren betrieblicher E-Commerce-Angebote im World Wide Web - eine vergleichende Untersuchung; *Wirtschaftsinformatik* 42, 2000, (3):222-232.
- [KST98] Kurbel, K.; Singh, K.; Teuteberg, F.: Search and Classification of Interesting Business Applications in the World Wide Web Using a Neural Network Approach. In: Forcht, K. (Hrsg.): *Proc. of the 1998 IACIS Conference*, 1998, S. 75-81.
- [MG09] Mell, P.; Grance, T.: The NIST Definition of Cloud Computing, Version 15. National Institute of Standards and Technology, 2009.
- [MS95] March, S. T.; Smith, G. F.: Design and Natural Science Research on Information Technology. *Decision Support Systems* 15, 1995, (4):251-266.
- [MT11] Martens, B.; Teuteberg, F. (2011): Towards a Reference Model for Risk and Compliance Management of IT Services in a Cloud Computing Environment Application Management. Eds.: Keuper, F.; Oecking, C.; Degenhardt, A.; Gabler 2011, S. 135-163.
- [Ne11] Neuroph Dokumentation: <http://neuroph.sourceforge.net/documentation.html>.
- [OBM10] de Oliveira, D.; Baião, F. A.; Mattoso, M.: Towards a Taxonomy for Cloud Computing from an e-Science Perspective. In: *Cloud Computing: Principles, Systems and Applications*, Springer-Verlag, London, 2010; S. 47-62.
- [OO10] Oesterle, H.; Otto, B.: Consortium Research. *Business & Information Systems Engineering* 2, 2010, (5):283-293.
- [RW10] Rey, G. D.; Wender, K. F.: Neuronale Netze: Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung. Huber, 2. Auflage, Bern, 2010.
- [RWZ11] Repschläger, J.; Wind, S.; Zarnekow, R.: Klassifikationsrahmen für die Anbietersauswahl in der Cloud. Jahrestagung der GI/ Workshop: Neue Aspekte der zwischenbetrieblichen Integration durch Enterprise 2.0 (IOS 2.0). Berlin, 2011.
- [Wei09] Weinhardt, C.; Anandasivam, A.; Blau, B.; Borissov, N.; Meinel, T.; Michalk, W.; Stöber, J.: Cloud Computing – A Classification, Business Models, and Research Directions. In: *BISE* 1, 2009, (5):391-399.
- [YBD08] Youseff, L.; Butrico, M.; Da Silva, D.: Toward a Unified Ontology of Cloud Computing. *Grid Computing Environments Workshop*, Austin, TX, 2008, S. 1-10.