

It's Getting Crowded! Verbesserung der Effektivität von Microtask Crowdsourcing

Ujwal Gadiraju¹

Abstract: Microtask Crowdsourcing hat sich als gut geeignete Methode zur Erwerbung von menschlichem Input auf Abruf hervorgerufen und findet verbreitete Anwendung für die Lösung zahlreicher Probleme. Bekannte Beispiele sind unter anderem Umfragen, das Erstellen von Inhalten und das Beschriften von Bildern. In den letzten zehn Jahren gab es bereits zahlreiche Microtask Crowdsourcing Anwendungen in verschiedenen Gebieten sowohl in der Forschung (von Sozialwissenschaften bis hin zur Informatik) als auch für praktischen Nutzen in anderen Fachrichtungen. Dies hat ohne Frage die Grenzen von qualitativen und quantitativen Studien überschritten durch die Möglichkeit zuvor eingeschränkte Laborstudien und kontrollierte Experimente auszuweiten [HRZ11, PCI10]. Heutzutage lässt sich in kurzer Zeit und mit einfachen Mitteln ein Goldstandard für Evaluationen erstellen [GL10] und potentielle Teilnehmer mit unterschiedlichen Demographien sind rund um die Uhr erreichbar [Ip10, Di15]. Daraus ergeben sich jedoch auch eine Vielzahl von Herausforderungen, insbesondere im Hinblick auf die fehlende Kontrolle von Teilnehmern und die Qualität der erhobenen Daten. In dieser Arbeit beschäftigen wir uns mit einigen dieser Herausforderungen. Dabei liegt der Fokus jedoch nicht auf Anwendungen des ehrenamtlichen Crowdsourcing, wie etwa Citizen Science [Bo14], "Serious Games" oder "Games with a Purpose" [VAD08], Wikis [DRH11] oder ähnlichen Ansätzen. Stattdessen konzentrieren wir uns auf das Lösen einiger kritischer Probleme im Bereich des bezahlten Microtask Crowdsourcing, die überwunden werden müssen, um das volle Potential des Modells auszuschöpfen. Unsere Arbeit wird durch die Überzeugung beeinflusst und angetrieben, dass es immer Arbeitsströme geben wird, die von verschiedenen Teilen der Gesellschaft benötigt werden und nicht mittels ehrenamtlicher Teilnahme oder Gamification bearbeitet werden können und für die bezahlte Kanäle am besten geeignet sind, um die Anforderungen zu erfüllen [Ki13].

Einige der wesentlichen Herausforderungen im Bereich des Microtask Crowdsourcing, zu dem jeweiligen Zeitpunkt, an dem die verschiedenen Arbeiten in dieser Arbeit vorgestellt wurden, wurden durchgeführt und veröffentlicht wie im Folgenden beschrieben. Wir haben uns mit jeder dieser Herausforderungen methodisch auseinandergesetzt, wobei wir jeweils das aktuelle Verständnis von Crowd Work erweitert haben oder neue Lösungen vorgestellt haben, die die zu dieser Zeit existierenden Methoden leistungsmäßig übertreffen haben. Ziel dieser Arbeit ist die *'Verbesserung der Effektivität des Microtask Crowdsourcing Modells'*. *Effektivität* ist in diesem Kontext definiert als Grad zu dem die Crowdworker hochqualitative Antworten beitragen und die Auftraggeber die gewünschten Ergebnisse erhalten, wobei die Kosten (Aufgabenbearbeitungszeit, Bezahlung) für alle beteiligten Akteure optimiert werden sollen. Wir identifizieren in diesem Zusammenhang bislang offene Schlüsselherausforderungen und stellen Methoden vor, mit denen sich die durch die Probleme entstehenden Mängel überwinden lassen. Wir decken außerdem weitere Faktoren auf, die die von Crowdworkern produzierte Arbeit qualitativ beeinflussen und bislang nicht bekannt waren.

¹ L3S Research Center, Leibniz Universität Hannover, gadiraju@L3S.de

- **Herausforderung #1** *Eingeschränktes Verständnis von Crowdsourcing-Tasks und Worker Charakteristiken* — Mit einem Alter von etwa zehn Jahren ist das Microtask-Crowdsourcing Feld noch sehr jung. Zu verstehen, welche Aufgabentypen für Crowdsourcing infrage kommen, wird beim Entwickeln besserer Plattformen helfen und die Marketplace-Dynamiken zwischen Aufgabenerstellern und Crowdworkern verbessern. Als unsere Arbeit zu diesem Thema vor einigen Jahren durchgeführt wurde [Ga15], war wenig über das Verhalten von Crowdworkern auf Microtask Crowdsourcing Plattformen bekannt. Zu verstehen, wie sich Crowdworker in ihrem Verhalten unterscheiden und wie die Qualität der Arbeit von diesem Verhalten abhängt, kann den Aufgabenentwurf beeinflussen und zu effektiveren Mechanismen zur Qualitätskontrolle führen.
- **Herausforderung #2** *Mangelhafte Methoden zur Vorauswahl von Crowdworkern* — Oft existieren nur wenig oder gar keine Daten über die bisherige Leistung von Crowdworkern, die einen Hinweis auf die Qualität der Arbeit geben könnten. Um in Abwesenheit dieser Indikatoren Crowdworker mit den gewünschten Fähigkeiten und erwiesener Leistung auszuwählen, verwenden Aufgabenersteller üblicherweise die Leistung von Crowdworkern in Qualifizierungstests oder in einem kleinen Teil der eigentlichen Aufgabe während der Pre-Screening Phase. Dies ist jedoch nur eine Annäherung an die Auswahl geeigneter Crowdworker. Benötigt werden deutlichere Indikatoren für die Kompetenz von Crowdworkern sowie effektivere Mechanismen für die Vorauswahl.
Wie von Barry Schwartz in dessen einflussreichen Arbeiten im Bereich der Psychologie und Gesellschaftstheorie herausgestellt, führt ein Überangebot oft zu nachteiligen Effekten auf den Entscheidungsprozess von Menschen [Sc04, SW04]. Die große Auswahl an Aufgaben, die für einen erfahrenen Crowdworker auf einer großen Crowdsourcing Plattform (wie etwa Amazons Mechanical Turk (AMT)³ oder Crowdflower⁴) zur Verfügung stehen, macht es schwierig, die Aufgaben zu finden, die am besten zum jeweiligen Crowdworker passen. Da viele Crowdworker sich für Aufgaben entscheiden, die nicht optimal zu ihnen passen, haben geeignete Crowdworker aufgrund der Beschränkungen in der Anzahl der Teilnehmer nicht die Möglichkeit, an diesen Aufgaben zu arbeiten. Crowdworker nehmen oft an Aufgaben teil, die jenseits ihrer Fähigkeiten liegen, obwohl sie Interesse daran haben, ihren Ruf zu erhalten. Dadurch sinkt die Effektivität des Crowdsourcing-Ansatzes. Die Vorauswahl von Crowdworkern ist eine verbreitete Methode, um das Problem der Teilnahme ungeeigneter Crowdworker an einer Aufgabe zu lösen [Ol11].
- **Herausforderung #3** *Unvollständige Betrachtung von Faktoren, die qualitätsbezogene Ergebnisse beeinflussen* — Qualitätskontrolle ist vermutlich das meist erforschte Thema im Bereich des Microtask Crowdsourcing. Trotzdem werden die Aspekte, die die Qualität der produzierten Arbeit beeinflussen nicht vollständig betrachtet. Um eine faire und gerechtfertigte Behandlung der von auf einer Microtask Crowdsourcing Plattform erstellten Arbeit zu gewährleisten, ist es wichtig, alle qualitätsbeeinflussenden Faktoren zu verstehen und einzubeziehen. Dies

³ <http://www.mturk.com/>

⁴ <http://www.crowdflower.com/>

wird gegenwärtig wenig bis überhaupt nicht getan. Aufgabensteller und Plattformen berücksichtigen überlicherweise nur die von Crowdworkern produzierten Endergebnisse, ohne darauf zu achten, *wie* die Ergebnisse produziert wurden. So haben beispielsweise Arbeiten in der Ethnographie, die sich mit dem Crowdsourcing-Prinzip auseinandersetzen, deutliche Unterschiede in den Umgebungen, in denen sich verschiedene Crowdworker befinden, festgestellt [Gu14, Ma14].

Eine mangelnde Beschreibung der Aufgabe hat klare Konsequenzen: aus Mangel an Alternativen im Marketplace versuchen sich Crowdworker oft an Aufgaben, für die sie kein optimales Verständnis besitzen. Auf der anderen Seite sind Aufgabensteller sich oft nicht den Schwachstellen ihres Aufgabendesigns bewusst und halten unzufriedenstellende Ergebnisse als Beweise für bewusst schädliches Verhalten, so dass die Crowdworkern die Bezahlung für ihre Arbeit verweigern. Daraus resultierend verlieren Crowdworker ihre Motivation, die Gesamtqualität der Arbeit sinkt und alle Akteure verlieren ihr Vertrauen in den Marketplace. Obwohl die Klarheit der Aufgabenstellung offensichtlich wichtig ist für das Microtask Crowdsourcing, gibt es kein klares Verständnis darüber, in welchem Umfang Unklarheit in Beschreibung und Anweisungen die Leistung der Crowdworker beeinflusst.

Wir haben wichtige Beiträge zu jeder der genannten Herausforderungen gebracht. Unsere Beiträge sind in Abbildung 1 dargestellt und weiter unten beschrieben.

- **Beitrag #1** *Erweiterung des aktuellen Verständnisses von Aufgabentypen, Crowdworker-Verhalten und Qualitätskontrolle* — Wir beschäftigen uns zunächst mit zwei zentralen Aspekten, die die Effektivität von Microtask Crowdsourcing beeinflussen: Aufgabendesign und das Verhalten von Crowdworkern. Um das aktuelle Verständnis von Microtasks und Crowdworker-Verhalten zu erweitern, haben wir eine weitfassende Studie mit 1000 Crowdworkern auf CrowdFlower durchgeführt. Basierend auf zuverlässigen Daten über die Aufgabentypen, die Crowdworker durchgeführt haben, haben wir ein zwei-Level Kategorisierungsschema für Microtasks vorgestellt und Einblicke in die Aufgabenbeliebtheit, den Aufwand von Crowdworkern für das Abschließen von verschiedenen Aufgabentypen und ihre Zufriedenheit mit der Bezahlung erhalten. Auf der obersten Ebene ist das Kategorisierungsschema *zielorientiert* (d.h. die Kategorien orientieren sich an den Aufgabenzielen) und die Sub-Kategorien sind *arbeitsflussorientiert* (d.h. die Sub-Kategorien orientieren sich an den Schritten, die durchgeführt werden müssen, um die Aufgabe erfolgreich abzuschließen). Die genaue Kategorisierung von Crowdsourcing-Aufgaben hat deutliche Auswirkungen auf die Nutzermodellierung von Crowdworkern und die Empfehlung von Aufgaben [GKD14]. Die vorgeschlagene Aufgabekategorisierung, inklusive der Ergebnisse unserer Analyse, unterstützt Aufgabensteller bei Design und Durchführung der Aufgaben. Beispielsweise haben Difallah et al. in ihrer Analyse des AMT-Marketplace untersucht, wie verschiedene Typen von "human intelligence tasks"(HITs) sich im Hinblick auf die von uns eingeführte Taxonomie über die Zeit weiterentwickelt haben [Di15]. Auf Basis unserer Ergebnisse im Zusammenhang mit aufgabenabhängigen Charakteristiken, wie etwa Aufgabenbeliebtheit, Aufgabenaufwand, benötigte Bezahlung usw. können Aufgaben-

Main Streams of Contributions

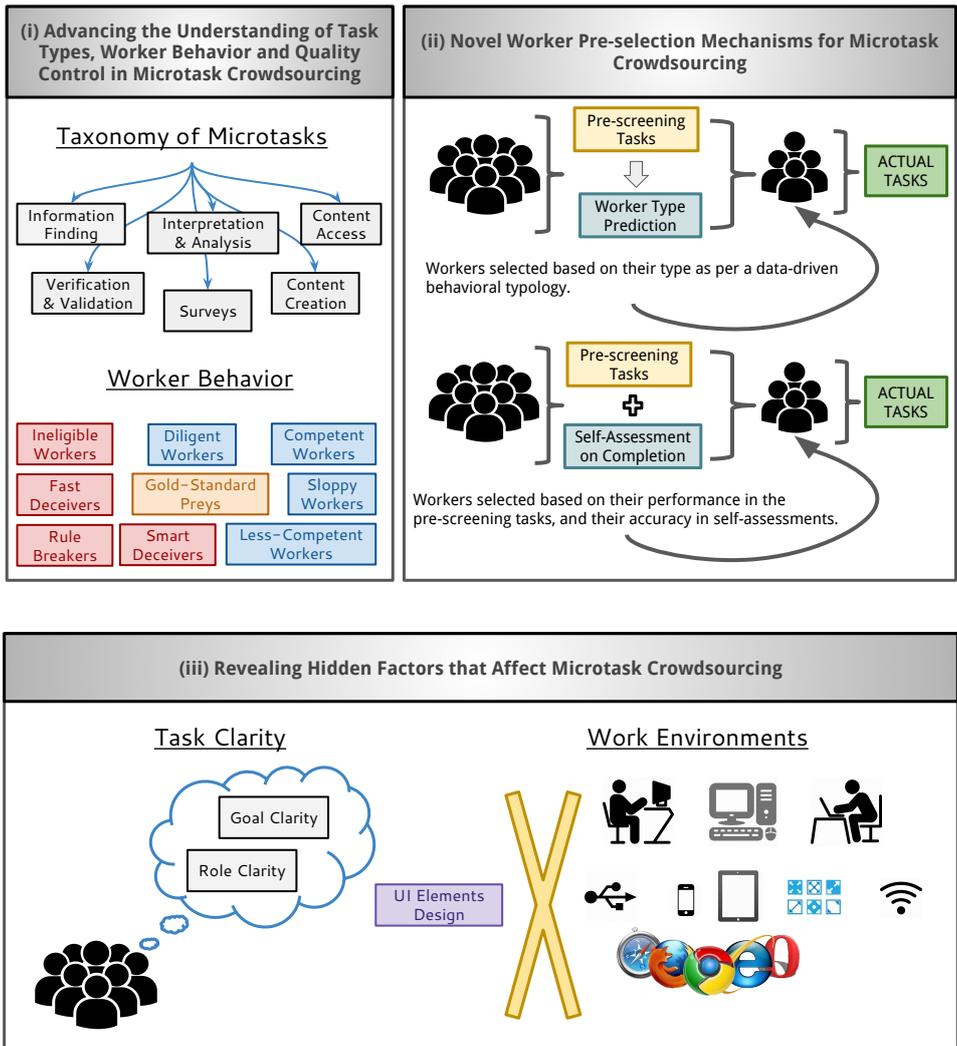


Abb. 1: Überblick über die wesentlichen Beiträge dieser Dissertation.

steller Aufgaben mit höherer erfolgsrate erstellen, d.h. die Qualität der Ergebnisse mit Rücksicht auf die Kosten maximieren. Die Mechanismen zur Qualitätskontrolle müssen eine weite Gruppe unterschiedlicher Crowdworker und unterschiedlichen Verhaltens abdecken. Ein zentraler Schritt zum betrugsfreien Aufgabendesign ist das Verstehen der Verhaltensmuster von Microtask Crowdworkern. Wir haben die verbreiteten bewusst schädlichen Aktivitäten

auf Crowdsourcing-Plattformen analysiert und das gezeigte Verhalten von vertrauenswürdigen und nicht vertrauenswürdigen Crowdworkern studiert, insbesondere für Umfragen. Basierend auf unserer Analyse haben wir verschiedene Typen von schädlichem Verhalten identifiziert (*ineligible workers*, *fast deceivers*, *rule breakers* und *smart deceivers*), die über das hinaus gehen, was in bisherigen Arbeiten gezeigt wurde. Das Verstehen dieser Aspekte hilft uns, Aufgaben zu entwerfen, die schädlichem Verhalten entgegenwirken und somit den Aufgabenstellern sowie den Crowdsourcing-Plattformen nutzen.

Um die Qualität der Ergebnisse zu verbessern, haben wir Maße vorgestellt, die auf dem Verhalten der Crowdworker basieren und zum messen und entgegenwirken von ungewollten oder potentiell schädlichen Aktivitäten in Crowdsourcing-Aufgaben genutzt werden können. Wir haben eine detaillierte Analyse über schädliches Verhalten von Crowdworkern im Verlauf der Aufgabe vorgestellt und einen ‘*Wendepunkt*’ definiert, der den Zeitpunkt markiert, ab dem ein Crowdworker dazu tendiert, schwache Antworten zu geben. In Anbetracht dieser Aspekte haben wir eine Anleitung zum effektiven Design von Crowdsourcing-Aufgaben eingeführt. Die Nutzung der Dynamik von Crowdsourcing-Aufgaben und das Prüfen von Crowdworker-Verhalten ist beim Design von besseren Aufgaben hilfreich.

- **Beitrag #2 *Neue Mechanismen für die Vorauswahl von Crowdworkern*** — Wir stellen zwei verschiedene neue Methoden zur Vorauswahl von Crowdworkern vor, die die State-of-the-Art Ansätze für verschiedene Aufgabentypen leistungsmäßig übertreffen.

Existierende Arbeiten haben die Aktivitäten von Crowdworkern anhand der jeweiligen Leistung binär in *gut* und *schlecht* eingeteilt [RK11, DHL16]. Die Autoren dieser Arbeiten haben die Vorteile, die sich aus ihrem Ansatz verglichen mit anderen Qualitätskontrollmechanismen ergeben vorgestellt. Aspekte wie etwa Aufwand, Fähigkeiten und Verhalten können anhand der Aktivitäten der Crowdworker interpretiert werden und können bei der Einschätzung der Qualität der Arbeit des Crowdworkers helfen [RK11, RK12]. Obwohl es mit Sicherheit sinnvoll ist zwischen guter und schlechter Arbeitsqualität zu unterscheiden, sind wir der Meinung, dass eine feinere Unterscheidung der Aktivitäten von Crowdworkern weitere Vorteile mit sich bringt. Zum Beispiel führt das Wissen, dass selbst gute Crowdworker verschiedene Abläufe zur Bearbeitung von Aufgaben verwenden zu der Frage, ob diese Unterschiede praktische Auswirkungen haben können. Mit der zunehmenden Verwendung von Crowdsourcing-Lösungen in Form von menschlichem Input mittels Microtask-Marketplaces haben sich auch neue Anforderungen ergeben. Wenn verschiedene Beschränkungen der Kosten existieren (etwa bei Zeit oder Geld), genügt es oft nicht, die Qualität der Arbeit alleine vorherzusagen. Ein besseres Verständnis davon, wie gute Crowdworker sich in komplexen Aufgaben von anderen Crowdworkern abheben, kann zu Verbesserungen wie einem weiterentwickelten HIT-Design oder verbesserten HIT-Zuweisungsmodellen führen. Um das aktuelle Verständnis verschiedener Crowdworker-Typen auf einer Plattform zu vergrößern und dieses Verständnis für die Crowdworker-Vorauswahl bei gegebener Aufgabe zu nutzen, haben wir Daten über die Aktivitäten von Workern in 1800 HITs mit variierender Länge, Schwierigkeit und Typ erhoben. Wir

haben das Verständnis von Crowdworke-Typen neu definiert und es um multi-dimensionale Definitionen innerhalb der Typology von Crowdworkern erweitert. Wir haben experimentell gezeigt, dass es möglich ist, Crowdworker automatisch in Klassen einzuteilen, basierend auf Machine Learning-Modellen, die die Verhaltensmuster von Crowdworkern bei der Arbeit an HITs verwenden. Die Nutzung einer Typ-Klassifizierung von Crowdworkern kann die Qualität von Crowdsourcing-Aufgaben durch die Vorauswahl von Workern für eine bestimmte Aufgabe verbessern. Unsere Vorauswahl-Methode führt zu einer Verbesserung von 10% verglichen mit bestehenden Vorauswahl-Verfahren. Wir stellen Verhaltensmerkmale für die Modellierung und Vorauswahl von Crowdworkern vor, basierend auf low-level Verhaltensmustern. Durch das Aufzeigen der praktischen Vorteile der Einteilung von guten Crowdworkern (*fleißige Worker* und *kompetente Worker*) nach unserer vorgestellten Typologie, haben wir gezeigt, dass eine Einteilung, die über *gut* und *schlecht* hinaus geht, wirksam ist. Dieses Ergebnis hat bedeutsame Auswirkungen auf Crowdsourcing-Systeme, in denen der Verhaltenstyp eines Crowdworkers vor der Teilnahme an einer Aufgabe nicht bekannt ist.

Auf der Basis von Selbsteinschätzungstheorien in der Psychologie, wie etwa dem Dunning-Kruger-Effekt, zeigen wir, dass Crowdworkern oft das Bewusstsein für ihr eigenes Kompetenzlevel fehlt und plädieren für kompetenzbasierte Vorauswahl in Crowdsourcing-Marketplaces. Der Dunning-Kruger-Effekt bezeichnet die kognitive Voreingenommenheit von weniger kompetenten Individuen, die zu einer überhöhten Selbsteinschätzung und illusionärer Überlegenheit führt [Du11]. Wir zeigen die Auswirkungen von mangelhafter Selbsteinschätzung auf reale Microtasks und stellen eine neue Vorauswahl-Methode für Crowdworker vor, die die Genauigkeit der Selbsteinschätzung von Crowdworkern miteinbezieht. Weiterhin haben wir verdeutlicht, dass die Einschätzung der Kompetenz eines Crowdworkers innerhalb einer Aufgabe durch die objektive Schwierigkeit der Aufgabe beeinflusst wird. Die Fähigkeit eines Crowdworkers, sich korrekt selbst zu beurteilen ist ein Teil der Kompetenz des Crowdworkers. Mittels gründlicher Bewertung für Tagging, Sentiment-Analysis und Image-Validation Aufgaben haben wir beobachtet, dass Aufgabensteller profitieren, indem sie die Selbsteinschätzung eines Crowdworkers mit in dessen Kompetenzbewertung miteinbeziehen, anstatt nur auf die Leistung des Workers in der Vorauswahl-Phase zu vertrauen. Unsere Ergebnisse zeigen, dass Worker, die mit unserem vorgestellten Ansatz ausgewählt werden, eine signifikant höhere Genauigkeit erreichen, verglichen mit Workern, die mittels herkömmlicher Vorauswahl-Methoden ausgewählt werden [Ga17b].

- **Beitrag #3** *Versteckte Crowdwork-Faktoren: Aufgabenklarheit und Arbeitsumgebung* — Worker auf Microtask-Crowdsourcing-Marketplaces streben nach einem Gleichgewicht zwischen finanziellem Einkommen und der Vergrößerung des eigenen Ansehens (Reputation). Dieses Gleichgewicht wird oft durch schlecht formulierte Aufgaben bedroht, da Crowdworker versuchen, diese durchzuführen, obwohl ihr Verständnis der Arbeit nicht ausreichend ist. Wir haben 100 Crowdworker auf der CrowdFlower-Plattform befragt, um das Vorhandensein von Problemen mit der Aufgabenklarheit in Crowdsourcing-Marketplaces zu verifizieren. Dabei haben wir herausgefunden, dass Crowdworker sich mit derartigen Problemen beschäftigen

müssen, was das Vorhandensein eines Mechanismus zur Vorhersage und Messung von Aufgabenklarheit motiviert. Inspiriert durch existierende Arbeit in Organisationspsychologie haben wir ein neues Modell für Aufgabenklarheit als Kombination von zwei Fragen vorgestellt: Wie klar ist das angestrebte Ziel einer Aufgabe (*goal clarity*)? Wie klar sind die durchzuführenden Aktivitäten (*role clarity*) [RN77]? Um besser zu verstehen, wie Klarheit von Crowdworkern wahrgenommen wird, haben wir die Einschätzungen von Crowdworkern für 7100 Aufgaben aus einem 5 Jahre umfassenden Datensatzes des AMT-Marketplaces gesammelt. In einer umfassenden Studie haben wir gezeigt, dass Klarheit von Workern zusammenhängend verstanden wird und vom Aufgabentyp abhängt. Zusätzlich haben wir Beweise dafür gefunden, dass Klarheit und Komplexität nicht direkt zusammenhängen. Dies zeigt eine komplexe Beziehung, die weiter untersucht werden sollte. Wir haben Features zum Erfassen von Aufgabenklarheit vorgestellt und die erhobenen Daten für das Trainieren und Validieren eines überwachten Machine Learning-Modells für die Vorhersage von Aufgabenklarheit verwendet. Anhand dieses Modells haben wir gezeigt, dass Aufgabenklarheit präzise vorhergesagt werden kann. Schließlich haben wir mittel einer temporalen Analyse gezeigt, dass Klarheit keine Macro-Eigenschaft eines AMT-Systems ist, sondern vielmehr eine lokale Eigenschaft, die durch Aufgaben und Aufgabensteller beeinflusst wird. Unsere Ergebnisse erweitern das aktuelle Verständnis von Crowdwork und haben wichtige Auswirkungen auf die Strukturierung von Arbeitsflüssen [GYB17]. Die Vorhersage von Aufgabenklarheit kann Crowdworkern bei der Aufgabenauswahl und Aufgabenstellern beim Design der Aufgaben helfen.

Ein weiterer Aspekt von Microtask-Crowdsourcing, der bisher wenig beachtet wurde, ist die *Arbeitsumgebung*, definiert als Hardware und Software, die von Crowdworkern für die Arbeit an Microtasks auf Crowdsourcing-Plattformen verwendet wird. Zunächst haben wir eine Pilotstudie zu den guten und schlechten Erfahrungen, die Crowdworker mit UI-Elementen in Crowdwork gemacht haben durchgeführt. Dabei haben wir die typischen Probleme aufgedeckt, mit denen sich Worker auseinandersetzen müssen. Wir haben herausgefunden, dass insbesondere folgende, schlecht designte UI-Elemente die Leistung von Crowdworkern negativ beeinflussen: große Eingabefelder, unproportional kleine Textfelder und Multiple-Choice Fragen mit zu vielen Radio- bzw. Check-Boxen. Um den Einfluss verschiedener Designmöglichkeiten im Hinblick auf UI-Elemente, die Leistung von Crowdworkern und den Zusammenhang mit variierenden Worker-Umgebungen zu analysieren, haben wir eine zweite Studie mit über 125 verschiedenen Microtasks auf CrowdFlower durchgeführt, die sich in zwei identischen Teilen an Worker in Indien und den USA richtete. Diese Aufgaben haben die guten und schlechten Designs von UI-Elementen in Crowdsourcing-Aufgaben imitiert. Wir haben Hardwaredetails, wie CPU-Geschwindigkeit und Gerätetyp sowie Softwaredetails inklusive des verwendeten Browsers, Betriebssystems und weitere Eigenschaften, die die Arbeitsumgebung von Crowdworkern definieren festgehalten. Für Information-Finding und Content-Creation Aufgaben benötigen Crowdworker, die mobile Geräte benutzen, deutlich mehr Zeit als die Vervollständigung der Aufgabe im Vergleich zu Workern mit Laptops oder Desktop-PCs. Worker aus den USA waren im Durchschnitt

schneller und haben eine bessere Leistung bei Aufgaben mit schlecht designten UI-Elementen gezeigt, verglichen mit Workern aus Indien für alle Aufgabentypen und Aufgabenumgebungen. Außerdem haben Worker aus den USA indische Worker in Audi-Transcription Aufgaben leistungsmäßig übertroffen (bei einer gleichzeitig guten Leistung bei Aufgaben mit schlechter Audioqualität). Die Varianz an Arbeitsumgebungen ist bei US-Workern größer als bei indischen Workern und US-Worker verwenden öfter aktuelle Technologien (z.B. aktuelle Betriebssysteme und Browser). Um den Einfluss der Arbeitsumgebung auf Crowdsourcing-Microtasks besser zu verstehen, haben wir semi-strukturierte Interviews mit CrowdFlower-Workern, die alle Aufgaben bearbeitet haben, durchgeführt.

Durch unsere Studien haben wir die bedeutende Rolle von Arbeitsumgebungen auf Crowdwork offengelegt [Ga17a]. Unsere Ergebnisse deuten an, dass Crowdworker eine Vielzahl unterschiedlicher Arbeitsumgebungen verwenden, welche die Qualität der Arbeit beeinflussen. Wir haben herausgefunden, dass einige Arbeitsumgebungen Crowdworkers besser unterstützen als andere, in Abhängigkeit von den verwendeten UI-Elementen. Unser vorgestelltes Tool *ModOp* hilft beim Entwurf von Crowdsourcing-Microtasks, welche für unterschiedliche Arbeitsumgebungen geeignet sind. Wir haben empirisch gezeigt, dass der Einsatz von *ModOp* die kognitive Belastung von Workern reduziert und somit die Erfahrung der Worker verbessert. Die Nutzung reaktiver Nutzerschnittstellen, welche in der Lage sind, sich an verschiedene Arbeitsumgebungen anzupassen, würde vielen Crowdworkern die Möglichkeit geben, effektiver zu arbeiten. Unsere Ergebnisse haben außerdem wichtige Auswirkungen auf die Aufgabenzuteilung in Crowdwork. Aufgaben, die eine schnelle Durchführung voraussetzen oder von bestimmten Arbeitsumgebungen profitieren, können gezielt bestimmten Crowdworkern mit geeigneten Umgebungen zugewiesen werden.

Mit der Durchführung diverser interdisziplinärer Studien, der Vorstellung verschiedener Methoden zum Umgang mit den wesentlichen von uns identifizierten Herausforderungen und mit ausführlichen Evaluationen haben wir die folgenden nennenswerten Beiträge für die Verbesserung der Effektivität von Microtask-Crowdsourcing erbracht: (i) Wir haben das Verständnis von Aufgabentypen, Crowdworker-Verhalten und Qualitätskontrolle erweitert. (ii) Wir haben neue Mechanismen zur Vorauswahl von Crowdworkern vorgestellt, welche die existierenden Methoden leistungsmäßig übertreffen. (iii) Wir haben den Einfluss versteckter Faktoren, wie Aufgabeklarheit und Arbeitsumgebungen auf die Qualität der Arbeit herausgestellt.

Literaturverzeichnis

- [Bo14] Bonney, Rick; Shirk, Jennifer L; Phillips, Tina B; Wiggins, Andrea; Ballard, Heidi L; Miller-Rushing, Abraham J; Parrish, Julia K: Next steps for citizen science. *Science*, 343(6178):1436–1437, 2014.
- [DHL16] Dang, Brandon; Hutson, Miles; Lease, Matt: MmmTurkey: A Crowdsourcing Framework for Deploying Tasks and Recording Worker Behavior on Amazon Mechanical Turk. arXiv preprint arXiv:1609.00945, 2016.

- [Di15] Difallah, D. E.; Catasta, M.; Demartini, G.; Ipeirotis, P. G.; Cudré-Mauroux, P.: The Dynamics of Micro-Task Crowdsourcing – The Case of Amazon MTurk. In: 24th International Conf. on World Wide Web (WWW). 2015.
- [DRH11] Doan, Anhai; Ramakrishnan, Raghu; Halevy, Alon Y: Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.
- [Du11] Dunning, David: The Dunning-Kruger Effect: On Being Ignorant of One’s Own Ignorance. *Advances in experimental social psychology*, 44:247, 2011.
- [Ga15] Gadiraju, Ujwal; Kawase, Ricardo; Dietze, Stefan; Demartini, Gianluca: Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. In: Proceedings of SIGCHI’15. S. 1631–1640, 2015.
- [Ga17a] Gadiraju, Ujwal; Checco, Alessandro; Gupta, Neha; Demartini, Gianluca: Modus operandi of crowd workers: The invisible role of microtask work environments. *Proceedings of the ACM on IMWUT*, 1(3):49, 2017.
- [Ga17b] Gadiraju, Ujwal; Fetahu, Besnik; Kawase, Ricardo; Siehndel, Patrick; Dietze, Stefan: Using Worker Self-Assessments for Competence-based Pre-Selection in Crowdsourcing Microtasks. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(4), 2017.
- [GKD14] Gadiraju, Ujwal; Kawase, Ricardo; Dietze, Stefan: A taxonomy of microtasks on the web. In: Proceedings of the 25th ACM conference on Hypertext and social media. ACM, S. 218–223, 2014.
- [GL10] Grady, Catherine; Lease, Matthew: Crowdsourcing document relevance assessment with mechanical turk. In: HLT-NAACL workshop on creating speech and language data with Amazon’s mechanical turk. *ACL*, S. 172–179, 2010.
- [Gu14] Gupta, Neha; Martin, David; Hanrahan, Benjamin V; O’Neill, Jacki: Turk-life in India. In: Proceedings of the 18th Int. Conf. on Supporting Group Work. S. 1–11, 2014.
- [GYB17] Gadiraju, Ujwal; Yang, Jie; Bozzon, Alessandro: Clarity is a Worthwhile Quality – On the Role of Task Clarity in Microtask Crowdsourcing. In: Proceedings of the 28th ACM, HT’17, Prague, Czech Republic, July 4-7, 2017. ACM, 2017.
- [HRZ11] Horton, John J; Rand, David G; Zeckhauser, Richard J: The online laboratory: Conducting experiments in a real labor market. *Exp. Economics*, 14(3):399–425, 2011.
- [Ip10] Ipeirotis, Panagiotis G: Demographics of mechanical turk. 2010.
- [Ki13] Kittur, Aniket; Nickerson, Jeffrey V; Bernstein, Michael; Gerber, Elizabeth; Shaw, Aaron; Zimmerman, John; Lease, Matt; Horton, John: The future of crowd work. In: Proceedings of CSCW’13. ACM, S. 1301–1318, 2013.
- [Ma14] Martin, David; Hanrahan, Benjamin V; O’Neill, Jacki; Gupta, Neha: Being a turker. In: Proceedings of ACM CSCW’14. ACM, S. 224–235, 2014.
- [OI11] Oleson, David; Sorokin, Alexander; Laughlin, Greg P; Hester, Vaughn; Le, John; Biwald, Lukas: Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. *Human computation*, 11:11, 2011.
- [PCI10] Paolacci, Gabriele; Chandler, Jesse; Ipeirotis, Panagiotis G: Running experiments on amazon mechanical turk. 2010.
- [RK11] Rzeszutarski, Jeffrey M; Kittur, Aniket: Instrumenting the crowd: using implicit behavioral measures to predict task performance. In: Proceedings of the 24th annual ACM UIST. ACM, S. 13–22, 2011.

- [RK12] Rzeszotarski, Jeffrey; Kittur, Aniket: CrowdScape: interactively visualizing user behavior and output. In: UIST' 12. ACM, S. 55–62, 2012.
- [RN77] Ruch, Libby O; Newton, Rae R: Sex characteristics, task clarity, and authority. Sex Roles, 3(5):479–494, 1977.
- [Sc04] Schwartz, Barry: The paradox of choice: Why less is more. New York: Ecco, 2004.
- [SW04] Schwartz, Barry; Ward, Andrew: Doing better but feeling worse: The paradox of choice. Positive psychology in practice, S. 86–104, 2004.
- [VAD08] Von Ahn, Luis; Dabbish, Laura: Designing games with a purpose. Communications of the ACM, 51(8):58–67, 2008.



Ujwal Gadiraju hat seinen Bachelor-Abschluss in Informatik 2010 an der VIT University in Tamil Nadu, Indien erhalten. Seinen Master-Abschluss in Informatik hat er 2012 an der Delft University of Technology in den Niederlanden erhalten. 2017 erhielt er seinen Doktor an der Fakultät für Elektrotechnik und Informatik an der Leibniz Universität Hannover. Zur Zeit arbeitet er als Postdoc am L3S Forschungszentrum an der Leibniz Universität Hannover. Seine wesentlichen Forschungsinteressen umfassen Human Computation, Crowdsourcing, Information Retrieval und Social Computing. Er erhielt den Douglas

Engelbart Best Paper Award auf der ACM Conference on Hypertext and Social Media (HT) 2017, den Best Poster Award auf der ACM Web Science Conference (WebSci) 2016 und den Best Poster Award auf der International Semantic Web Conference (ISWC) 2014. Weiterhin wurde er 2017 mit dem Outstanding Reviewer Award auf der World Wide Web Conference (WWW) und der Excellent Reviewer Recognition auf der ACM CHI Conference on Human Factors in Computing Systems (SIGCHI) ausgezeichnet. Er hat über 40 wissenschaftliche Artikel veröffentlicht, darunter Papiere bei hochrangigen Konferenzen und einflussreichen Journalen wie ACM SIGCHI, ACM TOCHI, ACM HT, ACM UbiComp, WWW, ACM SIGIR, ACM CIKM, ISWC und anderen. In den vergangenen Jahren hat Ujwal aktiv zu europäischen Projekten beigetragen, wie etwa DURAARK und AFEL.