

# Sag's mit einem Lächeln - Emotionsbasierte Qualitätsverbesserung

Simon André Scherr

Fraunhofer IESE  
Kaiserslautern, Germany  
simon.scherr@iese.fraunhofer.de

## ABSTRACT

„Was haben wir denn jetzt verbockt?“ Vor dieser Frage stehen App-Entwickler immer wieder. Typisch für Apps sind eine kurze Time-to-Market, ständige Updates und geringe Budgets. Nutzer löschen Apps sofort, wenn sie nicht so funktionieren wie erwartet. Was nun? Die gute Nachricht ist: Nutzer geben Feedback! Daraus kann man wertvolle Schlüsse ziehen. Heute lassen sich große Mengen von Feedback nicht vollautomatisch auswerten. Veränderungen der Stimmung können erkannt werden, ohne alle Details zu verstehen. Um das zu erreichen, analysieren wir die darin enthaltenen Emojis, die die Stimmung des Nutzers widerspiegeln. Dafür haben wir Emojis auf Emotionen bei etwa 2000 Menschen untersucht, mit dem Ergebnis, dass Menschen Emojis ähnlich wahrnehmen. Damit können sie sehr gut dazu verwendet werden, Texte auf bestimmte Gesichtspunkte hin zu analysieren. Die Analyse von Emojis in Nutzerfeedback ist eingebettet in einen Prozess zur systematischen Integration von Nutzerfeedback in die Produktverbesserung. Dieser Artikel zeigt unser Vorgehen, unsere Ergebnisse und was man aus Emojis im Feedback lernen kann, um Apps zu verbessern.

## KEYWORDS

Feedback, Nutzer, Emotion, Emoji, Produktverbesserung

## 1 Einführung und Motivation

Ein großer Teil der vernetzten Welt wird durch Apps auf Mobilgeräten bestimmt. Diese sind zunehmend intelligenter, aber auch komplexer. Im Hinblick auf die Entwicklung solcher Apps stehen wir vor großen Herausforderungen. Essentiell für die Nutzerakzeptanz ist ein fehlerfreier Ablauf, gute Qualität und innovative Ideen. Typisch sind auch eine kurze Time-to-Market, Updates im Wochenrhythmus sowie geringe Budgets. Darüber hinaus neigen Nutzer schnell dazu, Apps wieder zu löschen, wenn

sie nicht das machen, was von ihnen erwartet wird. Damit entsteht ein Spannungsfeld zwischen Qualität, Geschwindigkeit und Funktionalität. Die Frage: „Was haben wir denn jetzt verbockt?“, ist für Entwickler mit höchster Eiligkeit zu beantworten. Der Markt bringt aber auch eine zweite Frage mit sich, die immer wieder einer Antwort bedarf: „Was müssen wir an unserem Produkt verändern?“. Jederzeit muss auf Trends und neue Konkurrenz reagiert werden. Damit ist eine permanente Überprüfung der Anforderungen nötig. Dies alles muss während der regulären Produkt-Entwicklung passieren. Es ist notwendig diese Prüfungen sowohl effizient, als auch effektiv durchführen zu können und dabei weder hohe Kosten zu verursachen, noch lange zu brauchen. Die Herausforderungen sind damit enorm.

## 2 Nutzer geben Feedback

Wie können Entwickler auf diese Situation reagieren? Die gute Nachricht ist: Nutzer geben Feedback! Das Feedback der Nutzer spiegelt die Wahrnehmung von Änderungen am Produkt oder der Einsatzumgebung wieder [1]. Damit bietet es die Chance, seine Nutzer besser zu verstehen. So erfährt man nicht nur wie Änderungen aufgefasst werden, sondern auch was die nervigsten Probleme an einem Produkt sind. Darüber hinaus liefern Nutzer auch selbst Ideen für Änderungen oder komplett neue Funktionen. Dieses Feedback ist über eine große Menge an Kanälen wie App Stores, soziale Medien oder Kundenforen verteilt. Die breite Streuung von Feedback sorgt dafür, dass eine Vielzahl an Reviews pro Tag im Playstore, im App Store, auf Twitter, in Nutzerforen wie User Voice oder als in-App-Feedback verteilt entstehen und das auch noch in vielen Sprachen.

In diesem Feedback finden sich zahlreiche Anhaltspunkte für die Verbesserung von Produkten. Wir können Nutzerfeedback in drei Arten aufteilen: Fehlerberichte, Feature-Wünsche und Lob [1]. Während man mit letzterem hauptsächlich eine bestehende Qualität wahren oder die entsprechenden Inhalte für Marketingzwecke nutzen kann, sind für die Weiterentwicklung und Verbesserung vor allem Fehlerberichte und Wünsche der Nutzer interessant.

Durch die Streuung des Feedbacks über zahlreiche Quellen ist es nötig, Feedback ganzheitlich zu betrachten und einzusammeln. Wie eine Studie von Nayebi et al [2] deutlich macht, bei der App Store- und Twitter-Feedback verglichen wurde. So erreicht man beim zusätzlichen Betrachten von Twitter zum App Store 22.4% mehr Ideen für Funktionen und identifiziert mehr 12.89% Bugs.

Veröffentlicht durch die Gesellschaft für Informatik e.V. und die German UPA e.V. 2019 in S. Hess & H. Fischer (Hrsg.): *Mensch und Computer 2019 – Usability Professionals*, 08.-11. September 2019, Hamburg  
Copyright © 2019 bei den Autoren.  
<https://doi.org/10.18420/muc2019-up-0340>

Gleichzeitig haben Deshpande und Rokne [3] deutlich gezeigt, dass auf Twitter Fehler tendenziell schneller rückgemeldet werden im Vergleich zum App Store.

Sammelt man also all dieses Feedback kontinuierlich aus den verschiedensten Quellen ein, entsteht eine große und umfassende Datenmenge, die das Puzzle der Nutzermeinung mehr und mehr vervollständigt. In diesem Kontext hat sich in den letzten Jahren das Konzept CrowdRE etabliert. Hier wird versucht, gezielt dem Nutzer eine Stimme im Anforderungs- und Verbesserungsprozess zu geben. Damit erhalten die Nutzer ganz gezielt ihre Stimme im Entwicklungsprozess [4]. Ziel hierbei ist es, zusätzliche Anforderungen, geänderte Bedürfnisse oder auch falsche Annahmen zu identifizieren, sowie Probleme bei der Verwendung des Produktes aus dem Weg zu räumen. Durch das gezielte Einbeziehen der Nutzermeinung ist es möglich, ein noch nutzerzentrierteres Produkt zu erlangen.

Daher gilt es, die Meinungen der Nutzer zu analysieren, Trends zu identifizieren und Prioritäten abzuleiten. Dies ist allerdings manuell häufig aus zeitlichen Gründen nicht möglich. Vollautomatisch lassen sich größere Mengen von Nutzerfeedback aktuell noch nicht auswerten. Die Entwicklung der „Natural Language Processing“ Techniken ist noch nicht fortgeschritten genug. So variiert beispielsweise das Ergebnis von unterschiedlichen Techniken zur Sentimentanalyse, bei denen die Texte zwischen positiv und negativ klassifiziert werden, sehr stark. [5] [6] [7].

Dennoch steckt Nutzerfeedback voller Potenziale und kann auch mit den heute verfügbaren Mitteln für die Produktverbesserung genutzt werden. Was wir sehr wohl erkennen können, sind Themen und Trends [8]. Genauer gesagt können Veränderungen der Stimmung im Feedback, ohne genaues Verstehen von Details, hergeleitet werden. Um den zahlreichen Quellen von Feedback Herr zu werden sowie um an verwendbare Ergebnisse zu kommen, benötigt man ein Maß, das intuitiv verstanden wird und überall anzutreffen ist. Beschränkt man sich nur auf Feedback, das in App Stores abgegeben wird, hat man die Fünf-Sterne-Bewertungsskala. Diese kann man dort nutzen, um eine Einstufung des Beitrages und die damit verbundenen Themen abzuleiten. Leider ist diese Skala nicht in allen Quellen, wie z. B. Soziale Medien, Feedbackforen oder Supportanfragen, vorhanden. Darüber hinaus hat ein Eintrag im App Store üblicherweise genau eine Bewertung. Dies macht es schwer Feedback zu klassifizieren, das sowohl positive als auch negative Aspekte enthält.

Wünschenswert wäre also ein Maß, das man ähnlich leicht auswerten kann wie Sternbewertungen und das idealerweise mehrfach innerhalb des Feedbacks vorhanden ist. So können die im Beitrag enthaltenen Themen jeweils unabhängig voneinander bewertet werden. Ist ein solches Maß gefunden, kann man Entwickler unterstützen und hierbei vor allem Produktmanagern die Möglichkeit geben, auf die Nutzermeinung fundierte Entscheidungen zur Qualitätssteigerung der eigenen Produkte zu fällen.

### 3 Emojis in Online-Texten

Wie wir gesehen haben, funktioniert zur Auswertung von Feedback die Berücksichtigung von Bewertungssternen nur eingeschränkt.

Betrachtet man textuelles Feedback näher, stellt man fest, dass es voller Emotionen ist. Ein Weg wie Nutzer ihre Emotionen deutlich machen, ist die Verwendung von Emojis.

Die Historie von Emojis und ihrer Vorgänger zeigt, dass diese schon immer das Ziel hatten, Emotionen innerhalb von Texten auszudrücken. Emotionen in Texten begreifbarer zu machen, ist auch gar nicht verwunderlich. Für uns Menschen ist es schwer, rein auf der Wortebene Emotionen in Texten zu erkennen [9]. Der Grund dafür ist, dass uns in Texten all die nonverbalen Hinweise fehlen, die wir sonst in der täglichen direkten Kommunikation unter uns Menschen haben. [10]. Schon im 19. Jahrhundert wurden in gedruckten Texten mit Hilfe von kleinen Druckzeichen Emotionen deutlich gemacht [11]. Zu dieser Zeit nannte man diese Urahnen der Emojis noch „Setzerscherze“. Mit der Entwicklung von Schreibmaschinen und später Computern wurden diese Bemühungen auf die dort verfügbaren Zeichen eingeschränkt. So wurde mit Aufkommen des Internets den Emoticons zur Berühmtheit verholfen. Emoticons sind Kombinationen aus ASCII Zeichen, die Gefühle und Emotionen deutlich machen. In Asien entwickelten sich danach deutlich komplexere Zeichenkombinationen Namens „Kaomojis“, die auf asiatische Zeichensätze oder gar Unicode setzen. Durch die Verfügbarkeit von Internetanschlüssen in Privathaushalten und den immer mehr genutzten Instant Messaging Diensten, wurden grafische Emoticons oder Smileys eingeführt, die im Wesentlichen aus einer Umwandlung der Emoticons in Bildchen bestanden, aber auch Neuschöpfungen enthielten.

In Japan designte schließlich Shigetaka Kurita 1999 für das aufkommende mobile Internet des Telekommunikationsanbieters NTT DoCoMo einer der ersten Emoji-Sammlungen [12]. Hier wurden zusätzliche Zeichen eingeführt, die zunächst als 12x12 Pixel große Grafiken vorlagen. Auch weitere Firmen setzen diese und ähnliche Zeichen ein. 2010 begann das Unicodekonsortium offiziell Emojis zu standardisieren und zu veröffentlichen [13]. Seitdem brachte dieses Konsortium immer wieder neue Versionen heraus. Ihren weltweiten Durchbruch hatten die Emojis mit der Einführung der Emoji-Tastatur 2011 mit iOS 5 [13]. Andere Systeme zogen nach und so sind Emojis aus der modernen Kommunikation nicht mehr wegzudecken. Hunderte von Milliarden Emojis werden täglich verschickt. Emojis werden nicht an beliebigen Stellen in Texten verwendet, sondern zur Unterstreichung von wichtigen Dingen [14].

Darüber hinaus findet man sie auch an Flughäfen zur Bewertung von Sicherheitskontrollen und Toiletten wieder, in Werbeplakaten [15] [16] sowie in einem eigens nach den Bildzeichen benannten Kinofilm [17]. Der Wunsch, Emotionen in Texten ausdrücken zu können, ist in der breiten Gesellschaft angekommen. Dies wurde auch dadurch honoriert, dass ein Bericht des Emoji Research Team [18] 2015 Emoji zur der schnellst wachsenden Sprache der Welt erklärte und im selben Jahr das Oxford Dictionary mit „Gesicht mit Freudentränen“ 🥳 ein Emoji zum Wort des Jahres wählte [19].

Da Emojis uns erlauben in Texten Emotionen auszudrücken, helfen sie daher in Feedback heraufzufinden, wie ein Nutzer zu bestimmten Aspekten eines Produktes steht. Auf die Emotionen der Nutzer zu hören, bietet die Möglichkeit stetig Produkte weiter zu

verbessern, aber die von den Nutzern geliebten Funktionalitäten trotzdem beizubehalten oder ihnen die notwendigen Prioritäten zuzuordnen. Da es inzwischen jedoch Tausende dieser Zeichen gibt, stellt sich die Frage, wie einheitlich Emojis verstanden werden. Dazu haben wir Emojis auf Emotionen untersucht.

#### 4 Wie nehmen Menschen Emojis wahr?

Um Emojis aus Nutzerfeedback dazu zu verwenden Produkte zu verbessern, stellt sich natürlich zunächst die Frage: Wie nehmen Menschen Emojis wahr? Dazu ist es notwendig, ein Modell zu erarbeiten, um Emojis kategorisieren zu können, sowie eine Liste von Emojis zu sammeln, welche berücksichtigt werden sollen. Zum Zweck der Einordnung von Emojis in unser Modell, haben wir eine Umfrage durchgeführt.

Es existieren bereits unterschiedliche Klassifikationen von Emotionen, die sich sowohl in der Art wie sie Emotionen kategorisieren, beispielsweise über hormonelle Vorgänge oder Gesichtszüge, als auch in der Granularität unterscheiden. Grundlegend kann man Emotionen zwischen Basisemotionen oder Primäremotionen [20] und sekundären Emotionen [21] unterscheiden. Bei den Basisemotionen geht man davon aus, dass sie kulturunabhängig sind, wohingegen sekundäre Emotionen über soziale Prägung und das Umfeld gelernt werden müssen. Allerdings gibt es keine übereinstimmende Klärung, welche Emotionen zu welcher Gruppe gehören [20] [21].

Für die Nutzung von Emotionen im Rahmen der Qualitätssicherung haben wir uns für ein Modell entschieden, das ähnlich zu den Basisemotionen nach Paul Ekman ist [22] und wie bei Feldmann [23] Charakteristiken der Sentimentanalyse, also der Analyse der Grundstimmung nach positiv, negativ oder neutral, enthält. Unser Emotionsmodell sieht wie folgt aus (siehe Abbildung 1): Es gibt zwei Ebenen. Zuerst findet eine Einteilung in die Grundstimmung statt. Es wird zwischen „positiv“, „neutral“ und „negativ“ unterschieden. Die zweite Ebene ist die Emotionsebene und fungiert als Unterkategorie der Stimmung oder des Sentiments. Sie enthält im Wesentlichen Basisemotionen, welche einfach zu verstehen und zu unterscheiden sind.

Die Stärke des verwendeten Modells ist die Verwendung von Basisemotionen sowie die Verbindung von Sentiments – also, ob

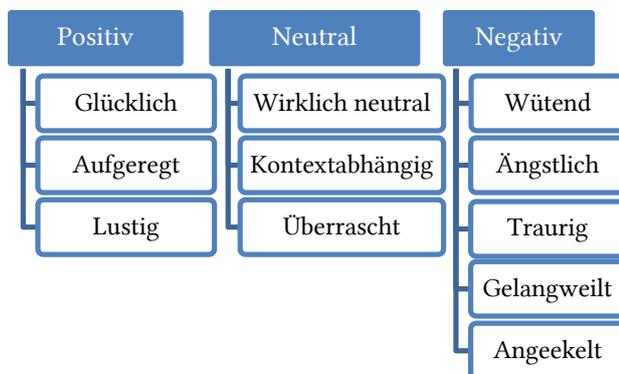


Abbildung 1: Emotionsmodell

etwas positiv, negativ oder neutral ist - und Emotionen. Anschließend haben wir Emojis und Emoticons gesammelt. Hierbei haben wir sowohl Listen mit Emoticons als auch den Emoji Standard 5.0 verwendet. Dies war zum Zeitpunkt der Konzeption der Studie der neuste Standard, der auf den gängigen mobilen Plattformen verbreitet war [24], [25]. Emoji 5.0 bietet 2667 Emojis [26]. Dazu kommen potentiell hunderte Emoticons.

Da Tausende dieser Zeichen zu viele sind, galt es sie sinnvoll zu reduzieren. Wir haben alle Emojis, die keine Emotion übermitteln, wie Gegenstände, Buchstaben, Wortersetzungen oder Flaggen, gestrichen. Die Ausgangsmenge reduzierten wir dadurch auf 717. Da es für viele Emojis Synonyme, sowie alternative Darstellungen gibt, haben wir anschließend die verbliebenen Emojis in Gruppen eingeordnet. Dabei wurden Emojis mit gleicher Bedeutung, wie den „Konfettiball“ 🎉 und den „Party Popper“ 🎊, zusammengefasst. Außerdem haben wir Emojis, die sich nur in ihrer Haut-, Haarfarbe oder dem Geschlecht unterscheiden, gruppiert. Jede Gruppe erhielt einen Repräsentanten, dessen Kategorisierung auf die restlichen Emojis der Gruppe übertragen werden kann. In der Regel wurde so vorgegangen, dass als Repräsentant das neutrale Geschlecht und die neutrale Farbe gewählt wurde. War dies in der Gruppe nicht möglich, wurde das bekannteste Emoji gewählt. Die Wahl der Repräsentanten und der Gruppen wurde unabhängig von drei Mitarbeitern geprüft. Durch die Gruppierung der Emojis kamen wir auf 132 Gruppen mit jeweils einem Repräsentanten.

Um eine Zuordnung zwischen dem Emotionsmodell und den gesammelten Emojis zu erhalten, haben wir eine großflächige Umfrage durchgeführt. Die Umfrage erlaubt uns, ein allgemeines Bild der Wahrnehmung von Emojis zu erlangen. Hierfür sollten Teilnehmer Emojis nach Stimmung und nach Emotion gemäß dem erwähnten Emotionsmodell einteilen (siehe Abbildung 2).

Abbildung 2: Ausschnitt aus dem Emoji Fragebogen



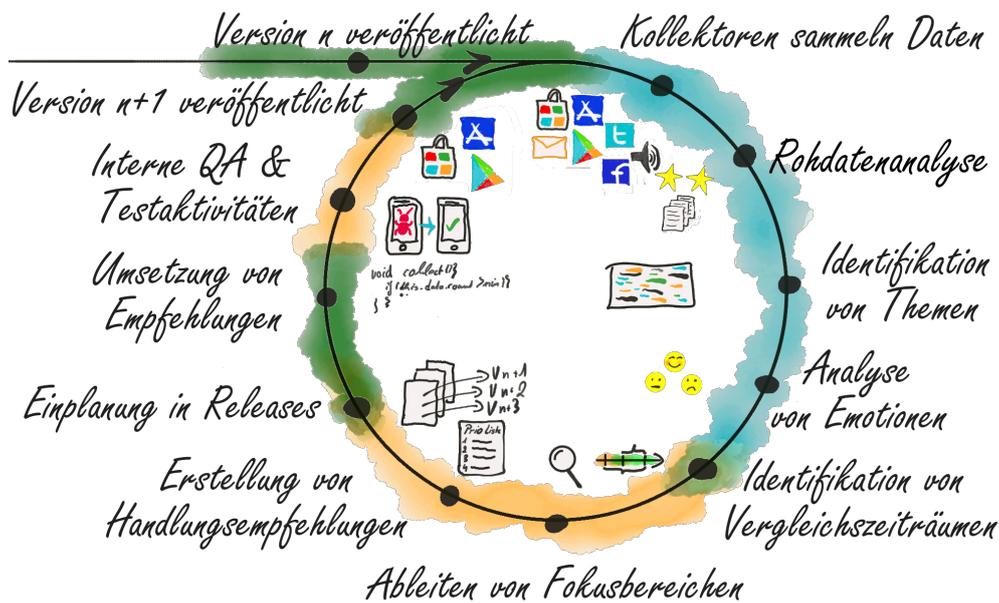


Abbildung 4: Prozess zur Qualitätsverbesserung durch Trenderkennung in Nutzerfeedback

Um sicherzustellen, dass Emojis wie die Sternebewertung ein gutes Mittel sind, Feedback zu klassifizieren, haben wir für 178.848 Feedbackbeiträge zu diversen Apps Sternebewertung und Sentiment der Emojis im Feedbackbeitrag korreliert. Hieraus ergab sich eine sehr starke Korrelation beider Werte. Damit sind Emojis als Bewertungsindikator vergleichbar zur Sternebewertung und können somit ergänzend oder auch alternativ genutzt werden.

## 5 Produktverbesserung durch Analyse von Nutzerfeedback

Basierend auf unseren Erkenntnissen zu Emojis haben wir einen Analyseprozess für Nutzerfeedback definiert (Siehe Abbildung 4). Zentrales Ziel ist es, die Entwicklungsorganisation möglichst wenig durch zusätzliche Aufgaben zu belasten, jedoch die Qualität der Software zu sichern. Kernidee ist, dass Nutzerfeedback parallel zur Entwicklung neuer Versionen der Produkte gesammelt und ausgewertet wird. Hierbei verwenden wir eine Mischung aus vollautomatisierten (blau), teilautomatisierten (beige) und manuellen (grün) Schritten. Insgesamt lässt sich der Prozess in drei Phasen einteilen: (1) Sammeln von Feedback, (2) Datenanalyse von Feedback und die anschließende (3) Produktverbesserung. Der Prozess wird durch unsere Software-Lösung *User Echo Service* unterstützt (siehe Abbildung 5). Mit Hilfe von diesem Werkzeug sammeln wir Feedback, werten es aus und binden Cloud-Dienstleister an um deren Analysen zu verwenden. Darüber hinaus enthält es Funktionen zur Verwaltung von Produkten und den zugehörigen Feedbackprojekten sowie den eigentlichen Feedbackdaten.

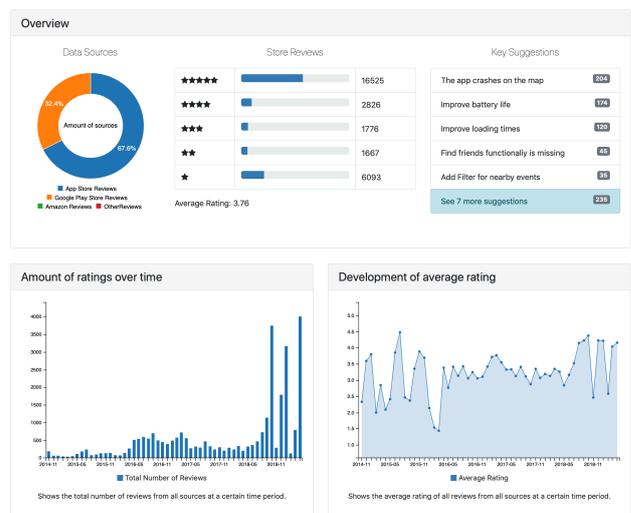


Abbildung 5. Auszug aus der Ergebnissicht des User Echo Services

### 5.1 Sammeln von Feedback

Kollektoren sammeln Feedback aus zahlreichen Datenquellen wie Stores, sozialen Medien, Foren etc. ein. Konkret setzen wir hierbei derzeit auf den App Store, Google Play, Amazon Market Place, Instagram und Twitter. Entsprechende Kollektoren für YouTube und Facebook sind in der aktiven Entwicklung. Darüber hinaus haben wir einen Mechanismus entwickelt, bei dem beliebiges Feedback z.B. aus dem Kundensupportbereich oder E-Mail-

Verkehr eingefügt werden kann. Damit sind wir in der Lage, dieses auch ohne Datenkollektor zu integrieren.

Es ist wichtig, dass sowohl Feedback aus unterschiedlichen Quellen hinterher miteinander verglichen und vernetzt werden kann, als auch keine speziellen Eigenheiten der Quelle durch Abstraktionen zu verlieren. Hierfür nutzen wir ein Datenstruktur, die die Charakteristiken von Feedback allgemein abbilden kann [8]. Damit ist es möglich, Feedback über Quellen hinweg zusammen zu betrachten, trotzdem aber keine Daten zu verlieren

## 5.2 Datenanalyse

In der Rohdatenauswertung wird bestimmt, welches Feedback aus welcher Quelle zu welchen Zeitpunkten vorkommt und ob eventuell im Feedback selbst schon Bewertungsindikatoren, wie Bewertungssterne oder Zustimmungen anderer Nutzer, vorhanden sind. Je nach Datenquelle werden aber auch Konfidenzmaße ausgewertet, wie Hilfreich-Bewertungen, Kommentare oder Likes. Diese Maße sagen nichts über die Bewertung des Produktes aus, sondern drücken den Grad des Interesses anderer Menschen zu dem gegebenen Feedback aus. Damit ist es möglich, einzelnen Beiträgen ein höheres oder geringeres Gewicht zu geben.

Im nächsten Schritt identifiziert der Prozess automatisch Themen innerhalb des Feedbacks. Themen können beispielsweise einzelne Features, Fehler, Anwendungsfälle oder Screens der Anwendung sein. Zur Themenerkennung verwenden wir standardisierte Ansätze des so genannten Topic-Modelings wie die Erstellung von Bag-of-Words. Ziel dieses Schrittes ist es ein besseres Verständnis zu erlangen, auf welche Aspekte der App sich die Nutzer fokussieren.

Der nächste Schritt ist die Erkennung von Emotionen mittels Emojis innerhalb der Feedback-Einträge und innerhalb der zuvor erstellten Themen. Hierbei klassifizieren wir sowohl nach Emotionen, aber auch nach der Stimmung, um so möglichst breit bewerten zu können. Dazu nutzen wir die Ergebnisse aus den in Abschnitt 4 beschriebenen Vorarbeiten. Spätestens hiermit ist das Feedback mit entsprechend bewertenden Indikatoren versehen. Die Nutzung von Emojis als bewertender Charakter erlaubt uns auch, im Gegensatz zu Metriken wie der Sternbewertung, das Feedback in positive und negative Aspekte zu zerlegen.

Bis zu diesem Zeitpunkt läuft der Prozess vollständig automatisiert ab.

Danach wird das Feedback mit Blick auf frühere Produktversionen ausgewertet. Je nach Thema werden geeignete Vergleichszeiträume gewählt. Damit werden Stimmungstrends, -Wendungen sowie Veränderungen der Menge innerhalb des Feedbacks erkannt. Dadurch werden die von den Nutzern diskutierten Themen um Trends angereichert.

Basierend auf diesen Analyseschritten leitet unser Ansatz Fokusbereiche für den Produktmanager ab. Ein Fokusbereich ist eine Teilmenge des Feedbacks, welches Aufmerksamkeit bedarf. Fokusbereiche an sich sind wertneutral zu sehen, da sowohl positive als auch negative Wendungen im Feedback aufgezeigt werden. Insbesondere positive Wendungen sind essentiell, um zuvor vorgenommene Änderungen am Produkt einordnen zu können. Die Analyse von Feedback soll schließlich nicht nur

aufzeigen, was am Produkt geändert werden sollte, sondern auch, welche Änderungen von den Nutzern gut aufgenommen wurden oder welche Bereiche des Produktes die Nutzer mögen.

## 5.3 Produktverbesserung

Die anschließende Ableitung von Handlungsempfehlungen ist auf Probleme im Produkt ausgerichtet. Ziel ist es, Empfehlungen zur Produktverbesserung abzuleiten und diese zu priorisieren. Die Empfehlungen können zum einen potenzielle neue User Stories sein, Change Request zu bestehenden Stories oder auch Bug Reports. Für den Fall, dass eine Handlungsempfehlung die Beseitigung eines Fehlers im Produkt ist, wird der Fehler klassifiziert und festgehalten. Damit entsteht auf lange Sicht eine Sammlung von Fehlern und Fehlermustern zum Produkt. Ist die Handlungsempfehlung eine Abänderung des Produktes, werden neue Anforderungen für die Entwickler abgeleitet.

Basierend auf den Handlungsempfehlungen erfolgt ein Hinzufügen zum Release-Plan. Daraufhin findet eine Umsetzung der Empfehlungen statt. Bei der folgenden Qualitätssicherung des Produktes wird zum einen traditionelle Qualitätssicherung betrieben, aber auch ganz bewusst auf die abgeleiteten Handlungsempfehlungen geachtet. Zusätzlich werden die gesammelten Fehlermuster geprüft, um so das Wiederauftreten von alten Fehlern zu vermeiden. Wenn die Version den gewünschten Qualitätsstandards entspricht, wird sie veröffentlicht.

## 6 Beispiele für Erkenntnisse aus Nutzerfeedback

Erfahrungen mit unserem Prozess haben wir bei der Untersuchung von Apps aus dem App Store gesammelt. Hierbei war unser Ziel, kurzzeitige und spontane Wendungen im Nutzerfeedback zu erkennen. Es gibt zahlreiche Apps, bei denen plötzliche Bewertungseinbrüche existieren bzw. es spontane Anstiege in der Menge an Feedback gibt. Dies gilt auch für Themen, die unser Ansatz erkennt.

Die App Snapchat beispielsweise hatte am 27.7.2017 eine außergewöhnlich hohe Menge an Feedback. Ursache war eine temporäre Störung des Dienstes. Einen ähnlichen Bewertungsverlauf konnten wir Ende Juni 2017 beobachten, als die App ein Update erhielt, das es ermöglichte, seine Freunde auf einer Kartenansicht darzustellen. Die Nutzer schienen nicht zu wissen, wie die Funktion zu benutzen ist und es gab einen Einbruch der Durchschnittsbewertung sowie verstärkt negative Emojis mit negativer Stimmung.

Des Weiteren untersuchten wir Instagram, dessen Bewertungsverlauf zwei deutliche Einbrüche Anfang und Ende Juli 2017 zeigt, die gleichzeitig mit einer erhöhten Menge an Feedback und vielen negativen Emotionen einhergingen. Bei der ersten Situation handelte es sich um ein kurzzeitiges Problem des Dienstes, bei dem Benutzerkonten irrtümlich als gelöscht markiert wurden und so eine Nutzung des Dienstes nicht mehr möglich war. Die zweite Situation war eine etwa zwölf Stunden andauernde Downtime für einen Teil der Nutzer.

WhatsApp vollzog Februar 2017 ein größeres Update, das die Art, wie Statusnachrichten verwendet werden sollten, veränderte. Statt eines Langzeitstatustextes ermöglichte man den Nutzern nun Bilder als Kurzzeitstatus zu verwenden. Dafür wurde auch eine eigene Rubrik in der App eingerichtet. Eine ähnliche Funktion bot zuvor schon Snapchat und Instagram. Interessanterweise fassten die Nutzer diese Änderung im ersten Moment sehr negativ auf. Binnen kurzer Zeit meldeten sich zahlreiche wütende Nutzer zu den Themenfeldern „Status“, „Update“, „Snapchat und „Instagram zu Wort“. Hierbei wurde der Unmut über die neue Funktion deutlich gemacht, das Fehlen des alten Dauerstatus und dass man versuche, zwei Apps zu klonen. WhatsApp reagierte auf die heftige Kritik mit einer Kompromisslösung. Man behielt zwar die zunächst ungeliebte Änderung bei, führte aber den alten Status zusätzlich ein. Dies führte zu einem abrupten Anstieg von positivem Feedback sobald die Änderung angekündigt wurde. Die Nutzer gaben der App für diese Entscheidung Vorschusslorbeeren und begrüßten das Vorhaben, ehe es umgesetzt war.

Uber zeigte einen deutlichen Anstieg von Feedback im Januar 2017. Die Menge von Feedback erhöhte sich von 1.903 Einträgen im Dezember 2016 auf 3.984 Einträge im Januar 2017 und fiel dann wieder auf 1.840 Einträge im folgenden Monat. Die Ergebnisse im Dashboard zeigen in diesem Monat eine häufige Verwendung des Themas „löschen“. Tatsächlich waren viele Nutzer sehr verärgert über die Zusammenarbeit zwischen Ubers CEO und einem Politiker, weshalb Uber viele Nutzer verlor. Im Februar trat CEO aufgrund dieser Ereignisse zurück [27]. Wir haben auch beobachtet, dass Uber oft mit seinem Mitbewerber Lyft verglichen wird. Als Nutzer sich für die Löschung von Uber entschieden, gaben viele an, sie seien zu Lyft gewechselt. Feedback über den Wechsel zu Lyft oder den Vergleich zwischen Uber und Lyft hat eine Durchschnittsbewertung von 1,28 Sternen und enthält meist negative Emotionen. Dieses Beispiel zeigt, dass Nutzerfeedback Einblicke in die Beziehung zwischen Nutzern und dem Ruf und der Ausrichtung einer Firma gibt sowie in konkurrierende Produkte.

Pinterest erlebte einen Einbruch im Bewertungsverlauf von 3,7 auf 3,0, als der „like“-Button aus der App entfernt wurde [28]. Im Juni 2017 gab es 722 Feedback Einträge, die das Wort „Button“ enthielten und viele Nutzer beschwerten sich darin über den fehlenden Button. Die Emotionen in diesem Feedback waren meist negativ, wobei „traurig“ die häufigste Emotion war. Eine mögliche Handlungsempfehlung für dieses Thema wäre das Zurückbringen des „like“-Buttons.

Tinder war im März 2018 von einem Problem betroffen, als die durchschnittliche Sterne-Bewertung um 0,8 auf 2,5 sank. In diesem Monat war die Menge an Feedback mehr als dreimal so hoch wie im vorherigen. Unsere Analyse zeigt zu dieser Zeit eine häufige Verwendung der Themen „Profil“ und „fake“ und der Großteil des Feedbacks, das eines dieser Themen enthält, stammt vom März 2018. Darüber hinaus war „fake“ in vielen Einträgen mit „Profil“ kombiniert. In der Tat gab es eine große Menge an Feedback, in dem Nutzer sich über zu viele Fake-Profile auf Tinder beschwerten. Die entsprechende Handlungsempfehlung wäre die Implementierung eines Features zur Reduktion von solchen Profilen oder eine Echtheitsprüfung. Alternativ könnte ein Feature

entwickelt werden, das „verifizierte Profile“ zeigt. In den folgenden Monaten nahm die Menge an Feedback über das Thema „fake“ deutlich ab.

In April war Tinder mit einem Problem ganz anderer Natur konfrontiert. Die Bewertungen brachen plötzlich ein und die Emotionen des Feedbacks waren voller „trauriger“ und „wütender“ Emotionen. Die zugehörigen Rezensionen zeigten, dass dies auf ein Problem mit dem Login zurückzuführen war. Nutzer waren permanent mit einem aufploppenden Facebook-Login-Dialog konfrontiert. Die Ursache hierfür lag nicht etwa bei der App selbst, sondern bei Facebook. Facebook veränderte unerwartet seine API. Dies führte dazu, dass der Tinder-Login-Mechanismus nicht mehr richtig funktionierte [29]. Das Problem begann am 3. April aufzutreten und erst am 6. April war Tinder in der Lage ein entsprechendes Update zu veröffentlichen. Dieses Problem macht deutlich, dass in Zeiten von Software-Ökosystemen auch schnell auf die Änderungen von Dritten reagiert werden muss, die möglicherweise Kernfunktionalitäten des eigenen Produkts zerstören.

Die aufgezeigten Beispiele machen deutlich, dass Feedback sehr vielschichtig Probleme in Apps aufzeigen kann. Diese Probleme müssen nicht zwangsläufig rein in der Mobilanwendung zu tun haben. Dinge wie technische Infrastrukturen, Serververbindungen oder externe Partner dafür sind ein erheblicher Faktor, der nicht zur Entwicklung kontrolliert werden kann. Trotzdem sind Fehler für den Nutzer primär immer in der App zu erkennen und dorthin wird dann auch das entsprechende Feedback gesendet. Apps sind immer mehr Teil von größeren Systemen, die eine sofortige Reaktion auf Probleme erforderlich machen. Daher ist die unmittelbare Erkennung von Trends in Daten wichtig, bevor längere Auswirkungen auf die Nutzer aufgetreten sind.

## 7 Fazit

Die Qualität von Apps kann durch eine kontinuierliche Einsammlung, Analyse und Einarbeitung von Nutzerfeedback verbessert werden, wenn Feedback Teil des Produktentwicklungszyklusses wird. Da viele Apps kontinuierlich weiterentwickelt werden, ist es notwendig, mit entsprechender Agilität auf die Nutzerwünsche einzugehen. Um dies zu ermöglichen, haben wir die Notwendigkeit erkannt, leichtgewichtige Textanalysemethoden zu entwickeln. Dazu haben wir einen Qualitätssicherungsprozess definiert. Ein Kernbestandteil des Prozesses bildet die Analyse von Emojis innerhalb des Feedbacks und die schnelle Erkennung von Trends, um auf dieser Basis, Handlungsempfehlungen ableiten zu können. Damit dies möglich ist, haben wir ein Emotionsmodell erstellt und eine Umfrage zu Emojis durchgeführt. Kernergebnis der Umfrage ist, dass Menschen eine sehr ähnliche Wahrnehmung von Emojis haben. Folglich können große Mengen an Texten auf Grund der verwendeten Emojis schnell analysiert werden. Dies ermöglicht eine umfangreichere Kategorisierung von Texten, verglichen mit der regulären 5-Sternebewertung, die beispielsweise im Apple App Store üblich ist.

Die Ergebnisse flossen in unseren Prozess zur Trenderkennung in textuellem Feedback ein. Zentrale Anforderung an den Prozess

sowie an die technische Umsetzung ist eine einfache Bedienbarkeit und die Einsatzfähigkeit für hochiterative Produktentwicklung.

Die Analyse von Emojis erlaubt uns dort Feedback zu analysieren, wo keine bewerteten Metriken, wie Bewertungsterne, mit dem Feedback mitgeliefert werden. Dies trifft insbesondere auf Feedback im Supportbereich und soziale Medien zu.

Der Vergleich von Sternebewertung und des emotionalen Sentiments zeigt eine starke Korrelation der beiden Metriken auf. Damit können wir eine stärker automatisierte Auswertung von Nutzerfeedback in heterogenen Datenquellen realisieren. Zahlreiche Beispiele zeigen auf, dass unser Prozess bei realen Produkten und deren Feedback wirken kann. Insbesondere ist es auch möglich, kurzzeitig auftretende Ereignisse im Nutzerfeedback messen zu können. Damit entsteht nicht nur das Potenzial große und langanhaltende Anliegen der Nutzer zu erkennen, sondern auch schnell auf neue Trends reagieren zu können. Wenn man für sein Produkt Feedback kontinuierlich sammelt und auswertet, entsteht ein Datenstrom, der recht schnell zu einer umfassenden Wissensdatenbank über das Produkt wird. Hierbei können auch immer wieder die Anforderungen verifiziert und validiert werden. Entwickler werden in die Lage versetzt, von früheren Ereignissen zu lernen und in der Zukunft das Produkt in einer noch besser zu den Nutzern passenden Art und Weise zu entwickeln. Natürlich gilt es hierbei zu bedenken, dass nicht jedes Feedback-Element naiv und blind direkt in das Produkt fließen soll. Wie Studien zeigen, geben insbesondere zufriedene, aber auch unzufriedene Nutzer mehr Feedback als Personen im Mittelfeld [30] [31]. Daher ersetzt die Analyse von Feedback nicht das Durchführen von solider Anforderungserhebung und User Research.

Einer der nächsten Schritte ist Feedback über verschiedene Datenquellentypen wie App Stores, soziale Medien und Kundenforen hinweg, gezielt zu vergleichen. Ein weiteres spannendes Themenfeld ist die Erstellung von Marktübersichten durch den Vergleich verschiedener Konkurrenzprodukte.

Da die Menge an Emojis stetig weiterwächst, haben wir eine Folgeumfrage in Arbeit. Im Wesentlichen besteht diese aus der Klassifizierung der im Herbst 2018 neu eingeführten Emojis, sowie detaillierten Fragen zu den nach unserer Studie schwer einzuordnenden Emojis. Hierbei wollen wir explizit den Kontext der Nutzung dieser Emojis erfahren, um sie in ihrer Einordnung besser zu verstehen. Das Verstehen der Kontextabhängigkeit ist auch für die Emojis relevant, die von den Teilnehmern sowohl als positiv als auch negativ eingeordnet wurden. Dadurch können sich Texte noch genauer klassifizieren lassen und man kann damit Nutzerfeedback besser verstehen.

## ACKNOWLEDGMENTS

Dieser Beitrag wurde im Kontext der Projekte EnStadt: Pfaff (Förderkennzeichen: 03SBE112D und 03SBE112G) gefördert vom Bundesministerium für Wirtschaft und Energie und Opti4Apps, gefördert durch das Bundesministerium für Bildung und Forschung (Förderkennzeichen: 02K14A182) erstellt.

## REFERENZEN

- [1] S. Hassan, C. Tantithamthavorn, C.-P. Bezemer und A. E. Hassan, „Studying the dialogue between users and developers of free apps in the Google Play Store,“ *Empirical Software Engineering*, Bd. 23, Nr. 3, pp. 1275–1312 <https://doi.org/10.1007/s10664-017-9538-9>, 2018.
- [2] M. Nayebi, H. Cho und G. Ruhe, „App store mining is not enough for app improvement,“ *Empirical Software Engineering*, Bd. 23, Nr. 5, p. 2764–2794, 2018.
- [3] G. Deshpande und J. Rokne, „User Feedback from Tweets vs App Store Reviews: An Exploratory Study of Frequency, Timing and Content,“ in *2018 5th International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)*, Banff, Canada, 2018.
- [4] E. C. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol, A. Perini und M. Stade, „The Crowd in Requirements Engineering,“ *IEEE Software*, Bd. 34, Nr. 2, pp. 44–52 DOI: 10.1109/MS.2017.33, 28 03 2017.
- [5] A. Hogenboom, M. Bal, F. Frasinca und D. Bal, „Towards cross-language sentiment analysis through universal star ratings,“ *Advances in Intelligent Systems and Computing*, Bd. 172, pp. 69–79, 2013.
- [6] F. N. Ribeiro, M. Araújo, M. A. Gonçalves und F. Benevenuto, „SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods,“ *EPJ Data Sci*, Bd. 5, Nr. 1, 2016.
- [7] R. Santos, E. C. Groen und K. Villela, „An Overview of User Feedback Classification Approaches,“ in *Joint Proceedings of REFSQ-2019 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 25th International Conference on Requirements Engineering: Foundation for Software Quality*, Essen, Germany, 2019.
- [8] S. A. Scherr, F. Elberzhager und L. Müller, „Quality Improvement of Mobile Apps - Tool-Supported Lightweight Feedback Analyses,“ in *PROFES 2018*, Wolfsburg, Germany, 2018.
- [9] J. Park, V. Barash, C. Fink und M. Cha, „Emoticon style: Interpreting differences in emoticons across cultures,“ in *Seventh International AAAI Conference on Weblogs and Social Media*, Cambridge, Massachusetts, USA, 2013.
- [10] L. Rezabek und J. Cochenour, „Visual Cues in Computer-Mediated Communication: Supplementing Text with Emoticons,“ *Journal of Visual Literacy*, 1998.
- [11] Keppler & Schwarzmann, „Typographical Art,“ *Puck*, Bd. IX, Nr. 209, p. 65, 1881.
- [12] A. Sternbergh, „Smile, You're Speaking Emoji. The rapid evolution of a wordless tongue,“ 16 November 2014. [Online]. Available: <http://nymag.com/daily/intelligencer/2014/11/emojis-rapid-evolution.html>. [Zugriff am 28 March 2019].
- [13] A. Pardes, „The wired guide to emoji,“ 2018. [Online]. Available: <https://www.wired.com/story/guide-emoji/>. [Zugriff am 28 March 2019].
- [14] R. R. Provine, R. J. Spencer und D. L. Mandell, „Emotional Expression Online,“ *J. Lang. Soc. Psychol.*, Bd. 26, Nr. 3, pp. 299–307, 2007.
- [15] S. Haysom, „British graffiti artist pranks McDonald's emoji billboard,“ *Mashable*, 15 07 2015. [Online]. Available: <https://mashable.com/2015/07/15/mcdonalds-emoji-advert-trolled-graffiti-artist/?europa=true>. [Zugriff am 16 11 2018].
- [16] S. Spary, „Do you think this McDonald's ad that turns people into emojis is creepy?,“ *Campaign*, 14 08 2015. [Online]. Available: <https://www.campaignlive.co.uk/article/think-mcdonalds-ad-turns-people-emojis-creepy/1360121>. [Zugriff am 16 11 2018].
- [17] IMDb, „The Emoji Movie,“ [Online]. Available: <https://www.imdb.com/title/tt4877122/>. [Zugriff am 28 March 2019].
- [18] Emogi, „2015 Emoji Report,“ 2015.
- [19] Oxford University Press, „Word of the Year 2015,“ 2015. [Online]. Available: <https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2015>. [Zugriff am 16 11 2018].
- [20] American Psychological Association, „primary emotion,“ [Online]. Available: <https://dictionary.apa.org/primary-emotion>. [Zugriff am 12 11 2018].

- [21] American Psychological Association, „secondary emotion,“ [Online]. Available: <https://dictionary.apa.org/secondary-emotion>. [Zugriff am 12 11 2018].
- [22] P. Ekman und W. V. Friesen, „Constants across cultures in the face and emotion,“ *J. Pers. Soc. Psychol.*, Bd. 17, Nr. 2, pp. 124-129, 1971.
- [23] R. Feldman, „Techniques and applications for sentiment analysis,“ *Communications of the ACM*, Bd. 56, Nr. 4, 2013.
- [24] Emojipedia, „Google Android 9.0,“ [Online]. Available: <https://emojipedia.org/apple/ios-12.1/>. [Zugriff am 28 March 2019].
- [25] Emojipedia, „Apple iOS 12.1,“ [Online]. Available: <https://emojipedia.org/google/android-9.0/>. [Zugriff am 28 March 2019].
- [26] Emojipedia, „Emoji Version 5.0,“ [Online]. Available: <https://emojipedia.org/emoji-5.0/>. [Zugriff am 19 11 2018].
- [27] J. C. Wong, „Uber CEO steps down from Trump advisory council after users boycott,“ 3 2 2017. [Online]. Available: <https://www.theguardian.com/technology/2017/feb/02/travis-kalanick-delete-uber-leaves-trump-council>. [Zugriff am 02 07 2017].
- [28] K. O., „Goodbye, Like button,“ 20 4 2017. [Online]. Available: <https://newsroom.pinterest.com/en/post/goodbye-like-button>. [Zugriff am 5 07 2017].
- [29] J. Fingas, „Tinder suffers sign-in problems following Facebook's privacy changes,“ *engadget*, 04 04 2018. [Online]. Available: <https://www.engadget.com/2018/04/04/tinder-sign-in-problems-following-facebook-changes/>. [Zugriff am 30 05 2018].
- [30] J. A. Chevalier und D. Mayzlin, „The Effect of Word of Mouth on Sales: Online Book Reviews,“ *Journal of Marketing Research*, Bd. 43, Nr. 3, pp. 345-354, 2006.
- [31] N. Hu, P. A. Pavlou und J. Zhang, „Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication,“ in *Proceedings of the 7th ACM conference on Electronic commerce*, Ann Arbor, Michigan, USA, 2006.