

Erfahrungsberichte aus zweiter Hand: Erkenntnisse über die Autorschaft von Arztbewertungen in Online-Portalen

Michaela Geierhos & Frederik S. Bäumer, Universität Paderborn

Für den Internetnutzer¹ entstehen immer mehr Möglichkeiten, Bewertungen über eine Vielzahl an Produkten (z. B. Amazon Reviews), Leistungen (z. B. jameda) und Erlebnissen (z. B. TripAdvisor) abzugeben. Diese suchen Bewertungsplattformen auf, um aktiv ihre Erfahrungen mit Dienstleistungen wie z. B. im Versandhandel, mit Hotelurlaube oder von Arztbesuchen mit anderen zu teilen. Häufig bestehen diese Bewertungskommentare aus Freitexten (sog. User Generated Content), die in Struktur und inhaltlicher Fokussierung deutlich voneinander abweichen. Vor der wissenschaftlichen Interpretation dieser Erfahrungsberichte steht jedoch der Prozess der Datenerhebung – die Entscheidung darüber, was die eigentlichen *Daten* für bestimmte Forschungszwecke sind. Was wir als Computerlinguisten unter empirischen Daten verstehen, und wie wir und andere Disziplinen den Prozess der Datenbeschaffung bezeichnen (z. B. Datenakquise vs. Datenkonstruktion), unterscheidet sich stark davon, wie Daten z. B. in der Soziologie erhoben und wie dieser Prozess dort verstanden wird.

Daten – Informationen – Erkenntnisse

Daten sind für uns nur Angaben über Sachverhalte und Vorgänge, die in einer Datenbank gespeichert werden. So liegen dieser Studie zufällig ausgewählte Datensätze der Portale jameda und DocInsider aus den Jahren 2009 bis 2015 zugrunde. Das Korpus umfasst 860.000 individuelle Erfahrungsberichte, die allesamt nach einem Arztbesuch in Deutschland verfasst und online gestellt wurden. Jede Arztbewertung besteht dabei aus einer Überschrift, dem eigentlichen Text und bis zu 17 verschiedenen numerischen Bewertungskategorien (u.a. Behandlung, Vertrauensverhältnis, Wartezeit, Barrierefreiheit). Bei der computergestützten Datenakquise wurden diese authentischen, online-verfügbaren Datensätze (sog. empirische Daten) originalgetreu zur weiteren Auswertung lokal bereitgestellt.

Erst durch die kognitive Verarbeitung des Lesers können aus diesen Daten *Informationen* werden. Hierfür setzt die Computerlinguistik Methoden zur Informationsextraktion ein, um sowohl Informationen aus den Meta-Daten (z. B. Erstellungsdatum, Name und Alter der Autoren) als auch aus dem Inhalt von Datenformaten (z. B. HTML) zu gewinnen. Anfangs wird definiert, welche Arten von Informationen extrahiert werden sollen. Auf diese Weise ist es möglich, aus Online-Arztbewertungen beispielsweise auszulesen, wer behandelt wurde, was die Symptome waren oder die Krankheitsgeschichte ist, wie (un)zufrieden derjenige war (z. B. „lange Wartezeit“; „kompetente Behandlung“), und ob er den Arzt weiterempfehlen würde. Hierzu müssen einerseits Zuordnungsregeln formuliert werden, die darüber entscheiden, welche „Daten“ welche *Informationen* repräsentieren, wie beispielsweise „Kassenpatient“ → *Gesetzliche Krankenversicherung*. Andererseits ist zu beachten, dass der quantitative Teil der Daten in Abhängigkeit der Ursprungsplattform (jameda oder DocInsider) anders zu interpretieren ist. Während jameda mit Schulnoten arbeitet, setzt DocInsider Sterne zur Bewertung der unterschiedlichen Aspekte beim Arztbesuch ein. Aufgrund dessen ist eine portalspezifische Abbildungsfunktion von Zahlen auf Denotationen zu definieren, so dass sichergestellt wird, dass eine 1,0 bei jameda der Bedeutung von „sehr gut“ entspricht, wohingegen derselbe numerische Wert bei DocInsider ein „ungenügend“ repräsentiert.

Anschließend können die Bewertungstexte erst maschinell analysiert werden, um *Erkenntnisse* über deren Inhalte zu gewinnen. Dies birgt zwar einige Herausforderungen hinsichtlich der Subjektivität (Wiebe, 2004; Bruce & Wiebe, 1999), Polarität (Bakliwal et al., 2012; Kim & Hovy, 2006; Qin et al.,

¹ Aus Gründen der leichteren Lesbarkeit wird auf eine geschlechtsspezifische Differenzierung verzichtet. Entsprechende Begriffe gelten im Sinne der Gleichbehandlung für beide Geschlechter.

2008; Pang et al., 2002) und der Semantik (Hu & Liu, 2004b; Kim & Hovy, 2006; Turney, 2002) der Aussage(n), führt aber nach Anwendung von automatischen Textanalyseverfahren zu qualitativ hochwertigen, interdisziplinär verwertbaren Ergebnissen. Zwar untersucht die computerlinguistische Forschung bereits automatisiert die Zufriedenheit von Konsumenten hinsichtlich bestimmter Produkte oder Dienstleistungen (Gamon, 2004; Hu & Liu, 2004a), unterstützt aber bisher nicht die Identifikation von thematischen Rollen in Bewertungstexten. Die Forschung im Bereich der Bewertungskultur im Web 2.0 deckt bereits eine Vielzahl unterschiedlicher Disziplinen (u.a. Soziologie, Marketing, Psychologie, Linguistik, Informatik) und verschiedene Aspekte (u.a. Bewertungsgegenstand, demographische Einflüsse, Emotionserkennung) ab. Selbst im Bereich des Textmining in Sozialen Medien werden Bewertungen und Kommentare vorwiegend auf Polarität (He et al., 2011; Jiang et al., 2011) und/oder Inhalt (Kushal et al., 2003) automatisch analysiert. Hierbei werden in der Regel die offensichtlichen sprachlichen Indikatoren (z. B. Emoticons, positive und negative Wörter) zur Identifizierung der Grundstimmung in den einzelnen Texten berücksichtigt (sog. Sentiment Analysis). Vereinzelt gibt es jetzt auch Arbeiten, welche sich mit der Stilistik von Bewertungstexten (Ludwig et al., 2013) auseinandersetzen. Während Autoreninformationen meist zur Erkennung von gefälschten Bewertungen (Liu, 2012:127; O'Connor, 2010; Hu et al., 2011) eingesetzt werden, konzentrieren wir uns in diesem Beitrag auf die Identifikation von Akteuren (sog. Rollen) im Online-Ärztewertungsprozess.

Rezensenten – Patienten – Ärzte

Vermutlich sind es Patienten, die ihre Ärzte bewerten und die eigenen Erfahrungen nach dem Arztbesuch weitergeben möchten. Allerdings finden sich auch einige Belege im Korpus, dass Kinder über den Arztbesuch ihrer alternden Eltern berichten oder Eltern für ihre minderjährigen Kinder Bewertungen vergeben. Damit wird nicht nur die Behandlung, sondern ebenfalls der Behandelte Gegenstand des Erfahrungsberichtes, so dass Autor und Patient nicht mehr in persona auftreten.

Der damit einhergehende Perspektivwechsel spiegelt sich in den morpho-syntaktischen Strukturen der Texte wider. So entwickelte Geierhos (2007) mit der Grammatik der Menschenbezeichner eine Methode, um Prädikat-Argument-Strukturen mittels lokaler Grammatiken (Gross, 1997) zur Identifikation von Personennennungen und ihrer Funktion (semantische Rolle) im unmittelbaren Kontext zu modellieren. Als Menschenbezeichner sind diejenigen Lexeme oder Mehrwortlexeme zu verstehen, welche auf Personen referieren. Hierbei handelt es sich aus syntaktischer Sicht um Eigennamen (z. B. Dr. med. Dieter Rempe), Personalpronomen (z. B. er, ihm, sein) oder Nomen (sog. „allgemeine Menschenbezeichnungen“ nach Geierhos, 2007:69), die u.a. Berufe, Nationalitäten oder Verwandtschaftsverhältnisse denotieren (Geierhos, 2007:74).

In Erfahrungsberichten über Arztbesuche konnten wir folgende *semantische Rollen* identifizieren:

- (1) Wenn der Rezensent selbst der Patient ist, so nehmen beide in persona die Rolle des *Experiencers* (Wahrnehmenden) ein.
- (2) Ist der Patient nicht der Rezensent, wird er zum *Instrument* der Handlung. Er ist damit der Anlass bzw. das Mittel zum Zweck, um eine Arztbewertung zu verfassen. Der Rezensent nimmt dabei die Rolle des *Agens* ein, der willentlich handelt, indem er die Rezension schreibt.
- (3) Unabhängig davon, ob Rezensent und Patient in gleicher oder unterschiedlicher Person auftreten, ist der Arzt und gegebenenfalls sein Praxisteam der Bewertungsgegenstand, der sogenannte *Patiens* des Erfahrungsberichts.
- (4) Zum *Thema* eines Erfahrungsberichts werden sämtliche Qualitätskriterien (u.a. Behandlung, Freundlichkeit, Zufriedenheit und Weiterempfehlung, Aufklärung, Genommene Zeit, Patientenumgang) bei der Benotung niedergelassener Ärzte durch den Rezensenten.

Insbesondere konnte mithilfe der automatischen Erkennung von Menschenbezeichnern im Korpus identifiziert werden, wer Rezensent und Patient im Fall (2) sind. So handelt es sich in ungefähr 17.800

Bewertungen um ein Kind, das den Arztbesuch mit einem oder beiden Elternteilen schildert. Bei weiterer Textanalyse konnte ebenfalls das Alter der Patienten aufgedeckt werden, die in der Regel weit über 70 sind, wenn ihre Kinder für sie stellvertretend den Arztbesuch bewerten. Meist sind es Angehörige, die in der Rolle der Rezensenten schlüpfen. Ähnliches gilt für Eltern, die für ihre Kinder sprechen (29.990 Datensätze). In diesen Fällen sprechen wir von Fremdwahrnehmung, da die Erlebnisse der Patienten – der ursprünglichen *Experiencer* – durch Dritte wiedergegeben werden und damit nicht mehr sichergestellt wird, inwiefern diese Meinungsbilder nicht durch die Rezensenten beeinflusst wurden.

Die Mutter ist nicht immer die Patientin

Ausgehend von der These, dass die Nennung von Verwandtschaftsverhältnissen ein Indiz für eine fremde Autorschaft ist, kann vermutet werden, dass eine Grammatik zur Modellierung der Prädikat-Argument-Beziehungen von Menschenbezeichnern überflüssig ist, um die semantischen Rollen zu bestimmen. Das Korpus liefert jedoch Belege, die ihre Anwendung rechtfertigen:

- (a) Meine Ärztin sollte nicht wie meine Mutter sein.
- (b) [...], da ich mittlerweile die gleiche Erkrankung wie meine Mutter habe.
- (c) Ich bin seit Jahren, genau wie meine Mutter, bei Dr. Esser in Behandlung.

Stellvertretend für weitere allgemeine Menschenbezeichnungen (Geierhos, 2007) wird anhand der „Mutter“-Beispiele in (a)-(c) deutlich, dass ihre syntaktische Einbettung im Kontext von bestimmten prädikativen Ausdrücken ihnen andere semantische Rollen zuweist. Eine Muster-basierte Textsuche nach „meine/unsere Mutter“ würde jegliches Vorkommen im Korpus unabhängig von seiner semantischen Rolle im Erfahrungsbericht berücksichtigen und eine Differenzierung über den Gebrauch im Satz wäre nicht mehr möglich.

Informiertes Einverständnis im Web 2.0?

Im Korpus sind zwar die Rezensenten anonym (ohne Namen), aber nicht die Ärzte. Die Veröffentlichung dieser personenbezogenen Daten unterliegt damit der komplexen Abwägung zwischen dem allgemeinen Persönlichkeitsrecht der Ärzte und dem Grundrecht auf Meinungs- und Informationsfreiheit. Gerichtlich wurde für diesen Anwendungsfall entschieden, dass das Erheben, Speichern, Verändern oder Nutzen im Sinne des Bundesdatenschutzgesetzes grundsätzlich zulässig ist. Weiterhin werden im Fall (1), wenn Rezensent und Patient ein und dieselbe Person sind, keine Persönlichkeitsrechte verletzt, da dieser seinen Erfahrungsbericht freiwillig veröffentlicht hat. Aber wie verhält es sich nun im Fall (2) bei denjenigen Datensätzen, die wir bereits erhoben haben und nun neue Erkenntnisse über deren Autorschaft gewonnen haben? Wissen die betroffenen Patienten überhaupt, dass ihre Erfahrungen beim Arztbesuch online – für jeden lesbar – zugänglich sind? Kann man bei der Datenakquise das informierte Einverständnis der Behandelten voraussetzen? Erheben wir Daten, sehen wir nicht die Person(en) dahinter, sondern nur eine Fülle an Informationen, aus denen wir weitere Schlussfolgerungen ziehen können. Doch werfen diese Erkenntnisse unter Umständen weitere (interdisziplinäre) Fragestellungen auf, die Konsequenzen haben, ob und wie wir forschungsethische Diskussionen führen.

Literatur

- Bakliwal, A.; Arora, P.; Madhappan, S.; Kapre, N., Singh, M. & Varma. V. (2012): *Mining sentiments from tweets*. In Proceedings of the 3rd WASSA-Workshop 2012, Jeju, Island, Republic of Korea.
- Bruce, RF. & Wiebe, JM. (1999): *Recognizing subjectivity: a case study in manual tagging*. In Natural Language Engineering, 1(1), pp. 187-205.

- Gamon, M. (2004): *Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis*. In Proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004, pp. 611-617.
- Geierhos, M. (2007): *Grammatik der Menschenbezeichner in biographischen Kontexten*. Arbeiten zur Informations- und Sprachverarbeitung. Band 2. Centrum für Informations- und Sprachverarbeitung. LMU München.
- Gross, M. (1997): *The Construction of Local Grammars*. In E. Roche und Y. Schabès (Hrsg.): *Finite-State Language Processing*, pp. 329–354. Language, Speech und Communication, Cambridge, Mass.: MIT Press.
- He, Y.; Lin, C. & Alani, H. (2011): *Automatically extracting polarity-bearing topics for cross-domain sentiment analysis*. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oreg., USA, 19–24 June 2011, pp. 123-131.
- Hu, M. & Liu, B. (2004a): *Mining opinion features in customer reviews*. In Proc. of the 19th National Conference on Artificial Intelligence, San Jose, Kalifornien, 25.-29. Juli 2004, pp. 755-760.
- Hu, M. & Liu, B. (2004b): *Mining and summarizing customer reviews*. In Proceedings of SIGKDD, pp. 168–177.
- Hu, N.; Bose, I.; Koh, NS. & Liu, L. (2012): *Manipulation of online reviews: An analysis of ratings, readability, and sentiments*. In Decision Support Systems 52(3), pp. 674-684.
- Jiang, L.; Yu, M.; Zhou, M.; Liu, X. & Zhao, T. (2011): *Target-dependent Twitter sentiment classification*. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oreg., USA, 19.-24. Juni 2011, pp. 151-160.
- Kim, S.-M. & Hovy, E. (2006): *Automatic identification of pro and con reasons in online reviews*. In Proceedings of the Poster Session at the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia, 17.-21. Juli 2006, pp. 483–490.
- Kushal, D.; Lawrence, S. & Pennock, D. M. (2003): *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews*. In Proceedings of the 12th World Wide Web Conference, Budapest, Hungary, 20.–24. Mai 2003, pp. 519–528.
- Ludwig, S.; de Ruyter, K.; Friedman, M.; Brüggem, E. C.; Wetzels, M. & Pfann, G. (2013): *More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates*. Journal of Marketing 77:1, pp. 87-103.
- O'Connor, P. (2010): *Managing a Hotel's Image on TripAdvisor*. In Journal of Hospitality Marketing & Management. 19(7), pp. 754-772.
- Qin, B.; Zhao, Y.; Gao, L. & Liu, T. (2008): *Recommended or not? Give advice on online products*. In Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, Shandong, China, 18.-20. Oktober 2008, pp. 208-212.
- Turney, P. (2002): *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Penn., 7.-12. Juli 2002, pp. 417-424.
- Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004): *Learning subjective language*. In: Computational linguistics, 30(3), pp. 277-308.