

SocialML: EUD im Machine Learning zur Analyse sozialer Medien

Simon Scholl¹, Christian Reuter², Thomas Ludwig¹, Marc-André Kaufhold¹

Institut für Wirtschaftsinformatik, Universität Siegen¹

Wissenschaft und Technik für Frieden und Sicherheit (PEASEC), TU Darmstadt²

Zusammenfassung

Die Analyse sozialer Medien hat eine enorme Bedeutung für die verschiedensten Bereiche, wie beispielsweise die Lagebeurteilung, journalistische Recherchen oder unternehmerische Analysen. Dabei sind die entsprechenden Analysten unterschiedlichen Kontextfaktoren unterworfen. Sei es durch ihren organisationalen Hintergrund, durch tätigkeitsbezogene Rollen oder den geografischen und zeitlichen Rahmen. Um den ständig im Wandel begriffenen Kontexten gerecht zu werden, ergeben sich permanent neue Herausforderungen für die Anpassbarkeit der Analyse. Dieser Beitrag untersucht, wie ein auf Machine Learning (ML) basierendes Tool zur Social Media Analyse (SMA) im Sinne des End-User Development (EUD) auf verschiedenen Komplexitäts- und Schrittebenen umgesetzt werden kann.

1 Einleitung

Soziale Medien werden - jenseits der privaten Nutzung - seit einigen Jahren auch in professionellen Kontexten eingesetzt (Kaufhold, Reuter, Ludwig, & Scholl, 2017). Die stetig wachsende Datenmenge stellt dabei allerdings hohe Anforderungen an die Datenanalyse und die damit verbundenen Erkenntnisgewinne (Nasser & Tariq, 2015). Machine Learning (ML) umfasst automatische Berechnungsverfahren, die aus einer Reihe von Beispielen lernen, um Vorhersagen für neue oder unvollständige Datensätze zu treffen (Ayodele, 2010). Der Kontext der Datenanalyse hängt dabei stark vom jeweiligen Nutzern ab. Hier bietet End-User Development (EUD) interessante Anknüpfungspunkte (Reuter, Kaufhold, & Ludwig, 2017): EUD umfasst Methoden, Techniken und Tools, und ermöglicht Softwarenutzern (im Gegensatz zu Software-Entwicklern) Software zu modifizieren (Lieberman, et al., 2006). Menon (2013) zeigt für Such- und Sortierungsfunktionen in der Textverarbeitung, dass sich das EUD-Prinzip „Programming by Example“ (Lernen eines neuen Verhaltens auf Basis konkreter Beispiele) auf den Lernprozess von ML übertragen lässt. Nutzern von Social Media Analytics (SMA) Systemen kann dies die Möglichkeit bieten, die Ausrichtung der Analyse zu steuern.

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2018 in
R. Dachsel, G. Weber (Hrsg.):
Mensch und Computer 2018 – Tagungsband, 02.–05. September 2018, Dresden.
Copyright (C) 2018 bei den Autoren. <https://doi.org/10.18420/muc2018-mci-0347>

2 SocialML: EUD in Social Media Analytics

Das von uns entwickelte Tool SocialML stellt einen Versuch der Integration von EUD und ML im Kontext von SMA dar. Der Nutzer kann im Sinne einer **anpassbaren Erhebung und Modellbildung** plattformübergreifend (Facebook, Google oder Twitter) Daten erheben, die auf Suchbegriffen und zeitlichen Eingrenzungen basieren. Anschließend kann er definieren, welche Klassifikationen in seinem Kontext von Interesse sind. Zur Modellbildung muss ein ML-Algorithmus mit einer ausreichenden Menge an Trainingsdaten angelehrt werden. Hier setzt die Anwendung der EUD-Komponente „Programming by Example“ an. Der Endnutzer erstellt Instanzen, welche das Modellverhalten beeinflussen. EUD-Komponenten zur **Anpassung von Darstellungselementen** und Parametrisierung erlauben es dem Endnutzer, die Ergebnisse während der Laufzeit anzupassen. Damit ein Nutzer die Qualität individuell bewerten kann, werden Indikatoren benötigt, die die Evaluation des jeweiligen Modells anhand der gegebenen Trainingsdaten ermöglichen. Dies entspricht der Strategie des Aufbaus einer Feedbackschleife, wie durch Bertini und Lalanne (2009) beschrieben, welche die Rückintegration der selbstreflektierten Evaluationsergebnisse in die Modellkomponenten sicherstellt. Die **Erweiterbarkeit der Applikation** ist ebenfalls insofern gewährleistet, dass Komponenten, insb. der ML-Algorithmus, austauschbar sind.

Wird der **Trainer** aufgerufen, muss zunächst das Modell (Abb. 1, #1), welches trainiert werden soll, erstellt oder selektiert werden. In Falle eines bereits trainierten Modells sollen die beiden angezeigten Indikatoren dazu dienen, die generelle Genauigkeit (#2) des jeweiligen Modells sowie das Verhältnis bereits verwendeter und neuer Trainingsdaten (#3) darzustellen. Der Button „Trainieren“ (#5) initialisiert die Bildung des Klassifizierers anhand der gewünschten Trainingsmethode (#4), an deren Ende ein nutzbares Modell steht.



Abbildung 1: SocialML – Übersicht des Trainers

Um ein neues Modell zu trainieren, muss der Nutzer zunächst Daten sammeln: Sendet der Nutzer die **Abfrage** (Abb. 2) ab, wird eine einmalige schlüsselwortbasierte Suche je nach Auswahl in Facebook, Google+ oder Twitter durchgeführt. Danach wird der Nutzer auf eine neue Oberfläche zur Verarbeitung (Labeln) der Trainingsdaten geleitet (Abb. 4). Das trainierte Modell kann nun auf beliebige **Datensammlungen** aus sozialen Medien angewandt werden. Das **Erstellungsfeld** (Abb. 3) erlaubt dem Nutzer die Kombination von Datensammlung, Modell und Visualisierung, welche dann auf dem zentralen **Dashboard** dargestellt wird. Angereichert wird jede **Visualisierung** mit einer Live-Suche (Abb. 5, #1.1) sowie der Möglichkeit der Sortierung entlang des Erstellungsdatums (#1.2) der Nachrichten. Die aktive Klassifi-

kation kann jederzeit über eine Änderung der Selektion umgestellt werden (#1.3). Des Weiteren gibt es zur Herstellung des zeitlichen Bezugs die Möglichkeit der Erstellung von Balkendiagrammen (#2) oder Liniengraphen (#3). Die zeitliche Verteilung ist dynamisch einstellbar und umfasst den Zeitraum zwischen dem Start der Datenerhebung und dem aktuellen Datum. Der Nutzer kann den Fokus innerhalb dieses Zeitraums durch die Funktionen des Kalenders (#8) beliebig anpassen. Eine zeitliche Skalierung (#6) soll zudem helfen, kleinere oder größere Zeiträume effektiver zu betrachten. Zusätzlich kann der Fokus innerhalb der Grafiken durch das Ein- und Ausblenden (#7) relevanter oder nicht relevanter Klassifikationen variiert werden. Die anderen Darstellungsformen, das Kreis- sowie das Donut-Diagramm (#4-5), dienen der Gesamtübersicht der anteiligen Verteilung der Nachrichten auf die Klassifikationen. Auch hier kann der Zeitraum eingegrenzt werden.

Abbildung 2: Trainingsdatenerhebung der Social Media API

Abbildung 3: SocialML - Hinzufügen einer neuen Sicht

Abbildung 4: SocialML – Hinzufügen von Trainingsdaten der SMA

Um neue Erkenntnisse für SMA durch die Anwendung von ML mit dem speziellen Fokus auf EUD zu erhalten, wurde eine **Evaluation** durchgeführt. Die acht beteiligten Nutzer (Softwareentwickler, Studenten, wiss. Mitarbeiter, Alter 20-30 Jahre) sollten die Applikation entlang eines erstellten Szenarios zur Analyse sozialer Medien nutzen und in der Rolle eines Analysten vordefinierte Aufgaben bewältigen: Obwohl die Evaluation zeigte, dass SocialML das Potential hat, die Integration des eigenen Erkenntnisinteresses entlang der Anwendung von ML durch die Funktionalitäten einer Vorselektion - über die Bildung eigener Analysen, bis hin zur visuellen Auswertung - zu unterstützen, wurde die lose Kopplung der Komponenten sowie der komplexe Analyseprozess als nicht intuitiv beschrieben. Insbesondere die Erstellung eigener Analysen in Form von Modellen zur Textklassifizierung sowie das Training dieser Komponenten verlangen ein **besonderes Maß an Nutzbarkeit**. Zudem zeigte die Evaluation auf, welche Faktoren im Rahmen der Anwendung von Supervised ML (Identifikation bekannter

Muster auf neuen Datensätzen) Einfluss auf die von Endnutzern gesteuerte Analyse nehmen. So besteht ein erhöhter Informationsbedarf darüber, was eine solche Komponente leisten kann. Für die Durchführung des Trainings wurde die Aufwandsreduktion sowie eine Ausrichtung am Prozessfluss sowie ein Assistenzsystem, das unterschiedliche Qualifikationsniveaus berücksichtigt, als wesentlicher Bestandteil für eine praxisnahe Umsetzung gesehen.

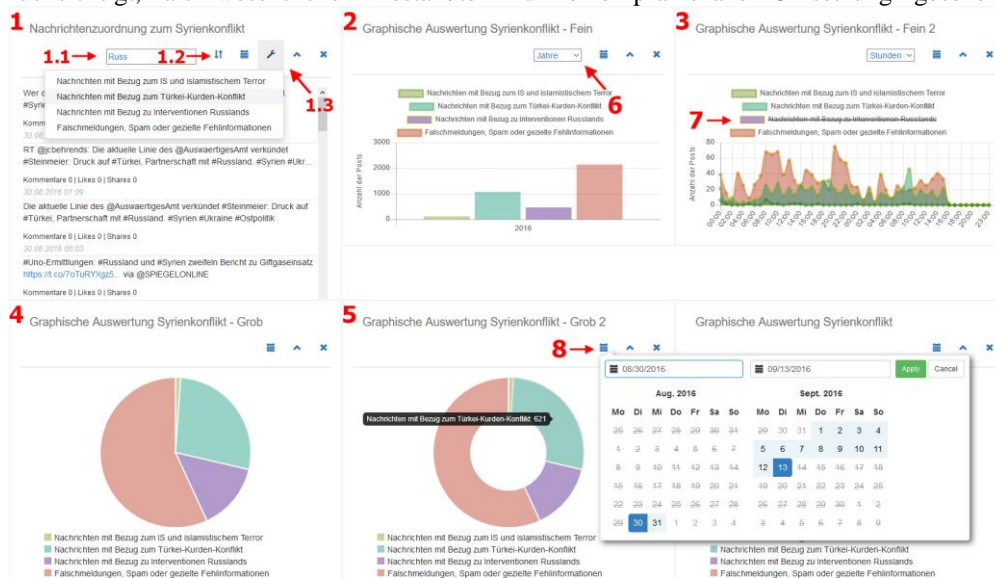


Abbildung 5: SocialML - Visualisierungssichten des Dashboards

Literaturverzeichnis

- Ayodele, T. O. (2010). Machine Learning Overview. In *Introduction to Machine Learning* (pp. 9–19). INTECH Open Access Publisher.
- Bertini, E., & Lalanne, D. (2009). Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. *Proc. Visual Analytics and Knowledge Discovery Integrating Automated Analysis with Interactive Exploration*, 12–20.
- Kaufhold, M.-A., Reuter, C., Ludwig, T., & Scholl, S. (2017). Social Media Analytics: Eine Marktstudie im Krisenmanagement. In *INFORMATIK 2017, LNI, Gesellschaft für Informatik*. Bonn.
- Lieberman, H., Paternò, F., Klann, M., & Wulf, V. (2006). End-user development: An emerging paradigm. *End User Development*, 9, 1–8. https://doi.org/10.1007/1-4020-5386-X_1
- Menon, A. K. (2013). A Machine Learning Framework for Programming by Example, 28.
- Nasser, T., & Tariq, R. (2015). Big Data Challenges. *Journal of Computer Engineering & IT*.
- Reuter, C., Kaufhold, M.-A., & Ludwig, T. (2017). End-User Development and Social Big Data - Towards Tailorable Situation Assessment with Social Media. In F. Paternò & V. Wulf (Eds.), *New Perspectives in End-User Development* (pp. 307–332). Springer.