

Komplexe Modellierung: Kann man mit Mathematik Wahlen gewinnen? Big Data Analysen von sozialen Netzwerken mit Schülerinnen und Schülern der Sek. II

Wie konnte Donald Trump die US-Präsidentschaftswahl gewinnen? Eine Teilantwort lautet: Die Firma Cambridge Analytica hat geholfen, adressatengerechte Botschaften zu verschicken, indem sie durch Datenanalysen psychologische Verhaltensmuster von Nutzern sozialer Netzwerke erstellt hat (vgl. Zastrow 2016, Cadwalladr & Graham-Harrison 2018). Dazu sind detaillierte, zuverlässige Informationen über Nutzer sozialer Netzwerke notwendig. Diesen Kontext erforschen Schülerinnen und Schülern der Sekundarstufe II und bereiten so den Übergang von der Schule zur Hochschule vor.

1. Modellierungsworkshop für Schülerinnen und Schüler

In einem eintägigen Workshop können Schülerinnen und Schüler im Schülerlabor CAMMP der RWTH Aachen und des KIT den Fragen „Kann man mit Mathematik Wahlen gewinnen? – Wie gut kann man Eigenschaften von Nutzern vorhersagen?“ nachgehen. Die Relevanz von Mathematik wird hierbei durch das Lösen dieser authentischen und relevanten Problemstellung mit Mathematik und Computereinsatz (MATLAB) deutlich. Im Workshop wird aufbauend auf der Forschungsarbeit von Garcia, Sarigol und Schweizer (2014) die Vorhersagbarkeit der sexuellen Orientierung fokussiert, einer sehr sensiblen Information. Der Ablauf des Workshops ist im Modellierungskreislauf des Schülerlabors (Abb. 1) dargestellt, der den bekannten Modellierungskreislauf von Blum und Leiss konkretisiert.

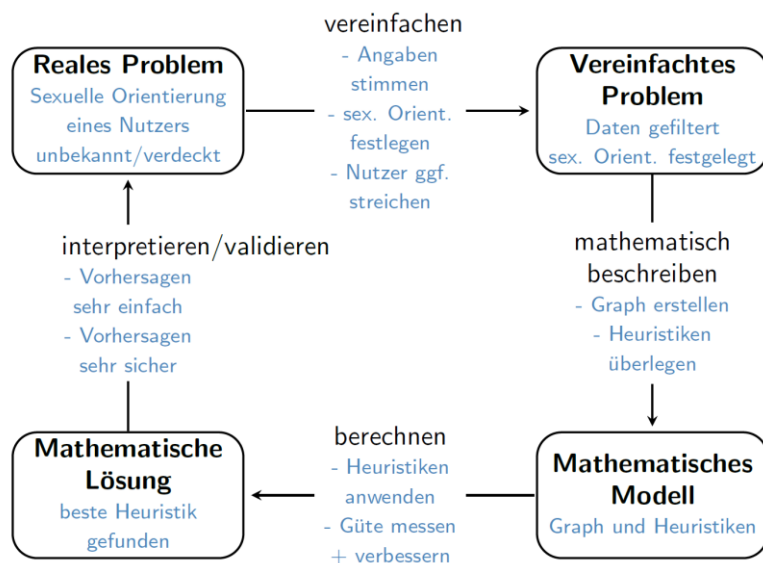


Abb. 1: Modellierungskreislauf im Workshop

2. Problemstellung und Problemlösung

Das reale Problem ist die Vorhersage der sexuellen Orientierung eines Nutzers, auch wenn diese aus den Angaben in seinem Profil nicht ablesbar ist. Dazu wird das soziale Netzwerk, Friendster, betrachtet und die Daten gefiltert. Der aufgearbeitete, echte Datensatz umfasst 56284 Nutzer, deren sexuelle Orientierung bekannt ist, und wird als ungerichteter Graph (Abb. 2) dargestellt. Die Information der sexuellen Orientierung wird bei der Vorhersage ausgeblendet und am Schluss zur Kontrolle der Vorhersage wieder eingeblendet. Das mathematische Modell wird danach

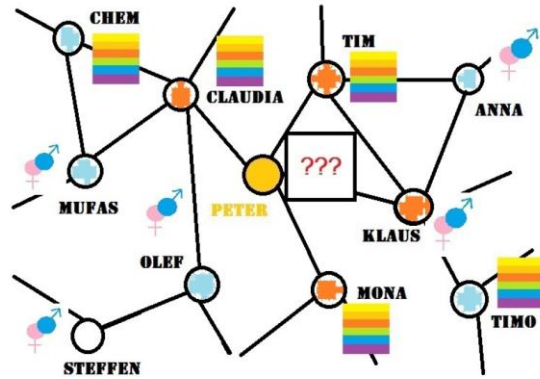


Abb. 2: Darstellung des Netzwerks

gebildet mit Hilfe von Heuristiken, die Vorhersagen über die sexuelle Orientierung machen. Wir arbeiten u.a. mit parameterabhängigen „Freunderegeln“, die zur Bestimmung der sexuellen Orientierung die Angaben der Freunde einbeziehen. Eine solche Regel: „Wenn der Nutzer mehr als 20% homosexuelle Freunde hat, dann ist er homosexuell. Wenn der Nutzer weniger als 10% homosexuelle Freunde hat, dann ist er heterosexuell. Sonst lasse den Zufall entscheiden.“

Das Bestimmen der mathematischen Lösung wird nun vollzogen, wobei in jedem Schritt gefragt wird „Welche Regel ist die beste?“:

Die Lernenden beurteilen die Güte zunächst anhand der Heuristiken durch erste Vermutungen. Dann wenden sie mit MATLAB alle Regeln auf einzelne Nutzer an. Danach wenden sie eine Regel auf mehrere Nutzer an und untersuchen verschiedene Stichprobengrößen. Sie variieren zudem die Parameter in den „Freunderegeln“ und beurteilen erneut die Güte. Hier ist die Vergleichbarkeit der unterschiedlichen Kriterien oft schwierig. Das motiviert die Einführung eines einheitlichen Gütekriteriums: Cohen’s Kappa. Die Schülerinnen und Schüler bestimmen zur Nutzung von Kappa die theoretische Häufigkeit

Cohen’s Kappa (Cohen 1960)

$$\kappa = \frac{p_H - p_Z}{1 - p_Z}$$

p_H : relative Häufigkeit richtiger Einschätzungen mit einer Heuristik;

p_Z : theoretische Häufigkeit richtiger Einschätzungen durch den datenbasierten Zufall (19,85% homosexuell, 80,15% heterosexuell, wirf „gefälschte“ Münze)

richtiger Einschätzungen durch den datenbasierten Zufall, interpretieren verschiedene Kappawerte und vergleichen die Heuristiken mit Kappa. Das Ergebnis ist, dass die oben genannte „Freunderegel“ den höchsten Kappawert, also das beste Ergebnis liefert. Nun stellt sich die Frage, ob die Güte der „Freunderegel“ durch eine geschickte Wahl von Parametern (Prozentwerte) gesteigert werden kann. Dazu wird eine Monte-Carlo-Optimierung durchgeführt und die Kappawerte können minimal gesteigert werden.

Die Fragestellung wurde gelöst: Durch recht einfache Regeln kann man die sexuelle Orientierung mit mehr als 92% Sicherheit vorhersagen. Durch andere Heuristiken oder einen anderen Datensatz könnte man die Modellierung zudem ggf. verbessern.

3. Schlussfolgerungen

Letztendlich kann gezeigt werden, dass mit Mathematik zusätzliche Informationen über Nutzer generiert werden können. Die Basis für adressatengerechte Botschaften in einem Wahlkampf kann geschaffen werden. Ob das nun Wahlkampf entscheidend ist, kann nicht direkt entschieden werden, aber es wird deutlich, dass die Mathematik ein mächtiges Werkzeug in diesem Kontext ist und eine hohe gesellschaftliche Relevanz hat.

4. Übergang Schule – Hochschule

Im Workshop wird damit durch eine komplexe Modellierung mit Inhalten der Schulmathematik (u.a. Häufigkeiten, Zufallsexperiment, Hypothesentests) und Inhalten der Hochschulmathematik (u.a. Datenfilterung, Graph, Heuristiken, Cohen's Kappa, Optimierung) der Übergang von der Schule zur Hochschule vorbereitet.

Literatur

- Blum, W. & Leiss, D. (2005). Modellieren im Unterricht mit der “Tanken”-Aufgabe. *Mathematik lehren*, 128, S. 18-21.
- Cadwalladr, C. & Graham-Harrison, E. (2018). How Cambridge Analytica turned Facebook ‘likes’ into a lucrative political tool. <https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm> (3.4.2018)
- CAMMP: Website. <http://www.cammp.rwth-aachen.de/> (3.6.2016)
- Cohen, J. (1960) A coefficient of agreement for nominal scales. In: *Educational and psychological measurement*, 20, S. 37-46.
- Garcia, D., Sarigol, E. & Schweitzer, F. (2014). Online Privacy as a Collective Phenomenon. COSN’14, Dublin. <http://arxiv.org/pdf/1409.6197v1.pdf> (6.6.2016)
- Sube, M. (2016) Wie sicher ist meine Privatsphäre in sozialen Netzwerken? ... und was hat das mit Mathe zu tun? Masterarbeit, RWTH Aachen. <https://blog.rwth-aachen.de/cammp/files/2016/10/thesis-soziale-netzwerke.pdf> (3.4.2018)

Zastrow, V. (2016). Wie Trump gewann. Frankfurter Allgemeine. http://www.faz.net/aktuell/politik/trumps-praesidentschaft/wie-der-wahlsieg-vondonald-trump-mit-big-data-gelang-14568868.html?printPageArticle=true#pageIndex_2 (20.12.2016)