

Komplexe Modellierung: Datensicherheit in sozialen Netzwerken

Der Workshop behandelt die aktuelle Fragestellung *Wie sicher ist meine Privatsphäre in sozialen Netzwerken?...und was hat das mit Mathe zu tun?* Dieser können Schüler/innen in einem Tagesprojekt an der RWTH Aachen im CAMMP (www.cammp.rwth-aachen.de) nachgehen. Im Rahmen einer Fortbildung konnten Interessierte auf der GDM 2017 ebenfalls einen mathematischen Blick auf die Fragestellung werfen, die nicht zuletzt mit Blick auf die Wahl von Donald Trump interessant ist. Analysen von Daten aus sozialen Netzwerken (z.B. facebook) und das Verschicken von personalisierten Nachrichten sollen maßgeblich an seinem Sieg beteiligt gewesen sein (Zastrow (2016)). Dazu sind detaillierte, zuverlässige Informationen über die Nutzer/innen der Netzwerke notwendig. Ob es möglich ist, die Informationen zu erhalten, wird im Workshop untersucht.

1. Problemstellung

Der Workshop basiert auf einem Artikel von Garcia, Sarigol und Schweitzer (2014). Betreiber von sozialen Netzwerken können sog. Schattenprofile erstellen. Gemeint ist ein Profil für eine Person des Netzwerks, das Informationen über diese enthält, die nicht angegeben wurden. Als zusätzliche Information über eine Person wird im Workshop die sexuelle Orientierung (homo- oder heterosexuell) betrachtet, eine sensible Information über deren Verfügbarkeit jeder selbst entscheiden können sollte. Im Workshop stehen Nutzerprofile des Netzwerks Friendster als authentische, echte Daten zur Verfügung. Hierbei wird angenommen, dass die Angaben in den Profilen wahrheitsgemäß sind. Vereinfachend werden von den ersten 1 Millionen Profilen gemäß der Beitrittsreihenfolge diejenigen betrachtet, denen eine sexuelle Orientierung aufgrund ihrer Angaben zugeordnet werden kann und die zusätzlich mindestens eine Freundschaft im Netzwerk aufweisen. So entsteht ein Datensatz von 56284 Personen, mit dem eine Vorhersage der sexuellen Orientierung kontrolliert werden kann. Er ist verfügbar unter <https://github.com/camminady/FriendsterDataSet>.

Als Modell des Netzwerks wird ein ungerichteter, endlicher Graph genutzt (Abb. 1). Mit Hilfe von Heuristiken (hier: Regeln zur Bestimmung der sexuellen Orientierung) wird untersucht, wie gut die sexuelle Orientierung vorhergesagt werden kann. Eine Heuristik als Beispiel:

Wenn die Person mehr als 55% homosexuelle Freunde hat, dann ist sie homosexuell. Wenn die Person weniger als 56% homosexuelle Freunde hat, dann ist sie heterosexuell. Sonst sage mit einer Wahrscheinlichkeit von

19,85%, dass die Person homosexuell und mit einer Wahrscheinlichkeit von 80,15%, dass sie heterosexuell ist.

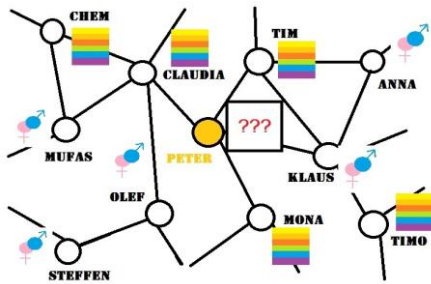


Abbildung 1: Soziales Netzwerk als Graph

Die letzten beiden Wahrscheinlichkeiten stammen aus dem Datensatz: Es sind 19,85% der Personen homo- und 80,15% der Personen heterosexuell. Die Vorgehensweise entspricht einem Bernoulli-Experiment, da die Vorhersage entweder richtig oder falsch sein kann.

Die Güte einer Heuristik wird mit verschiedenen Größen beurteilt. Zum einen kann mit Cohen's Kappa ein Vergleich zwischen der Vorhersagekraft der Heuristik und dem Zufall gezogen werden. Zum anderen können Trefferquote und Genauigkeit untersucht werden. Hier soll nur Cohen's Kappa betrachtet werden, das laut Rost (2004) definiert wird als

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

wobei p_H die Trefferwahrscheinlichkeit der Heuristik und p_Z die des Zufalls benennt. Beschreibt man die Bestimmung der sexuellen Orientierung durch den Zufall als zweistufiges Experiment mit den genannten Wahrscheinlichkeiten aus dem Datensatz, so ergibt sich mit Pfad- und Summenregel eine Trefferwahrscheinlichkeit von ca. 68,18%.

Neben der Messung der Güte der Heuristiken stehen im Workshop auch die Verbesserung der Güte und die Validierung der Ergebnisse im Fokus. In der angegebenen Heuristik sind die Parameter (55%, 56%) willkürlich gewählt. Durch systematisches Ausprobieren gemäß der Monte-Carlo-Methode werden sie optimiert und die Kappawerte verglichen. Mit der genannten Heuristik kann ein Kappawert von 0,75 erreicht werden, also eine Trefferwahrscheinlichkeit von ca. 92%. Die Validierung der Ergebnisse erfolgt mit Binomialtests, wobei die Trefferwahrscheinlichkeit von 92% zu einem Signifikanzniveau von 0,01 statistisch belegt wird. Die Heuristik liefert im Vergleich zu den anderen im Workshop untersuchten Heuristiken die höchste Trefferwahrscheinlichkeit. Die Schüler/innen entdecken somit, dass durch die Nutzung des Anteils an homosexuellen Freunden der Personen die

sexuelle Orientierung mit einer 92%-igen Trefferwahrscheinlichkeit vorhergesagt werden kann.

2. Didaktische Aspekte

Der didaktische Hintergrund des Workshops ist das realistische bzw. angewandte Modellieren (Kaiser, Sriraman (2006)). Im Zentrum steht die Frage nach der Datensicherheit in sozialen Netzwerken, die als Problemstellung gelöst werden muss. Der Modellierungskreislauf nach Blum (1985) wird dabei im Workshop durchlaufen. Die Lösung des Problems wird mathematisch mit Hilfe von Computereinsatz (MATLAB und bereitgestellten Worksheets) getätigt, indem Simulationen durchgeführt werden, was auch Greefrath und Siller (2009) nahelegen. Der Workshop ist zudem im Bezug zu den Kernlehrplänen des Landes Nordrhein-Westfalen zu sehen. In den Bereichen Problemlösen und Modellieren sowie in der Stochastik sind Aspekte des Workshops vorhanden. Als Beispiele sollen die Schritte des Modellierens, zweistufige Zufallsexperimente sowie Hypothesentests ausreichen.

Die Schüler/innen beantworten die Fragestellung in verschiedenen Erkenntnisschritten, die in Arbeitsphasen eingeteilt sind. Die Leitfrage ist: *Welche Regel ist die beste?* Zunächst lernen die Schüler/innen die Heuristiken im Detail kennen, indem sie nur durch Betrachtung der Regeln, die Güte bewerten. Danach wenden sie die Regeln auf einzelne Personen an, wobei sie entdecken, welche Regel richtige, falsche oder auch sich verändernde Vorhersagen macht. Es folgt die Anwendung der Regeln auf mehrere Personen. Die Schüler/innen können die Stichprobengröße und in einigen Regeln die Parameter variieren. Der Vergleich der Heuristiken erscheint an dieser Stelle aufwendig und uneinheitlich. Hier wird die Einführung eines Kriteriums motiviert: Cohen's Kappa. Hierzu wird zunächst die Trefferwahrscheinlichkeit mit Hilfe des Zufalls bestimmt und die Interpretation verschiedener Kappawerte geübt. Mit Hilfe des Kriteriums erfolgt eine erneute Güteuntersuchung der Heuristiken. Zusätzlich können die Schüler/innen auch die Trefferquote und die Genauigkeit der Regeln untersuchen. Danach werden die parameterabhängigen Heuristiken optimiert und abschließend Hypothesentests durchgeführt. Zudem können Rechercheaufgaben verfolgt werden zu den AGB von WhatsApp oder Datenschutzmöglichkeiten. Die Bearbeitung erfolgt in Partner-/Gruppenarbeit sowie in Besprechungsphasen.

Die beschriebenen Schritte bergen Differenzierungsmöglichkeiten: Die Schüler/innen können in vielen Schritten nur eine Regel oder mehrere Regeln betrachten. Arbeitsteiliges oder -gleiches Vorgehen ist möglich, da in Besprechungsphasen die Ergebnisse stets gebündelt werden. Der

Heterogenität wird zudem mit Hilfekarten und differenziertem Feedback bei der Lösungskontrolle durch MATLAB begegnet.

Der Workshop wird abgerundet mit der Diskussion verschiedener Fragen. Beispiele hierfür sind: *Sind die Ergebnisse realistisch oder übertragbar auf "moderne" Netzwerke? (Wie) Kann ich meine Daten schützen? Sind Nicht-Nutzer/innen in ihrer Privatsphäre sicher? Ändere ich mein Verhalten in sozialen Netzwerken? Ist es plausibel, dass Donald Trump vor allem wegen Datenanalysen und personalisierten Nachrichten gewonnen hat?*

3. Zusammenfassung und Ausblick

Erfahrungsgemäß bietet der Workshop Möglichkeiten, sich mit einem brisanten, aktuellen Thema aus mathematischer und persönlicher Sicht zu befassen. Durch das Material (u.a. MATLAB-Worksheets, Arbeitsblätter, Videos) meistern die Schüler/innen bekannte mathematische Inhalte aus der Schule und auch ihnen unbekannte und lernen so neue Aspekte der Mathematik kennen. Es werden Modellieren und Problemlösen geschult.

Der Workshop wird im Rahmen einer Dissertation weiterentwickelt. Aktuell wird eine Vorhersage des Alters von Personen in sozialen Netzwerken untersucht. Didaktisch-methodisch wird der Workshop für ein höheres Schwierigkeitsniveau ausgebaut. Zudem liegt ein Vorschlag für eine authentische Abituraufgabe bezogen auf die Fragestellung vor.

Literatur

- Blum, W. (1985). Anwendungsorientierter Mathematikunterricht in der didaktischen Diskussion. *Mathematische Semesterberichte*, 32(2), 195–232.
- Garcia, D., Sarigol, E. & Schweitzer, F. (2014). Online Privacy as a Collective Phenomenon. *COSN'14*, Dublin.
- Greefrath, G. & Siller, H.-St. (2009). Mathematical Modelling in Class Regarding To Technology. *CERME-post-conference-proceedings*, Lyon.
- Kaiser, G. & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM*, 38(3), 302-310.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2013). *Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen. Mathematik*. Verfügbar unter http://www.schulentwicklung.nrw.de/lehrpläne/upload/klp_SII/m/GOST_Mathematik_Endfassung.pdf [01.04.2017].
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Verlag Hans Huber.
- Zastrow, V. (2016). *Wie Trump gewann. Frankfurter Allgemeine*. Verfügbar unter http://www.faz.net/aktuell/politik/trumps-praesidentschaft/wie-der-wahlsieg-von-donald-trump-mit-big-data-gelang-14568868.html?printPagedArticle=true#page-Index_2 [20.12.2016].