Toxische Sprache im Internet – Erkennung mit Verfahren der computerlinguistischen Forensik

Seminar im Modul M-GSW-10 SoSe 2019

Prof. Dr. Udo Hahn

Lehrstuhl für Angewandte Germanistische Sprachwissenschaft / Computerlinguistik

Institut für Germanistische Sprachwissenschaft Friedrich-Schiller-Universität Jena

http://www.julielab.de

Allgemeine Hinweise

- Termin: Do, 16-18h (Johannisfriedhof,SR 2)
- Materialien im Netz
 - http://www.julielab.de
 "Students"
- Sprechstunde: Mi, 12-13h (nA) (FG 30, R 004)
- © Email: udo.hahn@uni-jena.de
- Fachliteratur: durchgängig in Englisch

Textuelle Gegenstände der Computerlinguistik: Fakten - Informationsextraktion



Polizei wegen Blitz-Marathon in der Kritik Automobilklubs: Nur symbolischer Charakter

D 200 EURO >

Nachsicht.

noch einmal

Wer hat mehr Bajonette?

Nach dem TV-Duell zwischen Obama und Romney steigen Historiker ins Archiv

Aps affer Web

Todliche Schlüserei

auf dem Alex:

Zwei Minne

stellen sich

Ausrottung von

Kinderlähmuns

ist bald möglid

Der Roboter

Comeback im Film

Sebe 11

Welche Firma

hat den

besten Service

Im Plus





Propensity Score – eine alternative Methode zur Analyse von Therapieeffekten

Oliver Kuss, Maria Blettner, Jochen Börgermann

ZUSAMMENFASSUNG

Hintergrund: Nur die Randomisierung parantiert in Theraniestudien eine gleichmäßige Verteilung aller bekannten und unbekannten Patientenmerkmale auf eine Interventions- und eine Kontrollgruppe und erlaubt dadurch kausale Aussagen über Therapieeffekte. Randomisierte kontrollierte Studien werden jedoch auch für ihre fehlende externe Validität kritisiert. Nichtrandomisierte Studien sind eine Alternative, allerdings besteht hier die Gefahr, dass sich die Interventions- und die Kontrollgruppe bezüglich bekannter und unbekannter Patientenmerkmale unterscheiden. Zur Analyse von nichtrandomisierten Studien werden in der Regel multiple Regressionsmodelle verwendet, immer häufiger wird aber auch auf die sogenannte Propensity-Score-Methode zurück-

Methode: Auf Basis einer selektiven Literaturrecherche und der wissenschaftlichen Erfahrung der Autoren wird die Propensity-Score-Methode anhand eines Beispiels aus der koronaren Bypass-Chirurgie ausführlich dargestellt und

Ergebnis: Der Propensity Score (PS) ist definiert als die Wahrscheinlichkeit, mit der ein Patient die zu prüfende Therapie erhält. Der PS wird in einem ersten Schritt aus den vorhandenen Daten geschätzt, beispielsweise in einem logisti schen Regressionsmodell. Im zweiten Schritt erfolgt die Schätzung des eigentlich interessierenden Therapieeffekts unter Zuhilfenahme des PS. Dabei stehen vier Methoden zur Verfügung: PS-Matching, "inverse probability of treatment weighting"(IPTW)-Schätzung, Stratifizierung nach dem PS oder Regressionsadjustierung für den PS.

Schlussfolgerung: Die Propensity-Score-Methode ist eine gute Alternative zur Auswertung von nichtrandomisierten Therapiestudien. Sie hat erkenntnistheoretische Vorteile im Vergleich zur herkömmlichen Regressionsanalyse. Der Propensity Score kann allerdings nur für die bekannten und tatsächlich gemessenen Störgrößen adjustieren. Die gleichmäßige Verteilung von unbekannten Störgrößen bleibt die Domäne randomisierter kontrollierter Studien.

Kuss O. Blettner M. Börgermann J: Propensity score: an alternative method of analyzing treatment effects—part 23 of a series on evaluation of scientific nublications Disch Arztebl Int 2016: 113: 597-603 DOI: 10.3238/arztebl 2016.0597

Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-He Universität Düsseldorf, Institut für Biernetrie und Epidemiologie und Centre for Health and Society (chs), Medizinische Fakultiit, Heinrich-Heine-Universität Düsseldorf; Prof. Dr. sc., hum. Kuss Institut für Medizinische Biometrie. Epidemiologie und Informatik (IMBEI). Universitätsmedizin de Johannes Gutenberg-Universität Mainz; Prof. Dr. rer. nat. Riettner

Künik für Thorax- und Kardiovaskularchinurgie, Herz- und Diabeteszentrum Nordmein-Westfaler Universitäskünik der Ruhr-Universität Bochum, Bad Oeynhausen: PD Dr. med. Börgermann

Deutsches Ärzteblatt | Jg. 113 | Heft 35-36 | 5. September 2016

an ist sich in der medizinischen Forschung weitgehend einig darüber, dass Therapien primär in randomisierten kontrollierten Studien genrüft werden sollten. Nur die Randomisierung garantiert eine gleichmäßige Verteilung aller bekannten und unbekannten Patientenmerkmale auf eine Interventions, und eine Kontrollgruppe und erlaubt dadurch kausale Aussagen über Therapiceffekte. Randomisierte kontrollierte Studien sind jedoch in manchen Fällen "unnötig, ungeeignet, unmöglich oder ungenügend" (1) und werden darüber hinaus immer wieder für ihre fehlende externe Validität kritisiert: Patienten in randomisierten kontrollierten Studien sind in der Regel jünger und gesünder

als der durchschnittliche Patient (2, 3). Nichtrandomisierte Studien können für die Evaluierung von Therapien eine Alternative sein, allerdings haben diese das Problem der fehlenden internen Validität: Die Therapiezuteilung erfolgt nicht randomisiert und die Interventions- und die Kontrollgruppe können sich systematisch bezüglich bekannter und (schlimmer noch) unbekannter Patientenmerkmale unterscheiden. Mögliche Unterschiede, die sich in den Gruppen im Verlauf der Studie ergeben, können daher nicht notwendigerweise auf die unterschiedliche Behandlung zurückgeführt werden. Diese Unterschiede könnten auch durch die systematischen Differenzen zwischen den Gruppen zustande gekommen sein.

Fine Reihe von statistischen Verfahren wurde entwickelt, um diese Unterschiede bei der Auswertung zu berücksichtigen. Standardverfahren sind dabei die multinlen Regressionsmodelle. Immer häufiger wird jedoch auch die sogenannte Propensity-Score-Methode angewendet (4). Im Folgenden wird der Propensity Score eingeführt und zuerst allgemein, dann anhand eines Beispiels aus der koronaren Bypass-Chirurgie ausführlich dargestellt und erklärt. In einem weiteren Abschnitt wird die Methode gegenüber den herkömmlichen Regressionsmodellen abgegrenzt. Der Artikel schließt mit einigen grundsätzlichen Bemerkungen zum Erkenntnisgewinn in der medizinischen Forschung.

Propensity-Score-Methode

Der Propensity Score (PS) ist die Wahrscheinlichkeit, mit der ein Patient die zu prüfende Therapie erhält. In einer 1:1-randomisierten Studie ist diese gerade 0,5. In einer nichtrandomisierten Studie ist diese Wahrschein-

Textuelle Gegenstände der Computerlinguistik: Subjektivität – Meinungsanalytik (opinion mining)

Users Speak Out: Canon PowerShot SX10 IS Submitted by Kamugin on 1/30/2009

BV: Allison J, DigitalCameraReview.com Editor PUBLISHED: 4/13/2009

photographers – since November. User reviews from ou place as batteries, efficient image stabilization.



featuring a CMOS sensor and HD video recording. Howe the most-viewed cameras on our website. It sports a 11 OVERALL SCORE: 5/10

one of the best ultrazooms on the market

PROS: Huge zoom, hot shoe for external flashgun, tilt LCD, AA batteries equipped, good ergonomic design with useful buttons and controls, plenty of user friendly The Canon PowerShot SX10 IS has been on store shely features and full manual mode, compartment for memory card isn't in the same

in ever since, and we've rounded up a small sample for CONS: Small LCD, heavy, cheap lens cover and hood, cheap cover for USB connector, hard to open battery hatch, not too good autofocus especially with artificial light, no RAW mode, no battery level indicator.

OVERALL SCORE: 8/10

Slow Lens, Clumsy Controls a Poor Mix

Submitted by Rachel*B on 11/7/2008

PROS: Excellent resolution, reasonably sharp images. Better noise control than Canon's previous "S" models, ISO usable through 400; 800 in a pinch. Image edges, corners are sharper. Lens aberration is less than average for this zoom range. Huge zoom range includes wide angle. Digital zoom images are surprisingly good. iContrast increases dynamic range, means more shadow detail, less highlight clipping. Bright vari-angle LCD with wider viewing.

CONS: Slow lens beyond the 100mm mark - image stabilization can't keep up unless ISO 1600 is employed, which is much too noisy. Need a monopod for much zooming below ISO 800 in less than bright light. Images have an unpleasant blue cast that is hard to stomach. Autofocus is very slow, difficult to achieve indoors and often fails altogether; and while autofocus outdoors in bright light is usually snappy. Since the release of the SX10, Canon added the SX1 to failure to focus also occurs sometimes even under the most optimal conditions. Control wheel is clumsy and frustrating.

Und dies ...

Directed Hate



@usr A sh*t s*cking Muslim bigot like you wouldn't recognize history if it crawled up your c*nt.You think photoshop is a truth machin



@usr shut the f*ck up you stupid n*gger I honestly hope you get brain cancer

Generalized Hate



Why do so many filthy wetback half-breed sp*c savages live in #LosAngeles? None of them have any right at all to be here.



Ready to make headlines. The #LGBT community is full of wh*res spreading AIDS like the Black Plague. Goodnight. Other people exist, too.

Was ist toxische Sprache?

- Sprachformen, die herabwürdigend bzw. beleidigend wirken und andere wegen ihrer Hautfarbe, sexuellen oder religiösen Orientierung, ihres Geschlechts oder ihrer körperlichen Versehrtheit/Besonderheiten (auch Behinderungen) in verletzender oder obszöner Form ansprechen
- Englische Termini: hate speech, harassment speech, offensive, obscene, abusive, derogatory language, cyber-bullying

Rechtliche und technische Aspekte

- Rechtliche Grundlagen
 - Grundgesetz: Recht auf freie Rede
 - Art 5. (1) "Jeder hat das Recht, seine Meinung in Wort, Schrift und Bild frei zu äußern und zu verbreiten und sich aus allgemein zugänglichen Quellen ungehindert zu unterrichten. Die Pressefreiheit und die Freiheit der Berichterstattung durch Rundfunk und Film werden gewährleistet. Eine Zensur findet nicht statt."
 - Bürgerliches Gesetzbuch: Beleidigung, üble Nachrede
- Digitale Kommunikationsformen
 - ... erleichtern Individuen, ihre Identität zu verschleiern (user names, camouflierte Email-Adressen usw.) und ermutigen sie dadurch, toxisch zu kommunizieren

Datensatz zu toxischer Sprache (Korpus-Ausschnitt)

- #illner. erst hieß es, es kämen nur top Arbeitskräfte. jetzt lese ich NUR von wichsenden, vergewaltigenden, betrügenden #asylanten. #merkel
- Erschreckend, daß es #Frauen sind, die 30 Jahre #Emanzipation zugunsten islamistischer #Vergewaltiger in die Tonne treten! #rapefugees #AfD
- Die heutige Tagesscheisse heist #Hoaxmap. Die tätsächlichen Vergewaltigungen durch #Flüchtlinge und #Asylanten werden nicht erwähnt.
- Was die allermeisten immer noch nicht begriffen haben: Der #Islam hat seine eigenen #Menschenrechte, #scharia-basiert. Ihr Deppen!
- #schweinefleisch für alle #Flüchtlinge und #Asylanten. Wer saufen kann und Kinder vergewaltigen, braucht keine extra ""Wurst"" ..Fuck Refugges
- Nicht in Schwimmbäder scheissen zu dürfen ist ein Verstoß gegen die Menschenwürde. #Hungerstreik #Rapefugees
- Kinderfickende, schächtende #rapefugees! Wer Kinder-, Frauen-, oder Tierrechte verteidigt, darf nicht #gruene #spd oder #cdu wählen! #ltwbw

Lexikon zu toxischer Sprache (Wiktionary)



Lexikon zu toxischer Sprache (Ausschnitt)

- Abbumser
- Abfallecker
- Abfalleimervagina
- Abfallficker
- Abfallschlucker
- Abfalltonnenvollscheißer
- Abficker
- Abgefickter
- Abortschüsseltaucher
- Abschaum
- Abscheißer
- Abspritzer
- Abspritzmuschi
- Abspritzmuschie

- Druckfurzer
- Druffi
- Drüberrutscher
- Dubbel
- Dullhans
- Dulli
- Dumbatz
- Dumbo
- Dummarsch
- Dummbatz
- Dummbold
- Dummbolzen
- Dummbrot
- Dummdepp

Lexikon zu toxischer Sprache (Ausschnitt)





Wir haben Dein Lieblingsschimpfwort vergessen? Dann aber schnell eingeben!

Die beliebtesten Schimpfwörter 1. Steckdosenbefruchter 2. Sohn einer blutpissenden Hafenhure 3. Monsterbacke 4. Spermarutsche 5. Evolutionsbremse 6. Gabbafotze 7. Homo-Fürst der Finsternis 8. Fickfehler 9. Karussellbremser 10. Teflongesicht 11. Arschgeficktes Eichhörnchen 12. Puffgezeugte Arschgeburt 13. Fettgondel 14. Hodenkobolt 15. Eunuch im Neoprenanzug .schimpfv

Schimpfwörter und mehr...

n<mark>pfw</mark>oerter.de/schimpfwoerter

Abstufungen toxischer Sprache

- @anna_IIna Kann man diesen ganzen Scheiß noch glauben..?
- (a) Training sample categorized as PROFANITY
 - @AchimSpiegel "Sigmar Dumpfbacke Gabriel" gefällt mir richtig gut
 - (b) Training sample categorized as INSULT
 - @diMGiulia1 Araber haben schon ekelhafte Fressen....!!
 - (c) Training sample categorized as ABUSE

Forensik toxischer Sprache

- Forensik-Ressourcen
 - Korpora mit toxischer Sprache
 - Lexika zu toxischer Sprache
- Computerlinguistische Forensik
 - Toxik-Filter: Klassifikation "toxisch" "nicht toxisch"
 - Typologische Forensik
 - · Geschlecht, Religion, sexuelle Orientierung, Rasse
 - Shared Task
 - GermEval Task 2018 Shared Task on the Identification of Offensive Language

Seminarleistungen

- Vortrag (mündlich)
 - 1-stündig
 - Elektronische Version (PDF, PPT) verfügbar machen
- Referat (schriftlich)
 - 15-20 Seiten Kerntext (mit Standardformaten)
 - Elektronische Version (PDF, DOC) verfügbar machen
 - Eidesstattliche Erklärung zur Eigenautorenschaft
 - Wir prüfen mit Plagiatserkennungs-Software
 - Abgabe: Ende Juli 2019

Bemerkungen zu Referaten

• Aufbaumuster:

- Deck- bzw. Titelblatt mit vollständigen Angaben
- Inhaltsverzeichnis
- Einführung ins Thema, Motivation
- Themenabhandlung: grundlegende Formalisierungen, Verfahrensbeschreibungen (Algorithmen), Systemfunktionalitäten, Ressourcenmerkmale, Experimente/Evaluationen usw.
- Fazit mit kritischer Würdigung, offene Probleme ansprechen
- Bibliographie

• Zitationen:

- Alle verwendeten Quellen zitieren
 - Mit einem bibliographisch korrektem Zitat die jeweilige Quelle eindeutig beschreiben
 - Fachartikel <u>nicht</u> mit http://...foo.pdf-Link zitieren
 - Online-Quellen mit URLs und Datum des letztem Zugriffs
- Wikipedia ist keine zitierfähige wissenschaftliche Quelle!
- Eigenleistungen (Literatur, Beschäftigung mit konkreten Ressourcen/Systemen usw.) sind sehr erwünscht→unabdingbar!

Wege zum Vortrag und Referat

- Email: Anmeldung von drei nach fallender
 Priorität geordneten Themenwünschen
 - First-come, first-served
- Email: Themenvergabe durch Dozenten
- Erste Literaturhinweise als "Saat" nach Bestätigung der Themenauswahl
- Themenbearbeitung durch Referenten
 - Mündlicher Vortrag zum vereinbarten Termin
 - Schriftliches Referat (unter Einhaltung der organisatorischen Verabredungen) zum vereinbarten Termin

Machine Reading

Ablaufplan

```
18.4.
          Hahn
25.4.
          Hahn – Themenvergabe
02.5.
09.5.
16.5.
           Gesprächstermin
23.5.
30.5.
           Gesprächstermin
06.6.
13.6.
20.6.
27.6.
04.7.
11.7.
```