

Social Botnets als politische Meinungsmacher in der Ukraine

Big-Data-Methoden zur Analyse von Twitter-Manipulationen

PD Dr. Simon Hegelich

Der Ukrainekonflikt als Fall für Political Data Science

Umsturz auf dem Maidan



Die Bedeutung von Social Media

- Kommunikationsplattform getrennt von den staatlich kontrollierten Kanälen
- Propagandawerkzeug
- Cyber Warfare
 - Desinformation
 - Surveillance
 - Social Engineering

Social Botnets

Fake-Accounts online kaufen

Provider	Quantity	Rate for 1000
Twitter.com Profiled	19747	1K-10K: \$35 10K-20K: \$33 20K+: \$32
Twitter.com UK+Asarap	14191	1K-10K: \$28 10K-20K: \$27 20K+: \$26
Twitter.com+Asarap	0	1K-10K: \$25 10K-20K: \$24 20K+: \$23
Twitter.com UK	24692	1K-10K: \$22 10K-20K: \$21 20K+: \$20
Twitter.com	50348	1K-10K: \$20 10K-20K: \$18 20K+: \$17
Twitter.com AOL	530	1K-10K: \$18 10K-20K: \$17 20K+: \$16

Bulk Twitter accounts can be used for twitter marketing. We're the only bulk account provider, who sell 7 kinds of Twitter accounts. We have Mail.ru-Based Twitter Accounts, Yahoo.com-based, and AOL.com-based Twitter accounts. We also have **bulk twitter accounts with random avatar**, twitter accounts with bio and location, and twitter accounts with followers.

All our **twitter accounts** are verified by e-mail. This means you won't have any warning messages or e-mail confirmation requests: accounts are ready for immediate usage. All twitter accounts come with E-mail mailboxes included in the default format.

BuyAccs.com

BUY BULK ACCOUNTS AT BEST PRICES

Ein neues Phänomen?

- **Netzwerk von Fake Twitter-Accounts**
 - (Boshmaf et al. 2013)
- **Durch eine Software gesteuert**
 - (Hegelich/Janetzko 2015)
- **Sehr autonom**
 - (Hegelich 2015)
- **Dadurch schwer zu entdecken**
 - (Chanda 2014)
- **Herausforderung für die Twitterforschung**

Social Botnets: Der Fall Ukraine

Social Botnet Software



Aufspüren des Social Botnet

- **Ausgangssample**
 - Tweets vom 22.02.2014 mit Hashtag #Ukraine (1.3 Millionen Tweets)
- **Tweets mit identischem Text, aber kein Retweet**
- **Überprüfung der Metadaten**
 - Source URL: vk-prim.ru
- **Analyse des Friend/Follower Netzwerks**
- **18.000 eindeutig identifizierte Bots**

Das Twifarm-Botnet

Ярослав Гур
@gur16012
Для ревности нет ничего хуже смеха.
📍 Warren

TWEETS 2.525 FOLGE ICH 554 FOLLOWER 403 [Folgen](#)

Tweets Tweets & Antworten

Ярослав Гур @gur16012 · 10 Std.
Визитная карточка настоящего мужчины — женщина, которая счастлива от того, что он рядом

Ярослав Гур @gur16012 · 8. Apr.
День памяти погибших подводников: Когда наверх уже не всплыть…

Ярослав Гур retweetete
Вера Карпушенко @ECW97VzZj1KnDcr · 3. Apr.
Правий сектор» говорить, що для того, щоб повернути Криму та Донбасу українське

Ярослав Гур retweetete
Анатолий Сахно @sakhno_tolik · 2. Apr.
RT @akuma_no_ficl: Кравчук, Кучма и Ющенко вошли в состав Конституционной комиссии Украины

Verpasse keine Neuigkeiten von **Ярослав Гур**

[Registriere Dich bei Twitter!](#)

Trends · [Ändern](#)

- #PrayForMamaSwift
- #WalterScott
- #AnimalFacts
- #BadShowerGifts
- #willithotpocket
- Sabra
- Charley Hoffman
- Ben Crane
- Cherry Bomb

© 2015 Twitter Über uns Hilfe Werbung Info

Das Twifarm-Botnet



Twitter durchsuchen



patetpecoli
@patetpecoli
Черная полоса.Интересно,а я вдоль или поперек иду?
Казань

3 Fotos und Videos



Tweets Tweets & Antworten

patetpecoli retweetete
July Litovets @youlichka00 · 23 Std.
Мой твиттер превратился в страничку из ссылок на инстаграмм ...

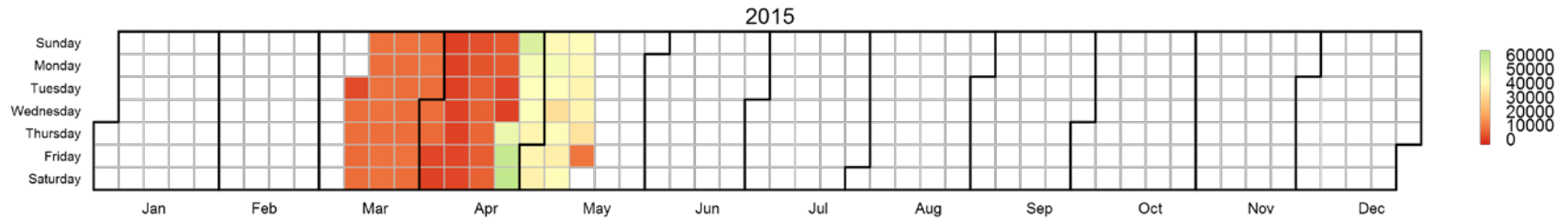
patetpecoli @patetpecoli · 3 Std.
Кличко пообещал не менять кресло мэра на депутатское

patetpecoli retweetete
Даша @cintisrsecwe · 6. Okt.
Тому, кто умеет ждать, всегда достается самое лучшее.

patetpecoli retweetete
Mura Inkles @quheqibyjiqy · 10 Std.
Принятый закон об очищении власти может испортить репутацию Украины,…

Deskriptive Statistik: Aktivitäten

Calendar Heat Map of Bot Activity



Beispiel-Tweets

- "Hollywood stars did not come to the party of the British Prime Minister"
- "Bermuda Triangle: sofa- kitchen-computer ... Oh, and a lot of people in it was gone !!!"
- "Who is to blame? Maidan? Putin? Yanukovych? right sector? Oligarchs? Russia? West? Tymoshenko? Kravchuk? Kuchma? Yushchenko ..."
- "RT @ ArtemD1114: RTsaiyawame: right quadrant Kherson: A policeman wished his colleagues from the battalion "Kherson" luck "in the fighting in the East"
- "RT @ ArkadijDR5C: Hercules beginning of legends download <http://t.co/VvSjK0d2YP>"

Text-Mining

Wordcloud

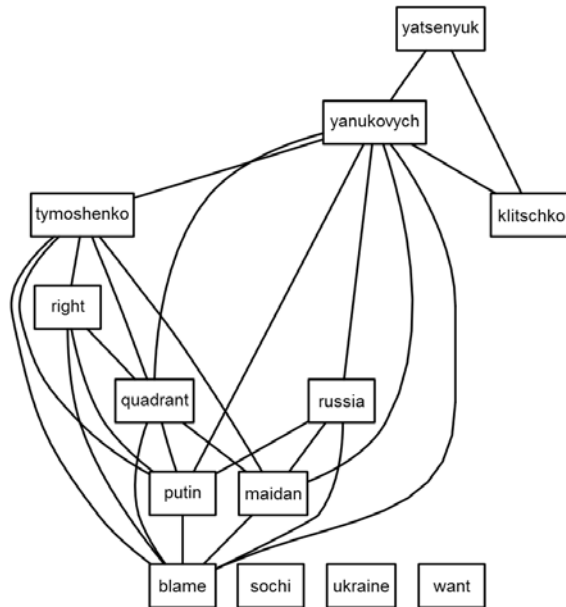


The Signal and the Noise

- Sample
 - 100.000 zufällig ausgewählte Tweets
 - Google-Translation API
- Bag of Words
- Häufige Worte
 - Maidan, Russia, Yanukovych, Putin
- Mehrzahl der Worte ist selten
 - Godzilla, Hercules, Beer

Text-Mining: Korrelationen

Korrelationsplot



Auswertung

- Worte mit Korrelationskoeffizient über 0.5
- Politische Begriffe überwiegen
- „right quadrant“ = Rechter Sektor

Unsupervised Learning: hierarchisches Clustering

Algorithmus

1. Begin with n observations and a measure (such as Euclidean distance) of all the $\binom{n}{2} = n(n-1)/2$ pairwise dissimilarities. Treat each observation as its own cluster.
2. For $i = n, n-1, \dots, 2$:
 1. Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters. The dissimilarity between these two clusters indicates the height in the dendrogram at which the fusion should be placed.
 2. Compute the new pairwise inter-cluster dissimilarities among the $i-1$ remaining clusters.

(James et al. 2013)

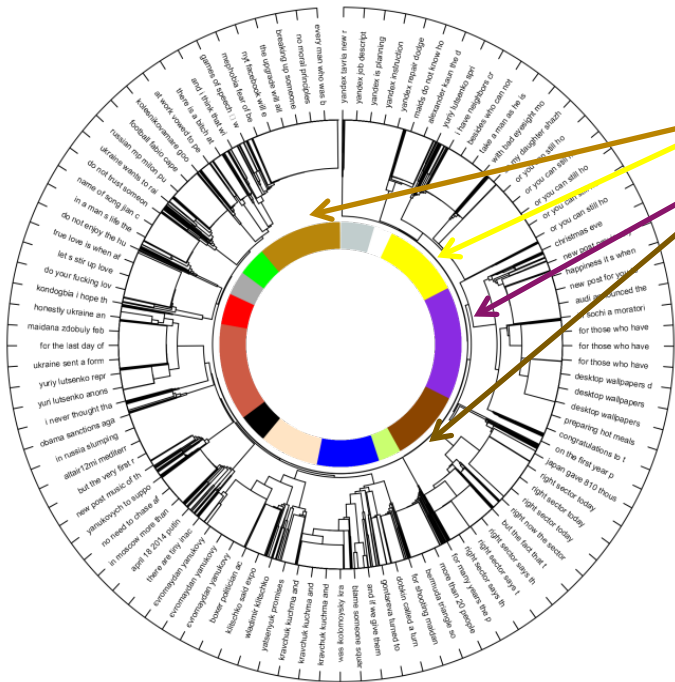
Bewertung

- Erläuterung:
 - Startet mit jeder Beobachtung in einem Cluster.
 - Verbindet je zwei Beobachtungen/Cluster, die besonders ähnlich sind.
- Pros:
 - Findet Ähnlichkeiten, ohne zu wissen, wonach gesucht wird.
 - Eindeutig!
 - Für GPU-Computing parallelisiert.
 - Erlaubt unterschiedliche Distanzfunktionen.
 - Kann durch Bootstrapping überprüft werden.
- Cons:
 - Nicht so gebräuchlich.
 - Komplexe Ergebnisse.
 - Keine definitive Clusteranzahl.
 - (Speicher Fehler in R bei sehr großen Datenmengen -> R-spezifische Big Data Probleme)

Vergleich der Algorithmen am konkreten Fall wird empfohlen. (Hastie et al. 2009)

Übertragung auf weitere Daten

Chorddiagramm



Erläuterung

- Anzahl der Cluster muss angepasst werden
- Signal
- Noise

3 Arten von Botverhalten

- **Verschleierung**
 - Scheinbar belanglose und normale Inhalte
- **Attraktion**
 - Inhalte für Zielgruppen (Nachrichten)
 - Inhalte von genereller Attraktivität (Downloads wegen Google)
- **Verstärkung**
 - Verbreitung von Hashtags
 - Propaganda
- **Political Data Science Methoden sind essentiell**

Each of your social media messages is strike to support Ukraine!

JOIN THE UKRAINE INFORMATION ARMY AND HELP UKRAINE DEFEND ITSELF ONLINE!

JOIN

Read more



WELCOME TO THE HEADQUARTERS OF THE INFORMATION ARMY!

You may know that Russia's war against Ukraine is a *"hybrid"*.
This is because the war is *real* but information about it is *false*.
There are a lot fronts opened against us and the *information* front is one of the most important.

In a year we have created a strong army wich courageously protects us in Donbass.
Now is time to fight back Russian occupants on the information front.

Every Ukrainian with Internet access can make his or her contribution in the information struggle. To do it you need:



JOIN



ACCURATELY PERFORM



DEVOTE TIME DAILY

PoliticalDataScience.blogspot.de

Political Data Science

suchen

Mosaic ▾ | Startseite Data Mining für SozialwissenschaftlerInnen Crashkurs R

#LinearRegression has no existence - #Hegel would say

In #datamining, we are bored by linear regression. It does not work very well and I have personally never seen a qqplot that was strictly on a line, anyway.

How to select specific random rows from #MySQL within #F

When you work with big data

ker friseur (p=0.8224)

hen correlate over that time only



#Hegel: Why there will never be artificial intelligence (#AI)

... but perhaps something more powerfull.

Lately, I have been thinking about AI a lot. Right now, I am reading Hegel again, and I am trying to do it seriously (sorry guys, I do not know if this could be done in English...).

What strikes me is this: There is no intelligence.



Handling #bigdata in R (1): data.table

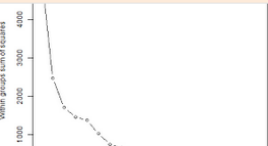
How big is big? Are you fit for real big data?



Using ngrams with #RTextTools

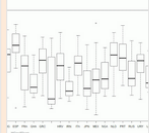
This little example shows a workaround for a bug in RTextTools.

Using ngramLengt would lead to an error.



RandomForest to predict the Soccer World Cup

I now present the R-code for the random forest prediction (the group-variable was added here):



#NRW - Tweets aus Nordrhein-Westfalen in #R saugen

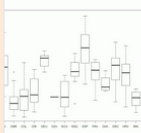
In diesem Post zeige ich ein kurzes Beispiel, wie man das streamR-Package nutzen kann, um Tweets aus NRW zu speichern.

Voraussetzung ist, dass man...

Love your #nearest neighbors

In this post the k-nearest neighbors algorithm is used to classify Twitter-data.

We have 1000 tweets about fracking which are labeled "c" = contra fracking, "d" = pro



Was @RitaOra #hacked? Probably not really... #Twitter-Forensics

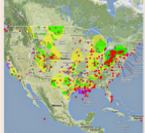
According to



#SimulatedData #R #caret

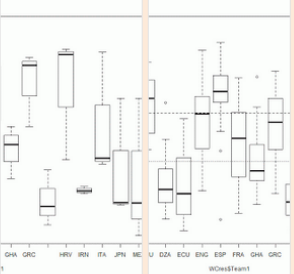
I just noticed a very cool function in the caret-package, I would like to share.

The package can produce simulated data, which is very useful for Monte Carlo Simulations. or



What is wrong with US #energy #budget 2000?

The policy agenda project (PAP) database shows for year 2000 an annual percentage change of MINUS 218 PER CENT (!!!). I was wondering all the



Lösungen zu Übungen I

...

Übungen zur Linearen Regression und Kreuzvalidierung

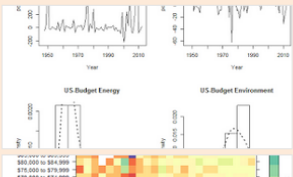
Hallo! Die folgenden Übungen fassen den Stand des bisherigen Seminars zusammen. Die Lösung wird dann nach Montag nachgereicht.

Getting Data

Ein kleines Beispiel, wie sich Eurostat-Daten direkt in R einlesen lassen...

Cross-Validation

Im Folgenden gehen es um drei Methoden der Cross-Validation hold-out, leave



Linear Regression in R

In diesem Script geht es um lineare Regressionen mit mehreren Prediktoren. Angelehnt ist da

Lineare Regression erklärt für Nicht-StatistikerInnen

Eine Regression beschreibt einen Zusammenhang zwischen einer abhängigen

Twitter dates

Wer Twitter-Date in R einliest, stößt auf Datumsvariablen ("created_at" un

Simulation zu

"") - d(adi

Literatur

- Boshmaf, Y., Muslukhov, I., Beznosov, K. and Ripeanu, M. Design and analysis of a social botnet. *Computer Networks*, 57, 2 2013, 556-578.
- Chanda, K. Hybrid Botnet Detection Mechanism. *International Journal of Computer Applications*, 91, 5 2014, 12-16.
- Grabau, M., Hegelich, S.: The Gas Game: Simulating Decision-Making in the European Union's External Natural Gas Policy. Under review at *Swiss Political Science Review 2015*.
- Hastie, T., Tibshirani, R. and Friedman, J. *The elements of statistical learning*. Springer, New York, 2009.
- Hegelich, S. Herrschaft – Staat – Mitbestimmung, Springer-VS, Wiesbaden, 2013.
- Hegelich, S. Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian/Russian Social Botnet, presented at *Midwest Political Science Association, Annual Conference 2015*.
- Hegelich, S. Decision trees and random forests: Machine learning techniques to classify rare extreme events, under review at: *European Political Science*, special issue: "Quantitative Approaches to Comparative Analyses: Data properties and their implications for theory, measurement and modeling", 2015.
- Hegelich, S. Simulations in Politics and Technology: Innovation policies in the field of photovoltaic cells, in: Hilpert, Ulrich (Hrsg): Handbook of Politics of Technology, Routledge 2015.
- Hegelich, S., Janetzko, D. Social Bot Behavior on Twitter, Submitted to: *The Fourth International Conference on Data Analytics, DATA ANALYTICS 2015*.
- Hegelich, S., Fraune, C. and Knollmann, D. Point Predictions and the Punctuated Equilibrium Theory: A Data Mining Approach—US Nuclear Policy as Proof of Concept. *Policy Studies Journal*, 2015.
- Hegelich, S., Shahrezeya, M. The communication behavior of German MPs on Twitter: Reaching to the converted and attacking opponents. Under review at European Policy Studies.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. *An introduction to statistical learning*. Springer, 2013.
- Jones, B. D., Baumgartner, F. R., Breunig, C., Wlezien, C., Soroka, S., Foucault, M., François, A., Green-Pedersen, C., Koski, C. and John, P. A general empirical law of public budgets: A comparative analysis. *American Journal of Political Science*, 53, 4 2009, 855-873.
- Sabatier, P. A. and Weible, C. M. *The Advocacy Coalition Framework*. Westview Press Boulder, City, 2007.
- Schelling, T. C. Dynamic models of segregation. *Journal of mathematical sociology*, 1, 2 1971, 143-186.

Vielen Dank für Ihre Aufmerksamkeit.