Time-aware online reputation analysis

Peetz, M.-H.

Publication date
2015

Document Version
Final published version

Social media has become an integral part of society. Omnipresent mobile devices allow for immediate sharing of experiences. Experiences can be about brands and other entities. For social media analysts a collection of posts mentioning a brand can serve as a magnifying glass on the prevalent opinion towards a brand: The overall estimation of a its reputation is increasingly based on the aggregation of a brand's reputation polarity in social media posts. This polarity of reputation is currently annotated manually. However, with the dramatic increase of social media, this is no longer feasible.

This thesis aims to facilitate and automate parts of the process to estimate the reputation of a brand. We motivate this by performing user studies with expert social media analysts. We analyse three resulting datasets: a questionnaire, log data of a manual annotation interface, and videos of annotating experts following the think-aloud protocol. We find the online and offline authority of the user posting online influences the annotation decision most. This online and offline authority is therefore a strong indicator for reputation polarity of this posting. Additionally, experts welcome automation of information retrieval and filtering tasks. For both, information retrieval and filtering, as well as for several indicators, the reputation analysts' background information proves vital.

Based on the indicators used for manual annotation, we proceed with the development of algorithms for the automatic estimation of reputation polarity. Unlike earlier, static evaluation scenarios, we follow a dynamic scenario, which mimics the daily workflow of social media analysts. Our algorithms are successful because we distinguish between reputation and sentiment.

The second part of this thesis is motivated by the analysts' desire for automation of retrieval and filtering of new media. For information retrieval, we present two improvements to existing algorithms. The first improvement is based on burst identification in time series of pseudo-relevant documents. We sample terms in those bursts for query modeling, thus improving effectiveness on news and blog corpora. Secondly, recency is an important aspect of relevance in social media. Inspired by memory models from cognitive science, we point out new (cognitive) document priors based on those models. We show that those priors are more effective, efficient, and plausible than commonly used temporal priors.

Background knowledge is essential for information filtering. Additionally, topics around an entity are dynamic. We find that for a consistently strong performance of filtering algorithms, the expertise of social media analysts is needed. The algorithms for filtering presented in this thesis are based on active learning: For documents where the algorithm cannot classify with high certainty, manual labels are requested from the analyst. Using intuitions about bursts and cognitive priors for sampling difficult to classify documents, we need very little help from annotators to reach high effectiveness.

We conclude that many aspects of the annotation of reputation can be automated – using in particular time series analysis, memory models, and low-impact help from expert social media analysts.

ISBN 978-94-6182-540-7

MARIA-HENDRIKE PEETZ TIME-AWARE ONLINE REPUTATION ANALYSIS

TIME-AWARE ONLINE REPUTATION ANALYSIS

MARIA-HENDRIKE PEETZ

Time-Aware Online Reputation Analysis

Maria-Hendrike Peetz

Time-Aware Online Reputation Analysis

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
Prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in
de Agnietenkapel
op dinsdag 24 maart 2015, te 10:00 uur

door

Maria-Hendrike Peetz

geboren te Soltau, Duitsland

Promotiecommissie

Promotor:
Prof. dr. M. de Rijke

Co-Promotor:
Prof. dr. W.M. van Dolen

Overige leden:
Prof. dr. F.M.G. de Jong
Prof. dr. M. Welling
Dr. F. Diaz
Dr. J. Gonzalo

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

SIKS Dissertation Series No. 2015-07
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

The research was partially supported by the Dutch national program COMMIT.

The research was partially supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 288024.

Cover by David Graus
Printed by Off Page, Amsterdam

ISBN: 978-94-6182-540-7

Für meinen Vater, Heiner Peetz

Dankwoord

Maarten, hi. Without you, this book wouldn't exist. You knew when to push me and when to stop me. Unlike your own prognosis, I never hated you. Thank you.

Edgar, Wouter, and Manos. You taught me the nitty bits at the beginning, with more patience than I had myself.

Daan, David, Richard, Evgeny, Anne, and Zhaochun. We grew together as PhD students, what a ride. There is much I would like to take from each of you: Open arms from Daan, humor from David, calmness and deep problem solving from Richard, elephant skin from Evgeny, *yes, but* from Anne, and motivation from Zhaochun.

Willemijn. You inspired me to think differently, not from a computer science perspective. You are a great role model.

Martin. I will always remember *my spot*, hacking and drinking away. Thank you for telling me I could program, it took me years to believe you.

Sara. You are the smartest and kindest woman I know. Spending time with you goes easy.

Shinbukan, my beloved dojo. Thank you for throwing me, spending wonderful saturday afternoons in cafés, and (training) holidays at Balaton. You all taught me something. Ferenç-sensei showed me my force-power. Joost-sensei taught me how to fly. Gerald-sensei convinced me that there is a sweet-spot between fear and suicidal bravery. Edo, you believed in me, in academia and aikido. Grainne, you found my grounding. And Desmond, soulmate, you found my mirror.

Jessie und Wendy. Danke für die Stunden die wir lachend, weinend, diskutierend, organisierend, streitend (über likelihoods und andere semantischen Differenzen), einander stützend verbracht haben. Danke Dir besonders, Jessie, für die Inspiration zu Kapitel 7.

Mama. Danke dass Du mir gezeigt hast, dass man alles erreichen kann, aber die Liebe immer das wichtigste im Leben bleibt.

Thomas. In diesem Buch steckt mehr von Dir als Du denkst. Ich liebe Dich.

Contents