Course Desription

Risks and Opportunities of Social Media Data

Contents

The Big Idea

In today's online society, social media gives users a wealth of opportunities to be connected with other people, to express themselves to a bigger audience or to gain all kind of information from others. Thereby, all these activities produce data, that is collected by the service providers. It is often not clear to what extent a service provider makes use of this data. On the one hand, personal data collections can serve a useful and necessary purpose, such as the improvement of the user experience or the facilitation of core functionality of the service. It can give also the opportunity for new businesses and research fields to emerge. On the other hand personal data collections has led to a widely applied practice of data analysis and personal information trading, which can highly compromise the users' privacy and lead to many further problems[1].

In face of these contradictions, this course deals with the implications for the society from personal data collections. Therefore the main subjects are the discussion of business models based on personal data[2], common practices to analyse personal data collections and alternatives to the current data driven social media landscape.

Intended learning outcomes

By the completing this course the participants will:
- be familiar with data and ad-based revenue models
- be aware of threats which exist due to the collection of personal data in social media and implications of these threats.

- be able to lead a discussion about different infrastructures for social media services and will know their strengths as well as shortcomings.
- Know the idea personal data markets and be able to lead a discussion about the matter and valuation of personal data
- know methods for protecting users' privacy in social network services with a single service provider and their benefits and limitations.
- have fundamental knowledge about data mining, machine learning and its usage.
- be able to explain basic algorithms, which are used for data mining.
- know ethical issues and problems, that arise with the usage of data mining when processing personal data.

Structure of the course

The Social Media Industry

Regarding today's social media services an overwhelming number of the popular services are provided by commercial companies. [3] Those companies have got expenses for maintaining and extending their server infrastructure, implementing new features in their services, employing qualified staff and so forth. Even though, many of these services remain free to use and therefore the providers have the need to establish other revenue streams than charging their users. Out of this need several business models have been evolved, which rely on the collection of user data, which will be a central part of this lecture.

With the aid of case studies this course raises critical questions about the modern data driven business models and gives the motivation for this subject. Then centralised infrastructures are introduced and its inherent difference to open technologies such as email are explained. Finally, the course summarised major problems of centralised social media services and its personal data collections.

The value of personal data and data markets

When it comes to the usage of social media services, users are not seldom confronted with a trade-off decision between privacy and the benefits of using such services. Revealing personal information to other parties can be beneficial for users (e.g. for convenience reasons or receiving discounts, interaction with other users), but it can also be a risk, when these data get known by other parties or used in harmful ways[4]. There are many ways how data can change the owner. These are for example data breaches, which happen frequently[5] and contribute to the distribution of personal data on black markets. First party service provider, which offer an incentive for users to share their data in the first place, also have the possibilities to trade their data collections or share these with their affiliates. Another source for personal can be third parties, which previously bought this data from several sources or collect personal data through tracking. Then the data gets aggregated and sold to any other interested party[6].

However, these markets give an indicator of a financial value of personal data in regards of the price a buyer is willing to pay. From a user's perspective, it might be more difficult to estimate the value of personal data and if it is worth to give up privacy for using a service in return[7].

This course explores questions regarding the value of personal data and looks at markets, where user data is traded. The course also approaches the question of the value of personal data from the user side. Therefore, personal data categorisations are examined, which demonstrate the granularity and origin of personal information. To round up this subject, the course discusses new developments in this area, such as personal data markets.

Decentralisation of Social Network Services (SNS)

Due to privacy concerns and other disadvantages many researchers have attempted to solve the problem of massive personal data collections by single authorities. One major direction is to decentralised the infrastructure of social media services, which is the subject of this lecture.

Decentralisation can be achieved by the usage of server federations or peer-to-peer networks. Services with decentralised infrastructure unburden single service provider by distributing the network load over several parties. Another main reason for decentralisation is to reduce economic incentives; to offer more scalability, openness and therefore stand out for the protection of users' privacy and freedom of speech[8].

However, services driven by a single service provider come with a lot of advantages in relation to security, which cannot be archived by decentralised services in the same manner. Therefore, new challenges need to be addressed in order to make these services open, secure, reliable, scalable and easy to use[1][9].

In order to get into this topic, several varieties of decentralized infrastructures, such as peer-to-peer systems and server federations, will be introduced and compared with a centralised infrastructure in the context of SNS. Afterwards, the course discusses which problems need to be addressed in such solutions and how privacy can be prevailed. At the end, the course looks also at strengths and limitations of these solutions in regards of security, availability, performance and scalability.

Data Mining and Machine Learning

The mass data from users' online activities, which is collected by social media services, contains useful information for the service providers, advertisers and so on. However, these collections are often by far too huge for analysing them manually. This is where data mining comes into play. With this technology, it is possible to discover not already known patterns in the data, which are also difficult to discover. However, data mining is the process to analyse the data, but other steps (like preparation, interpretation, selection and evaluation) are needed as well.[10].

This process is known as Knowledge Discovery in Databases (short KDD) and is the topic for this workshop. Furthermore, this course will also look at the related field machine learning. In order to gain an understanding of machine learning, the course will deal with the functioning of classifiers, pattern mining and clustering. With the knowledge of these algorithms in mind, the participants will explore ethical issues, which appear personal data is processed[11].

Didactic Concept, Schedule and Assignments

This course encompasses online workshops, exercises, case studies, basic readings and seminar papers. Foundations are built by discussions and talks during the workshops. Two basic readings will be given to the participations, which need to be prepared before the workshops. These readings will be use within the online workshops to address topics beyond the foundations. Additionally, case studies will be given during the workshops and used to point out critical issues.

In order to dive deeper into specific subjects advanced seminar papers are written by the participants. The subjects can be suggested by the students or taken from this list. The seminar papers are also the foundation for students' grades. Further information and requirements can be found on this page as well.

Introductory lecture

The first lecture begins with a recent case study to motivate the topic and give an overview of the state of the art as well as ethical issues and conflicts in the field of social media data. Afterwards the participants and the lecturer will explore the notions of technical terms and revenue models in an interactive way. The session ends with a short discussion about conflicts and alternatives of advertisement based revenue models and organisational aspects.

1st online Workshop

The course participants will work in groups and do a practical example to explore data collections in social media services. Afterwards the lecture present data markets and discuss these approaches with the participants. For the next workshop the paper from Schwittmann et al.[12] will be prepared by the students.

2nd online Workshop

At first the paper from Schwittmann et al.[12] is going to be discussed to study different types of decentralised SNS. Additionally the participants will experience use cases and user needs in several contexts for decentralised services. The students need to prepare chapter 6 & 8 & 10 from the book by Han et al.[10] for the next workshop.

3rd online Workshop

The course participants will delve into basic concepts of data mining and the KDD process on the basis to the paper from Han et al.[10]. Afterwards the students are going to work in groups examining and discussing three case studies to elaborate ethical issues in data mining and machine learning.

Wrap up session (on-site)

Students give presentations about their written papers. The presentation should build a foundation for a fruitful discussion with the course participants.

Examination

During the semester the participants will write a paper about one specific subject. A short presentation to this subject must be given within the final session and counts 20% to the final grade.

References

↑ 1.0 1.1 Zhang, Chi; Xiaoyan, Zhu; Yuguang, Fang, eds (2010) (in English). *Privacy and security for online social networks: challenges and opportunities*. Network, IEEE. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5510913&isnumber=5510907/.

↑ Albarran, Alan, ed (2013) (in English). *The Social Media Industries*. Routledge. ISBN 978-0-415-52319-6.

↑ Google+ and Youtube ⇒ Google, Skype ⇒ Microsoft, Facebook ⇒ Facebook Inc, Twitter ⇒ Twitter Inc., Xing ⇒ Xing AG>

↑ Honan, Matt, ed (2012) (in English). *How Apple and Amazon Security Flaws Led to My Epic Hacking*. www.wired.com. http://www.wired.com/2012/08/apple-amazon-mat-honan-hacking/all/.

↑ The Privacy Rights Clearinghouse, ed (Last seen November 2017) (in English). *Data breaches*. https://www.privacyrights.org/data-breaches.

↑ Dewes, Andreas, ed (2016) (in English). Build your own NSA - How private companies leak your personal data into the public domain, and how you can buy it.. 33C3 - Chaos Computer Club. https://media.ccc.de/v/33c3-8034-build_your_own_nsa.

↑ Acquisti, Alessandro; Wagman, Liad, eds (2015) (in English). *The Economics of Privacy*. Conditionally Accepted at the Journal of Economic Literature. http://dl.acm.org/citation.cfm?id=2392630/.

↑ Anwitaman, Datta; Strufe, Thorsten; Rzadca, Krzysztof, eds (2010) (in English). *Decentralized online social networks*. In Handbook of Social Network Technologies and Applications,. http://link.springer.com/book/10.1007%2F978-1-4419-7142-5.

↑ Paul, Thomas; Sonja, Buchegger; Thorsten, Strufe, eds (2012) (in English). *Exploring decentralization dimensions of social networking services: adversaries and availability*. New York, NY, USA: In Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, HotSocial '12. http://dl.acm.org/citation.cfm?id=2392630/.

↑ 10.0 10.1 10.2 Han, Jiawei; Pei, Jian, eds (2011) (in English). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann. ISBN 978-0-12-381479-1.

↑ Custers, Bart; Schermer, Bart; Zarsky, Tal, eds (2013) (in English). *Discrimination and Privacy in the Information Society - Data Mining and Profiling in Large Databases*. **3**. Heidelberg: Springer

Berlin Heidelberg. ISBN 978-3-642-30486-6. http://link.springer.com/book/ 10.1007/978-3-642-30487-3/.

↑ 12.0 12.1 Schwittmann, Lorenz; Boelmann, Christopher; Weis, Torben, eds (2014) (in English). *Privacy Preservation in Decentralized Online Social Networks*. IEEE Internet Computing, vol. 18. http://www.computer.org/csdl/mags/ic/2014/02/mic2014020016-abs.html.

27.02.2019