Sample Chapter LNCS 8401: Interactive Knowledge Discovery and Data Mining: State-of-the-Art and Future Challenges in Biomedical Informatics

Andreas Holzinger[1]

[1]Medical University Graz, Institute for Medical Informatics, Statistics and Documentation
Research Unit HCI, IBM Watson Think Group, Auenbruggerplatz 2/V, A-8036 Graz, Austria
a.holzinger@hci4all.at

Abstract. One of the grand challenges in our networked world are the large, complex, multi-dimensional and weakly structured data sets (big data) and the masses of unstructured information. These increasingly enormous amounts of data require new, efficient and user-friendly solutions for knowledge discovery. This challenge is most evident in the field of Biomedicine: the trend towards personalized medicine has resulted in a data tsunami in the amount of generated (-omics) data. This is the template for papers contributing towards a synergistic combination of methodologies and approaches of two areas may offer ideal conditions towards unraveling these problems: Human-Computer Interaction (HCI) and Knowledge Discovery and Data Mining (KDD). The idea of this book is to provide a rapid, short and concise overview about current state-of-the-art research and open questions and future research challenges to stimulate debate and to be used as input for joint grant proposals. Although the samples and cases are presented mainly from life sciences, the approaches shall be seen very generic and ensure added values in other domains as well.

Keywords: Methods, Knowledge Discovery, Interactive Data Mining, Future Challenges, Research routes, … (Minimum of three Keywords)

1 Introduction

This first chapter shall provide a short and concise introduction and motivation on why and how this topic is important and for whom.

One of the grand challenges in our networked world are the large, complex, and often weakly structured data sets, or massive amounts of unstructured information. This “big data” challenge (V4: Volume, Variety, Velocity, Veracity) is most evident in the biomedical domain [1]: the trend towards personalized medicine (P4 Medicine: Predictive, Preventive, Participatory, Personalized) has resulted in an explosion in the amount of generated biomedical data sets – in particular Omics data (e.g. from

genomics, proteomics, metabolomics, lipidomics, transcriptomics, epigenetics, microbiomics, fluxomics, phenomics, etc.).

A major challenge is in how to deal with this data? Or more concise: How can we discover knowledge out of these massive data sets?

A synergistic combination of methodologies and approaches of two areas may offer ideal conditions towards unraveling these problems: Human-Computer Interaction (HCI) and Knowledge Discovery and Data Mining (KDD), with the goal of supporting human intelligence with machine intelligence – to discover new, previously unknown insights into the data [2]. To form a network of excellence, to bring passionate researchers from diverse areas in an highly interdisciplinary manner together, to stimulate fresh ideas without any boundaries and encourage multidisciplinary work, the expert network HCI-KDD has been established [3].

A grand goal is to bundle synergies, to participate in joint project proposals for getting funding on various levels, inclusive travel funds and student exchange to move forward solutions in the above mentioned challenges.

The first goal is to support our network of excellence with the COST program [4], which is a flexible, fast, effective and efficient tool to network and coordinate nationally funded research activities, bringing excellent scientists together under light strategic guidance. COST is based on networks, called COST Actions, centered around research projects in fields that are of interest to at least five COST countries.

The COST program is an ideal basis for the development of joint projects in HORIZON 2020, which is the largest EU Research and Innovation program ever since, with nearly 80 BEUR of funding available over 7 years (2014 to 2020) – in addition to the private investment that this money will attract. It promises more breakthroughs, discoveries and world-firsts by taking great ideas from the lab to the market [5].

HORIZON 2020 offers a couple of interesting opportunities, including:

1) Excellent Science > ERC – FET – Marie Curie – Infrastructure with free choice of topic

2) Industrial Leadership

3) Societal Challenges > Health, Demographic Change and Wellbeing

Information and Communication Technologies (ICT) samples include:

ICT 15 – 2014: Big data and Open Data Innovation and take-up

Specific Challenge: The activities supported under this topic address the general technological and systemic data challenges that concern entire value chains and/or bridge across borders, languages, industries and sectors. The aim is to improve the ability of European companies to build innovative multilingual data products and services, in order to turn large data volumes into semantically interoperable data assets and knowledge.

ICT 16 – 2015: Big Data research: The activities supported within LEIT under this topic contribute to the

Big Data challenge by addressing the fundamental research problems related to the scalability and responsiveness of analytics capabilities (such as privacy-aware machine learning, language understanding, data mining and visualization). Special focus is on industry validated, user-defined challenges like predictions, and rigorous processes for monitoring and measurement.

2 Glossary and Key Terms

This chapter shall define the most important or ambiguous terms used in the paper to avoid any danger of misinterpretation and to ensure a common understanding.

Example:

Transdisciplinary research: the latin prefix “trans” denotes transgressing the boundaries defined by traditional disciplinary modes of enquiry. Whilst interdisciplinary is a “mix of disciplines”, transdisciplinary tend towards a “fusion of disciplines”, i.e. team members from various disciplines work together by application of a shared conceptual framework, which draws together concepts, theories, and approaches from multiple disciplines [6], [7], [8].

3 State-of-the-Art

This chapter is the main part and may be divided into subchapters, but remember to be concise and straightforward.

Establishment, development and management of successful research and development require systematic knowledge and skills and a target-oriented process model. It begins with a vision and requires a clear mission and accordant strategy in order to achieve these goals. The people involved in the team are of primary importance; everything depends on the interaction of this team. To create this team, to develop, scaffold, advance and lead this team is a grand challenge. However, even the best team is ineffective if there is no funding. Money is not everything but without money everything is nothing. By all love to science, a substantial budget is required to cover staff costs, premises and basic equipment, travel costs, computing facilities and basic software, a scientific software portfolio, hosting, special equipment, literature, the organization of workshops, hosting visiting researchers, financing open access publications, etc. In an environment of decreasing public budgets, external funding becomes increasingly important in order to sustain international competitiveness, top-quality and to maintain excellence. Ultimately, the team is assessed by output, which is composed of measurable, published "items" [9].

3.1 Subchapter

Ideally you should provide references to related work including at least three sources: the oldest reference, the most cited reference, and the most recent reference.

3.2 Advantages and Disadvantages

The paper shall contain an overview on advantages and disadvantages, which is a useful quick reference for the reader.

4 Open Problems

The paper shall describe the most pressing open problems and unsolved aspects in this area.

5 Future Outlook

The main issue is to describe interesting future research avenues and possible approaches which can be directly used in research proposals. Here it shall be clearly determined which part can be addressed from the authors and which expertise from different areas is necessary.

Acknowledgements

<if applicable>

References

1. Holzinger, A.: Biomedical Informatics: Computational Sciences meets Life Sciences. BoD, Norderstedt (2012)
2. Holzinger, A.: Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Alfredo Cuzzocrea, C.K., Dimitris E. Simos, Edgar Weippl, Lida Xu (ed.) Multidisciplinary Research and Practice for Information Systems, Springer Lecture Notes in Computer Science LNCS 8127, pp. 319-328. Springer, Heidelberg, Berlin, New York (2013)
3. hci4all.at, http://www.hci4all.at/expert-network-hci-kdd/
4. http://www.cost.eu/about_cost
5. European Commission, http://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020
6. Mobjörk, M.: Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. Futures 42, 866-873 (2010)
7. Wickson, F., Carew, A.L., Russell, A.W.: Transdisciplinary research: characteristics, quandaries and quality. Futures 38, 1046-1059 (2006)
8. Lawrence, R.J., Després, C.: Futures of Transdisciplinarity. Futures 36, 397-405 (2004)
9. Holzinger, A.: Successful Management of Research and Development. BoD, Norderstedt (2011)