

## Editorial

Johannes Fürnkranz

Published online: 8 January 2015  
© The Author(s) 2015

You hold in your hands the first issue of *Data Mining and Knowledge Discovery* in 10 years that does not bear the name of Geoffrey Webb as its editor-in-chief on the cover. It is a great honor but a much greater challenge for me to try to fill his shoes. And those are really big shoes to fill.

As the editor-in-chief of this journal, Geoff has brought an enormous commitment and dedication to his role. His primary concern has always been the authors of the submitted papers, who should receive fast feedback of the highest quality, and, in case of acceptance, support for making the best out of their work. Most importantly, they should always be treated with the highest respect. As an editorial board member and action editor for this journal, I knew exactly that no review and no decision letter that I write will be unread by Geoff, and more than once he got back to me and pointed out shortcomings and improvements in my work for the journal. It is a daunting task to continue at this level, but I will do my best to live up to it.

Thanks to his efforts, the journal is now in a healthy state. When Geoff took over in 2005, the journal published six issues in two volumes with a total of 22 papers on a bit more than 600 pages. In 2014, the journal has returned to publishing a single volume per year, but this volume consisted of four regular issues and a double special issue with a total of 48 papers on more than 1,600 pages. From 2005 onwards, the impact factor of the journal had continuously risen to its highest level of 2.95 in 2009. Since then, it has taken some ups and downs but essentially maintained its high level, which is the highest for journals that focus specifically on data mining. I am glad that Geoff will continue to serve on the Advisory Board of the journal.

Where will the journal be heading in the next 10 years? One of the big challenges is open access publishing. The journal continues to be a commercial publication, but

---

J. Fürnkranz (✉)  
TU Darmstadt, Darmstadt, Germany  
e-mail: juffi@ke.informatik.tu-darmstadt.de

it tries to accommodate the needs of the research community. The journal does not charge mandatory publication fees, and allows authors to retain the copyright on their articles. Many researchers have unlimited free access to all articles through institutional subscriptions, and an increasing number of papers is published under Springer's Open Choice model which grants unlimited free access for a per-article publication fee. Timely reviews for the journal are acknowledged with a complementary book. As much as I value free access to academic publications as a researcher, I also see a continuing role for commercial publication, and am convinced that this journal will continue to strike the best balance between both worlds.

It is also interesting to speculate where our field will be heading. From its inception in the 1990s, knowledge discovery in databases has emphasized the scalability of learning and discovery algorithms, long before *big data* became a media buzzword. There is still no end in sight to the exponential growth of data in business, manufacturing, science, and, of course, the internet and social media. In the future, we can expect to see many success stories of data mining in these areas, and I will keep a particular eye on application papers describing advances in application areas that would not have been possible without data mining technology.

We are, however, also on the verge of the age of *big knowledge*. It will become increasingly important to not only be able to cope with huge volumes of data, but to operate in the context of massive knowledge bases. The Semantic Web and Linked Open Data are slowly but steadily permeating into everyday life. Systems like IBM's Watson, which won the Jeopardy! quiz show in a spectacular media event, demonstrate the potential of having large semantic databases at your fingertips. It is not so important that these knowledge bases are carefully crafted, consistent, and error-free, as was the goal of the expert systems of the 1980s and 1990s, as long as they are based on large, multiple, and redundant information sources, so that inaccuracies can be corrected through the power of multiple, independent evidence. Nevertheless, in the light of our large ever-growing body of existing knowledge and of the field's classic definition by Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy, namely the discovery of "valid, novel, potentially useful, and ultimately understandable patterns in data", we will need to pay increasing attention to the novelty of our discoveries.

Will our methods continue to scale up with the rapid growth of data volumes? Will we be able to cope with massive knowledge bases? Will we manage the integration of knowledge into the data mining process? I think so, but there is still a long road ahead. And I am certain that this journal will play a key role in the publication and proliferation of the results obtained on the way.