Changing the Focus of the IDA Symposium

Niall M. Adams[1], Paul R. Cohen[2], and Michael R. Berthold[3]

[1] Department of Mathematics, Imperial College London, UK
[2] Department of Computer Science, University of Arizona, USA
[3] Dept. of Computer and Information Science, University of Konstanz, Germany

Abstract. The IDA Symposium series had evolved into a meeting apparently concerned with routine data mining. This report describes our attempts to re-focus the ninth IDA symposium on objectives derived from the genesis of the series. These attempts included a dramatically modified call for papers, and a drastically changed reviewing procedure.

1 Background

Since 1995, The Intelligent Data Analysis (IDA) Symposium series has been a biennial event at locations around Europe. The series has produced a selection of high quality proceedings ([1,2,3,4,5,6,7,8]).

Originally conceived with the objective of promoting work at the intersection of data analytic disciplines (including statistics, machine learning, and artificial intelligence) for dealing with challenging real world problems, the Symposium has lost some of this ambition and focus. Instead of encouraging papers on complex, real-world problems, there was a tendency to favour papers looking much like mainstream data mining or machine learning. A rather unkind characterisation of such material is that it often appears to represent incremental advances on existing methodology and algorithms for *well defined* learning tasks: classification, regression, clustering, etc. While there is absolutely nothing wrong with such fare, there are already ample outlets for it, and there are not enough outlets for research that engages statistics, machine learning and artificial intelligence for modeling complex, real-world systems.

For some time, the IDA council had been keen to orient the IDA series in the direction of the original ambitions, updated according to the tremendous advances in data collection, data storage and analysis methodology that have occurred in the interim. [9] gives some ideas about the modern ambitions for IDA, including the importance of data challenges and the complexity of data analysis *processes* in modern real world problems. In order to address these new ambitions, we convened this extra Symposium (the ninth, in 2010), which broke the biennial cycle and was held for the first time in the USA.

There seem to be a number of possible reasons why the focus of IDA had slipped. First, many authors around the world are keen to submit work about vanilla data mining – a very rich and important subject. Second, aspects of the standard reviewing process may have acted to favour such material. Based on these conjectures, we decided to revise the call for papers and reviewing process.

P.R. Cohen, N.M. Adams, and M.R. Berthold (Eds.): IDA 2010, LNCS 6065, pp. 1–5, 2010.

These revisions are discussed in the following sections, followed by a discussion of what happened, and some recommendations for the future.

We note that the reinvigoration of both conferences and journals seems to be part of the zeitgeist. Many recent conferences are including breaking news sections, or otherwise encouraging different types of submission.

2 Call for Papers

It is interesting to speculate about the the role and importance of the Call for Papers (CfP). Notably, the CfP had perhaps not changed sufficiently between IDA Symposia over recent years. IDA has become a regular part of the conference calendar and developed a specific identify associated with this CfP.

We produced a *completely* rewritten CfP for this symposium. This CfP included the following

> "IDA 2010 ... particularly encourages papers about
> 1. end-to-end software systems that incorporate several of these [*data processing and analysis*] technologies;
> 2. applications of these technologies to modeling complex systems such as gene regulatory networks, economic systems, ecological systems, resources such as water, and dynamical social systems such as online communities; and
> 3. robustness, scaling properties, and other usability issues that arise when these technologies are put into in practice."

Note the special emphasis on applications, processes, and practicality. The particular tastes and interests of the conveners placed great stress on modeling complex and dynamic systems.

In addition, and notably, the CfP explicitly *discouraged* certain types of submission, with

> "Papers about established technology will only be accepted if the technology is embedded in intelligent data analysis systems, or is applied to analysing and/or modeling complex systems. IDA 2010 does not encourage papers about isolated algorithms that refine, polish, extend slightly, or offer minor variants of established ideas; and it will not publish papers about classification, clustering, dimension reduction, and other conventional data mining topics unless they clearly contribute to problems of modeling complex systems."

Of course, a modified CfP is insufficient: the submitting population may not respond, or the programme committee may not take the details of the CfP into account. This, and other aspects of reviewing, are discussed in the next section.

3 Reviewing

There was a perception that the reviewing of the IDA Symposia (and perhaps many other meetings) had adopted a specific character, perhaps in response to

the deluge of data mining papers. Essentially, this reviewing appeared to look for easy rejects, by emphasizing what was wrong with a submission, rather than what was right. This approach often manifested as a narrow criticism of empirical experiments or the presence of free parameters. While these criticisms occasionally cut at the heart of a research effort, they more frequently cite technical infractions. They are the standard criticisms of unengaged reviewers who approach papers looking for reasons to reject them. Narrow, incremental algorithmic papers can be protected from these criticisms, but these are not the kinds of papers we wanted to publish in IDA.

There are two interesting aspects to the IDA view. First, very complicated problems are not going to be handled by single algorithms, but by cleverly constructed processes using a set of carefully designed algorithms. In this context, it is difficult to analyse performance over a collection of public domain data sets. Arguing about empirical performance of algorithms necessarily reduces the problem to pre-processed data sets, with algorithms as filters. This in turn leads to incremental refinements to existing algorithms and increasing abstraction from real problems. Again, we would like to stress that *there is nothing wrong with this approach*, it is simply not what IDA was intended to address. We thus modified the reviewing process for this IDA symposium to prevent easy rejects based on the standard criticisms of narrow, incremental work, and to encourage reviewers to focus on the new CfP.

Our proposed solution was rather ambitious. On the premise that it is easier to reject a paper than accept – which often leads to a specific sort of accepted paper – we felt that the relative anonymity of reviewing was not helping. While not wanting to disregard rigourous technical reviewing, we did want to emphasise the CfP. To this end, we introduced a *senior* PC, in addition to the *regular* PC, whose role was to *overrule* the regular PC if it was felt that a paper had merit with respect to the CfP despite technical shortcomings. However, to make such a judgment, members of the senior PC *had* to be willing (in principle) to publish a short summary in support of the paper *with their name on it*. This is a realisation of the idea that it is easier to trust a reviewer willing to publicly support a paper, than an anonymous reviewer who might be taking the path of least resistance. The senior PC was composed of members of the IDA council, and senior colleagues elected by recommendation.

The reviewing process worked as follows. Each member of the regular PC delivered a report with a recommendation, as in previous years. The regular PC were requested to pay attention to the new CfP, but to provide rigourous review. The *only* thing the senior PC was able to do was overrule a reject decision from the regular PC, at the risk of possibly publishing a short document in support. In this way, either regular or senior PC could recommend acceptance, but only the regular PC could reject. The senior PC were not asked to make reject decisions - they either accepted, or made no comment. This mechanism was simply an attempt to support interesting papers that did not meet the technical standards of narrow, incremental work.

4 What Happened

There were only 40 submissions to the Symposium. This may seem few, but a number of things need to be taken into account, including the new CfP, and a new slot in the calendar that overlapped with big meetings.

The submissions were a mixture of standard material and papers aligned with the new CfP. Each submission was reviewed by members of the regular PC, and at least one member of the senior PC, in the manner described above.

The senior PC recommendations were broadly in line with those of the regular PC. Very few rejects were overruled by the senior PC. We do not have a complete explanation for this – perhaps publicly supporting a paper is too intimidating, or perhaps the papers were not sufficiently exciting to energise the senior PC. On the other hand, many of the regular PC engaged fully with the revised call for papers, and toned down negative reviews, and specifically cited the new CfP in their reviews. When it came to the final decisions, such remarks were valuable and carried significant weight.

In the end, 21 papers were accepted. Five of these were identified as *application* papers, which were accepted because they tackled interesting and challenging application domains, even if their analysis was in some cases preliminary.

5 Where Next?

We were somewhat successful in attracting submissions matching the new CfP. The effectiveness of the revised reviewing process is more equivocal. Requiring the senior PC to "put their money where there mouth is" might have discouraged aggressively supportive reviewing (although this remark lives in the context of the submitted papers). However, the regular PC was sympathetic to the new CfP and reviewed accordingly, for the most part, so there were few opportunities for the senior PC to overrule the regular PC.

What have we learned? The CfP is important, and we think it should change frequently. Reviewing is important, but the two-tier system proved unnecessary because the regular PC generally did what we hoped the senior PC would do, leaving few opportunities for the senior PC to accept papers that it felt were wrongly rejected. It is possible that this role of the senior PC will become more important in future symposia, where we receive more papers. However, encouraging all reviewers to be guided by the CfP is crucial. In future, we might suggest a modified reviewer form, where reviewers are asked explicitly to comment (and score) the extent to which a submission has matched the CfP.

References

1. Liu, X., Lasker, G.E. (eds.): Advances in Intelligent Data Analysis, IIAS, Canada (1995)
2. Liu, X., Cohen, P.R., Berhold, M.R. (eds.): IDA 1997. LNCS, vol. 1280. Springer, Berlin (1997)

3. Hand, D.J., Kok, J.N., Berthold, M.R. (eds.): IDA 1999. LNCS, vol. 1642. Springer, Berlin (1999)
4. Hoffmann, F., Hand, D.J., Adams, N.M., Fisher, D.H., Guimar aes, G. (eds.): IDA 2001. LNCS, vol. 2189. Springer, Berlin (2001)
5. Berthold, M.R., Lenz, H.-J., Bradley, E., Kruse, R., Borgelt, C. (eds.): IDA 2003. LNCS, vol. 2810. Springer, Berlin (2003)
6. Famili, A.F., Kok, J.N., Peña, J.M., Siebes, A., Feelders, A.J. (eds.): IDA 2005. LNCS, vol. 3646. Springer, Berlin (2005)
7. Berthold, M.R., Shawe-Taylor, J., Lavrac, N. (eds.): IDA 2007. LNCS, vol. 4723. Springer, Berlin (2007)
8. Adams, N.M., Robardet, C., Boulicault, J.-F., Siebes, A. (eds.): IDA 2009. LNCS, vol. 5772. Springer, Berlin (2009)
9. Cohen, P.R., Adams, N.M.: Intelligent Data Analysis in the 21st Century. In: [7]