

Novel Algorithms and the Benefits of Comparative Validation

Rob Smith^{1*}, Dan Ventura¹ and John T. Prince²

¹Department of Computer Science, Brigham Young University Provo, UT 84602

²Department of Chemistry, Brigham Young University Provo, UT 84602

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

ABSTRACT

Contact: 2robsmith@gmail.com, jtprince@chem.byu.edu

Bioinformatic research has produced a large volume of proposed algorithmic solutions to a host of problems. Whether presented as a processing step in a clinical experiment or treated in a stand-alone publication, novel bioinformatic algorithms are often not subjected to the thorough comparative evaluation endured by their counterparts in other closely related fields—such as computer science—where an algorithm unevaluated against extant methods is considered unpublishable. Two audiences are interested in algorithmic publications: the practitioner, who may use the algorithm, and the researcher, who will work to develop solutions superior to those extant. We argue that failure during the review/publication process to require comparative evaluation for novel algorithms is detrimental to both parties.

To demonstrate the dilemma, we conducted a case study of novel LC-MS alignment algorithms. Of the 48 publications from 2001 to 2012 that present alignment algorithms of which we are aware, 60% include no comparison to other methods. Another 20% compare their method to one or two others (see Figure 1). Only two papers compare performance against the state-of-the-art methods available at the time of publication. Interestingly, both of these, with 6 and 7 comparisons respectively, reuse comparative evaluation performance data and data sets from a stand-alone review paper of 6 methods (Lange *et al.* (2008)).

It is natural to wonder if publication year correlates to the number of comparisons made. After all, earlier papers would have less methods to compare against. We found no correlation ($r=0.397$) between year of publication and number of comparisons (see Table 1). Again, the correlation number would be even lower if it weren't for the fact that someone published a comparative evaluation of at least some of the extant alignment methods. Without the reuse of that survey paper data, the correlation coefficient would drop to 0.313. These data reinforce the prevailing paradigm that comparative performance of a new algorithm to existing ones is too time consuming for the author and reviewers and ought to be the subject of dedicated research (Ballardini *et al.* (2011)). At least for alignment, such dedicated comparison studies are few and far between—we are aware of only one such comparative survey paper,

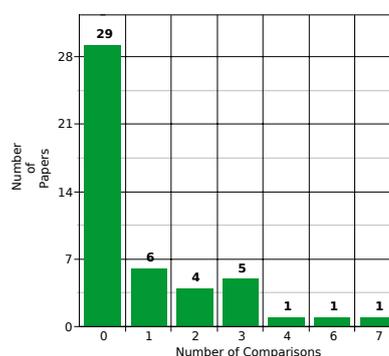


Fig. 1. A comparison of the number of papers presenting MS alignment algorithms and the number of competing algorithms against which they compare. The majority of novel alignment method papers fail to compare against even one extant method.

even though almost 50 new algorithm papers have been published over the last 11 years (see Lange *et al.* (2008)). Even if these evaluative review papers were more numerous, there are many reasons why these evaluations ought to be primarily provided in the novel algorithm publications themselves.

A practitioner relies on the peer review process to ensure that the methods they are choosing have met a minimum standard of quality. Though a new method's description or performance may be convincing, these qualities alone are insufficient to weigh the usefulness of an algorithm. Without comparative evaluation, algorithms that under-perform against existing ones can easily flood a domain, making the practitioner's task of selecting an algorithm more difficult with every additional publication. Besides an extensive literature review caused by the inundation of papers on the subject, the practitioner must also perform a comparative evaluation of the existing algorithms since they have no mechanism for quantifying the comparative strengths or weaknesses of the methods from the publications themselves. As pointed out by a recent paper, this process is as time consuming as it is difficult, given the oft-encountered difficulties of obtaining and then successfully running someone else's software (Ballardini *et al.* (2011)). Extensive comparative analysis reduces the practitioner's overall time commitment by reducing the number of algorithms under consideration as well as by providing a realistic expectation

*to whom correspondence should be addressed

Table 1. A list of papers presenting novel -omics alignment algorithms. The data has a correlation coefficient of 0.397, suggesting there is no trend towards comparison against extant algorithms.

Publication	#Comp	Year	Venue
Fraga <i>et al.</i>	0	2001	Anal Chem
Hastings <i>et al.</i>	0	2002	Rapid Com in MS
Bylund <i>et al.</i>	1	2002	J Chrom A
Torgrip <i>et al.</i>	2	2003	J Chemometrics
Åberg <i>et al.</i>	0	2004	J Chemometrics
Lee <i>et al.</i>	0	2004	Anal Chim Acta
Tomasi <i>et al.</i>	0	2004	J Chemometrics
Eilers	0	2004	Anal Chem
Vorst <i>et al.</i>	0	2005	Metabolomics
Pierce <i>et al.</i>	0	2005	Anal Chem
Walczak <i>et al.</i>	4	2005	Chem Intel Lab Sys
Baran <i>et al.</i>	0	2006	BMC Bioinformatics
Smith <i>et al.</i>	0	2006	Anal Chem
Sadygov <i>et al.</i>	0	2006	Anal Chem
Fischer <i>et al.</i>	0	2006	Bioinformatics
Jaitly <i>et al.</i>	0	2006	Anal Chem
Prince <i>et al.</i>	1	2006	Anal Chem
Skov <i>et al.</i>	0	2007	J Chemometrics
Yao <i>et al.</i>	0	2007	J Chrom A
Kirchner <i>et al.</i>	0	2007	J Stat Software
Palmblad <i>et al.</i>	0	2007	ASMS
Lange <i>et al.</i>	0	2007	Bioinformatics
Wang <i>et al.</i>	0	2007	Biostatistics
Mueller <i>et al.</i>	0	2007	Proteomics
Listgarten <i>et al.</i>	0	2007	Bioinformatics
Fischer <i>et al.</i>	2	2007	BMC Bioinformatics
Csenki <i>et al.</i>	3	2007	Anal Bioanal Chem
Åberg <i>et al.</i>	0	2008	J Chrom A
De Groot <i>et al.</i>	0	2008	Proteomics
Suits <i>et al.</i>	0	2008	Anal Chem
Shinoda <i>et al.</i>	0	2008	Bioinformatics
Christin <i>et al.</i>	1	2008	Anal Chem
Podwojski <i>et al.</i>	2	2009	Bioinformatics
Befekadu <i>et al.</i>	3	2009	IEE EMBS
Christin <i>et al.</i>	3	2010	JPR
Daszykowski <i>et al.</i>	0	2010	J Chrom A
Tomasi <i>et al.</i>	1	2010	J Chrom A
Bloemberg <i>et al.</i>	1	2010	Chem Intel Lab Sys
Eliasson <i>et al.</i>	0	2011	Curr Pharm Biotech
Sinkov <i>et al.</i>	0	2011	Anal Chim Acta
Befekadu <i>et al.</i>	3	2011	IEEACM TCBB
Tang <i>et al.</i>	3	2011	Prot Science
Ballardini <i>et al.</i>	6	2011	J Chrom A
Voss <i>et al.</i>	7	2011	Bioinformatics
Zhang	0	2012	ASMS
Struck <i>et al.</i>	1	2012	J Chrom A
Hoekman <i>et al.</i>	2	2012	ASBMB
Kaya <i>et al.</i>	3	2012	Inform Sciences

of performance, hopefully justifying the inevitable inconvenience of obtaining and operating new software. Often, evaluation is made much more difficult (if not impossible) when open source code is omitted in submission. While English descriptions and pseudocode assist in building intuition about an algorithm, they are lossy definitions that leave out essential details needed for code implementation. Besides time savings, requiring source code facilitates more expansive comparison through automation as well as providing the reviewers an easy metric to determine whether the method is suitably formally defined to be distributed and replicated or whether it is an ad-hoc agglomeration.

There are also secondary consequences to consider. Publication is an incentive that can drive innovation. If novel algorithms are not required to outperform extant ones, then innovation—true forward progress not necessarily achieved by mere invention—is less likely to occur. Finding the best choice in an expanding sea of mediocre choices then becomes a Herculean task sure to exhaust any practitioner. The practical result is that practitioners stop short of exhaustively evaluating all the possible options and choose based on some other criteria (e.g., popularity, ease of use, or familiarity). The inevitable outcome of the algorithm selection crapshoot are results poorer than what may otherwise have been.

Researchers (the algorithm makers) also suffer when comparative evaluation is neglected. In the face of burgeoning publication numbers, they encounter the same exhaustive search problem faced by the practitioner, but they also face a moral dilemma—the current environment makes it easy to generate many publications, yet very difficult to perform the sort of due diligence comparison advocated in this letter. A good comparison requires choosing among the several existing evaluation methods, each of which highlight only specific behavior. The choice is non-trivial—in alignment, metrics include metrics that evaluate the alignment in isolation (Christin *et al.* (2010, 2008); Van Nederkassel *et al.* (2006)), in combination with other data processing steps (Lange *et al.* (2008); Ballardini *et al.* (2011)), globally, and locally. One must also find data sets, which should include sufficient data representative of the different typical performance-affecting real-world characteristics (e.g., complexity of the data, variability of peptide concentration, number of unique and common peptides, extent and form of retention time shift in the data, etc.). What's more, there is no disincentive provided for publishing work untested against existing methods. Thus, left to their own devices, will the researcher ever behave in a manner that is not in his best interest, though it is in the best interest of the field? Apparently, not very often. Our experience suggests that the pattern we found in alignment algorithms applies to algorithmic approaches in proteomics and metabolomics generally, and it may extend to other bioinformatics subfields where we have less experience.

So what is the solution? The problem, we have found, does not lie in the lack of venue requirements for performance demonstration against state of the art algorithms. Interestingly, many of the papers with zero-comparisons came from journals that explicitly require authors to provide quantitative comparison with state-of-the-art methods. Similarly, though an openly available group of standard data sets and metrics as described here would greatly facilitate the evaluations petitioned for, authors in other fields manage to provide comparisons even without standardized metrics or open frameworks for evaluation.

We suggest that greater care be taken by editors and reviewers to require novel algorithmic contributions to contain a reasonable comparative quantitative evaluation with existing methods. New contributions should also include necessary elements to facilitate future comparisons with other algorithms such as source code and parameter setting guidance. Such an effort will inevitably maximize the outcome of practitioner results, encourage the widespread use of the highest-quality tools, and provide researchers an incentive to truly innovate.

Funding: This work was supported by the National Science Foundation Graduate Research Fellowship [DGE-0750759].

REFERENCES

- Ballardini, R., Benevento, M., Arrigoni, G., Pattini, L., and Roda, A. (2011). Massuntangler: A novel alignment tool for label-free liquid chromatography–mass spectrometry proteomic data. *Journal of Chromatography A*, **1218**(49), 8859–8868.
- Christin, C., Smilde, A., Hoefsloot, H., Suits, F., Bischoff, R., and Horvatovich, P. (2008). Optimized time alignment algorithm for lc- ms data: Correlation optimized warping using component detection algorithm-selected mass chromatograms. *Analytical chemistry*, **80**(18), 7012–7021.
- Christin, C., Hoefsloot, H., Smilde, A., Suits, F., Bischoff, R., and Horvatovich, P. (2010). Time alignment algorithms based on selected mass traces for complex lc-ms data. *Journal of proteome research*, **9**(3), 1483–1495.
- Lange, E., Tautenhahn, R., Neumann, S., and Gröpl, C. (2008). Critical assessment of alignment procedures for lc-ms proteomics and metabolomics measurements. *BMC bioinformatics*, **9**(1), 375.
- Van Nederkassel, A., Daszykowski, M., Eilers, P., and Heyden, Y. (2006). A comparison of three algorithms for chromatograms alignment. *Journal of Chromatography A*, **1118**(2), 199–210.