PG Spin on Koopman's Checklist for Writing an Abstract

April 30, 2015

Why Koopman's checklist? A bit of Googling will reveal a lot of advice on writing abstracts. Some I found way too formulaic. For instance http://www.wikihow.com/Write-a-Scientific-Abstract takes the form: X goes in the first sentence, Y goes in the second, Z in the third… Other advice seemed too generic, particularly sub-rubrics within "how to write a good paper" rubrics, with items like: make sure the abstract is well-written. To me, Koopman's piece hits a sensible middle ground. I also noticed that some current websites on scientific writing cite Koopman, 18 years on.

How to adapt from computer architecture to statistical science? Pretty easily. For instance, swap out "simulation, analytic models, prototype construction, or analysis of field data" with something like "theoretical derivation, simulation, analysis of real data, ..." Swap out "faster, cheaper, smaller, or otherwise better than something else" with "faster, lower MSE, more interpretable, less reliant on assumptions, ...".

Considerations for matching your abstract to the intended venue? Statisticians end up writing for journals (and other venues) ranging from the very theoretical (e.g., *Annals of Statistics*) to the very applied (subject journals in life-sciences, health-sciences, social-sciences).

- Technical terms (not explicitly defined in the abstract) and even math notation would be fine in an *Annals of Statistics* abstract, not in an *American Journal of Epidemiology* abstract.
- Some subject-area journals demand structured abstracts with pre-defined paragraph headings (e.g., Background, Methods, Results, Conclusions).

What in Koopman's rubric particularly resonates with me?

- Only five items on checklist.
- "...full self contained..."
- Stress *implications* of your work.
- Truth in advertising! ("weasel words")
- "...room for creativity..."

How to Write an Abstract

Philip Koopman, Carnegie Mellon University
October, 1997

Abstract

Because on-line search databases typically contain only abstracts, it is vital to write a complete but concise description of your work to entice potential readers into obtaining a copy of the full paper. This article describes how to write a good computer architecture abstract for both conference and journal papers. Writers should follow a checklist consisting of: motivation, problem statement, approach, results, and conclusions. Following this checklist should increase the chance of people taking the time to obtain and read your complete paper.

Introduction

Now that the use of on-line publication databases is prevalent, writing a really good abstract has become even more important than it was a decade ago. Abstracts have always served the function of "selling" your work. But now, instead of merely convincing the reader to keep reading the rest of the attached paper, an abstract must convince the reader to leave the comfort of an office and go hunt down a copy of the article from a library (or worse, obtain one after a long wait through inter-library loan). In a business context, an "executive summary" is often the *only* piece of a report read by the people who matter; and it should be similar in content if not tone to a journal paper abstract.

Checklist: Parts of an Abstract

Despite the fact that an abstract is quite brief, it must do almost as much work as the multi-page paper that follows it. In a computer architecture paper, this means that it should in most cases include the following sections. Each section is typically a single sentence, although there is room for creativity. In particular, the parts may be merged or spread among a set of sentences. Use the following as a checklist for your next abstract:

- **Motivation:**
  Why do we care about the problem and the results? If the problem isn't obviously "interesting" it might be better to put motivation first; but if your work is incremental progress on a problem that is widely recognized as important, then it is probably better to put the problem statement first to indicate which piece of the larger problem you are breaking off to work on. This section should include the importance of your work, the difficulty of the area, and the impact it might have if successful.
- **Problem statement:**
  What *problem* are you trying to solve? What is the *scope* of your work (a generalized approach, or for a specific situation)? Be careful not to use too much jargon. In some cases it is appropriate to put the problem statement before the motivation, but usually this only works if most readers already understand why the problem is important.
- **Approach:**
  How did you go about solving or making progress on the problem? Did you use simulation,

analytic models, prototype construction, or analysis of field data for an actual product? What was the *extent* of your work (did you look at one application program or a hundred programs in twenty different programming languages?) What important *variables* did you control, ignore, or measure?

- **Results:**
  What's the answer? Specifically, most good computer architecture papers conclude that something is so many percent faster, cheaper, smaller, or otherwise better than something else. Put the result there, in numbers. Avoid vague, hand-waving results such as "very", "small", or "significant." If you must be vague, you are only given license to do so when you can talk about orders-of-magnitude improvement. There is a tension here in that you should not provide numbers that can be easily misinterpreted, but on the other hand you don't have room for all the caveats.
- **Conclusions:**
  What are the implications of your answer? Is it going to change the world (unlikely), be a significant "win", be a nice hack, or simply serve as a road sign indicating that this path is a waste of time (all of the previous results are useful). Are your results *general*, potentially generalizable, or specific to a particular case?

Other Considerations

An abstract must be a fully self-contained, capsule description of the paper. It can't assume (or attempt to provoke) the reader into flipping through looking for an explanation of what is meant by some vague statement. It must make sense all by itself. Some points to consider include:

- Meet the word count limitation. If your abstract runs too long, either it will be rejected or someone will take a chainsaw to it to get it down to size. Your purposes will be better served by doing the difficult task of cutting yourself, rather than leaving it to someone else who might be more interested in meeting size restrictions than in representing your efforts in the best possible manner. An abstract word limit of 150 to 200 words is common.
- Any major restrictions or limitations on the results should be stated, if only by using "weasel-words" such as "might", "could", "may", and "seem".
- Think of a half-dozen search phrases and keywords that people looking for your work might use. Be sure that those exact phrases appear in your abstract, so that they will turn up at the top of a search result listing.
- Usually the context of a paper is set by the publication it appears in (for example, *IEEE Computer* magazine's articles are generally about computer technology). But, if your paper appears in a somewhat un-traditional venue, be sure to include in the problem statement the domain or topic area that it is really applicable to.
- Some publications request "keywords". These have two purposes. They are used to facilitate keyword index searches, which are greatly reduced in importance now that on-line abstract text searching is commonly used. However, they are also used to assign papers to review committees or editors, which can be extremely important to your fate. So make sure that the keywords you pick make assigning your paper to a review category obvious (for example, if there is a list of conference topics, use your chosen topic area as one of the keyword tuples).

Conclusion

Writing an efficient abstract is hard work, but will repay you with increased impact on the world by enticing people to read your publications. Make sure that all the components of a good abstract are included in the next one you write.

Further Reading

Michaelson, Herbert, *How to Write & Publish Engineering Papers and Reports*, Oryx Press, 1990. Chapter 6 discusses abstracts.

Cremmins, Edward, *The Art of Abstracting 2nd Edition*, Info Resources Press, April 1996. This is an entire book about abstracting, written primarily for professional abstractors.

Embedded system designers may be interested in my blog.

The Annals of Statistics
2004, Vol. 32, No. 2, 407–499

LEAST ANGLE REGRESSION

BY BRADLEY EFRON,[1] TREVOR HASTIE,[2] IAIN JOHNSTONE[3] AND ROBERT TIBSHIRANI[4]

Stanford University

The purpose of model selection algorithms such as *All Subsets*, *Forward Selection* and *Backward Elimination* is to choose a linear model on the basis of the same set of data to which the model will be applied. Typically we have available a large collection of possible covariates from which we hope to select a parsimonious set for the efficient prediction of a response variable. *Least Angle Regression* (LARS), a new model selection algorithm, is a useful and less greedy version of traditional forward selection methods. Three main properties are derived: (1) A simple modification of the LARS algorithm implements the Lasso, an attractive version of ordinary least squares that constrains the sum of the absolute regression coefficients; the LARS modification calculates all possible Lasso estimates for a given problem, using an order of magnitude less computer time than previous methods. (2) A different LARS modification efficiently implements Forward Stagewise linear regression, another promising new model selection method; this connection explains the similar numerical results previously observed for the Lasso and Stagewise, and helps us understand the properties of both methods, which are seen as constrained versions of the simpler LARS algorithm. (3) A simple approximation for the degrees of freedom of a LARS estimate is available, from which we derive a C_p estimate of prediction error; this allows a principled choice among the range of possible LARS estimates. LARS and its variants are computationally efficient: the paper describes a publicly available algorithm that requires only the same order of magnitude of computational effort as ordinary least squares applied to the full set of covariates.

1. Introduction. Automatic model-building algorithms are familiar, and sometimes notorious, in the linear model literature: Forward Selection, Backward Elimination, All Subsets regression and various combinations are used to automatically produce "good" linear models for predicting a response y on the basis of some measured covariates $x_1, x_2, \dots, x_m$. Goodness is often defined in terms of prediction accuracy, but parsimony is another important criterion: simpler models are preferred for the sake of scientific insight into the $x - y$ relationship. Two promising recent model-building algorithms, the Lasso and Forward Stagewise lin-

Received March 2002; revised January 2003.
[1]Supported in part by NSF Grant DMS-00-72360 and NIH Grant 8R01-EB002784.
[2]Supported in part by NSF Grant DMS-02-04162 and NIH Grant R01-EB0011988-08.
[3]Supported in part by NSF Grant DMS-00-72661 and NIH Grant R01-EB001988-08.
[4]Supported in part by NSF Grant DMS-99-71405 and NIH Grant 2R01-CA72028.
AMS 2000 subject classification. 62J07.
Key words and phrases. Lasso, boosting, linear regression, coefficient paths, variable selection.

Does Food Vendor Density Mediate the Association Between Neighborhood Deprivation and BMI?

A G-computation Mediation Analysis

Y. Tara Zhang,[a] Barbara A. Laraia,[b] Mahasin S. Mujahid,[a] Aracely Tamayo,[a] Samuel D. Blanchard,[c] E. Margaret Warton,[d] N. Maggi Kelly,[c] Howard H. Moffet,[d] Dean Schillinger,[e] Nancy Adler,[e] and Andrew J. Karter[d]

Background: In previous research, neighborhood deprivation was positively associated with body mass index (BMI) among adults with diabetes. We assessed whether the association between neighborhood deprivation and BMI is attributable, in part, to geographic variation in the availability of healthful and unhealthful food vendors.

Methods: Subjects were 16,634 participants of the Diabetes Study of Northern California, a multiethnic cohort of adults living with diabetes. Neighborhood deprivation and healthful (supermarket and produce) and unhealthful (fast food outlets and convenience stores) food vendor kernel density were calculated at each participant's residential block centroid. We estimated the total effect, controlled direct effect, natural direct effect, and natural indirect effect of neighborhood deprivation on BMI. Mediation effects were estimated using G-computation, a maximum likelihood substitution estimator of the G-formula that allows for complex data relations such as multiple mediators and sequential causal pathways.

Results: We estimated that if neighborhood deprivation was reduced from the most deprived to the least deprived quartile, average BMI would change by -0.73 units (95% confidence interval: -1.05, -0.32); however, we did not detect evidence of mediation by food vendor density. In contrast to previous findings, a simulated reduction in neighborhood deprivation from the most deprived to the least deprived quartile was associated with dramatic declines in both healthful and unhealthful food vendor density.

Conclusions: Availability of food vendors, both healthful and unhealthful, did not appear to explain the association between neighborhood deprivation and BMI in this population of adults with diabetes.

(*Epidemiology* 2015;26: 344–352)

Neighborhood deprivation indices are composite measures of area socioeconomic status commonly used in early neighborhood effects research as crude proxies for area-level deprivation and as predictors of health access and outcomes.[1] Much of our understanding of the relevance of place to health comes from these early ecologic and multilevel studies of the relation between neighborhood deprivation and disease risk.[2] Our previous research found that, independent of personal characteristics, neighborhood deprivation index had a significant positive and monotonic relation with body mass index (BMI) and cardiometabolic risk factor control among adults with diabetes.[3] Diabetes is a chronic disease influenced by health-related behaviors including diet and exercise, and thus place-based interventions that promote weight loss or simply weight maintenance may improve long-term diabetes outcomes.[4]

The pathways through which neighborhood deprivation index affects BMI are not well understood, but the food retail environment has been proposed as an important mediator and has been shown to have strong cross-sectional associations with BMI in both healthy and chronically ill populations.[5–7] Our previous analysis of the Diabetes Study of Northern California (DISTANCE) found that among moderate to high-income subjects, greater neighborhood healthful food retail density was associated with lower obesity prevalence.[8] However, no studies to date examine whether and how much geographic variation in food retail density accounts for neighborhood-level socioeconomic disparities in BMI.

Is neighborhood density of retail food outlets a major contributor to the BMI disparities we observed between more- and less-deprived neighborhoods in this population

Submitted 7 April 2014; accepted 25 January 2015.

From the [a]School of Public Health, Division of Epidemiology, [b]School of Public Health, Division of Community Health and Child Development, [c]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA; [d]Kaiser Permanente Division of Research, Oakland, CA; and [e]School of Medicine, University of California, San Francisco, CA.

This study was supported by National Institute of Diabetes and Digestive and Kidney Diseases (R01-DK-080744); "Ethnic disparities in diabetes complications" (PI Andrew Karter, R01 DK065664-01-A1); and "Neighborhood Effects on Weight Change and Diabetes Risk Factors (PI Barbara Laraia, R01 DK080744).

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the authors.

Correspondence: Y. Tara Zhang, School of Public Health, 525 University Hall, University of California, Berkeley, CA 94720–7360. E-mail: tarazhang@berkeley.edu.

ISSN: 1044-3983/15/2603-0344
DOI: 10.1097/EDE.0000000000000271