Elsevier Editorial System(tm) for Information Fusion
Manuscript Draft

Manuscript Number: IF11S10-PHI-CFMR2

Title: Guest Editorial - Special Issue on Information Fusion Applications to Human Health and Safety

Article Type: Editorial

Keywords: public health, surveillance, comparative methods

Corresponding Author: Dr. Paul Snow,

Corresponding Author's Institution:

First Author: Al Ozonoff

Order of Authors: Al Ozonoff; Fabio Cuzzolin; Paul Snow

Guest Editorial

Special Issue on Information Fusion Applications to Human Health and Safety

This special issue presents articles submitted in response to calls for papers in the domains of public health and of comparisons of fusion methods in real-world applications. The fruits of the two searches were combined for presentation under the new rubric, in one issue. We aim to increase awareness of and involvement in the discipline of information fusion among researchers who are unfamiliar with the discipline and might benefit from new approaches. We also hope to refocus the attention of those with experience and expertise in information fusion on new application areas, in the hopes that new partnerships and collaborations might result across disciplines..

The research presented in this special issue samples the range of applied problems in human health and safety that have already seen benefits from an information fusion approach. The studies presented here offer a glimpse of large expanses of applied scientific territory still waiting to be explored. In welcoming people who might be newly aware of the field, it may be appropriate to begin with a definition.

Perhaps because of its relative novelty, and because of the diversity of people working in it, finding a single consensus definition for the field is not trivial. Bostrom and colleagues surveyed published definitions of *information fusion* as a field of research [1]. They identified dozens of instances where ‘information fusion’ or a similar term appears in the literature. Of course, there is widespread consensus that information fusion works on multiple sources of information, but to what end, and how?

Bostrom *et al.* conclude with a new definition adapted from the various statements they found, an information fusion in its own right. Their definition focuses on the transformative nature of information fusion: to paraphrase, our end result should improve upon that with which we began.

> Information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points of time into a representation that provides effective support for human or automated decision making.

Each of the papers in this special issue not merely combines information sources but transforms them, to improve decision making that supports human health and safety.

To illustrate the rapidly changing landscape of the science of human health and safety, we note that disease surveillance has undergone a major transformation in the past decade. Information fusion seems certain to play an increasing role in disease surveillance [2]. The problem of counting who is born, who gets sick, and who has died is part of the public health infrastructure that we sometimes take for granted. We rely on organs of the state to maintain accurate vital statistics, births and deaths, and to monitor illness and cause of death in the interests of the public health. What we might call 'classical' surveillance comprises these basic functions. The practice of surveillance has traditionally involved systematic and methodical, sometimes painstaking, collection of specific data related to health events or outcomes.

As our world has changed, so too has the environment in which we conduct surveillance. We often have relatively unfettered access to electronically collected data, and thus we can transmit and manipulate those data nearly instantaneously. This has led to a more modern concept of surveillance, with multiple non-specific data sources, each of which may reveal some partial information about a disease or event of public health interest.

Electronic data on large portions of the population have moved disease surveillance much closer to the field of public health or medical informatics [2, 3]. The convergence of epidemiology, statistical methods, and information technology have opened up new scientific avenues to improve the human condition. The Centers for Disease Control and Prevention (CDC), which supports surveillance at the national and state levels, has invested heavily in information technology in recognition of this changing landscape.

The CDC approach to information fusion was nicely summarized by CDC's Rolka et al. [4]. There they argue that information fusion will be an important part of future CDC capabilities to detect localized outbreaks of disease, to manage the spread of emerging infectious diseases, and to respond and mitigate natural disasters with their resulting public health calamities. Magruder [5] from CDC's National Center for Public Health Informatics recently commented on the potential public health impact on public health surveillance of Health Information Exchanges (HIEs), and recent literature on the topic has been nicely summarized by Revere and Stevens [6].

We believe disease surveillance is one potential area where the process, methods, and scientific rigor of information fusion research might make a valuable contribution. Recalling the old fable of the blind man and the elephant, these disparate data sources may often appear to be unrelated or even contradictory, but when placed in context, they may provide a more revealing picture of a complex problem. The following collection of papers offers more applications and avenues for further research.

The papers in this special issue

Banks et al. introduce Bayesian conditional auto-regressive (CAR) models to combine count data from sites distributed over both space and time. This paper begins with a general and flexible log-linear framework to model counts of events. By specifying functional forms for various parameters, the model specializes to a number of interesting variants which are appropriate in certain settings.

To illustrate the value of their approach, the authors fit a relatively parsimonious model to a nationwide data set on opioid drug abuse. Simulations support the claim that these models represent an advance over the current methods used in disease surveillance. The simple and elegant statement of the model belies its potentially powerful capabilities. The authors discuss important extensions or applications which will soon be feasible with faster processor speeds

Fricker et al. pursue a very different approach to monitor multiple sites over time. By framing the problem in terms of sensors with a threshold for detection, the authors consider how to adjust a sensor network to provide optimal detection capabilities using a non-linear programming approach. They also consider relevant applications to nationwide disease surveillance at the city or county level. What is especially notable about their approach is that it leverages existing methodologies, like the Shewhart control chart processes, first used in industrial quality control, while fusing data in a novel way to enhance the information and decision-making capacity already available from existing public health surveillance systems.

Mnatsakanyan et al. describe an implementation of a Bayesian network design to link individual surveillance streams into a robust Intelligent Decision Support Network. The authors demonstrate that such a network offers more sensitivity to detect possible events of interest compared to an approach which considers sites individually, i.e. which ignores available information from across the network. However, the challenge for such systems, and for the field of biosurveillance in general, is to achieve simultaneous improvements in sensitivity and specificity, thereby detecting more events but also reducing the false alarm rate.

Ye et al. describe a very different sort of information fusion from the previous papers discussed, but one which we should consider carefully nonetheless as an advance in meta-analysis and fusion of prior data. Combining or pooling data from several similar studies to improve precision of an estimate is a biostatistical approach which is commonly practiced both in public health and clinical research settings. In this sense, a classical meta-analysis is quite literally a fusion of data from several studies, with special attention paid to the appropriate recalculation of the standard error i.e. the precision (or imprecision) of the resulting estimate.

The authors show that models of chronic disease using a Markov state-change approach may incorporate a wider range of prior data despite differences in study design or theoretical constructs. They apply this approach to the Michigan Model of Diabetes, and incorporate data from several studies to improve estimates of state-change probabilities and other probabilities. Their simulation results make a strong case for the potential of this method to enhance our ability to model chronic disease progression, which offers some promise of further demonstrations, particularly with models of chronic disease besides diabetes.

Marhic et al. describe a smart home scenario in which a health care surveillance system monitors the behavior and activities of an impaired or elderly person. Emergency personnel are automatically notified if the system decides that there is a health or safety concern. Supervised learning approaches based, for instance, on stochastic dynamical models are inadequate for detecting such unusual events, as the number of samples is very limited or altogether lacking. In addition, sensors may fail or be unreliable, so that faults need to be detected and isolated while not being confused with emergency anomalies.

The novelty in the authors' approach is a two-stage architecture design to detect both unreliable sensors and abnormal subject behavior using the mathematical framework of the Transferable Belief Model, a variant of Dempster-Shafer theory. Unusual activities are detected by analyzing the conflict between the predicted and the observed motion state in a Markov chain model framework. Unreliable sensors are isolated by analyzing the degree of conflict among the data provided by different sensors. The paper explores experimental scenarios using three heterogeneous sensors and concerning persons monitored with or without fault detection procedures.

Frigui et al. offer a comprehensive survey and evaluation of different fusion algorithms for the detection of anti-tank land mines. Finding abandoned mines to remove them is a significant research problem. There are an astounding 100 million landmines estimated to be left behind in over 80 countries.

Ground Penetrating Radar sensors offer the best performance when detecting land mines that contain little metal, but false alarm rates remain an issue. In addition, the performance of detection algorithms can vary greatly under different geographical and environmental conditions. That is why information fusion schemes which take advantage of the stronger algorithms without suffering from the effect of the weaker ones are sorely needed.

The authors use cross-correlation to compare seven of the most popular algorithms for combining several classifiers: Bayesian, distance-based, Dempster-Shafer, Borda count, decision template, Choquet integral and context dependent fusion. The latter method assigns variable weights to different classifiers based upon clusters of features which predict which classifiers will work best in the current detection task. In general, increasing the number of classifiers to be merged does not necessarily improve detection performance. Overall, among the fusion methods tested, context dependent fusion seems to have the best performance, which is not degraded by the addition of new sources to be fused.

Conclusions

As Thomas Friedman [7] reminds us, we live in an age where 'the world is flat'. He was of course referring to the economic globalization of the early Twenty-first century, but his words describe equally well to the rise of broadly interdisciplinary and translational research., reversing a long trend of ever narrower specialization in the physical, biological, and social sciences. Likewise the advent of electronic collection of data on a massive scale has flattened our world with respect to data access, availability, and ease of manipulation. The interaction between information fusion and human health and safety exemplify this new paradigm.

We described our aims for this special issue as two-fold: introduce new researchers from other disciplines to information fusion, and introduce those already familiar with the area to new areas for potential collaboration. By presenting applications that range widely across the fields of health and safety, we aim for progress in both directions. We hope that those with expertise in information fusion will recognize the problems and challenges facing researchers in these fields and reach out with potential solutions or assist with novel applications.

Likewise we expect those in other disciplines, and the disease surveillance community in particular, will find valuable methodology and approaches from the information fusion literature to improve and advance their own work. For example, we are intrigued by the potential to leverage existing infrastructure for syndromic surveillance as described in this issue, as well as the distribution of surveillance functions to the public via mobile technology [8,9]. We hope that new methodology might better fuse these new streams of information on human health in a systematic and scientific fashion with existing data collection systems.

We look forward to future applications in public health, clinical medicine, and the several engineering disciplines that safeguard people. In order to explore these new frontiers, we must familiarize ourselves with new fields, develop new competencies, and learn new techniques. As described above, information fusion by its very nature is a critical part of the process by which this new knowledge will be synthesized. We hope that this special issue helps specialist in these domains to see the value of making new connections with information fusion specialists. Conversely, the issue may help the information fusion community to realize the potential of alliances with those working in these crucial application areas.

References

[1] H. Boström, S.F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. van Laere, L. Niklasson, M. Nilsson, A. Persson, T. Ziemke, On the definition of information fusion as a field of research, Technical Report HS-IKI-TR-07-006, School of Humanities and Informatics, University of Skövde, 2007. Available online:
http://his.diva-portal.org/smash/get/diva2:2391/FULLTEXT01

[2] J.S. Lombardo, D.L. Buckeridge, Disease surveillance: A public health informatics approach, Wiley, Hoboken, NJ, 2007.

[3] M.M. Wagner, A.W. Moore, R.M. Aryel (Eds.), Handbook of Biosurveillance, Elsevier Academic Press, Burlington, MA, 2006.

[4] H. Rolka, J.C. O'Connor, D. Walker, Public health information fusion for situational awareness, Lecture Notes in Computer Science Volume 5354, Springer-Verlag, Berlin, 2008, pp. 1-9.

[5] C. Magruder, Public health/Health Information Exchange collaborative: a model for advancing public health practice, J Public Health Informatics 2(2), 2010. Available online: http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/ojphi/article/view/3217/2639

[6] D. Revere, K. Stevens, Accelerating situational awareness through Health Information Exchanges: An annotated bibliography. J Public Health Informatics 2(2) (2010). Available online
:http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/ojphi/article/view/3212/2636

[7] T.L. Friedman. The world is flat: A brief history of the Twenty-first Century, Farrar, Straus and Giroux, New York, 2005.

[8] J.S. Brownstein, C.C. Freifeld, B.Y. Reis, K.D. Mandl, Surveillance sans frontieres: internet-based emerging infectious disease intelligence and the HealthMap project. J PLOS Medicine 5(7) (2008). Available online:

http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0050151

[9] C.C. Freifeld, R. Chunara, S.R. Mekaru, E.H. Chan, T. Kass-Hout, A.A. Iacucci, J.S. Brownstein, Participatory epidemiology: use of mobile phones for community-based health reporting. PLOS Medicine 7(12) (2010). Available online:

http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000376

Acknowledgments

The guest editors wish to thank this issue's authors who submitted their work and worked patiently with us through the editorial process. We also wish to thank the many anonymous reviewers who freely gave of their time and effort to bring this issue to press. Special thanks go to Belur V. Dasarathy, the Editor in Chief of the journal, for his encouragement and perseverance in seeing this project through to completion.

Al Ozonoff

Clinical Research Program
Children’s Hospital Boston
300 Longwood Avenue
Boston, MA 02215
al.ozonoff@childrens.harvard.edu

Fabio Cuzzolin

Department of Computing and Communication Technologies
Oxford Brookes University
Oxford UK OX33 1HX
Fabio.Cuzzolin@brookes.ac.uk

Paul Snow

P.O. Box 6134
Concord, NH 03303-6134
paulusnix@gmail.com