

Foreword

by William H. Hsu

A TAKE ON INFORMATICS AND KNOWING

It is my pleasure to introduce this volume edited by Dr. Hai-Jew, whose previous books I have served twice on as a contributor: one on open-source technologies for education and one on techniques for the presentation of digital information, including information visualization. This foreword briefly outlines the convergence of technologies and user needs that makes the present book timely and useful.

DATA SCIENCE OF, BY, AND FOR HUMANS

The field of qualitative and mixed methods for data analysis has expanded significantly in the past decade due to an influx of contributions from researchers working at the interfaces of informatics, especially between the humanities and data science. This includes both fields traditionally classified as digital humanities, such as digital ethnography, and applications of quantitative (especially statistical) computational methods to problems that involve a “human in the loop” and intrinsically subjective elements, such as user experience.

There is a fine line between these two aspects of qualitative data analysis as a result of technological convergence, but the primary distinguishing characteristic of the first category is that it involves data science and other information technology as assistive tools for activities that are traditionally carried out using analog media. For example, dialogue and discourse analysis are aspects of the theory of communication that originate from linguistics, anthropology, and sociology, pre-dating computer-mediated communication and electronic representation of text. By contrast, the second category involves tasks that have always been defined within a computing context but admit an irreducible aspect of individual aesthetics and qualitative judgment.

That is, both types of problems are defined of, by, and for humans, but the difference lies in the motivation for and direction of synthesis: introducing new informatics tools for a study of human culture, especially natural language communication, or incorporating subjectivity into “hard” data science. In the intersection of these categories lie tasks subjectivity, sentiment, and discourse analysis from text, traditionally considered the purview of soft computing. As a topic of knowledge representation and automated reasoning, soft computing comprises uncertain reasoning using probability and other representations cf. fuzzy logic and vague domains.

Meanwhile, subjective topics such as human factors and ergonomics have traditionally been treated using the methods of quantitative science or qualitative study, rather than the empirical synthesis that underlies many of the chapters of this book. This is especially relevant to applied areas of the study of User Experience (UX) that have undergone rapid growth, such as user modeling, adaptive hypermedia, and personalization of user interfaces.

THE NEW DATA ANALYTICS: RELEVANCE AND SCOPE

“Big data” and qualitative data analytics are two highly popular terms that do not seem to mix very much or very well at the time of this writing. The term “big data” refers to size and complexity at a minimum of upper terascale to lower petascale as of 2014, while qualitative methods such as field notes and human-produced annotations, except for digital video and other high-volume media, represent bandwidth use that is at least several orders of magnitude below this in scale. Thus, the conventional wisdom is that qualitative and mixed methods simply do not constitute big data at present and will not do so for the foreseeable future.

Some evidence runs counter to this notion: namely, the global scale of social networks and media and the explosion in development of new media. The observed growth rate of digital hypermedia such as the Web and post-Web environments (dynamic content and data grid services *aka* Web 2.0, the semantic Web *aka* Web 3.0), even during the worldwide recession of the past decade, indicates a demand for UX technology that matches that for physical and natural sciences.

Beyond simple considerations of scale and bandwidth, however, there are some basic trends in methodology to observe. The first of these is an accelerating convergence of information visualization as an area of data science and engineering with the social science of applied analytics. This brings together machine learning and data mining, which are themselves also motivated by performance elements such as Decision Support Systems (DSS), Business Intelligence (BI), Customer Relationship Management (CRM), and the above-mentioned UX areas of personalization and user modeling, with a powerful delivery mechanism that we are continuing to understand in terms of graphical design principles. The second of these is organic growth of the search engine industry in service to existing and emergent user needs: ubiquitous computing in the form of new user interfaces (e.g., wearable computing devices and augmented reality systems) and resources for analytics such as the gigaword parallel corpora developed primarily for research in machine translation and multi-lingual text mining. The third is the parallel convergence of assistive and augmentative technologies, such as Unmanned Aerial Vehicles (UAVs) for videography, 3-D printing, medical devices, and cyber-physical systems. These are technologies that serve humans in new ways but are based on underlying principles that have been studied for much longer than the new synthesis.

THE FUTURE OF QUALITATIVE AND MIXED METHODS: AN INFORMATICS PERSPECTIVE

Informatics is a broad academic field encompassing information science and technology, algorithms, specific areas of applied mathematics, digital foundations in the humanities and social sciences, and applications of computation to many other disciplines, particularly the natural sciences and engineering. It describes computational, mathematical, and statistical tools for data integration, analytics, and information retrieval. Drawing on theory and research from systems, information processing, simulation and modeling, visualization, machine learning, data mining, and databases, informatics is a pervasive, crosscutting theme.

Some of the desiderata of qualitative and mixed methods for data science are identical or similar to those for quantitative methods: reproducibility and general applicability to previously unseen domains. This is a hallmark of informatics as an area of research and development – it is intrinsically applied but

rooted in the needs of empirical science. Most areas of informatics that serve scientists as purveyors and users of data resources are furthermore subject to the same challenges of usability, validation, and adaptivity as other areas of UX for users in general.

The interrelated fields of predictive analytics, Computational Information and Knowledge Management (CIKM), and scientific, data, and information visualization continue to be recombined in significant ways through developments in fundamental theory and methodology. It is to be hoped that volumes such as this one will help to guide the way to realizing fielded systems from these advances.

*William H. Hsu
Kansas State University, USA*

William H. Hsu is an associate professor of Computing and Information Sciences at Kansas State University. He received a BS in Mathematical Sciences and Computer Science and an MSEng in Computer Science from Johns Hopkins University in 1993, and a PhD in Computer Science from the University of Illinois at Urbana-Champaign in 1998. His dissertation explored the optimization of inductive bias in supervised machine learning for predictive analytics. At the National Center for Supercomputing Applications (NCSA), he was a co-recipient of an Industrial Grand Challenge Award for visual analytics of text corpora. His research interests include machine learning, probabilistic reasoning, and information visualization, with applications to cybersecurity, education, digital humanities, geoinformatics, and biomedical informatics. Current work in his lab deals with: spatiotemporal mapping of opinions, crimes, and other events; data mining and visualization in education research; graphical models of probability and utility for information security; analysis of heterogeneous information networks; and domain-adaptive models of large natural language corpora and social media for text mining, link mining, sentiment analysis, and recommender systems. Dr. Hsu is editor of the forthcoming book *Emerging Methods in Predictive Analytics*, and has over 50 refereed publications in conferences, journals, and books, plus over 35 additional publications.