Ubiquitous Analytics: Interacting with Big Data Anywhere, Anytime

Niklas Elmqvist (Purdue University) and Pourang Irani (University of Manitoba)

Ubiquitous analytics amplifies human cognition by embedding the analytical process into the physical environment to enable sensemaking of big data anywhere and anytime.

With more than 4 billion mobile devices in the world today, mobile computing is quickly becoming the universal computational platform of the world (P. Baudisch and C. Holz, "My New PC is a Mobile Phone," *ACM XRDS*, 16(4):36-41, 2010). Building on this new wave of mobile devices are personal computing activities such as microblogging, social networking, and photo sharing, which are intrinsically mobile phenomena that occur while on-the-go. Mobility is now propagating to more professional activities such as data analytics, which need no longer be restricted to the workplace. In fact, the rise of big data increasingly demands that we be able to access data resources anytime and anywhere, whether to support decisions and activities for travel, telecommuting, or distributed teamwork.

In other words, it is high time to fully realize Mark Weiser's vision of ubiquitous computing in the realm of data analytics (M. Weiser, "The computer for the twenty-first century," *Scientific American*, 265(3):94–104, 1991). In particular, the quiet revolution in mobile computing has led to massive infrastructures of computational power embedded in our everyday surroundings that are just waiting to be harnessed. We call this new brand of situational sensemaking *ubiquitous analytics* (or ubilytics): the use of multiple networked devices in our local environment to enable deep and dynamic analysis of massive, heterogeneous, and multi-scale data anytime and anywhere.

Taking this new approach to data analytics not only liberates analysts from the confines of the office and leverages existing computer hardware that today is mostly used for personal and entertainment purposes, but also taps into deep processes in human cognition that may actually improve the analytical process. *Socially distributed cognition* (Figure 1) models human thought as an intrinsically system-level process that goes beyond information processing in the brain using sensory input, but also involves interactions with the body, the surrounding physical world, tools and artifacts, and other humans (E. Hutchins, *Cognition in the Wild*, MIT Press, Cambridge, MA, 1995). Current personal computers provide only small viewing portholes and limited input bandwidth into the digital universe, whereas embedding interactive analytics components into smartphones, tablets, laptops, large displays, and even tabletops enable a more natural and distributed approach to interacting with big and complex data.

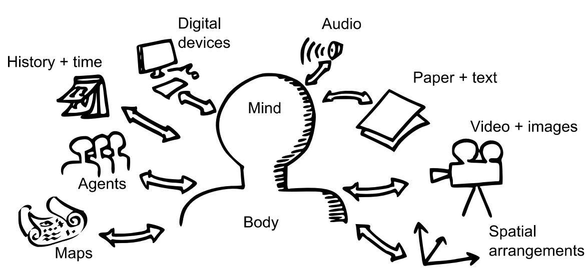

***Figure 1**: Human cognition can be seen as a system-level process involving not only the brain and its sensors, but also physical space, tools and artifacts, and other people.*

MOTIVATING EXAMPLE

Consider a public health scenario where local authorities are trying to contain a pandemic outbreak in an area. This is a real-time emergency situation where data will be coming in from multiple heterogeneous sources, including both historical information from existing databases, as well as dynamic feeds from hospitals, doctors' offices, pharmacies, etc. Furthermore, the data will have many forms, such as admissions data, health records, sales statistics, etc. Unexpectedly useful data might also come from social networks ("Who interacts with who?") and credit card data ("Who has gone where?"). All this information is temporal and spatial in nature and must all be considered to determine where the outbreak started and how it is spreading. Finally, agents, analysts, and decision makers in this rapidly evolving scenario are likely distributed both in time and in space.

Figure 2 shows a conceptual view of a ubilytics approach to supporting this pandemic outbreak scenario where an ensemble of networked devices—ranging from smartphones and tablets for agents in the field, laptops and desktops for analysts working individually, and large displays and tabletops for collaboration—are combined into a single reasoning environment. This will allow key stakeholders, such as scientists, doctors, health officials, epidemiologists, and policy makers, to collaboratively explore the data, and, drawing from their collective expertise, to make conclusions, verify them, and take appropriate action. The diversity of the participants' skills and backgrounds and the use of interactive analytics embedded in the physical environment to allow them to collaborate effectively will lead to decisions that are superior to those made by any single individual.

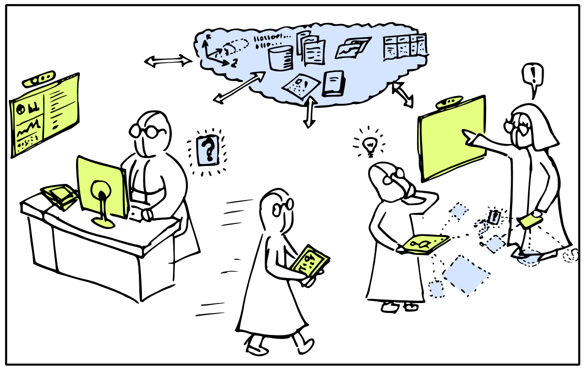

***Figure 2**: Multiple stakeholders collaborating over a shared data repository using a menagerie of devices for pandemic management.*

CHALLENGES

To make this ubiquitous computing approach to analytics possible, several challenges must be addressed. More specifically, instead of being restricted to single computers, ubilytics applications must be able to execute on an ecosystem of networked devices, each of which may join or leave the shared ubilytics space at any time. Furthermore, there exist analytics tasks—e.g., statistical, cluster, or probabilistic analyses—that are simply too computationally expensive to be performed on a mobile device ecosystem. Also, how can we harness the specific capabilities of each device, including varying display size, input modality, and computational resources? And, finally, tomorrow's analytics software must clearly be designed not only for individual work, but also for groups collaborating on big data problems.

Networked Device Ecosystem

Traditional data analytics applications are built for a single computer. While such software may use concurrency to take advantage of multiple cores, this is still a far cry from the distributed computational platforms required to enable seamless analytics across an ensemble of networked devices, each of which may join or leave the shared environment at any time.

Limited Computational Resources

While it is true that the capabilities of mobile computing is steadily increasing, mobile devices still have orders of magnitude less in computing power and memory compared to standard computers. This has practical impact on all mobile applications, particularly for analytics applications that often incorporate big data, complex algorithms, and rich and interactive visual representations.

Polymorphic Hardware

Supporting an effective analytical discourse in a ubiquitous setting requires taking full advantage of the specific hardware capabilities of each participating device and accommodating its weaknesses. Mobile computing in particular is plagued by an intrinsic conflict between minimizing device form factor and maximizing interaction and display surface. In other words, even if modern smartphones consist almost entirely of a multitouch display, these are still very small screens that require good eyesight and precise input and thus are not well-suited for spatially intensive tasks.

Individuals and Groups

While individual analysts will remain the core user group for most analytics software, big data problems often require correspondingly large teams to tackle efficiently. Groups contribute additional viewpoints, broader expertise, and multiple roles and authorities to the analytical process, which often leads to better results compared to work conducted by individual analysts. However, this also gives rise to issues in connecting collaborators across space and time, promoting awareness and consensus between team members, and resolving conflicts.

UBILYTICS EXAMPLES

By way of addressing the above challenges, we draw upon our own work designing, building, and evaluating three separate ubilytics systems: the Hugin toolkit, Branch-Explore-Merge, and Around-Device binning.

The Hugin Toolkit

Our Hugin toolkit is a mixed-presence framework for supporting settings when some participants are co-located in the same physical space and others are connected over the network (K. Kim, W. Javed, C. Williams, N. Elmqvist, P. Irani, "Hugin: A Framework for Awareness and Coordination in Mixed-Presence Collaborative Information Visualization," *Proc. ACM ITS*, pp. 231–240, 2010). Figure 3a shows Hugin connecting two multitouch tabletops in our respective labs, allowing geographically separated analysts to visually explore various multidimensional datasets together.

Hugin uses a scalable client/server architecture to connect multiple heterogeneous tabletops using a platform-independent network protocol. To accommodate different display sizes, the virtual space for each tabletop is stacked and scaled to fit the maximum physical size. Finally, the toolkit has native support for group awareness and conflict resolution, including mechanisms for visualizing remote touches, overview maps, and access control.

Branch-Explore-Merge

Conflicts and interference can arise even when analysts collaborate in the same physical space and at the same time. The Branch-Explore-Merge protocol applies ideas from source code revision control systems such as CVS and SVN to the analysis process by allowing individual analysts to branch off a current query, explore new parameters, and then merge new findings back into the shared state (W. McGrath, B. Bowman, D. McCallum, J. D. Hincapié Ramos, N. Elmqvist, P. Irani, "Branch-Explore-Merge: Facilitating Real-Time Revision Control in Co-Located Collaborative Visual Exploration," *Proc. ACM ITS*, pp. 235–244, 2012). Figure 3b

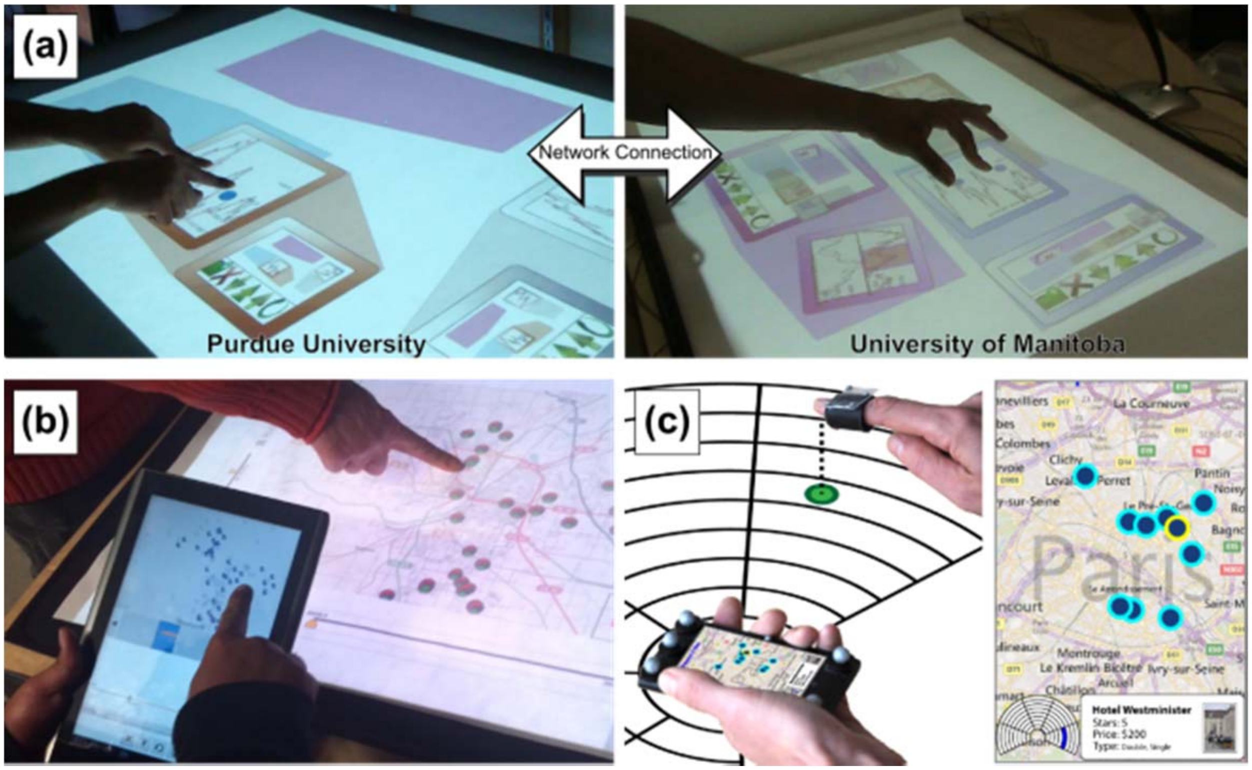

Figure 3*: Three ubilytics systems. (a) Collaborative data analysis using the Hugin mixed-presence toolkit. (b) Tablet and tabletop for collaborative search through Branch-Explore-Merge. (c) Appropriating physical space around a mobile device using AD-binning.*

shows our BEM prototype being used to search for real estate in the Greater Lafayette, IN area.

Our BEM prototype combines both Android smartphones and tablets with a digital tabletop. Android devices serve as clients that connect to the BEM server running on the tabletop. Costly computation and rich visualization is offloaded to the server since it is running on a personal computer, whereas the clients have limited computational resources. Such a collaborative platform facilitates fluid transitions between shared and personal workspaces.

Around-Device Binning

Traditional mobile interfaces are not designed to handle big or spatially intensive data. Thus, as mobile devices become the primary means for interacting with large datasets on-the-go, novel interfaces are needed to fully support the sensemaking process. We have found that mobile interfaces can be evolved to support these big data interaction tasks using minimal mobile sensing augmentation. Around-Device (or AD) binning is our initial realization for allowing users to interact with large search content without requiring minute and tedious operations, such as multi-flicking on a limited size display, as is common today (K. Hasan, D. Ahlström,P. Irani, "AD-Binning: Leveraging Around Device Space for Storing, Browsing and Retrieving Mobile Device Content," *Proc. ACM CHI*, 2013).

AD-binning equips a mobile device with the ability to sense interactions beyond its screen and in mid-air (Figure 3c). Users can interact with objects, either by placing them into spatial bins and later retrieving them, or by having the system automatically place items, in air and around the device. Such augmentations create new opportunities for transparent interactions with large data repositories. More importantly, it is a step toward mitigating the limitations imposed by device form factors and display sizes. This general approach has also been dubbed *appropriated interaction* because it borrows space from the surrounding physical environment to offset limited display and input sizes (C. Harrison, "Appropriated interaction surfaces," *IEEE Computer*, 43(6):86–89, 2010).

CONCLUSION

While still nascent, a ubiquitous approach to analytics will imbue a paradigm shift in our sensemaking methodology. Now, analysts will be able to interact with big and complex data on an individual basis or collaboratively, on-the-go or in the office, synchronously or not. We have outlined some of the major challenges needed to be addressed by the research community to advance the field. In an attempt to address some of these challenges, our own work only provides a glimpse of the full potential of ubiquitous analytics. Much work is indeed necessary to lay the groundwork for this growing and increasingly important discipline.

Niklas Elmqvist *is an assistant professor of electrical and computer engineering at Purdue University in West Lafayette, IN, USA. Contact him at elm@purdue.edu.*

Pourang Irani *is an associate professor of computer science at University of Manitoba in Winnipeg, MB, Canada. Contact him at irani@cs.umanitoba.ca.*

The work of the first author was partially supported by U.S. National Science Foundation award IIS-1253863. The work of the second author was partially supported by the Canadian NSERC Discovery program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the respective funding agencies.