

Social-Sensed Multimedia Computing

Peng Cui and
Wenwu Zhu
Tsinghua
University, China

Tat-Seng Chua
National
University of
Singapore

Ramesh Jain
University of
California, Irvine

Multimedia computing, which ultimately aims to deliver multimedia content to users according to their information needs (intents), can be decomposed into various stages, such as multimedia compression (for storage), multimedia communication (for delivery), and multimedia content analysis (for intelligence). The first two stages are comparatively well studied, while multimedia analysis started becoming mainstream in the multimedia community at the end of the last century, with related technologies advancing quickly. However, understanding and predicting what multimedia content users need in different situations and contexts—the last-mile technology for delivering multimedia services—has not been well studied (see Figure 1).

This negligence has resulted in a gap between multimedia data and the information needs of users, creating a bottleneck in advancing intelligent multimedia computing technologies for use in real-world applications. To bridge this gap, we need to invest more effort in understanding users, both individually and collectively. Fortunately, the explosive growth of social multimedia content on the Internet is revolutionizing the landscape of various multimedia applications. The new types of multimedia content, metadata, context information, and interaction behaviors in social multimedia present a significant opportunity to advance and augment multimedia and content-analysis techniques.

Existing research in this area viewed social multimedia as a new research objective and thus proposed methods to exploit the new data types or solve the new problems.¹ Here, we argue that the significance of social multimedia with respect to the field of multimedia goes far beyond the emergence of new types of data and problems.

User-Centric Multimedia Computing

Four years ago, at the ACM International Conference on Multimedia 2012, Klara Nahrstedt

and Malcolm Slaney organized a panel called “Coulda, Woulda, Shoulda: 20 Years of Multimedia Opportunities,” in recognition of the 20th anniversary of *ACM Multimedia*. They invited several leading researchers in the multimedia community to serve as panelists, including one of us (Ramesh Jain), along with Dick Bulterman, Larry Rowe, and Ralf Steinmetz. Among the topics discussed, one of the most thought-provoking and sobering questions was why, although multimedia analysis has been a hot topic in the community for dozens of years, the popular new multimedia systems and platforms (such as Flickr, YouTube, and Instagram) haven’t been founded by people in the multimedia community? And why haven’t these systems leveraged advanced multimedia analysis technologies?

Another panel at the conference, “Content is Dead; Long Live Content!,” further pushed attendees to be introspective about recent multimedia content-analysis research. There were various, sometimes contradictory, arguments during and after the conference, yet one of the well-accepted opinions was that researchers and engineers make simple hypotheses about user needs in terms of multimedia data. For example, in image retrieval, a user is assumed to be looking for images that include the query text as a tag or label; in video recommendation, a user is assumed to be interested in videos that are similar to what he or she has watched. However, rarely do we try to investigate and understand the real user needs. Although relevance feedback technology incorporates users’ interactions, the fact that it relies on users’ explicit feedback goes against the habits of typical (lazy) users, making it challenging to apply in practice.

We need to treat the understanding and prediction of user needs as a first-class citizen in the multimedia community. Now is the time to reconsider the traditional multimedia computing paradigm, which is either *data centric* or

content centric. As a discipline that mainly targets technologies and services for users, multimedia computing should pay more attention to user needs. How to transform data-centric or content-centric multimedia computing into *user-centric multimedia computing* is a great challenge and opportunity for both academia and industry.

The Semantic Gap vs. Need Gap

If we say that the major goal of content-centric multimedia computing is to bridge the semantic gap between low-level multimedia features and high-level semantics, then the ultimate goal of user-centric computing is to bridge the gap between multimedia content (represented by features and semantics) and the user needs.

Semantics is the study of meaning. The gap between the formal low-level representations in computational machines and the richness of high-level semantic meanings in the human mind is often referred to as the *semantic gap*. Bridging this gap—especially in unstructured multimedia data, such as visual and acoustic information—is the ultimate goal of content-centric computing technologies. Several research communities have devoted many efforts to understanding what the data represents—for example, recognizing objects in visual images, identifying targeting events in video sequences, and measuring textual semantic similarities at different levels (from words to documents). Although we still have a way to go in bridging the gap, the success of search engine and many vertical applications, such as face detection and recognition, have demonstrated rapid advancements toward this goal.

Yet once again, understanding user needs will be critical. In the scope of information science, user needs can be understood as a user's desire to obtain information to satisfy his or her conscious or unconscious needs, and they can be further specified as interests and intents, where interest represents long-term user needs and intent represents instantaneous user needs. However, these needs are often latent, so inferring them from observed data is challenging. A basic hypothesis is that user needs will be triggered in certain situations and manifested as behaviors. Thus, behaviors can be regarded as the reflection of user needs.

On the one hand, how users interact with multimedia (that is, their “interaction behavior”) depends on the semantics of the multimedia data, because users have different preferences

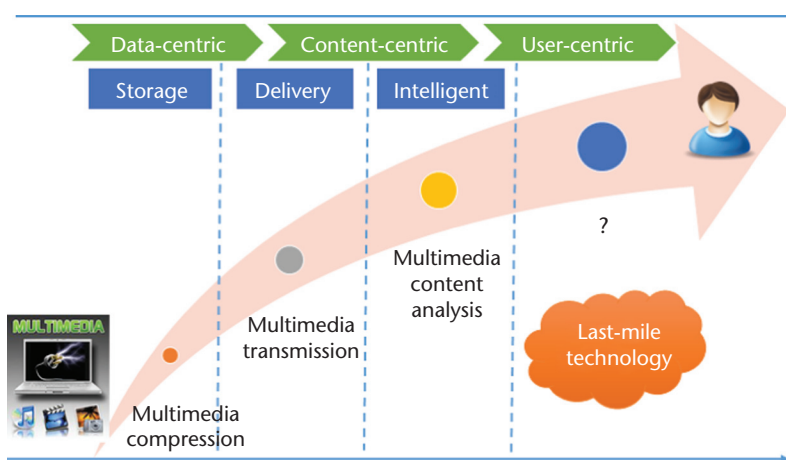


Figure 1. The multimedia technology lifecycle. Understanding and predicting which multimedia content users need in different situations and contexts—the last-mile technology for delivering multimedia services—has not been well studied.

for the semantics delivered by multimedia contents. On the other hand, the mapping from semantics to user needs is complicated, and the process for learning this mapping is ill-defined. Thus, between the semantics of multimedia data and the user needs for multimedia data, there also exists a “need” gap (see Figure 2). Both the semantic gap and need gap are critical in multimedia computing, but the semantic gap has been the focus of more research. So can we somehow straightforwardly derive user needs from semantics? The answer is no.

It is well accepted that user needs can be represented by a distribution over semantics, but the distribution is heavily dependent on the context. For example, the videos that a user wants to watch will depend on the user's mood and environment, and whether the viewing is for work or leisure. Considering the fact that user needs are implicit, and further considering the incompleteness and uncertainty of observed user behaviors, we need more comprehensive research into discovering the mapping mechanism between multimedia data and user needs and how this mechanism can be coupled with rich context information.

Another argument might be that if we can replace multimedia data with semantics, which are often textual words or sentences, then the need-gap problem would become irrelevant. We argue that multimedia data cannot be fully abstracted by using textual semantics. Such semantics cannot accurately or comprehensively represent visual styles, delivered visual effects, and psychovisual factors, all of which

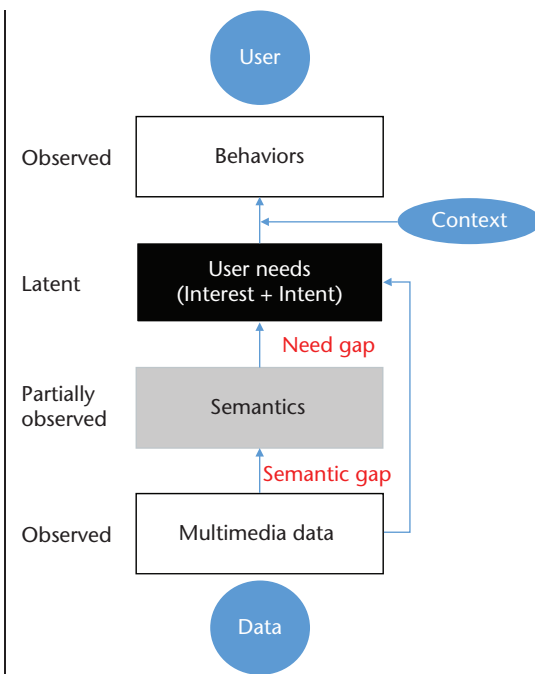


Figure 2. The gap between the formal low-level representations in computational machines and the richness of high-level semantic meanings in the human mind is often referred to as the “semantic gap.” Yet there is also a “need gap” between the semantics of multimedia data and the user needs for multimedia data.

play important roles in deciding users’ interaction behaviors with multimedia data, especially considering the fact that most users watch multimedia content for entertainment or exploratory goals. Therefore, both semantics and multimedia content should be jointly considered in bridging the need gap. The problem can be formulated as seeking a mapping function between multimedia content (represented by both textual semantics and visual factors) and the observed user behaviors.

Social-Sensed Multimedia Computing

Data specifying user needs is necessary to understand real user needs for multimedia. An alternative to explicitly surveying user needs, which is often costly and impractical, is to collect multimedia data, user data, and the interaction behaviors between users and multimedia data, from which we can implicitly and continuously discover users’ multimedia needs.

Although these data types were not readily available in the past, with the emergence of social media platforms (such as Flickr, Facebook, and YouTube), billions of users are now

proactively interacting with (generating, sharing, commenting on, and so forth) huge volumes of multimedia data. These interaction behaviors are being recorded at an unprecedented level. Thus, social media has formed a valuable pool of data about user needs, presenting a precious opportunity to bridge the need gap in multimedia computing.

More specifically, users’ need-related information (including long-term interests, instantaneous intents, and emotions of both crowds and individuals), their behavior patterns, and ultimately the common principles of user-multimedia interactions under different contexts can all be sensed from social media and summarized using social knowledge of user-multimedia interactions. This social knowledge reflects user needs and establishes a bridge for multimedia data and user needs.

Determining how to organically integrate multimedia data, user needs, and social knowledge into multimedia computing technology is a critical issue. Thus, we propose a new multimedia computing paradigm—*social-sensed multimedia computing*—to bring social media into the loop of multimedia computing (see Figure 3). This new paradigm should naturally transform the landscape of multimedia computing from traditional data-centric or content-centric multimedia computing to user-centric multimedia computing, which will improve users’ experiences with various multimedia applications and services.

Of course, compared to traditional multimedia computing, this paradigm will face new problems, which will be related to social multimedia representation, user modeling, and user-multimedia interaction analysis.

Socialized Representation of Multimedia Data

Social multimedia was first defined by Mor Naaman as “an online source of multimedia resources that fosters an environment of significant individual participation and that promotes community curation, discussion and re-use of content.”¹ The main characteristic that differentiates social multimedia from traditional multimedia is that the former boasts significant user participation and interactions with multimedia content. For example, users tag images in Flickr, add comments to videos in YouTube, and “like” or “dislike” videos and images on Facebook. These are all important resources for discovering patterns of user behaviors toward multimedia content.

However, current multimedia representation methods (such as low-level features, concepts, and visual attributes) are designed to bridge the semantic gap. There remains an obvious gap between semantics and user responses and behaviors—the need gap. Therefore, new representation methods for social multimedia must be found to simplify mapping between multimedia content and user responses and behaviors.

User Profiling and Social Graph Modeling

User profile inference, social graph analysis, and tie-strength measurement have become popular in the social network analysis field, but most related research has been based on text information. When we attempt to apply these findings and approaches to social-sensed multimedia computing, several fundamental issues arise: Can the user and social knowledge learned from text data be adapted to multimedia data? Is multimedia data able to tell us a different (or a more complete) story about the characteristics of users and social relations?

Intuitively, multimedia content has intrinsically different structures and feature spaces from text content and can typically provide much richer semantics and meanings. To guarantee that the learned user profiles and social graph models can be seamlessly bridged with multimedia data, we must revisit user profiling and social graph modeling in the context of the social multimedia environment.

User-Multimedia Interaction Behavior Analysis

A major goal of the sensing part of social-sensed multimedia computing is to discover user-multimedia interaction behavior patterns, from which user needs regarding multimedia data can be inferred. User-multimedia interaction behaviors should be investigated at different scales, depending on the level of support required by various multimedia applications. In particular, these interaction behaviors can be categorized into the following three levels: microscopic, mesoscopic, or macroscopic, which correspond to the interaction behaviors of individual users, groups of users, and global users, respectively.

Microscopic analysis can be specified as, but not limited to, user interest modeling and user sentiment analysis, which can support personalized search and recommendation for multimedia. Mesoscopic analysis includes collective behavior analysis, social influence modeling, and so on, which can support social multimedia marketing and socially aware multimedia com-

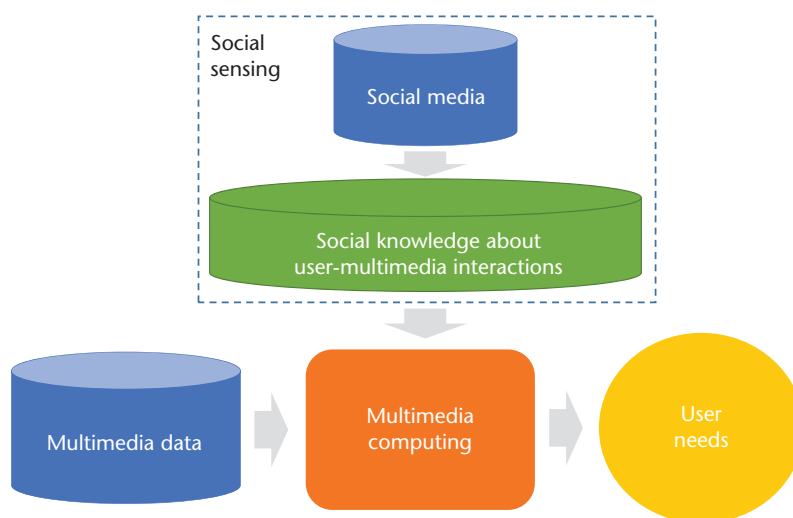


Figure 3. Illustration of the social-sensed multimedia computing paradigm. This new paradigm should naturally transform the landscape of multimedia computing from traditional data-centric or content-centric multimedia computing to user-centric multimedia computing.

munication. Macroscopic analysis broadly refers to multimedia propagation analysis and prediction in a social environment, which can support multimedia popularity prediction and social multimedia monitoring. These all make it possible to infer user needs with regard to multimedia data; consequently, they play significant roles in social-sensed multimedia computing.

The rise of social networks and social media platforms brought many people to the Internet for the first time, and observable user profiles and behaviors provide us with valuable resources to discover common principles of interactions between users and multimedia data. The research principles here include how and why users generate, share, and assimilate multimedia data, as outlined by Shih-Fu Chang and Alan Hanjalic.^{2,3} Notably, these common principles should not be limited to social multimedia platforms. Instead, they—and the knowledge sensed from social media—should be generalizable to other multimedia applications in which user profiles, behaviors, and social relations are not observable.

For example, can knowledge sensed from Flickr be used to improve Google image search? Can knowledge sensed from YouTube help design a recommendation system for TV programs? Can knowledge sensed from Flickr and Instagram be integrated to form a more comprehensive understanding of users?

The ultimate goal of social-sensed multimedia computing should be to sense transferable and interoperable common principles and knowledge from social multimedia and seamlessly integrate this knowledge with various multimedia services. This objective poses great challenges to those who seek to improve current techniques in the multimedia community and opens up a broad assortment of new research topics that are worth investigating. **MM**

References

1. M. Naaman, "Social Multimedia: Highlighting Opportunities for Search and Mining of Multimedia Data in Social Media Applications," *Multimedia Tools and Applications*, vol. 56, no. 1, 2012, pp. 9–34.
2. S.-F. Chang, "How Far We've Come: Impact of 20 Years of Multimedia Information Retrieval," *ACM Trans. Multimedia Computing, Communications, and Applications (TOMCCAP)*, Oct. 2013, article no. 42; <http://dl.acm.org/citation.cfm?id=2523001.2491844>.

3. A. Hanjalic, "Multimedia Retrieval that Matters," *ACM Trans. Multimedia Computing, Communications, and Applications (TOMCCAP)*, Oct. 2013, article no. 44; <http://dl.acm.org/citation.cfm?doid=2523001.2490827>.

Peng Cui is an assistant professor at Tsinghua University, China. Contact him at cui@tsinghua.edu.cn.

Wenwu Zhu is a professor at Tsinghua University, China. Contact him at wwzhu@tsinghua.edu.cn.

Tat-Seng Chua is a KITHCT Chair Professor at National University of Singapore. Contact him at chuats@comp.nus.edu.sg.


Ramesh Jain is a Bren Professor at the University of California, Irvine. Contact him at jain@ics.uci.edu.


cn Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.


stay connected.


Keep up with the latest IEEE Computer Society publications and activities wherever you are.

IEEE Computer Society

 | @ComputerSociety
 | @ComputingNow

 | facebook.com/IEEEComputerSociety
 | facebook.com/ComputingNow

 | IEEE Computer Society
 | Computing Now

 | youtube.com/ieeecomersociety