MAKING SENSE OF MEANING: LEVERAGING SOCIAL PROCESSES TO UNDERSTAND MEDIA SEMANTICS

Hari Sundaram

Arts Media Engineering, Arizona State University

Tempe, AZ 85281

Email: hari.sundaram@asu.edu

ABSTRACT

In this position paper, we propose the idea that emergent and evolutionary aspects of semantics, which are complementary to the problem of semantic detection, are foundational to multimedia computing. We show that media rich social networks reveal certain implicit assumptions in concept learning about semantics, including semantic stability, emergence, and stability of context. We study the problem of semantic evolution in the context of media rich networks – (a) since meaning is an emergent artifact of human activity, it is crucial to study how human beings interact with, consume and share media data. (b) The ready availability of large scale social interaction datasets of blogs including sites such as Flickr and YouTube, allows us to instrument the relationship between media and human activity at a scale not available to earlier researchers. We have identified three initial problem areas critical to evolutionary aspects of semantics – community discovery, information flow and semantic diversity. We shall present examples of research problems addressed in each of the three areas.

1 INTRODUCTION

A fundamental research challenge in the multimedia community is to understand the semantics of media, including images, video, sound and haptic data. The ability to interpret media with limited error, has significant, transformative outcomes including understanding human activity to take effective decisions and effective media search and organization.

While there has been considerable progress in the community to address problems of related semantic inference, much research remains to be done (e.g. see [9] for a discussion on recent developments in media search). In this paper, we would like to understand how social processes, instrumented via analysis of online media rich networks can provide additional insight into familiar multimedia research problems. In particular, we advance the idea that the emergent and evolutionary aspects of media semantics are as important as the issue of semantic detection.

Understanding media semantics has been posed as a learning problem in our community. Assume that we have set of media objects **O** (assumed to be images, for the sake of definiteness), feature space **X** that can be used to represent the images, labels **L** (e.g. "house", "cup") output semantics **S** and a space of learners **F**. By semantics, in this paper, we mean the *relationship* between the labels l and the feature distribution x. More specifically, semantics refer to the relation $\{l, P_l(x)\}$ where $P_l(x)$ is the feature distribution induced by the features corresponding to objects associated with the label l.

For example, the semantics $s \in \mathbf{S}$, of a media object $o \in \mathbf{O}$ with feature vector $x \in \mathbf{X}$, can be inferred using the discriminant function $f \in \mathbf{F}$, which has been learnt from prior training data to minimize training error. In other words, the media semantics problem is represented as $l=f(x)$, where the output symbol[1] l is assigned to the media object o via f. Learning algorithms including support vector machines and neural networks allow us to successfully learn f even when the feature space **X,** corresponding to different semantic relations, is not linearly discriminable.

A few of the implicit assumptions in the framework used for inferring multimedia semantics are revealed when dealing with data from media rich social networks. In particular, these implicit assumptions include semantic stability, a lack of semantic evolution, and a consistent context.

We typically assume that the relationship (concept semantics) between labels **L** and the underlying data **X** is fixed – i.e. the feature distributions that define a particular concept are stable – the main issue to learning this concept is additional training data. However, how we represent (or associate) familiar concepts changes with visuals changes over time – for example, for an abstract concept such as "beauty," visual representations from the early part of the 20th century will look very different from those from more recent times. This isn't an error (or inconsistency) in the training data – the relationship between the symbol "beauty" and the feature vectors have changed over time. Note that temporal variability is different from polysemy – where a concept can have multiple feature distributions[2]. Also note, that a concept can have temporal existence (a gesture that indicates greeting), and still have underlying feature distributions change over time.

Secondly, the learning framework excludes the possibility that *new* concepts emerge over time. Our work on image annotation [16], with image datasets from the popular photo sharing website Flickr, brought forth some interesting observations. We found stable concepts that couldn't have existed 20 years earlier (e.g. "pwn" a new word that means "to own"), transient concepts that were relevant only to a specific event (e.g. "BurningMan 2007") and were never used again.

Thirdly, learning a classifier on a set of images tagged with the same keyword implicitly assumes that the photos share the same context in which the keyword is appropriate. While this may be true in a carefully designed dataset (e.g. Corel dataset) this is a strong assumption on Flickr. For example, on Flickr, there are thousands of photographs labeled as "yamagata" – some are of the town, some refer to the visual artist (Hiro Yamagata), while still others refer to the singer (Rachel Yamagata). What is

---

[1] Can be assumed to be discrete, without loss of generality

[2] These can arise, for example, due to different cultural contexts.

missing is the *context* in which tag makes sense for the author / annotator of the photo. This lack of context makes it difficult to use one classifier per concept trained on all photographs in the Flickr data pool, to classify a photograph whose context is not known.

The idea of emergent semantics is also present in other disciplines including distributed cognition and sociology. Hutchins [7] observed that no single human being has ever been to every point in a map. Maps emerged in the middle ages, through work of cartographers. They collated observations from many different ship logs, resulting in emergent knowledge – a map of the world. The key idea here is that useful knowledge can emerge by the appropriate transformation and subsequent integration of partial observations from many observers; further note that no single observer had knowledge of the complete map. Garfinkel [5] suggested that the fundamental question in sociology should not be a top-down study of social order (capitalism, socialism etc.), but instead how these social systems arise out of ordinary interactions between people – i.e. he suggested that the social systems were emergent artifacts of mutually observable human interaction. This observation is central to our work on community discovery.

These observations reveal that "semantics" to be a vibrant, evolutionary aspect of our human experience. However, in the media computing community, the principal focus has been on detecting meaningful entities, and not in how these entities evolve over time, or the emergence of new concepts. The above observations indicate that integration of data from multiple sources can result in new knowledge and that new concepts emerge and evolve over time.

In this paper, we present an overview of our initial work on emergent semantics in media rich social networks. Analysis of interaction between people, in media rich social networks is important to the addressing the issues relating to semantic evolution. Media rich social networks are valuable in two ways: (a) since meaning is an emergent artifact of human activity, it is crucial to study how human beings interact with, consume and share media data. (b) The ready availability of large scale social interactions via analysis of blogs including sites such as Flickr and YouTube, allows us to instrument the relationship between media and human activity at a scale not available to earlier researchers. We have identified three initial problem areas – community discovery, information flow and semantic diversity. We shall present examples of research problems addressed in each of the three areas.

2 RESEARCH CHALLENGES

The emergent and evolutionary aspects of semantics, is a fundamental issue for the media computing community. We have identified three research challenges:

1. *What are meaningful human networks?* Since in our observations human networks have been the locus for the construction of meaning, we believe that it is crucial to find real human networks that collaborate with respect to some activity. This has spurred our work on community discovery.
2. *How does information flow in these networks?* Given a human network, how do people communicate and share artifacts with each other? What factors affect how and when two people communicate? This lead to the study of the flow of information.
3. *Semantic Diversity:* Our interaction with the physical world results in a rich, multimodal experience. The semantics present in this interaction is context dependent, and whose number seems to be exponentially large. The large number of meanings present in human interaction with the world, first poses a challenge due to the problem scale – (a) which concept detectors should we build? (b) How many concept detectors are needed? (c) How to account for diversity of user context? (d) is it possible for us to find training examples train each classifier?

For our work on emergent semantics, given the necessity of analysis of human networks, and the availability of large scale datasets, it was natural to begin exploring the problem within the context of media rich social networks.

3 COMMUNITY DISCOVERY

The goal is to extract human communities that collaborate around certain topics / activities. While there has been much prior work on community extraction in computer science, a key issue is limited problem formulation – a social network is seen as a graph (with people as nodes, and hyperlinks (or some other activity) as edges), and community extraction is typically viewed as a dense sub-graph extraction problem [6, 8, 14, 17]. While the formalism is neat, community formation is a social phenomena – an understanding of community discovery in computer science should incorporate the wealth of knowledge in sociology, *on how human networks form.* Specifically, this means that semantics of the connectivity (i.e. what does it mean to connect two people via an edge) between people needs to be carefully examined. This is important, since ground truth regarding the community is rarely available – we cannot "ask" the people if they "belong" to the community that we have discovered. Community discovery validation is a key research question in this problem area.

Our approach on community discovery in large scale blog networks has been primarily motivated by Garfinkel's work on the necessity of mutually observable actions, for the formation a strong connection between two people. A key idea in this work is that human communities emerge through observable actions by the community members. Members develop a common sense understanding through the reciprocal actions in a so called *action community*. To form a community, it is therefore critical that individual bloggers become aware of each other's presence through interaction (comments, hyperlinks, trackbacks etc.). We refer to this bi-directional property as "mutual awareness" of bloggers.

In our work [10-12], we initially focused on detecting communities that have a strong content related theme (i.e. query specific community discovery). We first construct a directed graph that is time-dependent, and weighted with respect to a specific query. Each node is a blogger, and the edges are the observable actions. We create a "symmetric social distance" (in Travers and Milgram's well-known small world experiment[3], social distance was one-way) which is estimated using a random walk process, to capture mutual awareness expansion within a community. We then extract communities by recursively maximizing the distance between two sets of bloggers. We developed an interaction space based representation to quantify

[3] This is the well known "six degrees of separation" experiment.

community dynamics, where each dimension represents a pair-wise interaction between two people, and each community is a vector in the interaction space. Two communities are close if they are close in the interaction space.

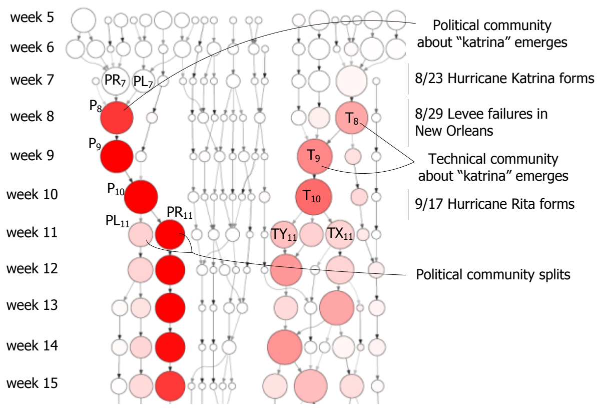

Figure 1: Community evolution graph for the query "Katrina" [11].

In Figure 1, each node represents a detected community where the communities detected during the same week are aligned horizontally and the communities at different time snapshots are connected by arrows through our evolution analysis. The grayscale of an arrow is proportional to the interaction similarity between the two communities. The node size reflects the number of community members and the reddish shade of node is proportional to the query relevancy for the keyword "katrina". More saturated node represents more relevant community. We observed emergence of two types of communities in our dataset, one with a focus on politics (shown on the left) and the other with a focus on technology (shown on the right).

More recently, we have extended this work, using language grammar as a metaphor [13] – we are now looking to extract *grammatical properties of interaction* within communities. Specifically, we are extracting the triples of *people*, *actions* and *media artifacts* with the natural analogy to subject, verb and object. Work on the grammar of social interaction, can lead to generalizable descriptors of communities, as well as develop into a rich vein of computational problems in social networks, akin to problems in computational linguistics.

Community discovery can have a huge impact on information search, particularly when dealing with costly semantic queries. For questions such as ("what is a good school district?," "should I buy a car?," "where should I take my spouse on our 10th wedding anniversary?") we rely on our real-world social networks, not Google. However, it is not always possible for our trusted network to be able to answer all of our questions. We believe that query-specific community discovery will play an important role in address these costly queries – query-specific communities will allow us to view and track a set of complementary opinions on a topic of great interest to us, thus allowing us to take meaningful action at an appropriate time.

4 INFORMATION FLOW

How do we characterize the dynamics of information flow within human networks [2, 4]? Understanding how information flows, and the different information roles that individuals play within these networks are important in problems relating to ranking of information sources, assessing information quality, as well as in network communication resource allocation strategies. Companies are also interested in how they are perceived and discussed, with respect to specific events.

We have developed a multi-scale (individuals, groups, communities) characterization of communication flow within online social networks [1]. The key questions that we have attempted to answer in this work are: (a) given a certain message, what is the probability that an individual will respond? What would be the delay? (b) What is the role of the network topology, social context (information role, strength of weak ties) and topic relevance in how an individual reacts to communication? (c) What are the key groups that emerge around specific topics, and how do they affect the community?

One of the most important aspects of this research has been our ability to *validate knowledge about the network.* As mentioned earlier, it is possible to generate any number of numerical metrics (conductance, cohesiveness etc.), to extract facts about online networks. Which of these facts are actually useful is not very clear, since we rarely have ground truth data about social networks.

In our work on the analysis of technology blogs (Engadget), we have cross-validated our analysis with the stock market activity of several popular companies [3]. We can show that communication dynamics in these blogs is *a strong predictor of future events in the stock market*. The work on validation social knowledge will have a significant impact on the community – it opens up possibilities to extract facts that are useful (in the sense that they enable further human action). For example, maps are important because they represent transformations of data into knowledge that enables future human action.

5 SEMANTIC DIVERSITY

We began this research in an effort to annotate media objects, such as photos [15, 16]. A solution to the annotation problem will have important consequences in media search. An examination of photo and tag data from online photo social networks (Flickr), revealed several interesting observations – (a) We found that the tag frequency followed a classic power-law distribution, implying that most tags do not have enough example images to build classifiers. (b) New concepts emerge ("pwn") (c) Some concepts highly specific to a small group of people – i.e. the semantics of the tags are not universally shared or known.

There are two fundamental issues here – extraordinarily large number of concepts, and concept diversity. Simply, there are not enough examples available to build a classifier for every concept (due to the power-law characteristic), and it would be computationally prohibitive to test each image against every classifier. Secondly, since semantics are not universally shared, one cannot simply pool all images with the same tag to build a classifier.

Our *central innovation* was to do away with classifiers altogether – i.e. stop thinking of annotation as a classification problem [18]. Instead we combined three forms of knowledge – global (feature based distances), personal (tag co-occurrence probability, per individual) and social trust (finding people whose experiences are correlated), within an iterative framework. The framework is based on work found in web search algorithms and handles the scale of the social network data efficiently. The issue of concept diversity is addressed by the use of social trust and personal knowledge – i.e. we should recommend annotations only from those people who typically agree with us, given the same experience. We believe that the methodology adopted to address concept scale will have a significant impact within the media computing community as it is yet to pay significant attention to issue of large semantic concept cardinality.

6 CONCLUSIONS

In this position paper, we have advanced the idea that the emergent and evolutionary aspects of semantics are foundational to multimedia computing. Media rich social networks reveal certain implicit assumptions in concept learning about semantics, including stability, emergence, and stability of context do not hold. We identified three problem areas – including community discovery, information flow within communities and the issue of semantic concept cardinality, and showed early work in each area. Social network datasets provide significant opportunities for multimedia researchers – fundamental questions of emergence (including cultural memes), rates of changes of semantics and their relationships to interaction context, remain unaddressed.

7 ACKNOWLEDGEMENTS

We (the author and his students: Yu-Ru Lin, Munmun De Choudhury, Bageshree Shevade and Amit Zunjarwad) are grateful to the generous support of NEC, IBM and Avaya Labs in facilitating this research. Our research collaborators include Aisling Kelliher (ASU), Belle Tseng, Jun Tatemura and Yun Chi (NEC), Lexing Xie (IBM), Ajita John and Doree Duncan Seligmann (Avaya Labs).

8 REFERENCES

[1] M. D. Choudhury, H. Sundaram, A. John, and D. Seligmann, "Dynamic prediction of communication flow using social context," Proceedings of the nineteenth ACM conference on Hypertext and hypermedia, Pittsburgh, PA, USA, 2008.

[2] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "Contextual Prediction of Communication Flow in Social Networks," Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, San Jose, CA, 2007.

[3] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "Can blog communication dynamics be correlated with stock market activity?," Proceedings of the nineteenth ACM conference on Hypertext and hypermedia, Pittsburgh, PA, USA, 2008.

[4] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "What Makes Conversations Interesting? Themes, Participants and Consequences of Conversations in Online Social Media," accepted to appear in Proc. Of the 18th Intl. World Wide Web Conference, Madrid, Spain, 2009.

[5] P. Dourish, *Where the action is : the foundations of embodied interaction*. Cambridge, Mass. ; London: MIT Press, 2001.

[6] T. Falkowski, J. Bartelheimer, and M. Spiliopoulou, "Mining and Visualizing the Evolution of Subgroups in Social Networks," International Conference on Web Intelligence, 2006., 2006.

[7] E. Hutchins, *Cognition in the wild*. Cambridge, Mass.: MIT Press, 1995.

[8] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, "On the Bursty Evolution of Blogspace," *World Wide Web*, vol. 8, pp. 159-178, 2005.

[9] M. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 2, pp. 1-19, 2006.

[10] Y.-R. Lin, H. Sundaram, Y. Chi, J. Tatemura, and B. Tseng, "Discovery of Blog Communities based on Mutual Awareness," Third Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, at the 15th Annual World Wide Web Conference - WWW 2006, also AME-TR-2006-03, Edinburgh, Scotland, 2006.

[11] Y.-R. Lin, H. Sundaram, Y. Chi, J. Tatemura, and B. L. Tseng, "Blog Community Discovery and Evolution Based on Mutual Awareness Expansion," Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, San Jose, CA, 2007.

[12] Y.-R. Lin, H. Sundaram, and A. Kelliher, "Summarization of social activity over time: people, actions and concepts in dynamic networks," Proceeding of CIKM, Napa Valley, CA, 2008.

[13] Y.-R. Lin, H. Sundaram, and A. Kelliher, "Summarization of Large Scale Social Network Activity," accepted to appear in IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, 2009.

[14] G. Palla, A. Barabasi, and T. Vicsek, "Quantifying social group evolution," *eprint arXiv: 0704.0744*, 2007.

[15] B. Shevade, H. Sundaram, and L. Xie, "Exploiting Personal And Social Network Context For Event Annotation," IEEE International Conference on Multimedia and Expo, 2007, Beijing. China, 2007.

[16] B. Shevade, H. Sundaram, and L. Xie, "Modeling Personal and Social Network Context for Event Annotation in Images," Proc. Joint Conf. on Digital Libraries 2007, Vancouver, Canada, 2007.

[17] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888-905, 2000.

[18] A. Zunjarwad, H. Sundaram, and L. Xie, "Contextual Wisdom: Social Relations and Correlations for Multimedia Event Annotation," Proc. of the 15th annual ACM international conference on Multimedia, Augsburg, Germany, 2007.