Detecting Fake News with Weak Social Supervision

Kai Shu*, Ahmed Hassan Awadallah †, Susan Dumais†, and Huan Liu*
*Arizona State University, {kai.shu, huan.liu}@asu.edu
†Microsoft Research, {hassanam, sdumais}@microsoft.com

Abstract—Limited labeled data is becoming the largest bottleneck for supervised learning systems. This is especially the case for many real-world tasks where large scale annotated examples are either too expensive to acquire or unavailable due to privacy or data access constraints. Weak supervision has shown to be a good means to mitigate the scarcity of annotated data by leveraging weak labels or injecting constraints from heuristic rules and/or external knowledge sources. Social media has little labeled data but possesses unique characteristics that make it suitable for generating weak supervision, resulting in a new type of weak supervision, i.e., weak social supervision. In this article, we illustrate how various aspects of social media can be used to generate weak social supervision. Specifically, we use the recent research on fake news detection as the use case, where social engagements are abundant but annotated examples are scarce, to show that weak social supervision is effective when facing the little labeled data problem. This article opens the door for learning with weak social supervision for other emerging tasks.

Index Terms: data mining, machine learning, social networking

Deep neural networks have shown great achievements in various application domains such as image classification, natural language processing, graph modeling. These models usually require huge amounts of annotated data for training. However, limited labeled training data is becoming the bottleneck for supervised learning systems especially those employing deep neural networks. In many real-world tasks, large scale annotated examples are either too expensive to acquire or unavailable due to privacy or data access constraints. To combat this, researchers have developed approaches based on weak supervision to leverage noisy signals for learning. Along this direction, training data is automatically generated from unlabeled text corpora using a heuristic labeling functions or regular expression patterns (e.g., “if the word ‘causes’ appears between the chemical and disease mentions, then it indicates a true pair of chemical-disease relation”). Similarly, sentences for relation extraction can be aligned to entities and relations in knowledge bases (KB) with heuristics like “if a sentence mentions two entities of a triple from KB then the sentence assets their relation”. Most of these approaches rely on heuristics from text information (in the forms of rules or regular expressions) or external sources (such as structural knowledge bases).

In recent years, social media has become an important means of large-scale information sharing and communication in all occupations, including marketing, journalism, public relations, and more. Social media data is seldom labeled and has unique properties that make it suitable for generating weak supervision. First, social media
data is big. We have small data for each of most individuals. However, the social property of social media data links individuals data together, which provides a new type of big data. Second, social media data is linked. The availability of social relations determines that social media data is inherently linked, meaning it is not independent and identically distributed. Third, social media data is noisy. Users in social media can be both passive content consumers and active content producers, causing the quality of user generated content to vary. Social networks are also noisy with the existence of malicious users such as spammers and bots. Therefore, social media data provides a new type of weak supervision, weak social supervision, which has great potentials to advance a wide range of application domains.

Recent research shows that disinformation and fake news have been widely disseminated on social media and caused detrimental societal effects. Detecting fake news on social media is much desired to avoid people to consume false information and cultivate a healthy and trustworthy news ecosystem. However, detecting fake news on social media poses several unique challenges [1].

- First, the data challenge has been a major roadblock for researchers in their attempts to develop effective defensive means against disinformation and fake news. This is because the content of fake news and disinformation is rather diverse in terms of topics, styles and media platforms; and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. Thus, obtaining annotated fake news data is non-scalable and data-specific embedding methods are not sufficient for fake news detection with little labeled data.

- Second, the evolving challenge of disinformation and fake news makes it non-trivial to exploit the rich auxiliary information signals directly. Fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases (KB) due to the lack of corroborating evidence or claims. In addition, social supervision is rich in many ways, such as user profiling, user generated contents, and network interactions, which provide useful information to build better systems to detect fake news and disinformation. To tackle these unique challenges, it is important to learn with weak social supervision on little labeled data for detecting fake news.

In this article, we discuss three major perspectives of the social media data to generate weak social supervision for fake news detection: users, posts, and networks. We further introduce recent work on exploiting weak social supervision for effective and explainable fake news detection. First, we illustrate how we can model the relationships among publishers, news pieces, and social media users with user-based and network-based weak social supervision to detect fake news effectively. Second, we show how to leverage post-based weak social supervision to discover explainable comments for detecting fake news. We finally discuss several open issues and provide future directions of learning with weak social supervision.

Characterizing on Weak Social Supervision

In this section, we first discuss various definitions of weak supervision. We then describe different aspects of weak supervision and the new patterns found on social media for weak social supervision.

Weak supervision

Learning with weak supervision is an important and newly emerging research area, and there are different ways of defining and approaching the problem. Therefore, we first discuss and compare some widely used definitions of weak supervision in the existing literature. One definition of weak supervision is leveraging higher-level and/or noisier input from subject matter experts (SMEs) [2]. The supervision from SMEs are represented in the form of weak label distributions, which mainly come from the following sources: 1) inexact supervision: a higher-level and coarse-grained supervision; 2) inaccurate supervision: a low-quality and noisy supervision; and 3) existing resources: using existing resources to provide supervision. Another definition definition categorize weak supervision into inexact supervision, inaccurate supervision, and incomplete supervision [3]. The incomplete supervision means that a subset of training data are given with labels, which essentially includes active learning and semi-supervised learning techniques. The incom-
plete supervision tries to avoid asking SMEs for additional training labels, by either leverage domain- and task-agnostic assumptions to exploit the unlabeled data (semi-supervised learning), or minimizing the queries to SMEs to specific data instances (active learning); while the key practical motivation for weak supervision is to effectively obtain weak training information from SMEs.

Weak supervision can be formed in deterministic (e.g., in the form of weak labels) and non-deterministic (e.g., in the form of constraints) ways [2]. First, weak labels can come from heuristic rules, crowd sourcing, distant supervision, or existing classifiers. The most popular form is distant supervision, in which the records of an external knowledge base (KB) are heuristically aligned with data points to produce noisy labels. Second, weak supervision can be used as optimization constraints on output prediction derived from domain knowledge [1]. Next, we briefly discuss the representative approaches for learning with weak supervision by incorporating weak labels and injecting constraints.

**Incorporating Weak Labels** Learning with noisy (inaccurate) labels has been of great interest to the research community for various tasks. Some of the existing works attempt to rectify the weak labels by incorporating a loss correction mechanism. Sukhbaatar et al. [5] introduce a linear layer to adjust the loss and estimate label corruption with access to the true labels. Patrini et al. [6] utilize the loss correction mechanism to estimate a label corruption matrix without making use of clean labels. Other works consider the scenario where a small set of clean labels are available [7]. For example, Veit et al. use human-verified labels and train a label cleaning network in a multi-label classification setting. Recent works also consider the scenario where weak signals are available from multiple sources [2] to exploit the redundancy as well as the consistency among weak labels.

In some cases, weak supervision is obtained with inexact labels such as coarse-grained labels. For example, object detectors can be trained with images collected from the web using their associated tags as weak supervision instead of locally-annotated data sets. Multi-instance learning has been developed to learn from labels of instances (e.g., tags) to infer the labels of the corresponding bag (e.g., the object). Most of existing algorithms aim to adapt single-instance supervised learning to multi-instance representation, and some approaches attempt to adapt multi-instance representation to single-instance representation through transformation.

**Injecting Constraints** Directly learning with weak labels may suffer the from the noisy label problem. Instead, representing weak supervision as constraints can avoid noisy labels and encode domain knowledge into the learning process of prediction function. The constraints can be injected over the output space and/or the input representation space. For example, Stewart et al. [4] model prior physics knowledge on the outputs to penalize “structures” that are not consistent with the prior knowledge. For relation extraction tasks, label-free distant supervision can be achieved via encoding entity representations under transition law from knowledge bases (KB). The constraints form of weak supervision are often based on prior knowledge from domain experts, which are jointly optimized with the primary learning objective of prediction tasks.

**Weak Social Supervision** With the rise of social media, the web has become a vibrant and lively realm where billions of individuals all around the globe interact, share, post and conduct numerous daily activities. Social media enables us to be connected and interact with anyone, anywhere and anytime, which allows us to observe human behaviors in an unprecedented scale with new lens. However, significantly different from traditional data, social media data is big, incomplete, noisy, unstructured, with abundant social relations. This new type of data mandates new computational analysis approaches that can combine social theories and statistical data mining techniques.

Many application domains like recommendation, information sharing and communication, news consumption have rich social media engagements that can provide addition signals for obtaining weak supervision. Social media engagements are referring to the involvement and interactions of users participating in activities through social networking systems. Generally, there are three
major aspects of the social media engagements: users, contents, and relations. First, users may have different characteristics or establish different patterns of behaviors. Second, users express their opinions and emotions through posts/comments. Third, users form different types of relations on social media such as communities. The goal of weak social supervision is to leverage signals from social media engagements to obtain weak supervision for various downstream tasks. Similar to weak supervision, we can utilize weak social supervision mainly in the form of weak labels and constraints.

Learning with Weak Social Supervision

In the previous section, we introduced the conceptual characterization of weak supervision and weak social supervision. In this section, we further discuss the problem definition and the representative approaches of learning with weak social supervision.

Problem Definition

A training example consists of two parts: a feature vector (or called instance) describing the event/object, and a label indicating the ground-truth output. Let $D = \{x_i, y_i\}_{i=1}^n$ denote a set of $n$ examples with ground truth labels, with $\mathcal{X} = \{x_i\}_{i=1}^n$ denoting the instances and $\mathcal{Y} = \{y_i\}_{i=1}^n$ the corresponding labels. In addition, there is a large set of unlabeled examples. Usually the size of the labeled set $n$ is much smaller than the unlabeled set due to labeling costs or privacy concerns.

For the widely available unlabeled samples, we generate weak social supervision by generating weak labels or incorporating constraints based on social media engagements. For weak labels, we aim to learn a labeling function $g : \tilde{\mathcal{X}} \rightarrow \tilde{\mathcal{Y}}$, where $\tilde{\mathcal{X}} = \{\tilde{x}_j\}_{j=1}^N$ denotes the set of $N$ unlabeled messages to which the labeling function is applied and $\tilde{\mathcal{Y}} = \{\tilde{y}_j\}_{j=1}^N$ as the resulting set of weak labels. This weakly labeled data is then denoted by $\tilde{D} = \{\tilde{x}_j, \tilde{y}_j\}_{j=1}^N$ and often $n << N$. For formulating constraints, we aim to model prior knowledge from social signals on the representation learning of examples with a constraint function $h : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$, to penalize structures that are not consistent with our prior knowledge. Note that $g$ can also be applied to $\tilde{\mathcal{X}}$ to regularize the representation learning. In despite of the different forms we model weak social supervision, we are actually aiming to estimate a label distribution $p(\tilde{y} | \tilde{x})$ from weak social supervision. We give the following problem formulation of learning with weak social supervision.

**Learning with Weak Social Supervision:**
Given little data with ground truth labels $D$ and a label distribution $p(\tilde{y} | \tilde{x})$ derived from weak social supervision, learn a prediction function $f : \mathcal{X} \rightarrow \mathcal{Y}$ which generalizes well onto unseen samples.

Exploiting Weak Social Supervision for Fake News Detection

In this section, we describe an application of exploiting weak social supervision for fake news detection. We choose fake news detection since research has shown social media enables the wide propagation of fake news, extra auxiliary information can be used to as weak social supervision to help detect fake news. We first discuss the unique aspects of unique features of social media and its implications on extracting weak social supervision. Then, we demonstrate representative approaches of exploiting weak social supervision for challenging problems of fake news detection.

Generating Weak Social Supervision

In this subsection, we discuss different sources for generating weak social supervision. Generally, there are three major aspects of the social media context: users, generated posts, and networks. We will introduce how to extract/create weak social supervision in the next subsection.

User-based: Fake news pieces are likely to be created and spread by non-human accounts, such as social bots or cyborgs. Thus, capturing users’ profiles and characteristics as weak social supervision can provide useful information for fake news detection. User behaviors can indicate their characteristics who have interactions with the news on social media. These features can be categorized in different levels: individual level and group level. Individual level features are extracted to infer the credibility and reliability for
each user using various aspects of user demographics, such as registration age, number of followers/followees, number of tweets the user has authored, etc. Group level user features capture overall characteristics of groups of users related to the news. The basic assumption is that the spreaders of fake news and real news may form different communities with unique characteristics that can be depicted by group level features. Commonly used group level features come from aggregating (e.g., averaging and weighting) individual level features, such as ‘percentage of verified users’ and ‘average number of followers’.

Post-based: Users who are involved in news dissemination process express their opinions, emotions via posts/comments. These user responses provide helpful signals related to the veracity of news claims. Recent research looks into stance, emotion, and credibility to improve the performance of fake news detection. First, stances (or viewpoints) indicate the users’ opinions towards the news, such as supporting, opposing, etc. Typically, fake news can provoke tremendous controversial views among social media users, in which denying and questioning stances are found to play a crucial role in signaling claims as being fake. The stance of users’ posts can be either explicit or implicit. Explicit stances are direct expressions of emotion or opinion, such as Facebook’s “like” actions. Implicit stances can be automatically extracted from social media posts. Second, fake news publishers often aims to spread information extensively and draw wide public attention. Longstanding social science studies demonstrate that the news which evokes high-arousal, or activating (awe, anger or anxiety) emotions is more viral on social media. Third, weak social supervision on post credibility aims to infer the veracity of news pieces from the credibility of the posts on social media through network propagation. The basic assumption is that the credibility of a given news event is highly related to the credibility degree of its relevant social media posts.

Network-based: Users form different networks on social media in terms of interests, topics, and relations. Fake news dissemination processes tend to form an echo chamber cycle, highlighting the value of extracting network-based weak social supervision to represent these types of network patterns for fake news detection. Different types of networks can be constructed such as friendship networks, diffusion networks, interaction networks, propagation networks, etc.

First, friendship network plays an important role in fake news diffusion. The fact that users are likely to form echo chambers, strengthens our need to model user social representations and to explore its added value for a fake news study.

Second, the news diffusion process involves abundant temporal user engagements on social media. Fake news may have sudden increase for the number of posts and then remain constant beyond a short time whereas, in the case of real news, the increase of the number of posts are more steady.

In addition, an important problem along temporal diffusion is the early detection of fake news with limited amount of user engagements. Third, interaction networks describe the relationships among different entities such as publishers, news pieces, and users. For example, the user-news interactions are often modeled by considering the relationships between user representations and the news veracity values. Intuitively, users with low credibilities are more likely to spread fake news, while users with high credibility scores are less likely to spread fake news.

Moreover, the propagation networks have a hierarchical structure, including macro-level and micro-level propagation networks. On one hand, macro-level propagation networks demonstrate the spreading paths from news to the social media posts, and those reposts of these posts. On the other hand, micro-level propagation networks illustrate the user conversations under the posts or reposts, such as replies/comments. Micro-level networks contain user discussions towards news pieces, which brings auxiliary cues to differentiate fake news.

Exploiting Weak Social Supervision

Earlier, we illustrate different aspects that we can generate weak social supervision from. Actually, the extracted weak social supervision can involve single or multiple aspects of the information related to users, content, and networks. In this section, we discuss learning with weak social
supervision for fake news detection in different settings including effective fake news detection and explainable fake news detection. Specifically first, we illustrate how we can model the user-based and network-based weak social supervision to detect fake news effectively. Second, we show how to leverage post-based weak social supervision for discovering explainable comments for detecting fake news.

Effective Fake News Detection We aim to leverage weak social supervision as an auxiliary information to perform fake news detection effectively. As an example, we demonstrate how we can utilize interaction networks by modeling the entities and their relationships to detect fake news (see Figure 1). Interaction networks describe the relationships among different entities such as publishers, news pieces, and users. Given the interaction networks the goal is to embed the different types of entities into the same latent space, by modeling the interactions among them. We can leverage the resultant feature representations of news to perform fake news detection, and we term this framework Tri-relationship for Fake News detection (TriFN) [9].

Inspired from sociology and cognitive theories, we derive the weak social supervision rules. For example, social science research has demonstrated the following observations which serves our weak social supervision: people tend to form relationships with like-minded friends, rather than with users who have opposing preferences and interests. Thus, connected users are more likely to share similar latent interests in news pieces. In addition, for publishing relationship, we exploit the following weak social supervision: publishers with a high degree of political bias are more likely to publish fake news. Moreover, for the spreading relation, we have: users with low credibilities are more likely to spread fake news, while users with high credibility scores are less likely to spread fake news. We utilize nonnegative matrix factorization (NMF) to learn the news representations by encoding the weak social supervision.

Empirical Results To illustrate whether the weak social supervision in TriFN can help detecting fake news effectively, we show some empirical comparison results in the public benchmark Politifact dataset from FakeNewsNet[1] as in Figure 2. We compare TriFN with baselines that 1) only extract features from news contents, such as RST, LIWC; 2) only construct features from social supervision, such as Castillo; and 3) consider both news content and social supervision, such as RST+Castillo, LIWC+Castillo. We can see that the proposed TriFN can achieve around 0.75 accuracy even with limited amount of weak social supervision (within 12 hours after news is published), and has as high as 0.87 accuracy. In addition, with the help of weak social supervision from publisher-bias and user-credibility, the detection performance is better than those without utilizing weak social supervision. Moreover, we can see within a certain range, more weak social supervision leads to larger performance increase, which shows the benefit of using weak social supervision.

Explainable Fake News Detection In recent years, computational detection of fake news has been producing some promising early results. However, there is a critical missing piece of the study, the explainability of such detection, i.e., why a particular piece of news is detected as fake. Here, we introduce how we can derive explanation factors from weak social supervision.

Figure 1. The TriFN model of learning with social supervision from publisher bias and user credibility for effective fake news detection [9].
We observe that not all sentences in news contents are fake, and in fact, many sentences are true but only for supporting wrong claim sentences. Thus, news sentences may not be equally important in determining and explaining whether a piece of news is fake or not. Similarly, user comments may contain relevant information about the important aspects that explain why a piece of news is fake, while they may also be less informative and noisy. We use the following weak social supervision: the user comments that are related to the content of original news pieces are helpful to detect fake news and explain prediction results. Thus, we aim to select some news sentences and user comments that can explain why a piece of news is fake. As they provide a good explanation, they should also be helpful in detecting fake news. This suggests us to design attention mechanisms to give high weights of representations of news sentences and comments that are beneficial to fake news detection. Specifically, we first use Bidirectional LSTM with attention to learn sentence and comment representations, and then utilize a sentence-comment co-attention neural network framework called dEFEND (see Figure 3) to exploit both news content and user comments to jointly capture explainable factors.

**Empirical Results** We show the empirical results on Politifact platform from FakeNewsNet as in Figure 4. We can see dEFEND achieves very high performances in terms of accuracy ($\sim 0.9$) and F1 ($\sim 0.92$). We compare dEFEND with three variants: 1) dEFEND\C not considering information from user comments; 2) dEFEND\N is not considering information from news contents; and 3) dEFEND\Co eliminates the sentence-comment co-attention. We observe that when we eliminate news content component, user comment component, or the co-attention for news contents and user comments, the performances are reduced. It indicates capturing the semantic relations between the weak social supervision from user comments and news contents are important.
Open Issues and Future Research

In this section, we present some open issues in weak social supervision and future research directions. Learning with weak social supervision is a newly emerging research area in data mining and machine learning, so we aim to point out promising research directions specifically for fake news detection, and for other applications in general.

Weak Social Supervision for Fake News Detection

Most of the current methods are trying to exploit weak social supervision as constraints to help fake news detection. We can also exploit generating weak labels from the aforementioned social signals (user-based, post-based, and network-based) as labeling functions for early fake news detection. The advantage of leveraging weak social supervision for early fake news detection is that we can jointly learn the feature representations from little labeled data and weak labeled data, and when predicting unseen news pieces, we can perform prediction with few/no social signals, which perfectly satisfy the requirement of early detection. In addition, in the extreme case when no labeled data in available, we can utilize weak social supervision for unsupervised fake news detection. One idea is to extract users’ opinions on the news by exploiting the auxiliary information of the users’ engagements from posts on social media, and aggregate their opinions in a well-designed unsupervised way to generate our estimation results.

Learning with Weak Social Supervision

We expect along the direction of learning with weak social supervision, more research will emerge in the near future. First, leveraging weak social supervision for computation social science research is promising. Since computational social science research usually relies on relatively limited offline survey data, weak social supervision can serve as a powerful online resources to understand and study social computing problems. Second, existing approaches utilize single or combine multiple sources of weak social supervision, while to what extent and aspect the weak social supervision helps is fairly important to explore. Third, the capacity of ground-truth labels and weak social supervision and the relative importance between the sources are essentials to develop learning methodology in practical scenarios. Moreover, the weak supervision rules may have complementary information since they capture social signals from different perspectives. An interesting future direction is to explore multi sources of weak social supervision in a principled way to model the mutual benefits through data programming.

Conclusion

In many machine learning applications, labeled data is scarce and obtaining more labels is expensive. Motivated by the promising early results of exploiting weak supervision learning, we propose a new type of weak supervision, i.e., weak social supervision. We specifically focus on the use case of detecting fake news on social media. Specifically, we demonstrate that weak social supervision provides a new representation to describe social information uniquely available where better warning is sought, which has promising results and great potentials toward detecting fake news, including challenging settings of effective fake news detection and explainable fake news detection. We also further discuss promising future directions in fake news detection research and expand the field of learning with weak social supervision to other applications.

REFERENCES

1. K. Shu and H. Liu, “Detecting fake news on social media,” Synthesis Lectures on Data Mining and Knowledge Discovery, vol. 11, no. 3, pp. 1–129, 2019.
2. A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré, “Snorkel: Rapid training data creation with weak supervision,” Proceedings of the VLDB Endowment, vol. 11, no. 3, pp. 269–282, 2017.
3. Z.-H. Zhou, “A brief introduction to weakly supervised learning,” National Science Review, vol. 5, no. 1, pp. 44–53, 2017.
4. R. Stewart and S. Ermon, “Label-free supervision of neural networks with physics and domain knowledge,” in Thirty-First AAAI Conference on Artificial Intelligence, 2017.
5. S. Sukhbaatar, J. Bruna, M. Paluri, L. Bourdev, and R. Fergus, “Training convolutional networks with noisy labels,” arXiv preprint arXiv:1406.2080, 2014.
6. G. Patrini, A. Rozza, A. Krishna Menon, R. Nock, and L. Qu, “Making deep neural networks robust to label noise: A loss correction approach,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1944–1952.

7. D. Hendrycks, M. Mazeika, D. Wilson, and K. Gimpel, “Using trusted data to train deep networks on labels corrupted by severe noise,” in Advances in neural information processing systems, 2018, pp. 10456–10465.

8. V. Subrahmanian and S. Kumar, “Predicting human behavior: The next frontiers,” Science, vol. 355, no. 6324, pp. 489–489, 2017.

9. K. Shu, S. Wang, and H. Liu, “Beyond news contents: The role of social context for fake news detection,” in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. ACM, 2019, pp. 312–320.

10. K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, “defend: Explainable fake news detection,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2019.

Kai Shu is a PhD student and research assistant at Data Mining and Machine Learning (DMML) Lab at Arizona State University. His research interests include artificial intelligence, social computing, data mining, in particular, learning with weak social supervision for fake news detection. He is the leading author of the book on Detecting Fake News on Social Media, and co-presented two tutorials in KDD 2019 and WSDM 2019. He also worked as a research intern at Yahoo! Research and Microsoft Research in 2018 and 2019, respectively. Contact him at kai.shu@asu.edu.

Ahmed Hassan Awadallah is a principle research manager at Microsoft Research (MSR). He leads the language and information technology team at MSR, focusing on creating language understanding and user modeling technologies to enable intelligent experiences in multiple products. He received his Ph.D. from University of Michigan Ann Arbor. He invented, published, and patented new approaches in language understanding, information retrieval and machine learning, and published 60+ peer-reviewed papers in these areas and he is an inventor on 20+ (granted and pending) patents. Contact him at hassanam@microsoft.com.

Susan Dumais is a the technical fellow and deputy managing director at Microsoft Research (MSR), and an adjunct professor in the Information School at the University of Washington. Her research interests are in algorithms and interfaces for improved information retrieval, as well as general issues in human-computer interaction. She was elected to the CHI Academy in 2005, an ACM Fellow in 2006, received the SIGIR Gerard Salton Award for Lifetime Achievement in 2009, was elected to the National Academy of Engineering (NAE) in 2011, received the ACM Athena Lecturer and Tony Kent Strix Awards in 2014, was elected to the AAAS in 2015, and received the Lifetime Achievement Award from Indiana University in 2016. Contact her at sdumais@microsoft.com.

Huan Liu is a professor of Computer Science and Engineering at Arizona State University. His research interests are in data mining, machine learning, social computing, and artificial intelligence, investigating problems that arise in real-world applications with high-dimensional data of disparate forms. He co-authored the textbook on Social Media Mining: An Introduction by Cambridge University Press. He is a founding organizer of the International Conference Series on Social Computing, Behavioral-Cultural Modeling, and Prediction, and Chief Editor of Data Mining and Management in Frontiers in Big Data. He is a Fellow of ACM, AAAI, AAAS, and IEEE. Contact him at huan.liu@asu.edu.