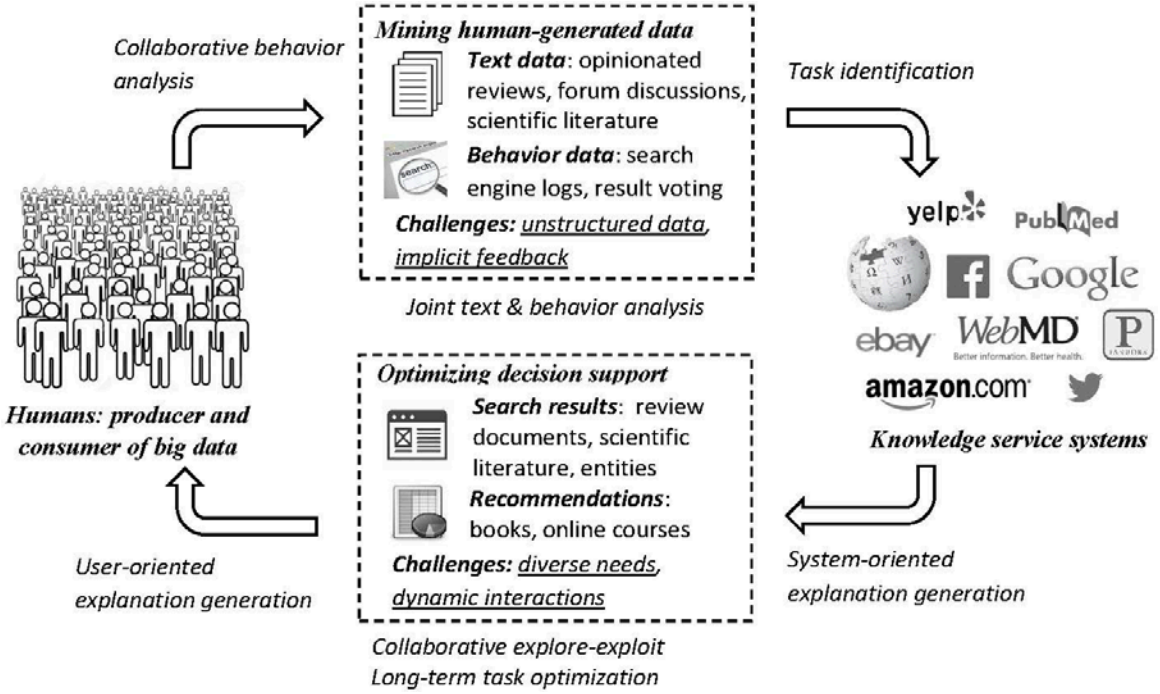

As data producers, humans constantly generate huge amount of text data and behavior data. Joint modeling of such diverse types of data ensures in-depth understanding of humans. On the other hand, as knowledge consumers, different users interact with various systems to fulfill their *idiosyncratic intents*. Traditional static, ad-hoc and passive machine-human interactions are inadequate to optimize such dynamic decision making process; and thus online adaptively learning driven by the feedback from humans is desired. As a result, only with such an integrated view, we can align our effort in knowledge discovery with that in optimizing humans' decisions, and identify opportunities to fundamentally improve existing algorithms' and systems' utilities.

The above figure of "human-centric knowledge discovery and decision optimization" illustrates the research in our group. In this loop, improved systems' utilities can be produced by in-depth understanding of humans (i.e., the flow from humans to systems); and optimized humans' decision making can be realized by customized knowledge services (i.e., the flow from systems to humans).

Human-Centric Data Mining Group

Hongning Wang

Assistant Professor
hw5x@virginia.edu
www.cs.virginia.edu/~hw5x/

Dept. of Computer Science
University of Virginia
Charlottesville, VA
434.982.2228

"Humans are Both Producers and Consumers of Big Data."

UNIVERSITY of VIRGINIA
SCHOOL of ENGINEERING & APPLIED SCIENCE

Contextual Bandits in a Collaborative Environment

Contextual bandit algorithms provide principled online learning solutions to find optimal trade-offs between exploration and exploitation with companion side-information. A common practice in bandits estimates the unknown bandit parameters pertaining to each user independently. This unfortunately ignores dependency among users and thus leads to suboptimal solutions, especially for the applications that have strong social components. In our work, we develop a collaborative contextual bandit algorithm, in which the adjacency graph among users is leveraged to share context and payoffs among neighboring bandits while online updating. We rigorously prove an improved upper regret bound of the proposed collaborative bandit algorithm comparing to conventional independent bandit algorithms. Extensive experiments on both synthetic and three large-scale real-world datasets verified the improvement of our proposed algorithm against several state-of-the-art contextual bandit algorithms.

Collaborative Model Adaptation for Personalized Sentiment Classification

Humans' opinions are idiosyncratic and variable: the same opinions can be expressed in various ways and the same expression can carry distinct sentimental polarities in different users. A global sentiment classification model is incompetent to distinguish such diverse opinions at an individual user level, and building personalized models is thus of great practical value. Inspired by the social norm and cognitive consistency theories that people tend to follow socially shared norms among their most similar peers about how one should feel and express such feelings, we propose to build personalized sentiment classification models via adapting a global model to individual users in a collaborative manner. The learned sentiment models are shared across neighboring users to conquer data sparsity challenge; and such model-based adaptation enables efficient online learning of personalized models.

Hidden Topic Sentiment Model

Various topic models have been developed for sentiment analysis tasks. But the simple topic-sentiment mixture assumption prohibits them from finding fine-grained dependency between topical aspects and sentiments. In this work, we build a Hidden Topic Sentiment Model (HTSM) to explicitly capture topic coherence and sentiment consistency in an opinionated text document to accurately extract latent aspects and corresponding sentiment polarities. In our solution, 1) topic coherence is achieved by enforcing words in the same sentence to share the same topic assignment and modeling topic transition between successive sentences; 2) sentiment consistency is imposed by constraining topic transitions via tracking sentiment changes; and 3) both topic transition and sentiment transition are guided by a parameterized logistic function based on the linguistic signals directly observable in a document.

RECENT RESEARCH DEVELOPMENTS

- Aspect-based opinion mining system: http://hcdm.cs.virginia.edu:8080/ReviewMiner

RECENT GRANTS

- NSF Faculty Early Career Development Program (CAREER) Award
- Yahoo Academic Career Enhancement Award

SEAS Research Information
Pamela M. Norris,
Executive Associate Dean for Research
University of Virginia
Box 400232
Charlottesville, VA 22903
EngineeringResearch@virginia.edu
434.243.7683

UNIVERSITY of VIRGINIA | SCHOOL of ENGINEERING & APPLIED SCIENCE