Opportunities for Bayesian Network Learning in Personal Informatics Tools

Alex Okeson
University of Washington
Seattle, WA 98195, USA
amokeson@cs.washington.edu

James Fogarty
University of Washington
Seattle, WA 98195, USA
jfogarty@cs.washington.edu

Accepted to CHI'20 Workshop on Artificial Intelligence for HCI: A Modern Approach.

Abstract

As AI becomes an increasingly ubiquitous component of end-user systems, questions of effective design of these systems should be situated in specific contexts. For example, personal informatics is an everyday context in which people encounter challenges in collecting, reflecting on, and learning from data, often mediated by statistical analyses and visualizations. In this workshop position paper, we consider the intersection of AI and *personal informatics*. Specifically, we present preliminary work highlighting opportunities for Bayesian network learning to support people in overcoming common challenges in: (1) answering questions people have for their data, (2) supporting goal evolution and iteration, and (3) learning more with less data.

Author Keywords

Bayesian networks; personal informatics; explainability.

Introduction

Li et al. define personal informatics systems as "*those that help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge*" [5]. This definition highlights the key activities of *collecting* and *reflecting* on personal data. Although many AI systems also aim to help people collect data, find insights, and make decisions, AI has been underutilized in many aspects of personal informatics. Specifically, AI is often featured in lower-level collection components of personal informatics

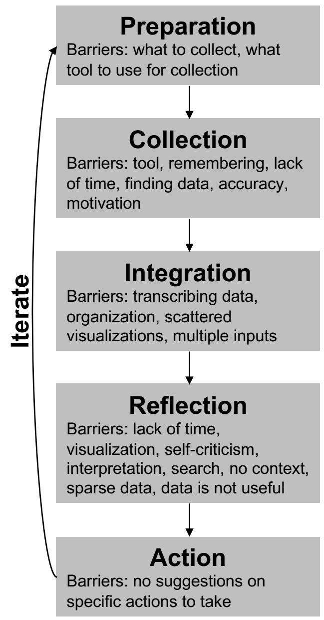

Figure 1: Li et al.'s stage-based model of personal informatics and barriers encountered in each stage.

tools (e.g., to categorize purchases in financial tracking apps, to determine duration and type of physical activity in fitness trackers, to infer quality of sleep in sleep trackers). However, higher-level support for reflection is often limited to simple statistical analyses and visualizations.

Two key reasons that AI systems have not often been used for higher-level reflection support are that: (1) they generally require significantly more data than is available for any single person, and (2) they are typically difficult to interpret, even for relative experts. In our research examining how to better integrate AI in support of personal informatics, we are examining how Bayesian network learning offers a potential solution to these problems because: (1) it can require far less data than other machine learning methods, and (2) it is not a black box method, but instead represents learning in a directly observable output network.

Challenges in Personal Informatics

Li et al. defined a 5-stage model of personal informatics systems that highlights barriers people encounter during each stage (Figure 1) [5]. During the reflection stage barriers include challenges presented by sparse data or data that lacks context, data being a poor fit to a person's question, and difficulty in appropriately interpreting or visualizing data. Epstein et al. expand this model and in the lived informatics model (Figure 2), which additionally highlights lapses in tracking, resumption of tracking, and changing goals over the course of tracking [2].

As part of diagnostic self-tracking, people often turn to self-experimentation to attempt to identify cause-and-effect relationships around health symptoms. However, Choe et al. found that even expert self-trackers in the Quantified Self community fall victim to common pitfalls in self-experimentation, including tracking too many things, not tracking triggers and context, and a lack of scientific rigor [1]. This has motivated tools that scaffold an end-to-end self-experimentation process [3], effectively guiding people through challenges in the 5-stage model. Schroeder et al. further proposed goal-directed self-tracking, emphasizing that personal tracking goals evolve over time and suggesting that personal informatics tools should explicitly support this goal evolution [6].

Opportunities for Bayesian Network Learning

Bayesian network learning (Figure 3) has several promising properties for use in personal informatics tools. First, in contrast to neural networks or even decision trees, Bayesian network learning requires significantly less data to learn relationships between variables. Second, Bayesian networks represent learning in a directly observable output network. Despite these promising properties, personal informatics tools tend to support reflection through frequentist statistics and through visualizations of different cuts of the data. This section considers three concrete opportunities to apply the potential benefits of Bayesian networks to specific challenges in personal informatics.

More Direct Answers to Questions

Schroeder et al. identified 9 classes of question related to self-experimentation (Table 1) and found that self-trackers want to ask these questions over their self-experimentation data [7]. Schroeder et al. then also showed that these questions are more closely answered by Bayesian statistics than by frequentist statistics. Unfortunately, a lack of familiarity with Bayesian methods means that personal informatics tools often provide only simple frequentist analyses or visualizations of raw cuts of collected data, leaving people to attempt to infer answers to their questions. Incorporating Bayesian network learning in personal informatics tools could therefore improve higher-level reflection by better aligning to the questions that people want to ask in their data.

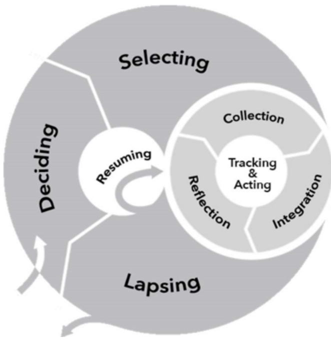

Figure 2: Epstein et al.'s lived informatics model places self-tracking in a larger cycle of deciding, selecting, and lapsing.

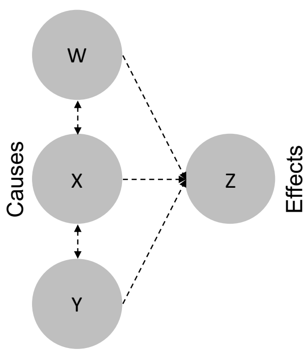

Figure 3: Example of a Bayesian network for self-experimentation. Dashed arrows indicate relationships to be learned and nodes indicate self-tracked causes (e.g., caffeine) and effects (e.g., IBS symptoms).

Support for Goal Evolution and Iteration
Goal evolution commonly occurs as people learn about themselves and ask new questions based on new understanding. The lived informatics model (Figure 2) highlights that such changes in a person's goal for tracking also require revisitation of the *Selecting* stage, in which people decide what tools to use in their tracking. Because Bayesian network learning can support many of the different questions self-trackers may want to ask in their data, personal informatics tools based on Bayesian network learning may be better able to adapt to goal evolution. For example, the frequentist analysis used in TummyTrials [3] is specific to only that self-experimentation goal, but a similar tool based on Bayesian network learning might support a variety of related questions.

Some new goals might even be answered without a need for additional tracking. For example, if a person transitions from one question in Table 1 to another without changing one of the causes (**W**, **X**, or **Y**) and one of the effects (**Z**), it may be possible to use existing data to immediately answer the new question. For example, a person might initially self-track to ask "*does caffeine affect IBS symptoms*" (i.e., tracking the amount of caffeine they consume and resulting symptoms). Upon reflecting and concluding that caffeine intake does appear to impact their symptoms, the person might next ask "*by how much does caffeine affect symptoms*". In this and many similar situations, a Bayesian network learned from the existing data can answer the new question.

Learning More with Less Data
If a new question cannot be answered by existing data, people may still benefit from an ability to reuse data from prior goals and tracking to more quickly learn the answers to related questions. For example, if a person has previously tracked at least one overlapping cause and effect, past data can be used to reduce the number of samples needed to learn the answer to the new question (Figure 3). This is because Bayesian networks can utilize data with missing values, inferring missing data through learned associations between variables.

Tracking "*burnout*" is a common reason people abandon self-tracking without reaching their goals, wherein a person tries to track too many things, becomes overwhelmed, then abandons tracking altogether. Because Bayesian network learning can accommodate many goals and missing data, people may feel less pressure to "*track everything*" for fear of not having the right data to answer their questions. People can instead track exactly and only enough to answer their current question and still have the ability to later reuse that data to answer different, related questions. Similarly, lapsing is also a common component of many self-tracking experiences (Figure 2) that is poorly supported by most personal informatics tools. Lapsing can be either complete (e.g., lapsing in all tracking while on vacation) or partial (e.g., reducing what data is tracked during an already stressful time). Because Bayesian network learning can better accommodate missing data, tools based on a Bayesian network can better support such partial lapsing while still learning relationships in available data.

Explaining Bayesian Network Learning Through Personal Informatics Questions
Although we have emphasized the statistical benefits of Bayesian network learning, frequentist analyses and simple visualizations remain common in part because their relatively simplicity is more straightforward to interpret. In contrast, Bayesian network learning and other AI systems are more difficult to correctly interpret (e.g., even for senior data scientists [4]). Although Bayesian network learning has the advantage that learning is directly observable in the output,

Does **X** have any effect on **Z**?
Does **X** have a noticeable impact on **Z**?
Do different things in combination with **X** affect the change in **Z**?
How does **X** affect **Z** differently depending on the time of day?
How much **X** is needed to see an impact on **Z**?
By how much does **Z** change with different amounts of **X**?
What will **Z** be like in the future if I avoid **X**?
What will **Z** be like in the future after my normal amount of **X**?
What will **Z** be like in the future after more than my normal amount of **X**?

Table 1: Classes of questions for self-experimentation data identified by Schroeder et al.

resulting networks are still difficult to understand. However, a key opportunity is to focus explanation on the specific personal informatics question that people have for a model. Instead of attempting to explain an entire Bayesian network, we are pursuing explanations that correspond to specific forms of questions common in personal informatics.

Workshop Participation

Although Bayesian network learning has many powerful advantages, challenges such as explainability remain a barrier to broader adoption in HCI. In our early and ongoing work, we are examining the potential for Bayesian network learning in personal informatics tools, including in more directly answering common questions, supporting goal evolution and iteration, and learning more with less data. We look forward to discussing both: (1) personal informatics as a promising area for AI in HCI, and (2) how approaches developed in personal informatics might be applied to end-user systems in other areas of AI and HCI.

Acknowledgements

Research was supported by the National Institutes of Health under award R01LM012810. Content is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health.

REFERENCES

[1] Choe et al. 2014. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. 1143–1152. DOI:http://dx.doi.org/10.1145/2556288.2557372

[2] Epstein et al. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. 731–742. DOI: http://dx.doi.org/10.1145/2750858.2804250

[3] Karkar et al. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. 6850–6863. DOI: http://dx.doi.org/10.1145/3025453.3025480

[4] Kaur et al. 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. DOI: http://dx.doi.org/10.1145/3313831.3376219

[5] Li et al. 2010. A Stage-Based Model of Personal Informatics Systems. (2010). DOI: http://dx.doi.org/10.1145/1753326.1753409

[6] Schroeder et al. 2019a. Examining Opportunities for Goal-Directed Self-Tracking to Support Chronic Condition Management. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article Article 151 (Dec. 2019), 26 pages. DOI: http://dx.doi.org/10.1145/3369809

[7] Schroeder et al. 2019b. A Patient-Centered Proposal for Bayesian Analysis of Self-Experiments for Health. *Journal of Healthcare Informatics Research* 3, 1 (2019), 124–155. DOI: http://dx.doi.org/10.1007/s41666-018-0033-x