SPHINX: Rich Insights into Evidence-Hypotheses Relationships via Parameter Space-based Exploration*

Abhishek Mukherji, Jason Whitehouse, Christopher R. Botaish,
Elke A. Rundensteiner and Matthew O. Ward
Dept. of Computer Science, Worcester Polytechnic Institute Worcester, MA, USA
(mukherab|jwhitehouse|cbotaish|rundenst|matt)@wpi.edu

ABSTRACT

We demonstrate our SPHINX system that not only derives but also visualizes evidence-hypotheses relationships on a parameter space of belief and plausibility. SPHINX facilitates the analyst to interactively explore the contribution of different pieces of evidence towards the hypotheses. The key technical contributions of SPHINX include both computational and visual dimensions. The computational contributions cover (a.) flexible computational model selection; and (b.) real-time incremental strength computations. The visual contributions include (a.) sense-making over parameter space; (b.) filtering and abstraction options; (c.) novel visual displays such as evidence glyph and skyline views. Using two real datasets, we will demonstrate that the SPHINX system provides the analysts with rich insights into evidence-hypothesis relationships facilitating the discovery and decision making process.

Categories and Subject Descriptors

H.5.0 [**INFORMATION INTERFACES AND PRESENTATION**]: General

Keywords

Hypothesis testing, evidence gathering, visual analytics

1. INTRODUCTION

Motivation. Confirmatory hypothesis testing is widely used in domains from scientific discoveries in physics, biology and medicine, forensic investigation to weather forecasting [2,3]. Yet, existing hypothesis analysis tools [2–4] have two key limitations.First, these tools are heavily dependent on the domain knowledge and expertise of analysts. Analysts drive all activities including formulating hypotheses, devising experiments, assigning weights to collected evidence and selecting the tests to apply. For large amounts of data observed or generated for such analysis, it becomes virtually impossible to manually inspect all the data to extract evidence supporting or refuting a particular hypothesis. Despite much domain knowledge and ample experience, the analyst may still not be aware of certain relevant evidence hidden within the abundant data.

Existing tools [2–4] also lack support for sense-making of the evidence-hypothesis relationships. Some *static* aspects may be available apriori, e.g., (a.) evidence collected for each hypothesis; and (b.) analyst's estimated belief in some evidence. Yet certain *dynamic* aspects are only learnt during the testing process including (a.) combined effect of multiple evidence on a hypothesis; (b.) correlation among multiple hypotheses. At times, the analysts may give undue importance to some evidence while overlooking other genuine pieces of evidence that may be hidden due to the sheer size or complexity of the datasets and models [3]. Insights, such as how the inclusion or exclusion of an evidence may influence the hypothesis strength computation, or which evidence group dominates the computation, may be particularly useful in complex decision support systems with a large number of hypotheses being tested, with an overwhelmingly large evidence set per hypothesis. Unfortunately, existing hypotheses analysis tools [2–4], that employ *black box* computations, are ill-equipped to handle dynamic hypotheses analysis scenarios as motivated below.

Example Scenario. An insurance analyst performs predictive analytics about potential health risks, e.g., collecting evidence towards frequent infections leading to Pneumonia. While analyzing reported cases of Pneumonia [6] she takes into account four singleton hypothesis, namely, (a.) bacterial; (b.) viral; (c.) fungal; and (d.) parasitic. For example, the analyst may infer that *patient X contracted Pneumonia from* $\{bacteria\}$. Further, Pnuemonia may be contracted due to a composite hypothesis (e.g., {bacteria, viral}). Collection of individual hypotheses forms the *frame of discernment*, denoted by θ = {bacterial, viral, fungal, parasitic}. Figure 1 depicts all possible hypotheses for Pneumonia.

Sources ranging from hospital reports, regional health alerts to past outbreak records can provide valuable evidence for this analysis. Yet another source is social media sites such as Twitter and Facebook [8]. A good number of messages on these sites may convey implicit or explicit health information. Posts such as "*I could not get out of bed today*", "*The pharmacy was out of flu shots*", and "*The doc says I have a stomach bug*" could be used as evidence towards a flu outbreak.

The analyst assigns belief to pieces of evidence of different types such as posts clustered on geolocation, and posts correlating symptoms (e.g., itchy eyes) with a viral infection (e.g., conjunctivitis). The overall likelihood of an outbreak hypothesis is computed by combining the belief values of all evidence confirming (or, disconfirming) the hypothesis. The analyst must tackle real-world ambiguities and noise. For example, the post "*I have itchy eyes*" does not distinguish "itchy eyes in conjunctivitis" from those "due to prolonged use of contact lenses". Support for computationally and visually analyzing multiple evidence in terms of their combined ef-

*Supported by National Science Foundation under grants IIS-0812027, CCF-0811510 and IIS-1117139.

CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.
ACM 978-1-4503-2263-8/13/10.
http://dx.doi.org/10.1145/2505515.2508202.

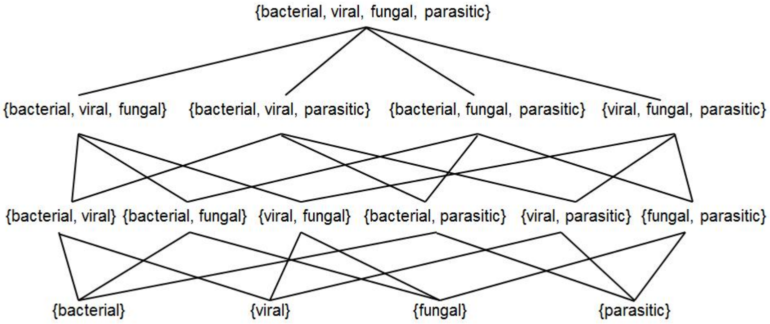

Figure 1: A Lattice of Causes (Hypotheses) of Pneumonia [6].

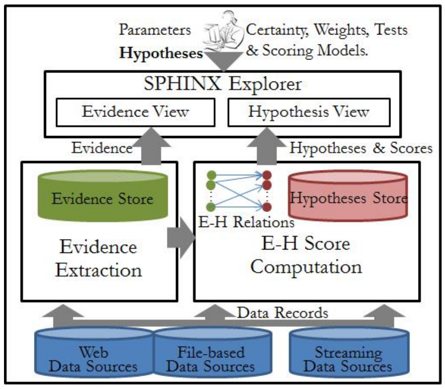

Figure 2: The SPHINX Framework.

fect on the hypotheses scores may provide valuable insights. In addition, features such as filtering, decluttering and evidence visualization are required to deal with a large set of evidence.

Demonstration. In this paper, we present the ***S****calable* ***P****attern and* ***H****ypothes****I****s a****N****alysis using* ***X****mdvtool* (**SPHINX**) system. The open-source XmdvTool [11] software is being developed at WPI for over 15 years with six NSF grants. SPHINX extends XmdvTool to support interactive parameter space exploration of Evidence-Hypotheses (E-H) relationships. It enables analysts to compare and contrast pieces of evidence towards multiple hypothesis and analyze their individual as well as combined effect on the hypotheses scores. We tackle both **computational** and **visual** challenges. The **computational** contributions cover (a.) flexible computational model selection; and (b.) real-time incremental strength computations. The **visual** contributions include (a.) sense-making over parameter space; (b.) multiple filtering and abstraction options; (c.) novel evidence glyph and skyline visual displays. We demonstrate SPHINX using real datasets emotions [7] and iris [9].

2. RELATED WORK

[4] proposes traditional hypothesis testing methods. [1, 2] emphasize on the evidence aggregation process to compute strength of a hypothesis using models such as Certainty Factor (CF) and Dempster-Shafer (DS) [2]. [3] developed an automated hypothesis generation and analysis system. While, these techniques provide only blackbox computation of strength scores; our SPHINX system, instead, provides an interactive parameter space view for rich insights into Evidence-Hypotheses relationships.

3. THE SPHINX FRAMEWORK

Figure 2 depicts the **SPHINX** Framework. **SPHINX Explorer** supports user interactions via **Hypotheses** and **Evidence** views. The **Hypothesis View** enables analysts to visually navigate through the evidence groups supporting different hypotheses distributed on the belief-plausibility parameter space (or the *bp-space*). The **Evidence View** displays the details of the selected evidence groups. The **Evidence Extraction** module extracts evidence from disparate sources, disambiguates them and stores them in the evidence store. The **Evidence-Hypothesis (E-H) Score Computation** module handles on-the-fly recomputation ofscores based on user selections.

The **E-H Score Computation module** establishes and maintains the evidence-hypotheses relationships. SPHINX uses well-defined measures such as *belief* and *plausibility* [1, 2]. Strength of a hypothesis is represented as a combination of the belief values of the evidence confirming or disconfirming it. The chosen model (e.g., CF or DS) combines two pieces of evidence based on their relationship (see [1,2] for details). The Evidence-Hypotheses relationships can be visually explored on the **SPHINX Explorer**.

4. SPHINX SYSTEM: KEY INNOVATIONS

SPHINX encompasses several computational and visual innovations to enable interactive exploration of E-H relationships.

4.1 Computational Contributions

The key computational contributions of SPHINX are as follows.

Flexible Computational Model Selection. Analysts may prefer to use alternate computational models that best fit their data and domain. Therefore, SPHINX allows flexibility to plug in any computational model for aggregating beliefs of evidence to compute hypotheses scores including Certainty Factor (CF) and Dempster-Shafer (DS) models [1, 2]. Further, in addition to belief and plausibility other measures can also be used for the parameter space instead of or in conjunction with our chosen measures.

Incremental Strength Computations. Rapid responsiveness is a key requirement for interactive systems. When analysts filter out evidence, recomputing hypothesis scores from scratch may have unacceptably high response times. We thus devise an incremental strength computation algorithm such that the response times are within acceptable time range even for huge numbers of evidence and hypotheses. The incremental algorithm uses the insight that if, for a hypothesis, out of 100 pieces of evidence, 10 are filtered out; instead of recomputing the new score from scratch using the 90 pieces, the old score (previously computed using 100 pieces) can be adjusted by considering the 10 filtered pieces with the opposite effect. Thus, previously confirming (disconfirming) filtered pieces are now accounted as disconfirming (confirming).

4.2 Visual Contributions

The SPHINX Explorer has the following contributions.

Sense-making over Parameter Space. SPHINX enables the analysts to interactively investigate the *bp-space* via a rich set of novel exploratory interactions. Different hypotheses are depicted with distinct colors, whereas the evidence types are shown on the bp-space in different shapes. In Fig. 3 *green triangles* denote evidence of type *eye aperture* (*triangle*) towards hypothized emotion *surprise* (*green*).

The supported interactions provide insights at two granularities, namely, *Hypothesis* and *Evidence* views (Fig. 3). For example, one interesting pattern could the analysts can observe are the distributions of evidence within the bp-space by hypothesis for different datasets. In Fig. 3 the overall scores of two hypotheses H_1 (e.g., *fear*) and H_2 (e.g., *anger*) are similar, yet in reality *fear* has several medium to low belief evidence and *anger* has fewer yet with high belief and plausibility evidence. The visual display helps the audience to visualize such subtle differences between such hypotheses. Similar insights on contributions of different evidence types are useful. For example, SPHINX discovers the significance of *petal width* in distinguishing different iris species.

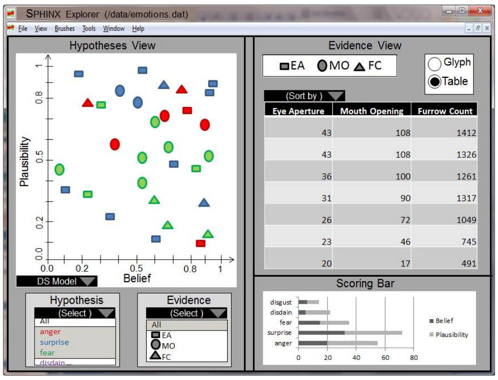

Figure 3: SPHINX Explorer (emotions [7]).

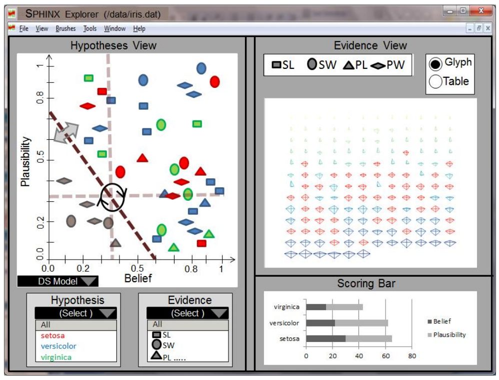

Figure 4: The bp-Slider (iris [9]).

Filter and Abstraction Options. Huge number of evidence pieces may clutter the bp-space and hinder sense-making. SPHINX provides multiple filter and zoom options based on hypotheses, evidence types, belief and plausibility. We also cluster *co-located evidence groups* into rectangular boxes to avoid cluttered view similar to our prior work [5]. Analysts can select certain evidence groups and explore their details in the tabular *Evidence* view (Fig. 3-RHS).

Evidence Glyph View. Beyond the straightforward tabular view (Fig. 3-RHS), we designed a novel *evidence glyph view* (Fig. 4-RHS) to facilitate efficient visual analysis of evidence sets. Glyphs are effective visualizations for shape comparisons as well as finding clusters or outliers by applying glyph placement strategies. A comprehensive taxonomy of glyph placement strategies has been studied in [10]. However, we are the first to employ glyph visualizations for evidence groups and explore placement strategies (Fig. 5) in the context of our proposed evidence glyph view. The glyphs are also color coded to match the corresponding hypotheses.

Evidence Skyline View. In Fig. 6, consider evidence E_1 and E_2. While E_1.plausibility $>$ E_2.plausibility, E_1.belief $<$ E_2.belief. Thus, neither of them fully dominates the other on both measures. Such pareto-optimal pieces of evidence can be displayed to analysts using the skyline view for them to explore.

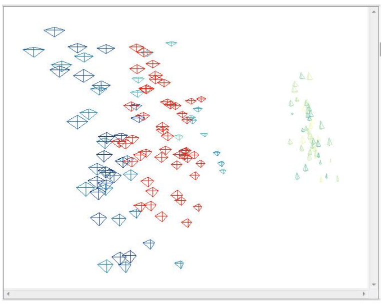

Figure 5: PCA-based Glyph Placement (iris data).

5. SPHINX DEMONSTRATION

The key innovations of SPHINX will be demonstrated using two popular real datasets emotions [7] (Fig. 3) and iris [9] (Fig. 4).

Hypothesis View. In the hypothesis view (LHS of Fig. 3) the pieces of evidence are plotted on the two-dimensional *bp-space*. The hypothesis scores are displayed on the *scoring bar* (Fig. 3-RHS). As described in Sec. 4.2, interesting patterns of the E-H relationships can be observed using different colors and shapes.

Using the hypotheses drop down menu, the audience may select a subset of hypotheses by control+click such that only the evidence corresponding to the selected hypotheses are displayed. In Fig. 3, only the emotions *anger*, *surprise* and *fear* are chosen. Further, the

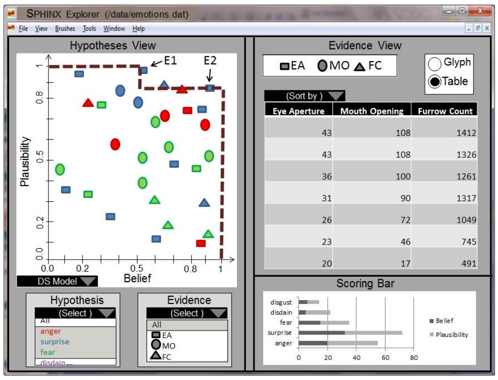

Figure 6: Evidence Skyline View.

evidence drop down menu is used to filter the evidence type(s) analysts want to view. These filtering features are useful in cases of overcrowding of evidence on the bp-space. We will also demonstrate the *evidence skyline view* as described in Section 4.2.

Evidence View. The RHS of Fig. 3 depicts the evidence view that lists the details of evidence valid within the selected hypothesis and evidence types via cross links between the views. The evidence can be viewed either in a table (Fig. 3) or as glyphs [10] (Fig. 4). The glyphs can further be clustered using layout strategies (Fig. 5).

Belief-Plausibility Slider. Fig. 4 depicts the bp-slider. The bp-slider is adjusted to filter out evidence (shown in grey) and the *scoring bar* is incrementally refreshed with the new scores.

Conclusion: We demonstrate our SPHINX system that provides rich insights into the E-H relationships.

6. REFERENCES

[1] Shafer: Belief functions. `www.glennshafer.com/assets/downloads/rur_chapter7.pdf`, June 2013.

[2] J. Gordon and E. H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artif. Intell.*, 26(3):323–357, 1985.

[3] G. Liu, M. Feng, Y. Wang, L. Wong, et al. Towards exploratory hypothesis testing and analysis. In *ICDE*, pages 745–756, 2011.

[4] `graphpad.com/www/book/choose.htm`, January 2013.

[5] A. Mukherji, X. Lin, C. R. Botaish, J. Whitehouse, E. A. Rundensteiner, M. O. Ward, and C. Ruiz. Paras: interactive parameter space exploration for association rule mining. In *SIGMOD*, pages 1017–1020, June 2013.

[6] Pneumonia wiki. `en.wikipedia.org/wiki/Pneumonia`, June 2013.

[7] N. Sebe, M. S. Lew, I. Cohen, A. Garg, and T. S. Huang. Emotion recognition using a cauchy naive bayes classifier. In *ICPR*, 2002.

[8] A. Thusoo, Z. Shao, S. Anthony, D. Borthakur, N. Jain, J. S. Sarma, R. Murthy, and H. Liu. Data warehousing and analytics infrastructure at facebook. In *SIGMOD*, pages 1013–1020, 2010.

[9] UCI machine learning repository, august 2013. `http://www.ics.uci.edu/~mlearn/MLRepository.html`.

[10] M. O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4):194–210, 2002.

[11] Xmdvtool. `http://davis.wpi.edu/~xmdv/`, June 2013.