In Papers from the AAAI Fall Symposium - Capturing and Using Patterns for Evidence Detection.
October 13–15, 2006, Washington D.C., USA.

A Position Paper: Value of Information for Evidence Detection

David L. Roberts*
College of Computing
Georgia Institute of Technology
Atlanta, GA
robertsd@cc.gatech.edu

Tina Eliassi-Rad
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
P.O. Box 808, L-560, Livermore, CA 94551
eliassi@llnl.gov

Abstract

In real-world applications, evidence detection involves evaluating a body of existing information from time-evolving multi-modal data sources. It seems obvious that approaches to evidence detection should consider the relative quality of data sources with respect to the *value of information* being produced over time. For instance, considering all data sources equally reliable can yield undesirable results. We highlight the distinction between the traditional value of information problem where information is *pulled from sources* and the value of information for evidence detection problem where information is *pushed from the sources*. We further comment on how this distinction enables new qualities of information to be measured and characterized. In this paper, we address the following questions: What should value of information mean for evidence detection? What are the components needed to characterize value of information? How should these components be measured and combined to compute a value for information? Finally, how should value of information be used in evidence detection? We develop a framework for implementing value of information for evidence detection and present the results of a preliminary feasibility study.

Introduction and Background

The sheer size and complexity of data sets in real-world applications have prompted many efforts on management and analysis of large complex networks. These networks store multi-source data with multi-modal and multi-relational properties in a human-understandable way. One of the associated challenges is to find evidence in support of or against a particular hypothesis. It is common to try to detect patterns in the data and use these patterns to perform inference or update beliefs in a set of hypotheses.

The published literature is full of algorithms for pattern matching, pattern mining, link analysis, and many others on such networks (Gallagher 2006; Getoor & Diehl 2005; Washio & Motoda 2003). A logical next step is to take these algorithms further and use their patterns/results as evidence. In this paper, we argue that any approach to evidence detection should consider the *value of information* (VOI) for the output of the data sources. Without such measurement, all sources and their data are considered "equal." This assumption is clearly false. For instance, using patterns that were generated from an unreliable data source is clearly not desirable. Our proposed VOI framework enables evaluation of detected patterns based on their qualities.

Throughout the remainder of this paper, we highlight related work from a number of fields. We provide an idealized definition of VOI as well as a technique for estimating it. We conclude by presenting some of the results obtained in our preliminary feasibility study and discussing the future directions for continuing this research.

The problem of determining VOI is well-studied in various fields dating back to the 1940s. All existing approaches solve a variant of the following problem. Given a set of sources, which is the best (or best set) to obtain an observation (or a set of observations) from? In other words, an agent must determine the optimal "activation schedule" for the sources of information to maximize (or minimize) some objective. There are a number of approaches to solving this problem, based mostly on decision theory and/or information theory. However, measuring VOI for evidence detection differs from previous work in several ways.

First, prior work typically makes inherent assumptions about reliability of information (Horvitz & Rutledge 1991; McCarthy 1956; Shannon 1948). In particular, the traditional approaches of information and decision theory assume sources to be fully reliable. Data for evidence detection is often not fully reliable or even relevant. It originates from multi-modal sources – each with varying characteristics that can change as the world evolves.

Second, previous work typically characterizes the value of querying an information source which is a *pull* problem (as data is pulled from sources) (Heckerman, Horvitz, & Middleton 1993; Zhang, Ji, & Looney 2002). For the purpose of evidence detection, we are interested in understanding how to interpret data that we have already obtained or has been *pushed* to us. To understand this distinction, consider the following situation. You are buying a new car from a company that is known to produce a high quality product. Unfortunately, the company has redesigned the car for this year and you do not know if it is up to the usual standards. There are two ways you can proceed. One is to proceed under the assumption that the company's reputation is sufficient and they will likely not produce a bad car. Alternatively, you

*Work performed at Lawrence Livermore National Laboratory.

can test drive the car and determine if the product is consistent with what you know about the company. The first case, when you make the decision solely based on reputation, is an example of a pull problem. The second case, when you decide after test driving the vehicle, is an example of a push problem. The key distinction is whether the value of what the source (or company) produces is determined prior to (pull) or after (push) inspection of the product (or information). Our position is that in detecting evidence within a body of existing information, push methods are what need to be utilized. To use the methods developed for the pull problem, where inspection of the current information does not occur when evaluating quality or value, imposes an artificial handicap on the detection of evidence. The pull approach is warranted when there is a cost for obtaining information as is typically assumed in decision theory. Since evidence is available without cost, we can exploit the opportunity to inspect the current information and develop a measure for VOI more attune to shifts in information quality.

Third, we found little work that attempted to learn VOI and/or its components across multiple sources over time. Generally speaking, the majority of work in this area comes from the Information Fusion community (Rogova & Nimier 2004). The goal in information fusion is to combine multiple sources of information into one coherent representation. Often, the pre-fusion information is missing values, pertains to disjoint concepts, or may be unreliable. All of these as well as other properties of the pre-fusion information must be taken into account when designing a fusion operator.

Moreover existing approaches in the fusion community, such as Delmotte, Dubois, and Borne's (1996), generally do not involve learning to characterize the quality of data. In this work, it is noted that the quality of knowledge produced by fusion is influenced by adequacy of the data, quality of the uncertainty model, and quality of the prior knowledge. Much of this work has focused on the improvement of an uncertainty model and has completely ignored the reliability of information. When it is considered, two measures of reliability are discussed: 1) The relative stability of the first order uncertainty; 2) The accuracy of the beliefs. It is assumed that the fusion operator will not introduce any residual uncertainty that is not due to the data itself. A fair amount of research has been devoted to the incorporation of reliability into fusion rules. In this research, the reliability measure comes in one of three forms: 1) It is encoded by external sources (*e.g.* context or an expert); 2) It is learned using training data; 3) It is constructed based on agreement of sources or consensus (Delmotte, Dubois, & Borne 1996; Parra-Loera, Thompson, & Salvi 1991). It is not, however, estimated prior to fusion. Historically, consensus models of reliability have taken one of the following two forms: 1) A degree of deviation between measurements of each source and the fusion result (*e.g.* posterior belief); 2) A measure of "inner trust" based on a pairwise degree of "likeliness" of agreement (or consensus) between sources. While this work is interesting and in some cases can improve inference performance, there is a notable problem with consensus measures of reliability. Specifically, lack of consensus is sufficient for low reliability but not necessary. Further, for VOI to

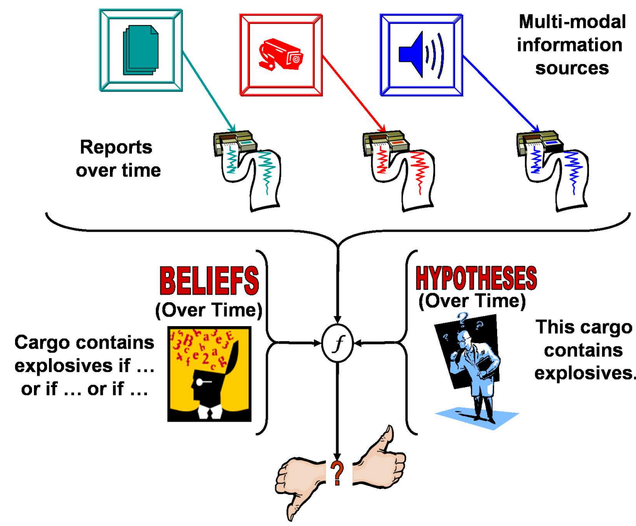


Figure 1: VOI Framework for Evidence Detection

be useful for evidence detection, it must be computed prior to fusion–and not as a function of its output.

Measuring VOI for evidence enables a solution to the problem of detecting information (*i.e.* evidence) with the highest *qualitative* value to either confirm a true hypothesis or disconfirm a false hypothesis.[1] VOI for evidence detection should: 1) capture information from possibly unreliable sources; 2) characterize the value of a body of existing information (a push problem); and 3) be tailored to improve inference and allow data triage.[2]

Figure 1 depicts our framework for VOI in evidence detection. At the top of the Figure there are a set of sources that are pushing reports (or information). The different icons in the sources are intended to represent the multi-modal data. The reports they generate evolve over time as the environment evolves. In the center of the Figure, there is an *f* inside a circle. To the left is a representation of a set of *beliefs* (*i.e.* prior knowledge or knowledge obtained earlier) that are evolving over time and to the right are the set of *hypotheses* that are (possibly) evolving over time. The *f* represents the combination of information, beliefs, and hypothesis into a quality metric. Note that this is not the same as fusion. In the fusion framework, the information from the sources would be combined to produce "manipulated" data (we return to this concept later). In our VOI framework, the information from the sources is combined to produce a measure of quality. This measure of quality can be used to detect evidence, inform inference, or even inform fusion. This quality (or VOI) measure is represented by the thumbs-up-thumbs-down icon at the bottom of the Figure and generally lies somewhere in between the extremes of useless and useful (*i.e.* it is not a binary measure of goodness).

We seek a learning algorithm (such as regression) to esti-

[1]For a discussion on the degree to which a hypothesis is confirmed or disconfirmed, see (Fitelson 2001b).

[2]Note that even with anonymized data, one needs to evaluate the source.

mate VOI based on data characteristics that can be used to inform inference. For example, if we are using probability as our model of uncertainty, the estimate of VOI can be interpreted as a likelihood factor that is used for belief update.

To recap, by removing the data-equality assumption, information can be judged based on its quality. This should improve our ability to detect evidence, test hypotheses, and perform data triage (Kent 1994; Schrage 2005). In the next section, we address the following questions: What should value of information mean for evidence detection? What are the basic parameters needed to characterize value of information? How should these parameters be measured and combined to compute a value for information? Finally, how should value of information be used in evidence detection?

VOI for Evidence Detection

Definition of VOI

To measure value of information, we first need to define what constitutes information. A piece of information is the smallest amount of data needed to update the probability distribution of a hypothesis within a database of hypotheses. Naturally, information is conditional on a chosen domain topic (e.g. Avian influenza) or a selected taxonomy (e.g. viral infections). This conditionality manifests itself in hypotheses and beliefs.

To reiterate, for evidence detection VOI should: 1) provide an accurate quality assessment; 2) improve performance of inference procedures; and 3) enable the triage of large amounts of information. The existing definitions of VOI from information theory and decision theory do not satisfy these requirements. In information theory, VOI is a measure of uncertainty reduction (entropy). In decision theory, VOI is a measure of influence over a decision or choice of actions (value for taking an action). In either case, however, the measures are traditionally defined to characterize the expected value of *querying* an information source and not to provide an interpretation of the information the source has *already produced*.

In contrast, for evidence detection, we define VOI to be a measure of the potential increase/decrease in believing a true/false hypothesis. Specifically, VOI for evidence detection is a measure of the ability for a body of information to *increase belief in a true hypothesis* (or set of hypotheses) or, conversely, *decrease belief in a false hypothesis* (or set of hypotheses). In the absence of an oracle that can inform us of the truth of hypotheses, we need algorithms to approximate VOI over time. Note that information in support of facts will have no value (since facts are hypotheses whose truth/falsehood is known with absolute certainty).

Two Notions of Temporal Tracking

There are two ways time is considered when it comes to VOI for evidence detection. The first notion of time computes VOI depending on when evidence is received—penalizing information that is saturated and/or arrives too late. For example, today's news report about a potential attack that happened yesterday is useless. This scenario is sometimes called the "overcome by events" phenomenon. Ideally, a "timely" piece of information dramatically reduces the gap between beliefs and ground truth.

The second (and arguably more important) way in which time is considered when it comes to VOI for evidence detection is one way in which our approach is set apart from existing work in VOI. Specifically, we use time to distinguish between a *process* across a range of multiple time steps and an *event* at a single point in time. When dealing with evidence, we have the ability to both consider the total (or partial) history of information produced by a source as well as the information it has produced most recently. This allows us to characterize the expected value of a source of information over time in a way similar to the information and decision theoretic measures where the value is based on the distribution of information to be produced by the source. Additionally, we can also characterize the quality of the information that has been produced by the source at the current time step without the need for an expert to encode contextual knowledge about the reliability and/or validity of the source. For example, even completely reliable news sources have to print a retraction on rare occasions. If we look beyond simply characterizing the reliability of the source and make use of our existing beliefs, then we may be able to tell ahead of time that information from a reliable source is potentially faulty and may later be retracted. Using our VOI for an event, the information that will eventually be retracted can hopefully be identified early, before it may be a detriment to inference. On the other hand, information and decision theoretic measures do not characterize the quality of the report, only the source, thus preventing the detection of mal-information prior to inference.

Components for Estimating VOI

We have identified three separate influences that we feel are responsible for how information is interpreted and that we will use as component measures to estimate VOI. They are *reliability* R, *independence* I, and *coherence* C.[3] Each of these can be measured in a variety of ways.

We first define some notation: $S = \{S_1, \ldots, S_n\}$ is a set of information sources. $T = \{t_x, \ldots, t_y\}$ is a time interval and t_i represents a particular point in time. $H = \{H_1, \ldots, H_n\}$ represents a set of hypotheses. $B = \{B_1, \ldots, B_n\}$ is a set of beliefs. Lastly, E represents evidence or the information produced by a source or set of sources at the given time step or time interval. This can be a single value (in the case of one source at one point in time) or as complex as a set of sets (in the case of a set of sources over a time interval) or anywhere in between. Figure 2 depicts the taxonomy to which our basic components for estimating VOI belong. As mentioned above, each of these components can be measured in a variety of ways. The categories are based on how memory/time (process vs. event) and beliefs/context (objective vs. subjective) are considered during the computation of the component measures. These

[3] Some might argue that there should be a fourth component measure—*relevance*. For example, measuring height is not relevant for inferring intelligence but measuring IQ is. For our purposes, we assume all information is at least relevant.

Measurements on Sources, S, over Time Interval, T				
Context →	Objective		Subjective	
Memory →	Process	Event	Process	Event
Reliability, R	$R(S_i \mid H, T)$	$R(S_i \mid H, t_k)$	$R(S_i \mid H, B, T)$	$R(S_i \mid H, B, t_k)$
Coherence, C	$C(S_i, S_j \mid H, T)$	$C(S_i, S_j \mid H, t_k)$	$C(S_i, S_j \mid H, B, T)$	$C(S_i, S_j \mid H, B, t_k)$
Independence, I	$I(S_i, S_j \mid H, T)$	$I(S_i, S_j \mid H, t_k)$	$I(S_i, S_j \mid H, B, T)$	$I(S_i, S_j \mid H, B, t_k)$

Figure 2: A Taxonomy of Basic Components for VOI

component measures are described in slightly more detail below. In some cases, we provide an illustrative example of how the measure may be computed.

We add a few comments on objective and subjective measures. The process of calculating a component measure of VOI (or VOI itself) conditioned on a belief set is known as a subjective measure. If the computation is performed without a belief set, it is known as an objective measure. This is reflected in Figure 2 where B appears in the subjective measures but not in the objective measures of the taxonomy. In either case, all computations are made with respect to a set of hypotheses (see Figure 2). This conditioning on hypotheses is to provide an appropriate frame of reference (*e.g.* the value of knowing that a car is blue is zero with respect to the hypothesis that the car is of high quality).

Reliability Reliability is intended to capture how frequently a source of information agrees with ground truth. In the context of news sources, reliability is interpreted as a measure of how frequently a particular news source reports information that turns out to be true. To assess the reliability of a source, we can consider statistical models that use the rate at which the source produces true positive or false positive information. Alternatively, in the absence of true/false positive rates, one approach is to estimate reliability using a function such as the following (Cronbach 1947):

$$R(S_i|H,B,T) = \frac{\sigma^2(B|H)}{\sigma^2(S_i|H)}$$

where σ^2 denotes variance. This is a measure of *subjective process reliability*.

Coherence Coherence captures agreement among sources. The interpretation of coherence among news sources is a measure of how frequently they agree. For example, take two conservative talk show hosts: Pat Robertson and Rush Limbaugh. As they are motivated by current events and politics, they frequently discuss many of the same topics on their shows. They are clearly independent as they are both free thinkers. However, they agree on many things, which makes them coherent sources of information. An example of an *objective event coherence* measure is (Olson 2002):

$$C(S_i, S_j|H, t_k) = \frac{P(S_i \cap S_j|H, t_k)}{P(S_i \cup S_j|H, t_k)}$$

Independence Independence captures the causal dependence between sources (rather than simple agreement). It has a very intuitive interpretation. For example, articles from the New York Times and Al Jazeera that both report the same thing are likely to be independent but two articles from the Associated Press on the same topic are likely not independent. In the context of the probability model of uncertainty, given a belief set, independence is a conditional measure that captures the degree of autonomy between sources. The standard way to measure independence is to use rules from probability such as the following *subjective event independence* measure:

$$\begin{aligned} I(S_i, S_j|H,B,t_k) &= \frac{P(S_i, S_j|H,B,t_k)}{P(S_i|H,B,t_k) \times P(S_j|H,B,t_k)} \\ &= \frac{P(S_i|S_j,H,B,t_k)}{P(S_i|H,B,t_k)} \end{aligned}$$

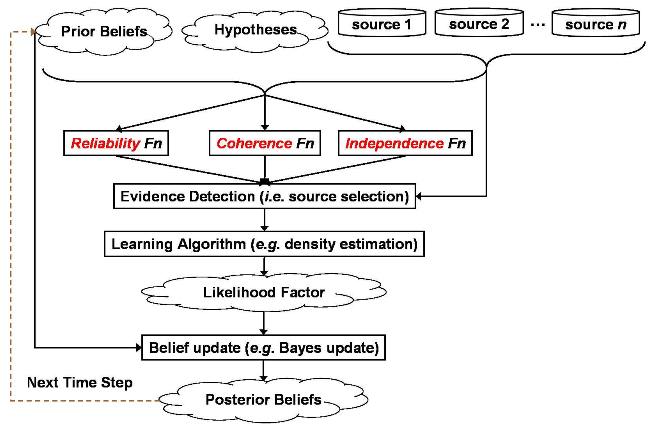


Figure 3: Flow Chart for Combining Components of VOI in the Context of Probability Theory

Learning VOI and its Uses

Our approach to VOI is to combine the aforementioned components using some (possibly learned) function. Figure 3 is a flow chart showing an abstract representation of how we may use these combined components in a probabilistic setting. At the top, prior beliefs, hypotheses, and a set of sources are represented. Each of these are used as input into individual reliability, coherence, and independence functions as described in the taxonomy in Figure 2. These components are combined to form an estimated VOI that is used to inform an inference or learning procedure such as density estimation. More specifically, VOI can be used to select the strongest evidence (or subset of information produced by the sources) at each time step to use as input to the learning or inference procedure. The output of the learning or inference procedure provides a likelihood factor that is used to update prior beliefs. The resulting posterior beliefs for the time step depicted in the figure then form the prior beliefs for the next time step (as depicted by the dashed line from the bottom of the figure to the top).

To learn VOI, we can use all available sources (passive selection) or select a subset of sources (active selection) (Liao & Ji 2006; Zhang & Ji 2006). For active fusion, we can use VOI estimates based on the previous time step (or steps) to help us select the most appropriate sources.

Lastly, we can also learn separate functions for reliability, coherence, and independence through various supervised learning approaches. This would be in contrast to using a predefined function for R, C, and I such as one of the examples given above.

To the best of our knowledge, no one has tried to develop a formal model that computes/learns VOI using a single function that captures reliability, coherence, and independence over time given a hypothesis and a belief set. The most relevant work has been in philosophy of science (Bovens & Hartmann 2001; 2003; Fitelson 2001a; 2001b). Bovens and Hartmann (2001) consider reliability and contextual fit[4] within the framework of a Bayesian network but sidestep the problem of independence between sources. In addition, Bovens and Hartmann (2003) address the issue of variety of evidence in support of a given hypothesis. This is different from determining whether two sources are independent. Fitelson (2001a) provides a Bayesian account for measuring independence across various sources. However, Fitelson does not consider reliability or coherence. Work on reliability has mostly been in the context of information fusion. Noble (2004) provides an overview of the literature on the reliability of open source information. However, he does not formally address independence or coherence.

Biases Introduced Through Manipulation

Real-world data comes at various levels of manipulation from multiple modalities of sources. The five most commonly used levels are: 1) Raw (data is collected at the source without any manipulation); 2) Calibrated (data is reduced through the process of feature selection); 3) Interpreted (data from the calibrated level is "interpreted" by a domain expert where the feature selection is more semantically oriented); 4) Extracted (data from the interpreted level is put in context with previously extracted data); 5) Exploited (data is converted into an "action report," where decisions are made). As data gets manipulated from one level to the next, biases get introduced. These biases must be accounted for. Evaluating information in data permits one way to measure such biases by highlighting the difference between values of information from different levels.

Preliminary Experiments

We ran some preliminary tests on a simplified model of macroeconomics to illustrate the feasibility of our ideas. The task we examined was characterizing the economy based on a number of indicators. We selected the indicators based loosely on a model proposed by Sondhauss and Weihs (1999). They are: 1) yearly growth rate of real GNP; 2) yearly growth rate of real private consumption; 3) government deficit as percent of GNP; 4) yearly growth rate of wage and salary earners; 5) net exports as percent of GNP; 6) yearly growth rate of M1 money supply; 7) yearly growth rate of real investment in equipment; 8) yearly growth rate of real investment in construction; 9) yearly growth rate of unit labor cost; 10) yearly growth rate of GNP price deflater; 11) rate of nominal short term interest rate; 12) real long term interest rate; The hypothesis we tested was: "The economy is currently growing." Data giving values for these indicators was collected quarterly from the first quarter of 1947 to the first quarter of 2006 and in each quarter it was noted whether or not there was economic growth. There were a number of missing values from some of the indicators. There are many challenges to using a data set such as this one. Particularly, the state of the world (economic growth or recession) changes over time. Further, the data's predictive power changes in response to the state of the world changing (*e.g.* leading vs. lagging indicators). In addition, because the data comes from a number of different government agencies as well as independent research firms, there can often be conflicts and/or inconsistencies. Further, there is a data sparsity issue as well. This is exactly the type of scenario where we expect to see the most benefit from using a VOI approach.

For this preliminary feasibility study we explored four concrete measures of VOI and compared the results. At each time step, we used VOI to select the strongest evidence from a fixed given number of sources and used it as the basis for inference and learning. We assumed the highest valued sources were producing the strongest evidence. In particular, we tested using mutual information between the source and the class label (Shannon 1948), objective event coherence, and objective process coherence as our estimates of VOI as well as using random source selection. As this is a position paper and we are only reporting on our preliminary feasibility study, we chose to omit independence and reliability.

The first 10 years of data were used to estimate the joint distribution $P(H, E)$ where E represents all possible values of each of the indicators. Then, at each time step, the next quarter was presented as a new instance. If the learned distribution was able to supply a probability estimate (*e.g.* $P(H|e) + P(\bar{H}|e) = 1.0$ where e is the instantiation of evidence from the current time step)[5], then it was output and later compared to ground truth to quantify error. Otherwise, the distribution was updated accordingly and we moved on to the next time step. We were interested in examining both the error rate and the number of estimates of $P(H|E)$ that were output during this process.

To illustrate one of the key challenges to a push problem, in Figure 4 we present the results of using objective source coherence as the criterion to select i sources for $i = 1, \ldots, 4$ sources. The important thing to notice in this plot is that when only one or two sources are selected for inference, there is a very smooth increase or decrease in belief that coincides nicely with the rising or falling of the economy. As we select three and four sources, we introduce the possibility of conflicting or noisy evidence and the smoothness of the belief curve erodes and becomes more erratic. This degradation illustrates the difficulty of a push problem where inference can be handicapped by information overload or poor

[4]Contextual fit measures the coherence of a piece of information with a given belief set.

[5]The learned distribution may be unable to estimate a probability if no similar combination of attribute values have been encountered. This is especially an issue in cases of relatively small high-dimensional data sets such as the economic indicators set.

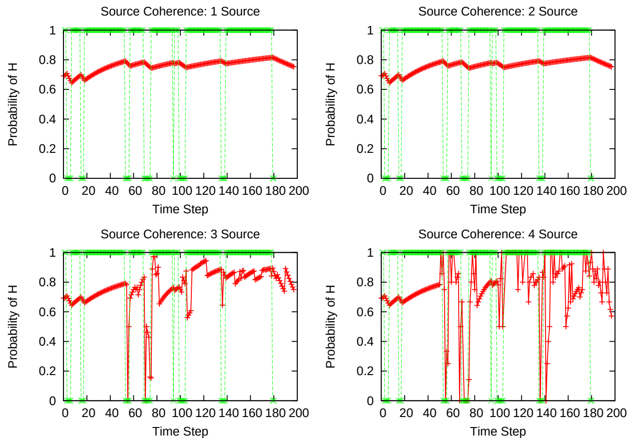


Figure 4: Belief Over Time Using Source Coherence to Select 1, 2, 3, and 4 Sources

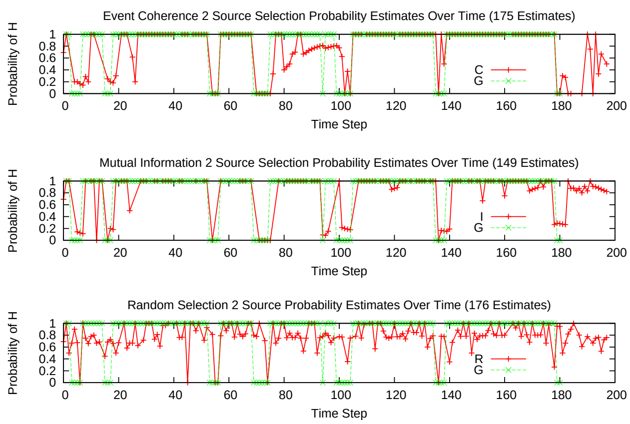
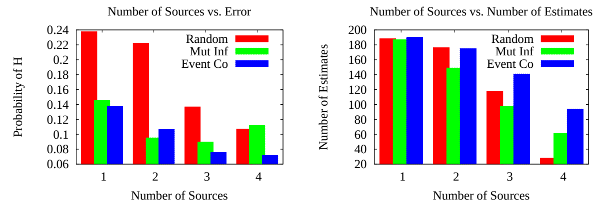


Figure 5: Belief Over Time Using 2 Sources Selected Randomly, Using Mutual Information, or Objective Event Coherence

quality information.

Representative results for this simplified macroeconomic model appear in Figures 5 and 6. Looking at Figure 5 where the hypothesis estimates over time are compared to ground truth for using three of the selection methods to select from two sources (the fourth appears in Figure 4, top right), we see that random selection is very erratic. More interestingly, we see that mutual information selection and event coherence selection perform similarly in a qualitative sense. These plots are a detailed representation of the summary statistics presented in column 2 of Figures 6(a) & 6(b).

What we begin to see when we look deeper into the details is that, with the exception of the four source selection case, mutual information and event coherence result in comparable error rates (Figure 6(a)). This is encouraging. However, the full picture is not revealed until we look at the number of estimates output (Figure 6(b)). In all cases, with the exception of selecting just one source, the number of estimates made when using coherence selection is notably higher than



(a) Error (b) Estimates

Figure 6: Performance of Random, Mutual Information, and Event Coherence Selection in Terms of Error Rate and Number of Estimates

the cases with mutual information based selection. We are hopeful that with further research and the incorporation of independence and reliability we can further decrease error and increase the number of predictions.

In addition to these general trends, there were a few interesting anomalies. The first has to deal with the relative error rate in prediction between mutual information selection and event coherence selection. Specifically, when two sources are selected, the error rate associated with objective event coherence selection is relatively higher than with mutual information whereas in all other cases it was the other way around. We have no good explanation for this. Additionally, in the four source selection case, the error rate for mutual information is higher than for the two and three source selection case (which further indicates the difficulty of the push problem).

In general, the trend we identified in most cases was a monotonic decrease in average error followed by a monotonic increase in average error as the number of sources selected increased. Figure 6(a) mostly shows the monotonic decrease in error. The trend is not surprising since we would expect that at first increasing the number of sources would lead to less average error while adding too many sources would eventually lead to increased average error (similar to overfitting in learning). Unfortunately, as the number of sources selected increased, the number of predictions output decreased too rapidly to draw a meaningful conclusion. While we do not present a formal explanation for our observation, the intuition seems to point to the nature of using process measures for the selection criteria. That is, when using a process measure, it is likely that as time goes on there will be one combination of sources that is consistently valued highest. Unless a sliding window is used for the calculation of the process measures, any short term degradation in the quality of information produced by those highly ranked sources will not get discovered. In contrast, an event based measure will be able to detect these degradations and suitably adjust the source values.

In the future, it could prove interesting to see how the set of selected sources evolves over time in response to these changes in quality. For example, in the top-most plot of Figure 5, somewhere just before the 80th time step, the certainty of prediction becomes somewhat erratic. This is a phenomenon that was discovered in almost all trials using objective event and process coherence as well as in some of the mutual information trials. While we do not have any research to support this, our intuition indicates there was a change in the way the statistics were reported that probably caused a change in the distribution of the classes with respect to the evidence. In the data mining community, this type of change is often referred to as concept drift.

Conclusion and Future Work

In this paper, we have argued for the importance of considering a qualitative value of information content for evidence detection. We outlined how to value information based on three time-dependent components: reliability of a data source, independence between data sources, and coherence among data sources. We also raised several issues with respect to temporal tracking, learning a value function, and biases in various levels of data. The bottom line is that evaluating information from real-world, dynamic data sources will provide us with a much-needed qualitative metric for both evidence detection and hypothesis testing.

We have also outlined a framework that allows the computation of VOI for both processes and events. This distinction between processes and events appears to be both novel and very powerful. The resulting potential increase in ability to inform inference is a fertile area for future research. Other work along these lines includes investigating appropriate learning methods for estimating each component function individually as well as learning how to best combine them. Additionally, comparing these learned component functions to some of the existing functions we surveyed will play an important role in characterizing performance. We also plan to conduct an extensive experimental study using additional data sets and begin looking at some of the alternate models of uncertainty (such as possibility theory or Dempster-Shafer theory).

Lastly, to fully enable triage, we have to look at how, specifically, VOI is used to select among the many possible combinations of sources. In our preliminary work, we assumed a fixed number of sources to be selected for inference. Ultimately, rather than selecting the strongest evidence produced by n sources, we would like this to be adapted at each time step to select the strongest evidence possible. Even with the limited experimentation described in this position paper, we see very encouraging results. We are excited to experiment with the full potential for this approach and pursue the many avenues of future research.

Acknowledgments

We thank Branden Fitelson for valuable insights.

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48 (UCRL-CONF-223973) and under the U.S. Department of Homeland Security (DHS) Scholarship and Fellowship Program, a program administered by the Oak Ridge Institute for Science and Education (ORISE) for DHS through an interagency agreement with the DOE. ORISE is managed by Oak Ridge Associated Universities under DOE contract number DE-AC05-06OR23100.

References

Bovens, L., and Hartmann, S. 2001. Belief expansion, contextual fit, and the reliability of information sources. In Akman, V.; Bouquet, P.; Thomason, R.; and Young, R. A., eds., *Modeling and Using Context: The Third International and Interdisciplinary Conference (CONTEXT)*, 421–432. Berlin: Springer.

Bovens, L., and Hartmann, S. 2003. *Bayesian Epistemology*. Oxford University Press.

Cronbach, L. J. 1947. Test reliability: Its meaning and determination. *Psychometrika* 1–16.

Delmotte, F.; Dubois, L.; and Borne, P. 1996. Context-dependent trust in data fusion within the possibility theory. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, volume 1, 538–543.

Fitelson, B. 2001a. A Bayesian account of independent evidence with applications. *Philosophy of Science* 68(3):S123–S140.

Fitelson, B. 2001b. *Studies in Bayesian Confirmation Theory*. Ph.D. Dissertation, Univ. of Wisconsin – Madison.

Gallagher, B. 2006. The state of the art in graph-based pattern matching. In *Papers from the AAAI Fall Symposium - Capturing and Using Patterns for Evidence Detection*.

Getoor, L., and Diehl, C. 2005. Link mining: A survey. *SIGKDD Explorations - Special Issue on Link Mining* 7(2):3–12.

Heckerman, D.; Horvitz, E.; and Middleton, B. 1993. An approximate nonmyopic computation for value of information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(3):292–298.

Horvitz, E., and Rutledge, G. 1991. Time-dependent utility and action under uncertainty. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 151–158. San Mateo, CA: Morgan Kaufman.

Kent, S. 1994. *Sherman Kent and the Board of National Estimates: Collected Essays*. Center for Intelligence Studies. chapter Words of Estimative Probability.

Liao, W., and Ji, Q. 2006. Efficient active fusion for decision-making via VOI approximation. In *Proc. of the 21th National Conference on Artificial Intelligence (AAAI-06)*.

McCarthy, J. 1956. Measures of the value of information. *Proceedings of the National Academy of Sciences* 42(8):654–655.

Noble, D. F. 2004. Assessing the reliability of open source information. In Svensson, P., and Schubert, J., eds., *Proceedings of the Seventh International Conference on Information Fusion*, volume II, 1172–1178. Mountain View, CA: International Society of Information Fusion.

Olson, E. 2002. What is the problem of coherence and truth? *Journal of Philosophy* 94:246–272.

Parra-Loera, R.; Thompson, W.; and Salvi, A. 1991. Adaptive selection of sensors based on individual performances in a multisensor environment. In Libby, V., ed., *Proceedings of SPIE – Volume 1470: Data Structures and Target Classification*, 30–36.

Rogova, G. L., and Nimier, V. 2004. Reliability in information fusion: Literature survey. In Svensson, P., and Schubert, J., eds., *Proc. of the 7th International Conference on Information Fusion*, volume II, 1158–1165. Mountain View, CA: International Society of Information Fusion.

Schrage, M. 2005. What percent is ‘slam dunk’? The Washington Post.

Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27:379–423.

Sondhauss, U., and Weihs, C. 1999. Dynamic bayesian networks for classification of business cycles. Technical Report SFB-475, University of Dortmund.

Washio, T., and Motoda, H. 2003. State of the art of graph-based data mining. *SIGKDD Explorations - Special Issue on Multi-Relational Data Mining* 5(1):59–68.

Zhang, Y., and Ji, Q. 2006. Active and dynamic information fusion for multisensor systems under dynamic bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics* 36(2):467–472.

Zhang, Y.; Ji, Q.; and Looney, C. G. 2002. Active information fusion for decision making under uncertainty. In *Proceedings of the Fifth International Conference on Information Fusion (Fusion 2002)*.