The use of Evidence Conflict to extend Diagnostic Models

John Mark Agosta
Edify Corporation
Santa Clara, CA 95051
johnmark@edify.com

Jonathan S. Katz
Department of Neurology and Neurosciences
Stanford University School of Medicine
Palo Alto, CA 94305
jskatz@stanford.edu

Abstract

Medical practitioners are pragmatic about defining new diseases to explain rarely occurring combinations of symptoms. This paper reviews a case from the literature that introduced a new disease by comparing two Bayes network diagnostic models, one that contains the disease, the other a sub-model without it. The analysis shows how the measure of evidence conflict proposed by Jensen 1990 applies in the case of a diagnosis where conflict appears as the disease progresses. We demonstrate how such a model with a known set of diseases can be extended to include a new disease.

I. Introduction: What constitutes a disease?

In much of medicine, the underlying physiological mechanisms of disease are unknown. Entities are defined not by the discovery of a root cause, but by studying the association of various clinical signs and symptoms that might predict an underlying biological process. Semiology is the term applied to this methodology. This approach leads to a medical nomenclature biased towards common or treatable clinical presentations. The nomenclature remains necessarily limited because accounting for every known clinical situation would create too much complexity. Gaps of knowledge exist where specific sets of signs and symptoms remain unnamed. Practically, clinicians are left with uncertainty when a clinical presentation or outcome varies too much from ideal descriptions of named disorders. In this situation, the diagnostician often becomes insecure and might suggest the symptom complex is best relegated to a category of a spurious coincidence of symptoms within one existing syndrome. In short, medical diagnosis does not work from an exhaustive enumeration of disorders.

Alternatively, an unknown presentation might suggest a condition for which the clinician is unaware or perhaps even a new syndrome that can be added to the medical nomenclature. If a review of the literature verifies that the syndrome is unnamed, the rationale for adding a new disorder to a semiotic group—that is, the set of disorders suggested by a common chief complaint—is for the most part a practical one. If there is a biological marker that can be easily measured, especially one that provides evidence of causation, the answer is simply to add the disease. Without a marker, there will be little motivation for creating a new disease based solely on semiotics unless there is a way for the clinician to intervene in its course. By the same token, with no marker and no way to intervene, there is little rationale to create new disorders. As such, entities with similar presentations but differing prognoses are often grouped together and individual cases that fall outside defined parameters tend to be lumped with closest defined matches. Any potential significance of such cases can be ignored until discovery leads to a rationale for their recognition.

In this paper, we explore the measure of evidence conflict in a diagnostic model as a means for determining when a Bayesian network model is inadequate. Specifically we consider the case of an unexpected temporal evolution of a disease following a common clinical presentation. Because we never can really work with an exhaustive set of diseases, we ask, “How do we know if this case lies in the set of diseases that have been left out of the nomenclature?” This is equivalent to recognizing when the observed outcome was not represented by any fault state within the existing model. We will argue that our use of conflict measures to decide when to extend a model is preferable to incorporating a fault state representing “other” to address gaps of medical knowledge and is consistent with a utility-driven value of
information (VOI) approach to the value of adding an additional disease state.

To test the applicability of this approach we built a diagnostic model specific to the case, which evaluates the small set of known disorders and supporting observations (the semiotic group) suggested by a common clinical presentation of progressive weakness (the chief complaint). This paper takes as its subject a multi-stage clinical case [Katz 1999] and shows how conflict measures apply in its analysis. The next section of this paper reviews the definition of conflict and some of its properties from the literature. Section three presents the clinical case. The model is presented in section four, and its results interpreted in light of the case. The last section discusses the possible responses to modifying the model.

II. Definition of evidence conflict

The specific measure of conflict we investigate has to do with conflict among a set \( J \) of observed evidence variables \( e_1,\ldots,e_j \) within a diagnostic model, as defined by Jensen 1990 and elaborated in Laskey 1991:

\[
c(J) = \log_2 \left( \frac{p(e_1)\ldots p(e_j)}{p(e_1\ldots e_j)} \right) \tag{1}
\]

Conflict is a relative measure, comparing the joint marginal of the evidence with the product of the individual evidence marginals. The product of evidence marginals serves as a hypothetical model of a case where there is no relationship among the evidence, as would be created in a diagnostic model by unobserved disease variables. When there is a set of evidence whose observations supports a disease state, it makes the joint marginal of the evidence more likely than when there is no relationship. This makes the denominator larger than the numerator, and \( c(J) \) negative. Conflict is indicated by the anomalous case where \( c(J) \) is positive. This intuition is formalized by the measure of evidence conflict. It’s one term in the Kullback-Leibler distance expectation between the two models. (See Laskey, 1991).

The term “conflict” has also been applied to agent’s differences between preferences and desires. In this paper we look strictly at evidence conflict, although it is an interesting question whether evidence conflict bears any relation to preference conflict. We will use the term “evidence conflict” when there is a possibility of confusion.

Equivalences in single fault node models

Conflict is computed on the un-instantiated Bayes network, using the prior marginals of the set of nodes that would be instantiated in the diagnosis. It’s a form of “pre-posterior” analysis.

Here are some of the basic properties of evidence conflict. They give intuition about how conflict depends on the conditional priors (the likelihoods) of the evidence variables.

Dependencies among evidence nodes

Proposition 1: For the findings of two evidence nodes disregarding the effect of other nodes, if the presence of one finding directly conditions a second finding, and makes the second finding less likely, then the presence of both findings will increase conflict.

Proposition 1 is clear if we consider the sign of \( \log(p(e_i) / p(e_i | e_k)) \) that appears as one term in the expansion of the quotient in the definition of conflict, Equation (1).

Conflict in naïve Bayes models

Proposition 2: For findings that are conditionally independent given the fault node with uniform priors, findings are in conflict if their log likelihoods have different signs.

This is shown by writing out each evidence term as:

\[
p(e_i) = p(e_i | N) p(N) + p(e_i | \bar{N}) p(\bar{N})
\]

\[
p(e_k) = p(e_k | N) p(N) + p(e_k | \bar{N}) p(\bar{N})
\]

\[
p(e_i, e_k) = p(e_i | N) p(e_k | N) p(N) + p(e_i | \bar{N}) p(e_k | \bar{N}) p(\bar{N})
\]

\[
+ p(e_i | \bar{N}) p(e_k | \bar{N}) p(\bar{N})
\]

Replacing \( p(N) \) by \( N \) and the likelihood terms by \( a, b, c \) and \( d \) we get

\[
p(e_i) = aN + c(1-N)
\]

\[
p(e_k) = bN + d(1-N) \tag{3}
\]

\[
p(e_i, e_k) = abN + cd(1-N)
\]

Conflict occurs if:

\[
p(e_i, e_k) < p(e_i) p(e_k) \tag{3}
\]

If we ignore the effect of priors, by setting \( p(N) = 1/2 \), then this inequality is equivalent to:
\[
\left( \frac{a}{c} - 1 \right) \left( \frac{b}{d} - 1 \right) < 0, \quad (4)
\]

which holds true when the log likelihood ratios \( \log(a/c) \) and \( \log(b/d) \) have opposite signs.

Dependence of evidence conflict measures on priors

Proposition 3: When prior probabilities tend to extremes for the same evidence likelihoods, conflict between evidence decreases. When priors are certain in Naïve Bayes models, there can be no conflict between evidence.

This is shown by substituting Equation (2) into (1) and letting \( N \to 0 \) or \( N \to 1 \). The conflict measure goes to zero.

Intuitively extreme priors remove the effect of the disorder node, and the models in the numerator and denominator of the conflict measure become essentially the same.

III. Case description

Initial presentation: A 53-year-old man presented with a 10-month history of progressive arm and shoulder girdle weakness (Figure 1) that began in the right arm and spread to the left arm 6 months later. An early examination showed severe muscle atrophy and weakness affecting both arms. The face, neck, and leg muscles were normal. Sensation was intact. This presentation led the clinicians to consider only two disorders: multifocal motor neuropathy (MMN), caused by blocking of the electrical impulse along the motor nerve; a treatable immune disorder that progresses very slowly or not at all; and amyotrophic lateral sclerosis (ALS; commonly known as Lou Gerhig Disease), a relentlessly progressive degeneration of motor nerve cells that invariably spreads throughout the body and leads to death in 1-5 years (Table 1).

Diagnostic testing: Electrical testing of the subject’s nerves and muscles showed a severe degeneration of the motor nerves in the arms but the legs and trunk were normal. There was no abnormality of electric conduction along the nerves traveling from the spinal cord to muscle (conduction block). Immunologic testing did not show anti-GM-1 antibodies in the serum.

Temporal Outcome: Over the next 2 years the arm weakness progressed until they hung limply at his sides. He had no ability to move them except to grip. He remained this way for the next 7 years but the disease never progressed to symptoms affecting his legs, breathing, or swallowing.

Conflicts among the diseases

The symptoms associated with each disease in the semiologic group are summarized in Table 1. A third condition named “brachial amyotrophic diplegia” (BAD) is “invented” to refer to cases similar to the patient presented above. That is a condition defined by severe degeneration of the nerves to the arm that never spreads outside the upper limbs. Common to the three diseases is “weakness in the arms,” the chief complaint, which is the symptom in the first column. The rest of the symptoms are specific to one or two of the three diseases. For instance “GM1 antibodies,” “Conduction block” and “Responds to therapy” differentiate MMN from the others. Similarly, the progression of weakness beyond the arms differentiates Lou Gerhig disease (or ALS) from BAD and MMN.

Proposition 2 suggests that when there is no disease state in the model for which log likelihood ratios of all evidence have a common sign, then the model will show conflict. We find this useful for understanding evidence conflict in a model based on the table of disease and symptom co-occurrences.
<table>
<thead>
<tr>
<th>Disorder</th>
<th>Weakness</th>
<th>Evolves in initial limb</th>
<th>Spreads to face, legs, breathing muscles</th>
<th>GM1 antibodies</th>
<th>Conduction block</th>
<th>Responds to therapy</th>
<th>Prognosis without therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lou Gehrig (ALS)</td>
<td>100%</td>
<td>100%</td>
<td>100% by 2 years</td>
<td>&lt;10%</td>
<td>&lt;2%</td>
<td>0%</td>
<td>Death in 1-5 years</td>
</tr>
<tr>
<td>MMN</td>
<td>100%</td>
<td>&lt;5%</td>
<td>&lt;5%</td>
<td>50-80%</td>
<td>70-80%</td>
<td>60-70%</td>
<td>Weakness in hands and ankles, never death</td>
</tr>
<tr>
<td>BAD</td>
<td>100%</td>
<td>100%</td>
<td>Never?</td>
<td>&lt;10%</td>
<td>&lt;2%</td>
<td>0%</td>
<td>Severe arm weakness</td>
</tr>
</tbody>
</table>

**Table 1:** Occurrence of symptoms for diseases that present initially with weakness in the arm

**Figure 2:** Diagnostic Bayes Network for ALS, MMN and BAD, with therapy and utility nodes.
V. Analysis

We built a multi-stage Bayes network diagnostic model incorporating the diseases and the progression of symptoms in Table 1 to compute evidence conflict under various circumstances. There’s a version of the model without BAD, representing medical belief about diseases in this semiologic group prior to the initial report of the condition, and one with BAD. As expected, the evidence conflict measure is positive when instantiating symptoms predicted by BAD and BAD has been left out of the model, but negative when BAD is included. Table 2 summarizes this.

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Disease node states</th>
<th>Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weakness:</td>
<td>ALS, MMN</td>
<td>0.171</td>
</tr>
<tr>
<td>Arms_only (T₀)</td>
<td>ALS, MMN, BAD</td>
<td>-1.10</td>
</tr>
</tbody>
</table>

Table 2: Conflict among the predicted symptoms and disease states

Medical Background:

Clinical viewpoint: The initial presentation, marked by severe muscle atrophy of both arms, is a relatively common presentation of ALS and there was no conflict associated with making this diagnosis. However, as the patient returned for follow-up visits, the expected spread to other regions of the body never occurred. For the first few visits, this was viewed as a low probability outcome for the most likely diagnosis. This assumption was associated with some increase in the measure of conflict but there was a tacit expectation that the conflict would be resolved once the invariable spread to other body regions was observed. The disorder was considered as a “benign” course of ALS with both clinician and patient still expecting death to be the ultimate outcome.

Initially the only other disorder on the fault list was MMN but this diagnosis was never attractive because initial testing had shown no signs of altered nerve transmission or immune abnormalities (conduction block and GM-1 antibodies). However, as the likelihood of ALS decreased with each subsequent clinic visit that showed no spread, clinicians tended somewhat paradoxically to increase the probability of MMN. MMN rarely spreads to other limbs and the initial model lacked other disorders to compete with ALS. The progressive increase in belief that MMN might be the cause was enough to justify on the basis of expected utility a “diagnostic” trial of expensive therapy that later turned out to be unwarranted.

Three possible explanations for conflict

Recognizing high conflict exists for certain clinical presentations called for a closer look at the assumptions underlying the diagnosis. Conflict can be explained either by:

- Lack of a disease state,
- Inaccuracy in evidence, or
- A rare random presentation within a known disease. We explore each of these in turn.

A. There is a disease state missing (the presentation falls into the realm of “other”)

Until 1986, when MMN was discovered, all similar presenting disorders were actually lumped under just a single heading of “ALS.” These patients developed progressive muscular weakness that could not be treated. A small percentage survived the disease but these cases were conveniently ignored in the nomenclature. The discovery of MMN in some of these patients partially explained this observation, and led to a great emphasis on finding measures that distinguish MMN from ALS.

Subsequent to 1986, a new diagnostic model contained two diagnoses; MMN to represent potentially treatable patients with slow progression, and ALS to represent untreatable relentless disease. However, the new model had some obvious gaps. For one, there was a small subset of patients that failed to progress but had negative testing for MMN. Because the concept of “non-progression” was unaccounted for in the new ALS fault, long durations of non-progression would drive up the posterior probability of MMN. Practically, clinicians would compare posterior probabilities of an atypical presentation of MMN versus an unusual outcome of ALS. We have shown here that as one updates the model after repeatedly observing “no spread to other limbs” there is a marked increase in the measure of conflict. Eventually, the probability of MMN becomes high enough to support the assumption that a patient should be treated, based on the expected utility of treatment. The treatment, which costs $10,000 to $20,000 for each patient, and is ineffective in these patients, could not be disregarded. Recognizing conflict and the utility of fixing it led to the creation of a new disease entity named BAD.

B. Unreliable measurement or observation of the evidence.

One possible explanation for conflict was that the lack of clinical progression could be explained away by suggesting there was a mis-measurement; perhaps earlier exams were done too quickly or noted incorrectly in the records. Clearly, this occurs often enough in medicine. We also observed instances where doctors had anchored on ALS as the disorder in the incorrect model and to reduce conflict they would overestimate degrees of weakness by unintentionally pushing too hard on patient’s legs to create
an appearance that the disorder was spreading into other limbs.  Our use of a node for “observed weakness” accounts for this potential of mis-measurement.  Clearly, by updating temporal information it became obvious that the disease was not spreading to other limbs and this measurement conflict could be ruled out. From a utility standpoint, the increase in conflict in a flawed model results in increased costs associated with retesting and repeat usage of costly tests.

The clinician can also explain away the lack of progression by holding to the belief that the condition is ALS and spread to other regions will occur if one can observe the patient indefinitely.  This is impractical and whether this belief turns out to be correct or not, it is still associated with costs of unneeded treatments and measurements along the way. By adding the “BAD” disease state the likelihood of evidence induced conflict was reduced.

C. “An atypical presentation of a common disease”

This explanation will turn out to be the cause of conflict in many cases.  Clearly, semiology is imperfect and even disorders with known underlying pathophysiological explanations can have differing and rare presentations. For those instances where pathophysiology is unknown, assuming one can make no further observations, one can never really know if they are witnessing a new disease or a rare presentation of a known entity. If temporal updating is possible, sooner or later the expected features of a known entity might suggest its presence and conflict should reverse.

Practically, once there is a rationale for recognizing atypical cases, holding to a single known clinical entity has several disadvantages.  Clinicians not only need to memorize the names of these conditions, but need to learn complex dependencies between various signs and symptoms. In the current model this would have led us to account for different risks of progression of ALS depending on the initial distribution of weakness, but this is akin to creating a new entity. We could also have simply allowed for some probability that ALS will not progress but this has never been observed in ALS cases that start in the throat or in the legs. In other models this shortcut might certainly suffice, especially if there are no issues regarding utility.

Here, the simplest way to resolve conflict was to add a new state to the model

**Conflict is a legitimate way to leave “other” out of the model states**

We argue that once questions about the accuracy of the evidence are resolved, it is better to modify the model by adding states as conflicts arise instead of trying to build an exhaustive model by adding a “catch-all” state for unknown diseases.

One can see how creating a state of *other* is not a practical approach in modeling this clinical situation. First, there is no clear definition of concept of *other* that passes the “clarity test.” Clinicians and experts alike were not aware of the relatively rare presentation of BAD and had no knowledge of its significance. And there is no practical way to incorporate probabilities of symptoms and outcome values for a state of *other*. Potentially, strict clinical definitions could have confined the parameters allowable for diagnosing ALS, and the remaining patients could have been placed in *other*, but there would be no way to recognize any need to limit the definition or acknowledge that these patients existed. Perhaps, if a database existed for all patients within the defined semiologic group, *other* could be used for cases that are unnamed and left out of fault states, but this requires clear clinical limits of diseases that would eliminate cases of *other* from the database to begin with. The use of *other* can be relegated to defined presentations left out of a model, perhaps for practical purposes such as limiting the size of a model. However, even with that approach, the state must represent an average of cases with different signs and symptoms and will lose some of its power to differentiate these cases from similar conditions that are included in the model. For example, in this model theoretical cases with severe progressive weakness and signs of conduction block would be averaged with cases of BAD, with no progression and no sign of conduction block. The ensuing entity would be meaningless. Instead, by conditioning the entire model on a known set of diseases, the accuracy of the model is not compromised if it implicates a disease it contains. In the case that the disease is missing there is a technique for modifying the model built around the interpretation of conflict that we have shown by example.
References


