Tenth International Symposium on Imprecise Probability: Theories and Applications (Lugano, 2017)

The Soft Propagation Algorithm: A Proposal for Responsive Belief Revision with Bayesian Networks

Sabina Marchetti

Department of Statistics
Sapienza University of Rome, Italy

sabina.marchetti@uniroma1.it

Due to a multiplicity of reasons, it is quite common to observe a number events of interest being affected by degrees of uncertainty, e.g. incomplete data, experts' opinions (epistemic peers: qualitative constraints, soft evidence), unreliable and/or coarse observations, etc.
As it is not always a Good idea to ignore information [8], uncertain evidence should also be accounted for by intelligent systems.

Background: Belief revision and Bayesian Networks

$(\Omega, \mathscr{A})$ - measurable space, $\mathscr{A}$ boolean σ-algebra
P - probability distribution (pd) defined on $(\Omega, \mathscr{A})$ s.t. $P(\alpha) > 0, \forall \alpha \in \mathscr{A}$
Belief revision as the process of changing a probability distribution based on new constraints posed by epistemic peers, by means of an appropriate Revision Rule (RR).
Distance between the original (P) and revised (R) probability distributions, measured by the **CD-distance** [3]:
Let R and P be two pds sharing the same support*, their CD-distance is

$$CD(P,R) = \log \frac{\max_{\omega \in \Omega} \frac{R(\omega)}{P(\omega)}}{\min_{\omega \in \Omega} \frac{R(\omega)}{P(\omega)}} \geq 0 \tag{1}$$

Probability Kinematics (PK) [3-5, 13] generalize ordinary conditionalization [4]. Pd R comes from P by PK if $\forall x \in Val(X)$ and $\forall \alpha \in \mathscr{A}$,

$R(x|c) \models \kappa(X|c)$ (Responsiveness)

$R(\alpha|x, c) = P(\alpha|x, c), \forall \alpha$ (Conservativeness or Rigidity Condition)

with $\kappa(X|c)$ **Conditional Bayes Factors** (CBF) [3,5,13]:

$$\kappa(X|c) = \frac{\pi(X|c)}{\pi(x|c)} = \frac{\lambda(X|c)}{\lambda(x|c)} \tag{2}$$

- x - fixed reference value of variable X in its possibility space $Val(X)$
- $\pi(.) = R(.)/P(.)$ relevance quotients (*soft evidence* [1,3-5,13])
- $\lambda(.)$ - likelihood ratios (*virtual evidence* [7])

Prop.1 (From [3]) Any RR based on Probability Kinematics minimizes the CD-distance

Bayesian Networks

A Bayesian Network (BN) [7] is a statistical model defined by the pair $(\mathscr{G}, P)$, where $\mathscr{G} = (V, E)$ is a Directed Acyclic Graph (DAG) with node and edge set, respectively, V and E. Each node $X_i \in V$ is associated to a random variable. P is a pd defined over $V = \{X_1, ..., X_n\}$.
P is *faithful* to $\mathscr{G}$ [10] if the conditional independence relationships among pair of variables in P are all and only those represented in the DAG. P faithful to $\mathscr{G}$, it factorizes as

$$P(V) = P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \tag{3}$$

$Pa(X)$ being the parent set of variable X (set of direct predecessors of node X in the DAG). We hereby consider the case of binary variables.
Evidence is propagated through the network via message-passing algorithms, such as the **Junction Tree** (JT) **algorithm** [6-7]. JT algorithm reduces the model's structure to a simplified graphical representation (a tree) and efficiently performs conditionalization on observed evidence.

The Soft Propagation algorithm

The Soft Propagation (SP) algorithm generalizes the JT to the case of *uncertain conditional observations*, minimizing the CD-distance, based on optimality axioms of belief revision [1,5].

The Soft Propagation Algorithm

1. Build a JT under the constraint $X_u \cup C \subseteq \mathscr{C}_R$, root-clique of the tree, as in [12].
   P_0 factorizes as [6]
   $$P_0(X_1, ... X_n) = P_0(\mathscr{C}_R) \prod_j P_0(\mathscr{C}_j)$$
   $V_T := \{\mathscr{C}_R, \mathscr{C}_1, ..., \mathscr{C}_J\}$ - set of cliques of the JT ($V \equiv V_T$);
2. Propagate hard evidence $H = h$ toward the root-clique; $P(\mathscr{C}_R) := P_0(\mathscr{C}_R|h)$;
3. Let $X_Q \subseteq \mathscr{C}_R \backslash \{X_u, C\}$, apply the Commutative Revision Rule (CRR) based on PK:
   $$R(X_Q) := P(\neg c')P(X_Q|\neg c') + \sum_i P(c'_i, \neg c'_{-i}) R^i(X_Q|c'_i, \neg c'_{-i}) +$$
   $$+ \sum_{i,j} P(c'_i, c'_j, \neg c'_{-i,j}) R^{ij}(X_Q|c'_j, c'_j, \neg c'_{-i,j}) + ... + P(c')R^{1...K}(X_Q|c') \tag{4}$$
   with $R^{1...K}(X_Q|c') := \frac{\sum_{x_u} P(X_Q, x_u, c') \prod_{i=1}^K \kappa(x_{ui}|c'_i)}{\sum_{x_u} P(x_u, c') \prod_{i=1}^K \kappa(x_{ui}|c'_i)}$ [5, 11, 13].
   $R(\mathscr{C}_R)$ follows from (4);
4. Back-propagate uncertain evidence; $R(V) := R(\mathscr{C}_R) P(V' \backslash \mathscr{C}_R | V' \cap \mathscr{C}_R)$;

- $X_u = \{X_1, ..., X_K\} \cup V$ - set of random variables/nodes on which uncertain evidence is asserted
- $\{C = c'\} = \{C_1 = c'_1, ..., C_K = c'_K\}$ - associated set (or subset) of conditioning events
- **Def.** ***Cross Context-Specific Independence*** (C-CSI): W.l.o.g., let P be a pd defined over variables X_i, C_i $i = 1, 2$, s.t. for fixed c'_1, c'_2, $P(c'_1, c'_2) > 0$. $(x_1, C_1 = c'_1) \perp\!\!\!\perp_c (x_2, C_2 = c'_{2'}), \forall x_1 \in Val(X_1), \forall x_2 \in Val(X_2)$

Theorem 1

Iff C-CSI holds for every pair in X_u and fixed $C = c'$, then (i) the SP algorithm is an exact belief revision procedure and (ii) it is invariant with respect to different revision schedules (commutativity). Let $R(V)$ be the resulting jpd, (iii) it is responsive to all uncertain instances provided and (iv) it minimizes the CD-distance with $P(V)$
Proofs of i-iv are based on previous results in [1,3,5,11,13]

*$P(A) = 0$ iff $R(A) = 0$
†A number of alternative heuristics may be used: a simple sketch is hereby illustrated to provide an intuition

SP.1 The Oedipus Strategy for singly connected root cliques (†): from (C-CSI) to $\mathscr{C}_R$ SCC-X_L

Def. ***U-Polytree***: A DAG $\mathscr{G}$ is a U-Polytree with respect to nodes A and B if the chain/fork path connecting the two may be broken by the removal of a single (not necessarily unique) edge
Def. ***Singly Connected Clique*** (SCC-U): A clique $\mathscr{C}$ is a SCC with respect to U, $U \subseteq \mathscr{C}$, whenever the pd defined over its elements is faithful to a U-Polytree with respect to all pairs in U.

The Oedipus Strategy

0 (Initialize) $X_L := \{(X_i, X_j) \subseteq X_u$: C-CSI fails $|C_i = c'_i, C_j = c'_j\}$, $|X_L| = L$, $1 \leq L \leq K(K-1)/2$ and $\mathscr{C}_R$ SCC-X_L; then $\mathscr{C}^{Oe} := \mathscr{C}_R$;
1 **For† every pair l, $l = 1, ..., L$ do**
2 Compute E_l, set of candidate edges of a SCC root clique to be broken, i.e. edges forming the path that connects pair l, with properly instantiated nodes
3 Let X, Y be endpoints of edge $(X, Y), \forall (X,Y) \in E_l$. $\forall (x, y) \in Val(X) \times Val(Y)$, compute Shogenji's measure [9]: $S(x, y) = \frac{P(x,y)}{P(x)P(y)}$
4 Break edge (X^*, Y^*) := $\text{argmin}_{E_l} \sum_{x,y} |1 - S(x, y)|$ and replace it with the binary non-probabilistic Oedipus node $O_{X^*Y^*}$, s.t. Oedipus node is a *collider*, [7]); update† $\mathscr{C}^{Oe}$ accordingly

$$P(O_{X^*Y^*} = o_{X^*Y^*}|x^*, y^*) := S(x^*, y^*)$$
$$\sum_{o_{X^*Y^*}} P(o_{X^*Y^*}|x^*, y^*) := 1, \forall (x^*, y^*)$$

5 **End for**
6 T^{Oe}'s clique set is $V_T \backslash \{\mathscr{C}_{\mathscr{R}}\} \cup \{\mathscr{C}^{Oe}\} \equiv V \cup O$, $O = \{O_{X^*Y^*}\}_{X^*Y^*} = \{O_1, ..., O_L\}$; run the SP algorithm on T^{Oe}; $R^{Oe}(V \cup O)$ results
7 $R(V) := R^{Oe}(V|O_1, .., O_L)$ is the revised jpd

▲ The Oedipus-SP algorithm is an approximate commutative revision procedure. Although R^{Oe} is fully responsive to uncertain evidence, R is not in the general case

▲ Step 7 is motivated by the *Extended Rigidity Condition*: $R^{Oe}(O|x, y) = P^{Oe}(O|x, y)$, $X, Y \in X_u$

▲ R was found to empirically outperform alternative revision rules in terms of CD-distance with P

SP.2 The Coherent Pooling Strategy for multiple overlapping peers (†): from single to multi-agent system

Based on previous results from [2,11,13-14], replace Step 3 of the SP algorithm with the following:

The Coherent Pooling Strategy

3a Generate *imprecise assessments* $K(X|C = c')$ as the convex hull of all uncertain assessments on the associate event, $X \in X_{Pool} \subseteq X_u$
3b Apply (4) to all precise uncertain assessments, i.e. $Y \in X_u \backslash X_{Pool}$; $P \to R$
3c Apply (4) to all combinations of extreme points of sets $K(.|.)$ generated in Step 3a; $R \to R_1, R_2, ...$
3d Take the lower/upper envelope of the pds generated in Step 3b: $\forall X \in X_{Pool}$
if† $R(x|c) \in [\underline{R}(x|c), \bar{R}(x|c)] \forall x$, *a form of dilation occurred [8] and* $R(X|c)$ *is not revised by Step 3c;* ***else*** $R(X|c) := K(X|c)$

Prop.3 $\mathscr{C}_R$ generalizes to the imprecise setting [14], based on the External Bayesianity property of the pooling operator we employed [11,13]. The associated jpd is coherent [2] while being consistent (valid) with both precise and imprecise uncertain evidence constraints, according to an exact routine [2,11].
▲ Step 3a refers to the case of overlapping epistemic peers whereas this procedure applies to the more general case of imprecise uncertain evidence, accounting, e.g., for incomplete information.

Forthcoming Research

The CRR: Absorption of conditional uncertain evidence based on Adam's conditioning [1,5] imposes heavy restrictions on the probability values. Forthcoming research will be oriented in extending the belief revision procedure on both epistemic and probabilistic bases.
Context-specific independence: Avaiable (from previous research) routines for the detection of context-specific independencies ought to be included; also, application of the procedure directly on suited graphical structures may be reveal efficient
The Oedipus strategy: Due to the questionability of the Extended Rigidity Condition and given the connections between Shogenij's measure and the concept of dilation [8], future effort will be spent on considering the epistemic and probabilistic implications of its usage. Also, much effort will be spent on further generalizing the procedure to increasingly dense graphical structures
The Coherent Pooling strategy: As a *Credal clique* would yield massive computational costs and render the strategy feasible only under restrictive sparsity conditions, last part of Step 3d is meant to address the precise setting; as an alternative, the criterion followed may be replaced by some other appropriate technique, such as that of choosing the pd in the credal set maximizing the Shannon Entropy. Additionally, imprecise extensions should be carefully handled based on their implications on the original independence pattern of the BN under study
It's a long way to the top (if you wanna rock 'n' roll)

Main References

[1] Bradley R. "Radical probabilism and Bayesian conditioning." Philosophy of Science 72.2 (2005): 342-364.
[2] Capotorti A, Regoli G, and Vattari F. "Merging different probabilistic information sources through a new discrepancy measure." UNCERTAINTY PROCESSING (2009): 35.
[3] Chan, H and Darwiche A. "On the revision of probabilistic beliefs using uncertain evidence." Artificial Intelligence 163.1 (2005): 67-90.
[4] Diaconis P, and Zabell SL. "Updating subjective probability." Journal of the American Statistical Association 77.380 (1982): 822-830.
[5] Douven I, and Romeijn JW. "A new resolution of the Judy Benjamin problem." Mind 120.479 (2011): 637-670.
[6] Jensen FV. An introduction to Bayesian networks. Vol. 210. London: UCL press (1996).
[7] Pearl J. "Morgan Kaufmann series in representation and reasoning. Probabilistic reasoning in intelligent systems: Networks of plausible inference." (1988).
[8] Pedersen AP, and Wheeler G. "Demystifying dilation." Erkenntnis 79.6 (2014): 1305-1342.
[9] Shogenji T. "Is coherence truth conducive?." Analysis 59.264 (1999): 338-345.
[10] Spirtes P, Glymour C, and Scheines R. Causation, prediction, and search. MIT press (2000).
[11] Stewart RT, Quintana IO. "Learning and Pooling, Pooling and Learning." Erkenntnis (2017). 1-21
[12] Valtorta M, Young-Gyun K, and Vomlel J. "Soft evidential update for probabilistic multiagent systems." International Journal of Approximate Reasoning 29.1 (2002): 71-106.
[13] Wagner CG. "Jeffrey conditioning and external Bayesianity." Logic Journal of IGPL (2009): jzp063.
[14] Walley P. "Statistical reasoning with imprecise probabilities." (1991).