Decomposed Utility Functions and Graphical Models for Reasoning about Preferences*

Ronen I. Brafman
Computer Science Department
Ben-Gurion University, Israel
brafman@cs.bgu.ac.il

Yagil Engel
Faculty of Industrial Engineering and Management
Technion–Israel Inst. of Technology
yagile@ie.technion.ac.il

Abstract

Recently, Brafman and Engel (2009) proposed new concepts of marginal and conditional utility that obey additive analogues of the chain rule and Bayes rule, which they employed to obtain a directed graphical model of utility functions that resembles Bayes nets. In this paper we carry this analogy a step farther by showing that the notion of utility independence, built on conditional utility, satisfies identical properties to those of probabilistic independence. This allows us to formalize the construction of graphical models for utility functions, directed and undirected, and place them on the firm foundations of Pearl and Paz's axioms of semi-graphoids. With this strong equivalence in place, we show how algorithms used for probabilistic reasoning such as Belief Propagation (Pearl 1988) can be replicated to reasoning about utilities with the same formal guarantees, and open the way to the adaptation of additional algorithms.

Introduction

The similarity between the decompositions of probability distributions and utility functions is well established. Probability distributions are usually decomposed as a product of factors using the chain rule, while utility functions are often assumed to have an additive structure, consisting of a sum of utility factors (Fishburn 1967; Bacchus and Grove 1995; Boutilier, Bacchus, and Brafman 2001). The obvious relation between the two via the log/exp transformation, allows adaption of algorithms that maximize the probability of a certain assignment (MPE), for solving the problem of finding the optimal (maximizing) assignment to a utility function. However, the similarity ends about there. Probability distributions have a much richer structure thanks to the notion of marginal and conditional probabilities, which reveal a finer notion of independence. For this reason, existing graphical utility models, such as GAI-nets (Gonzales and Perny 2004), UCP-nets (Boutilier, Bacchus, and Brafman 2001), and CUI-nets (Engel and Wellman 2008) which were motivated by their probabilistic analogues, cannot offer the same strength, intuitive appeal, and process for elicitation of structure.

Brafman and Engel (2009) introduce a simple mathematical definition of *marginal* utility using a fixed reference assignment. From this idea, a notion of conditional utility was derived, and from it a notion of conditional independence. This concept of conditional utility satisfies additive analogues of Bayes rule and the chain rule and can be used to define a directed model of utility that resembles a Bayes net. In this work we take this analogy a step forward: we show that this concept of utility independence is almost identical to the classical notion of probabilistic independence. In particular, it satisfies the semi-graphoid axioms (Pearl and Paz 1989), and that implies the existence of an additive utility decomposition that is perfectly analogous to Bayesian networks, supporting the same types of reasoning and algorithms, and where the notion of d-separation applies. We call the resulting utility model a *marginal utility network*. Using Pearl's well known belief propagation algorithm, we show how methods used in probabilistic inference can now be carried out in the area of preference reasoning, opening up a vast body of results which were not applicable to any existing graphical utility model.

Background

We assume familiarity with basic probability theory and review necessary background on multiattribute utility. Let Θ denote the space of possible outcomes, with $\preceq$ a preference relation (weak total order) over Θ. Let $\Gamma = \{a_1, \ldots, a_n\}$ denote a set of attributes describing Θ. Each attribute $a \in \Gamma$ has a domain $\mathcal{D}(a)$, so that $\Theta = \prod_{i=1}^{n} \mathcal{D}(a_i)$. We use prime signs and superscripts to denote specific assignments to an attribute. A concatenation of assignment symbols (as in $a_i' a_j''$) means that each of the attributes gets a respective value. We use γ and γ_i to denote subsets of Γ, with $\mathcal{D}(\gamma) = \prod_{a_i \in \gamma} \mathcal{D}(a_i)$, and the same notation as before to denote assignments to all the attributes in the set. For example, if $\gamma_1 = \{a_i, a_j\}$, then $\gamma_1^1 = a_i^1 a_j^1$.

The preference relation $\preceq$ over outcomes is usually represented numerically by a value function $v(\cdot)$.

Definition 1. $v : \Theta \rightarrow \Re$ *is a* value function *representing* $\preceq$

*The authors were supported in part by ISF grant 1101/07. Ronen Brafman was also supported by the Lynn and William Frankel Center for Computer Science, by the Paul Ivanier Center for Robotics Research and Production Management, and by COST Action IC0602. Yagil Engel was supported in part by an Ali Kaufman fellowship at the Technion.

if for any $\theta, \theta' \in \Theta$, $v(\theta) \leq v(\theta')$ *iff* $\theta \preceq \theta'$.

In many cases it is useful to represent, beyond a simple preference order over outcomes, a notion of *strength of preferences*. A value function that expresses strength of preferences is called a *cardinal utility function*, which we denote by $u(\Gamma)$. A cardinal utility function which represents preference for uncertain outcomes, or *lotteries*, is called an *expected utility function* or *von Neumann-Morgenstern (vNM)*. A *measurable value function (MVF)* is a cardinal utility function that represents order over preference differences, that is $u(\theta) - u(\theta')$ represents the strength of preference of $\theta \in \Theta$ over $\theta' \in \Theta$.

Previous Work on Graphical Models for Utilities

Graphical models have been employed for the representation of decomposed utility, as early as by Gorman (1968) However, the first representation that relies on *conditional independence*, and thus follows the footsteps of probabilistic models, is due to Bacchus and Grove (1995). In order to avoid expected utility theory, we provide a somewhat non-standard definition. We use attribute names (or sets) within functions for statements that hold for any value of that function, as in $u(\Gamma)$ below.

Definition 2. *Let* $\gamma_1, \gamma_2 \subset \Gamma$. γ_1 *and* γ_2 *are* conditionally additive independent (CAI) *given their complement* $\gamma_3 = \Gamma \setminus (\gamma_1, \gamma_2)$*, if there exist functions* u_1 *and* u_2 *such that* $u(\Gamma) = u_1(\gamma_1, \gamma_3) + u_2(\gamma_2, \gamma_3)$.

Bacchus and Grove (1995) show that conditional additive independence has a perfect map, meaning that given a set of attributes and a preference order, there exists a graph whose node separation expresses the exact set of independence conditions.[1] Further, they show that the utility function decomposes to a sum over lower dimensional functions, each defined over a maximal clique of the graph. This decomposition is a special type of *generalized additive independence (GAI)*, a global independence condition introduced originally by Fishburn (1967). A directional graphical model for GAI, named *UCP* nets, is introduced by Boutilier *et al.* (2001). The directionality is obtained by means exogenous to the GAI decomposition; by identifying preferential independence conditions over sets of attributes. Gonzales and Perny (2004) introduce GAI nets, a graphical representation for GAI, where nodes represent subsets of attributes.

Graphical model were also introduced for non-additive types of utility independence. CUI networks (Engel and Wellman 2008) rely on the concept of *conditional utility independence* (Keeney and Raiffa 1976), which requires the (cardinal) preference order over a subset of the attributes to be independent of another subset of attributes. Several other authors also sought utility representation that is similar to a probability distribution. Shoham (1997) proposes a redefinition of utility function as a set function, over additive factors in the domain that together contribute to the decision maker's well being. La Mura and Shoham (1999) propose a utility independence concept which is based on invariance of utility ratios. Abbas (2005) defines a subclass of utility functions in which a multiplicative notion of utility independence obeys an analog of Bayes's rule.

[1] We define perfect maps formally below, see Definitions 7,8.

Difference Independence

CAI and GAI require comparisons of probability distributions and preferences over lotteries. When uncertainty is not a crucial element (e.g., electronic commerce applications), it is not required and usually not desirable to involve probabilities in user interaction. Dyer and Sarin (1979) propose *difference independence*, an additive independence concept on which they build additive MVF theory. Engel and Wellman (2007) extend their work and introduce *conditional difference independence (CDI)*. Intuitively, attributes a and a' are CDI of each other if any difference in value among assignments to a does not depend on the current assignment of a', for any possible assignment to the rest of the variables.

Definition 3. *Let* $\gamma_1, \gamma_2 \subset \Gamma$. γ_1 *and* γ_2 *are* conditionally difference independent *given* $\gamma_3 = \Gamma \setminus (\gamma_1 \cup \gamma_2)$*, denoted as CDI*$(X, Y)$*, if* $\forall$ *assignments* $\hat{\gamma_3}, \gamma_1', \gamma_1'', \gamma_2', \gamma_2''$,

$$u(\gamma_1'\gamma_2'\hat{\gamma_3}) - u(\gamma_1''\gamma_2'\hat{\gamma_3}) = u(\gamma_1'\gamma_2''\hat{\gamma_3}) - u(\gamma_1''\gamma_2''\hat{\gamma_3})$$

CDI leads to a functional decomposition as in Definition 2, and has a perfect map as well. CDI can be applied to any value function, but it is mostly intuitive for MVF, and in particular when the preference difference carries particular meaning. In most purchase decisions, for example, utility can be measured in monetary terms, and the preference difference between θ and θ' represents the amount a buyer is willing to pay to get θ over θ' (Engel and Wellman 2007).

Reference and Conditional Utility

There are inherent differences between probability distributions and utility functions, which make any analogy between the two problematic. Most notably, there is no meaning for the utility of a set of atomic outcomes, and, thus, no obvious analogue to the idea of maginalization. A simple solution to this problem is presented by Brafman and Engel (2009), and we detail it in this section.

For probabilities, we ask the question *what is the probability of outcomes in* γ *when we don't know the value of* $\Gamma \setminus \gamma$. While we do not have an exact analogy for utilities, a similar effect can be achieved by fixing those parameters on some *reference assignment*. We can now ask: *what is the utility of outcomes in* γ *when the value of* $\overline{\gamma} = \Gamma \setminus \gamma$ *is fixed on the reference assignment*. The reference assignment is a predetermined complete assignment $a_1^r \ldots a_n^r \in \Theta$.

Definition 4. *The* reference utility function *is*

$$u_r(\gamma) = u(\gamma\overline{\gamma}^r)$$

The notion of conditioning, within a subspace of the domain, relies on preference differences.

Definition 5. *The* conditional utility function *is*

$$u_r(\gamma_1|\gamma_2) = u_r(\gamma_1\gamma_2) - u_r(\gamma_2)$$

This definition of conditional utility bears some similarity to the notion of conditional utility of La Mura and Shoham (1999), with two significant differences: (i) CDI_r

is additive rather than multiplicative, and (ii) $\gamma_1 \cup \gamma_2$ can be a proper subset of Γ. This definition is rooted in conditional probabilities; by exponentiating both sides we regain the standard, multiplicative definition of conditional probability. Moreover, this conditional utility function reveals an additive decomposition which is similar to the multiplicative decomposition of a probability function, via the *chain rule*.[2]

Theorem 1 (The chain rule).

$$u(\Gamma) = u_r(a_1) + \sum_{i=2}^{n} u_r(a_i|a_1, \ldots, a_{i-1})$$

It is also easy to see that this definition obeys an additive (logarithmic) adaptation of Bayes' rule.

Theorem 2 (Bayes' rule analog).

$$u_r(\gamma_1|\gamma_2) = u_r(\gamma_2|\gamma_1) + u_r(\gamma_1) - u_r(\gamma_2)$$

Note that in the chain rule, the last term ($i = n$) includes the left-hand side of the equation $u(\Gamma)$. However, the conditional utility function $u_r(a_i|a_1, \ldots, a_{i-1})$ may not depend on all of the attributes $a_1, \ldots, a_{i-1}$, but only on some subset of them, in which case the terms considered by the chain rule have lower dimensionality. This is formalized as follows.

Definition 6. *γ_1 is said to be* conditionally independent *of γ_2 given γ_3 ($CDI_r(\gamma_1, \gamma_2|\gamma_3)$) if*

$$u_r(\gamma_1|\gamma_2\gamma_3) = u_r(\gamma_1|\gamma_3)$$

CDI_r is a generalization of CDI; when $\gamma_3 = \Gamma \setminus \gamma_1 \cup \gamma_2$, then $\text{CDI}_r(\gamma_1, \gamma_2|\gamma_3)$ is equivalent to $\text{CDI}(\gamma_1, \gamma_2)$. Notably, in previous independence concepts (such as CDI) the conditional set must always be "the rest of the attributes". Here, in contrast, we specifically select a conditional set, and can *ignore* the rest of the attributes.

This "small" difference – the ability to define conditional independence on a strict subset of Γ – has considerable implications, paving the way to reduced conditioning sets in the subexpressions, $u_r(a_i|a_1, \ldots, a_{i-1})$, that comprise the additive decomposition provided by the chain rule, and to an elicitation process that mimics that of Bayes nets.

Graphical Models for CDI_r

In their seminal work on *Graphoids*, Pearl and Paz (1989) introduce an axiomatic characterization of probabilistic independence. In this section we show that CDI_r satisfies *all* of these properties, replacing the multiplication with summation where needed. This highlights the tight similarity between the two concepts, and allows us to introduce a corresponding undirected graphical model. To facilitate what follows, we divide the set of properties to two parts. [3]

[2] In contrast to Brafman and Engel (2009), the chain rule as adapted here *does not* require normalization of $u(\Gamma^r)$ to zero.

[3] *decomposition* is used by Pearl and Paz as definition of independence, whereas their first property is used here as definition.

Theorem 3. *The following properties hold for CDI_r:*

$$\begin{aligned}
&\textit{(decomposition) } CDI_r(\gamma_1, \gamma_2|\gamma_3) \\
&\Leftrightarrow u_r(\gamma_1, \gamma_2|\gamma_3) = u_r(\gamma_1|\gamma_3) + u_r(\gamma_2|\gamma_3) && \text{(1a)} \\
&\qquad CDI_r(\gamma_1, \gamma_2|\gamma_3) \\
&\Leftrightarrow u_r(\gamma_1, \gamma_3|\gamma_2) = u_r(\gamma_1|\gamma_3) + u_r(\gamma_3|\gamma_2) && \text{(1b)} \\
&\qquad CDI_r(\gamma_1, \gamma_2|\gamma_3) \\
&\Leftrightarrow u_r(\gamma_1, \gamma_2, \gamma_3) = u_r(\gamma_1|\gamma_3) + u_r(\gamma_2, \gamma_3) && \text{(1c)} \\
&\textit{(chaining) } CDI_r(\gamma_1, \gamma_2|\gamma_3) \quad \& \quad CDI_r(\gamma_1\gamma_3, \gamma_4|\gamma_2) \\
&\qquad \Rightarrow CDI_r(\gamma_1, \gamma_4|\gamma_3) && \text{(1d)}
\end{aligned}$$

Proof. We derive the first property as an example, and omit the rest for space considerations.

$$\begin{aligned}
u_r(\gamma_1|\gamma_3) + u_r(\gamma_2|\gamma_3) &= u_r(\gamma_1, \gamma_2^r, \gamma_3) - u_r(\gamma_1^r, \gamma_2^r\gamma_3) \\
&+ u_r(\gamma_1^r, \gamma_2, \gamma_3) - u_r(\gamma_1^r, \gamma_2^r, \gamma_3) = u_r(\gamma_1, \gamma_2, \gamma_3) \\
&- u_r(\gamma_1^r, \gamma_2\gamma_3) + u_r(\gamma_1^r, \gamma_2, \gamma_3) - u_r(\gamma_1^r, \gamma_2^r, \gamma_3) \\
&= u_r(\gamma_1, \gamma_2, \gamma_3) - u_r(\gamma_1^r, \gamma_2^r\gamma_3) = u_r(\gamma_1, \gamma_2|\gamma_3)
\end{aligned}$$

The first and third equalities are by definition of conditional utility. The second equality is obtained by applying $\text{CDI}_r(\gamma_1, \gamma_2|\gamma_3)$ on the term $u_r(\gamma_1, \gamma_2^r, \gamma_3) - u_r(\gamma_1^r, \gamma_2^r\gamma_3)$. For the other direction, if $\text{CDI}_r(\gamma_1, \gamma_2|\gamma_3)$ does not hold there must exist values for γ_1, γ_2 violating this equality. □

The second set of properties is of more interest to us, foremost because it is the basis of a *semi-graphoid*.

Theorem 4. *The following properties hold for CDI_r:*

$$\begin{aligned}
&\textit{(symmetry) } CDI_r(\gamma_1, \gamma_2|\gamma_3) \Rightarrow CDI_r(\gamma_2, \gamma_1|\gamma_3) && \text{(2a)} \\
&\textit{(subsets) } CDI_r(\gamma_1, \gamma_2\gamma_3|\gamma_4) \\
&\Rightarrow CDI_r(\gamma_1, \gamma_2|\gamma_4) \quad \& \quad CDI_r(\gamma_1, \gamma_3|\gamma_4) && \text{(2b)} \\
&\textit{(union) } CDI_r(\gamma_1, \gamma_2\gamma_3|\gamma_4) \\
&\Rightarrow CDI_r(\gamma_1, \gamma_2|\gamma_3\gamma_4) && \text{(2c)} \\
&\textit{(contraction) } CDI_r(\gamma_1, \gamma_2|\gamma_3\gamma_4) \ \& \ CDI_r(\gamma_1, \gamma_4|\gamma_3) \\
&\Rightarrow CDI_r(\gamma_1, \gamma_2\gamma_4|\gamma_3) && \text{(2d)} \\
&\textit{(intersection) } CDI_r(\gamma_1, \gamma_2|\gamma_3\gamma_4) \ \& \ CDI_r(\gamma_1, \gamma_3|\gamma_2\gamma_4) \\
&\Rightarrow CDI_r(\gamma_1, \gamma_2\gamma_3|\gamma_4) && \text{(2e)}
\end{aligned}$$

Proof. The properties are derived from the definition. As an example, we show the proof for (2b) and (2e). By definition of $\text{CDI}_r(\gamma_1, \gamma_2\gamma_3|\gamma_4)$,

$$\begin{aligned}
u_r(\gamma_1, \gamma_2, \gamma_3^r, \gamma_4) - u_r(\gamma_1^r, \gamma_2, \gamma_3^r, \gamma_4) = \\
u_r(\gamma_1, \gamma_2^r, \gamma_3^r, \gamma_4) - u_r(\gamma_1^r, \gamma_2^r, \gamma_3^r, \gamma_4)
\end{aligned}$$

hence $\text{CDI}_r(\gamma_1, \gamma_2|\gamma_4)$ holds. Similarly $\text{CDI}_r(\gamma_1, \gamma_3|\gamma_4)$ is proved. To prove (2e), we apply $\text{CDI}_r(\gamma_1, \gamma_2|\gamma_3\gamma_4)$ and then $\text{CDI}_r(\gamma_1, \gamma_3|\gamma_2\gamma_4)$:

$$\begin{aligned}
u_r(\gamma_1, \gamma_2, \gamma_3, \gamma_4) - u_r(\gamma_1^r, \gamma_2, \gamma_3, \gamma_4) = \\
u_r(\gamma_1, \gamma_2^r, \gamma_3, \gamma_4) - u_r(\gamma_1^r, \gamma_2^r, \gamma_3, \gamma_4) = \\
u_r(\gamma_1, \gamma_2^r, \gamma_3^r, \gamma_4) - u_r(\gamma_1^r, \gamma_2^r, \gamma_3^r, \gamma_4),
\end{aligned}$$

and hence $\text{CDI}_r(\gamma_1, \gamma_2\gamma_3|\gamma_4)$. □

Any relation for which (2a) - (2d) hold is called a semi-graphoid. Hence by Theorem 4, CDI_r is a semi-graphoid. The importance of this observation becomes evident below.

The graphoid concept is a mapping of an independence relation to an undirected graph by means of node separation. Let $\mathcal{R}$ be a conditional independence relation over a set Γ. With some abuse of notation, we refer to elements of Γ directly as nodes in a graph $G = (\Gamma, E)$. We use $\langle \gamma_1, \gamma_2 \mid \gamma_3 \rangle_G$ to denote *graph separation*, that is, the case where any path in G from a node in $\gamma_1 \subseteq \Gamma$ to a node in $\gamma_2 \subseteq \Gamma$ must include a node in $\gamma_3 \subseteq \Gamma$.

Pearl and Paz define the following types of mappings:

Definition 7. *A graph $G = (\Gamma, E)$ is a D-Map (dependency map) of $\mathcal{R}$ if for any non-intersecting subsets $\gamma_1, \gamma_2, \gamma_3 \subset \Gamma$, if $\mathcal{R}(\gamma_1, \gamma_2 | \gamma_3)$ then $\langle \gamma_1, \gamma_2 \mid \gamma_3 \rangle_G$.*

Definition 8. *A graph $G = (V, E)$ is a I-Map (independency map) of $\mathcal{R}$ if for any non-intersecting subsets $\gamma_1, \gamma_2, \gamma_3 \subset \Gamma$, if $\langle \gamma_1, \gamma_2 \mid \gamma_3 \rangle_G$ then $\mathcal{R}(\gamma_1, \gamma_2 | \gamma_3)$.*

Pearl and Paz go on to show that probability independence *does not* have a *perfect map*, that is a graph which is both D-Map and I-Map. Given the technical similarity expressed by Theorem 3, it is not surprising that the same holds for CDI_r.

Theorem 5. *There exist utility functions for which there exists no graph which is both a D-Map and I-Map for CDI_r.*

A graph $G = (V, E)$ is a *minimal I-Map* if for any $e \in E$, the graph $G' = (V, E \setminus \{e\})$ is not an I-Map.

Theorem 6. *For any $u_r(\cdot)$ and CDI_r, there exists a minimal I-Map $G = (V, E)$, where $V = \Gamma$, and E defined by: $(a, a') \in E$ iff $\text{CDI}_r(a, a' | \Gamma \setminus \{a, a'\})$ does not hold.*

That is, in this graph there is an edge between any two nodes that are not independent given all other nodes. This is exactly the method used by Pearl and Paz to define a *Markov-Net* of a probability distribution. Hence, an undirected model for CDI_r can be constructed in an analogous way to Markov networks. Results related to Markov Nets transform to CDI_r, and in particular the concept of *Markov Boundaries*: For a given attribute a, the minimal set of attributes that is required to render a independent of the rest coincides with the set of neighbors of a in its minimal I-Map. As argued by Pearl and Paz, this helps considerably with "intuitive reasoning", that is, testing independence relationship at the level of human knowledge.

Directed Model

As evident from Theorem 5, there exists no graphical model that reflects CDI_r *perfectly*. In fact, we observe that the minimal I-Map G of CDI_r is exactly the perfect map of CDI. To see that, consider any two sets of attributes γ_1 and γ_2. The pair is separated in G iff for any $a_1 \in \gamma_1$ and $a_2 \in \gamma_2$ holds $\text{CDI}_r(a_1, a_2 | \Gamma \setminus \{a_1, a_2\})$, which is equivalent to $\text{CDI}(a_1, a_2)$, hence γ_1 and γ_2 are separated iff $\text{CDI}(\gamma_1, \gamma_2)$.

Therefore, while any independence relation implied by the undirected model corresponds to a valid CDI_r condition, it turns out that many independencies are "lost", in the sense that they are not reflected by this model; whenever two (sets of) attributes are independent only given a strict subset of the complement, the undirected model does not reflect the separation between them. We seek to recapture those subtle independencies, and hence we keep following the probabilistic example by moving to *directed* graphical models. A *directed acyclic graph (DAG)* models a dependence relation via the concept of *d-separation*. Definitions 9 and 10 are adapted from Pearl (1988).

Definition 9. *Let $\gamma_1, \gamma_2, \gamma_3$ denote disjoint subsets of Γ, the nodes of a DAG. γ_1 is* d-separated *from γ_2 given γ_3 (denoted $\langle \gamma_1, \gamma_2 \mid \gamma_3 \rangle_D$) if the exists no path between a node in γ_1 and a node in γ_2 which is* **not blocked** *by γ_3. Such path is not blocked only if (i) every node with converging arcs, or a descendant of one, is in γ_3,* **and** *(ii) any other node is outside γ_3.*

I-Maps and D-Maps are defined with respect to DAGs using d-separation in an analogous way to the directed model. In particular, a DAG D is a minimal I-Map of $\mathcal{R}$, if it is a minimal DAG such that for any three disjoint sets of attributes γ_1, γ_2, γ_3 for which $\langle \gamma_1, \gamma_2 \mid \gamma_3 \rangle_D$, it holds that $\mathcal{R}(\gamma_1, \gamma_2 | \gamma_3)$. A *Bayesian Network* is a probabilistic directed graphical model which is defined as a minimal I-Map to probabilistic independence. The following definition establishes a simple procedure to create a Bayesian network.

Definition 10. *Let $d = (a_1, \ldots, a_n)$ denote a variable ordering. For $i = 1 \ldots, n$, let $\Gamma^i = a_1, \ldots, a_{i-1}$. A boundary DAG for independence relation $\mathcal{R}$ is defined by designating the set of parents $Pa(a_i)$ to be the minimal within Γ^i such that $\text{CDI}_r(a_i, \Gamma^i \setminus Pa(a_i) | Pa(a_i))$ holds.*

We are now ready to import the richness of probabilistic graphical models to the world of utility functions, exploiting the tight similarity we established between the axiomatic properties of CDI_r and that of probabilistic independence.

Definition 11. *A marginal utility network is a minimal I-Map of CDI_r.*

The following key result is used by Pearl (1988) to construct Bayesian Networks as Boundary DAGs.

Theorem 7 (Verma 86). *Let $\mathcal{R}$ be a semi-graphoid. D is a boundary DAG of $\mathcal{R}$ relative to any ordering d, iff D is a minimal I-Map of $\mathcal{R}$.*

Now due to Theorem 4,

Corollary 1. *A boundary DAG of CDI_r is a marginal utility network.*

A marginal utility network can therefore be constructed similarly to a Bayesian network. We iterate over the attributes according to some ordering $a_1, \ldots, a_n$, and for each a_i find the minimal set of attributes within Γ^i to be designated as the parents. It is sufficient to check independence separately for each $a_j \in \Gamma^i$: if $\text{CDI}_r(a_i, a_j | \Gamma^i \setminus \{a_j\})$, then $a_j \notin Pa(a_i)$. The boundary DAG property $\text{CDI}_r(a_i, \Gamma^i \setminus Pa(a_i) | Pa(a_i))$ then holds due to the intersection property (2e). Otherwise, we must set $a_j \in Pa(a_i)$, and this is due to the union property (2c).

Furthermore, because of d-separation, we can provide an equivalent, alternative definition for a marginal utility net.

Let $Dn(a)$ denote the descendants of a. Let $Co(a) = \Gamma \setminus (Dn(a) \cup Pa(a) \cup \{a\})$.

Corollary 2. *A* marginal utility network *is a DAG* $D = (\Gamma, E)$*, such that for any* $a \in \Gamma$*,* $CDI_r(a, Co(a)|Pa(a))$*, for an arbitrary, fixed* $a_1^r \dots a_n^r \in \Theta$*.*

That is, any attribute is independent of its non-descendants, given its parents. This result unites our directed graphical model with the one suggested by Brafman and Engel (2009), but here the model is based on the firm foundations of graphoid theory, and opens the way to reasoning algorithms as detailed in the next section.

The graphical structure is populated by numeric data, in the form of *local* functions, that is in any node we have a utility function, which represents conditional marginal utility over the node, and depends on the node and its parents.

Definition 12. *A conditional utility table (CUT) for an attribute* a_i *is the function* $u_r(a_i|Pa(a_i))$*.*

Retrieving the utility of a given instantiation of Γ from the marginal utility network is linear in the number of attributes; a direct result of the chain rule and Corollary 2 is,

Theorem 8. *The utility function can be computed from the marginal utility network as follows*

$$u(\Gamma) = u_r(a_1) + \sum_{i=2}^{n} u_r(a_i|Pa(a_i))$$

The CUTs can be obtained by querying a user for preference differences. For example, the node a with parents γ requires the function $u_r(a|\gamma)$, which is obtained by queries for differences of the form $u_r(a'\gamma') - u_r(\gamma')$, the marginal utility of a', compared to its reference value, given γ'.

It is important to note that the process above, and our results in general, apply to any value or utility function. That is, the fact that we define conditional utility using utility differences does *not* imply that we restrict our attention to measurable value functions; the only assumption is that these are real-valued functions. However, the process of structure elicitation requires the user to assess utility independence properties, which require the user to assess utility/value differences. This task is more natural when the differences carry intuitive meaning.

Reasoning

With analogous graphical models, vast literature on reasoning with Bayesian networks can now be adapted to choice and preference problems, in a straightforward way. In this section we demonstrate that with several examples.

Utility Maximization

It is well-known that the problem of finding an optimal assignment for a utility function is similar, in its combinatorial nature, to the problem of finding *Most Probable Explanation (MPE)* for a probability distribution. For this reason, systematic MPE algorithms, which perform combinatorial optimization (usually using dynamic programming techniques such as variable elimination (Dechter 1996)), can usually be adapted to utility maximization. This is not the case with some *non-systematic* techniques, most celebrated of which is *Belief Propagation (BP)* (Pearl 1988). BP is mostly known as an algorithm for the task of belief assessment, but can be applied to MPE by replacing summation operators with maximization (Marinescu, Kask, and Dechter 2003). BP is not guaranteed to achieve optimum for general graphs, but in practice it has been often shown to be more useful than systematic methods. In particular, BP performs better than systematic optimization for graphs of large tree-width (Marinescu, Kask, and Dechter 2003), and is also much more amenable to distributed computing (Pearl 1988).

The basic operations of belief updating (or belief revision for MPE) in BP rely on the graphical independence structure, and on Bayes' rule. With the mathematical analogy of independence and Bayes' rule for marginal utility networks, BP can be applied to marginal utilities, rather than probabilities. We do not provide full exposition of BP messages, but exemplify the analogy for the simple case of a tree. In what follows, we use prime signs to indicate specific attribute values (as a' for a), use a^- as shorthand for $Dn(a)$, and use short form for maximization over domain, as in $\max_a u(a) = \max_{a' \in \mathcal{D}(a)} u(a')$. Let $c_1, \dots, c_k$ denote the children of a, and assume that the descendants of each child c_i are already "maxed out", that is c_i "knows" $\max_{c_i^-} u_r(c_i^-|c_i')$, for any $c_i' \in \mathcal{D}(c_i)$. Each c_i computes the message $\lambda_{c_i}(a')$ it sends to a, for each a':

$$\lambda_{c_i}(a') = \max_{c_i^-, c_i}(u_r(c_i'|a') + u_r(c_i^-|c_i')) = \max_{c_i}(u_r(c_i'|a') + \max_{c_i^-} u_r(c_i^-|c_i'))$$

This message is essentially the message used in Pearl's belief revision, simplified for trees, and replacing product with summation. Note that the first part of the term is available from the CUT of c_i, and the second part is the information we assumed c_i to have. We now show that a can compute its own information based on the messages from its children.

Lemma 9.

$$\max_{a^-} u_r(a^-|a') = \sum_{i=1}^{k} \lambda_{c_i}(a')$$

Proof. By Bayes' rule: $u_r(a^-|a') = u_r(a'|a^-) + u_r(a^-) - u_r(a')$. Since c_i separates a from c_i^-, and then using Bayes' rule again $u_r(a'|a^-) = u_r(a'|\{c_1, \dots, c_k\}) = u_r(\{c_1, \dots, c_k\}|a') + u_r(a') - u_r(\{c_1, \dots, c_k\})$.

By independence of the children given their parent, and property (1a), $u_r(\{c_1, \dots, c_k\}|a') = \sum_{i=1}^{k} u_r(c_i|a')$. By definition of conditional utility,

$$u_r(a^-) = u_r(\bigcup_{i=1}^{k} c_i^-|\{c_1, \dots, c_k\}) + u_r(\{c_1, \dots, c_k\}) = \sum_{i=1}^{k} u_r(c_i^-|c_i) + u_r(\{c_1, \dots, c_k\}).$$

The second equality is by property (1a), and using the fact that c_i separates c_i^- from c_j and c_j^-. Putting these pieces together, we get

$$u_r(a^-|a') = \sum_{i=1}^{k} (u_r(c_i|a') + u_r(c_i^-|c_i))$$

Taking max on both sides, and switching the order of max and sum on the right one (note that each element of the sum is maxed independently), we get the desired result. □

Hence a can compute $\max_{a^-} u_r(a^-|a')$. If a is a leaf, then $\max_{a^-} u_r(a^-|a') = 0$. By induction, at the root we obtain $\max_\Gamma u(\Gamma)$, solving the utility maximization problem. We note that the maximizing assignment (the $\arg\max$) must also be propagated with λ.

Slightly more generally, consider a *singly-connected graph (polytree)*. Here a^- denotes the set of nodes accessible to a through its children, and a^+ those accessible through its parents. In a polytree, $a^+ \cap a^- = \emptyset$. In addition to λ messages from children to parents, the algorithm requires messages from parents to children, denoted π. $u_r(a^-|a')$ is computed as before from λ messages, and $u_r(a'|a^+)$ is computed from π messages. Whenever a node a is propagated new information from its parents and children, this information can be combined to represent marginals on a, due to the following key property, which is again log-analogous to the case of probabilities.

Lemma 10.

$$u_r(a'|a^+, a^-) = u_r(a^-|a') + u_r(a'|a^+) + \alpha,$$

where α is a normalizing constant.

The messages used for the polytree MPE algorithm are also translated to utility maximization by replacing products with summations. For general graphs, BP is applied with iterative and loopy protocols, essentially using the same kind of messages, but no longer with guaranteed convergence to the optimal solution. However, those protocols are shown to perform well in practice as approximation (Yedidia, Freeman, and Weiss 2001) and anytime (Dechter, Kask, and Mateescu 2002) algorithms. Because, as shown above, the basic operations of BP can be performed on marginal utility networks, this vast literature on belief propagation methods can be imported to the world of utilities, where this richness has not been available before.

Constrained Utility Maximization

Pearl (1988) describes how Bayes nets can be leveraged for reasoning over a constrained environment. For example, a constraint might rule out a particular combination of two variables, a_1 and a_2. To accommodate that, one can create a dummy node $\hat{a}$, whose parents are a_1 and a_2, and define $p(\hat{a} = true \mid a_1, a_2) = 0$ for the forbidden joint assignment of a_1 and a_2. Inference can then be done with the additional evidence that $\hat{a} = true$.

Constrained utility maximization is an important reasoning task for preferences; the decision maker can rarely choose freely among all the outcomes in the cartesian product of the attribute domains. With marginal utility networks, we can employ the same method; we create a similar dummy node $\hat{a}$, with $u_r(\hat{a} = true|a_1, a_2)) = -\infty$ in the CUT for any forbidden combination of values to a_1, a_2. Now, we can apply efficient methods such as BP to solve this problem.

Conclusion

We show that the recently introduced concept of a reference-dependent utility function satisfies the semi-graphoid axioms and leads to a directed model that tightly resembles Bayes nets. This enables us to use the methodology developed for eliciting structure and values in Bayes nets – an appealing possibility for utility elicitation. We briefly discussed two possible applications of this idea, both demonstrate adaptation of Bayesian reasoning algorithms, but we believe that more exist, in the areas of inference and learning. It is not clear that all concepts that exist in the probabilistic realm apply in the case of utilities, but understanding the implications of such concepts can lead to significant results in the area of reasoning about preferences and utilities.

References

Abbas, A. 2005. Attribute dominance utility. *Decision Analysis* 2:185–206.

Bacchus, F., and Grove, A. 1995. Graphical models for preference and utility. In *UAI*.

Boutilier, C.; Bacchus, F.; and Brafman, R. I. 2001. UCP-networks: A directed graphical representation of conditional utilities. In *UAI*.

Brafman, R., and Engel, Y. 2009. Directional decomposition of multiattribute utility functions. In *Algorithmic Decision Theory*.

Dechter, R.; Kask, K.; and Mateescu, R. 2002. Iterative join-graph propagation. In *UAI*.

Dechter, R. 1996. Bucket elimination: A unifying framework for probabilistic inference. In *UAI*.

Dyer, J. S., and Sarin, R. K. 1979. Measurable multiattribute value functions. *Operations Research* 27:810–822.

Engel, Y., and Wellman, M. P. 2007. Generalized value decomposition and structured multiattribute auctions. In *ACM-EC*.

Engel, Y., and Wellman, M. P. 2008. CUI networks: A graphical representation for conditional utility independence. *J. of Artificial Intelligence Research* 31:83–112.

Fishburn, P. C. 1967. Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review* 8:335–342.

Gonzales, C., and Perny, P. 2004. GAI networks for utility elicitation. In *KR*.

Gorman, W. M. 1968. The structure of utility functions. *Review of Economic Studies* 35:367–390.

Keeney, R. L., and Raiffa, H. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley.

La Mura, P., and Shoham, Y. 1999. Expected utility networks. In *UAI*.

Marinescu, R.; Kask, K.; and Dechter, R. 2003. Systematic vs. non-systematic algorithms for solving the MPE task. In *UAI*.

Pearl, J., and Paz, A. 1989. Graphoids: A graph based logic for reasoning about relevance relations. In Du Boulay, B., ed., *Advances in Artificial Intelligence II*.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

Shoham, Y. 1997. A symmetric view of probabilities and utilities. In *IJCAI*.

Yedidia, J.; Freeman, W.; and Weiss, Y. 2001. Generalized belief propagation. In *NIPS*.