

RESEARCH ARTICLE

Reconciliation of Decision-Making Heuristics Based on Decision Trees Topologies and Incomplete Fuzzy Probabilities Sets

Karel Doubravsky^{1*}, Mirko Dohnal²

1 Department of Informatics, Faculty of Business and Management, Brno University of Technology, Brno, Czech Republic, **2** Department of Economics, Faculty of Business and Management, Brno University of Technology, Brno, Czech Republic

* doubravsky@fbm.vutbr.cz



OPEN ACCESS

Citation: Doubravsky K, Dohnal M (2015) Reconciliation of Decision-Making Heuristics Based on Decision Trees Topologies and Incomplete Fuzzy Probabilities Sets. PLoS ONE 10(7): e0131590. doi:10.1371/journal.pone.0131590

Editor: Yongtang Shi, Nankai University, CHINA

Received: February 13, 2015

Accepted: June 2, 2015

Published: July 9, 2015

Copyright: © 2015 Doubravsky, Dohnal. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This paper was supported by grants FP-S-15-2825 "Economic Determinants of Competitiveness of Enterprises in Central and Eastern Europe" and FP-S-15-2787 "Effective Use of ICT and Quantitative Methods for Business Processes Optimization" from the Internal Grant Agency at Brno University of Technology. Support was also received from the Internal Grant Agency at Brno University of Technology Grant number: FP-S-13-2148 "The Application of ICT and Mathematical Methods in Business Management" Grant Recipient: Institute of Informatics, Faculty of Business and Management

Abstract

Complex decision making tasks of different natures, e.g. economics, safety engineering, ecology and biology, are based on vague, sparse, partially inconsistent and subjective knowledge. Moreover, decision making economists / engineers are usually not willing to invest too much time into study of complex formal theories. They require such decisions which can be (re)checked by human like common sense reasoning. One important problem related to realistic decision making tasks are incomplete data sets required by the chosen decision making algorithm. This paper presents a relatively simple algorithm how some missing III (input information items) can be generated using mainly decision tree topologies and integrated into incomplete data sets. The algorithm is based on an easy to understand heuristics, e.g. *a longer decision tree sub-path is less probable*. This heuristic can solve decision problems under total ignorance, i.e. the decision tree topology is the only information available. But in a practice, isolated information items e.g. some vaguely known probabilities (e.g. fuzzy probabilities) are usually available. It means that a realistic problem is analysed under partial ignorance. The proposed algorithm reconciles topology related heuristics and additional fuzzy sets using fuzzy linear programming. The case study, represented by a tree with six lotteries and one fuzzy probability, is presented in details.

Introduction

There is a broad spectrum of decision-making tasks, e.g. engineering, economics, sociology, ecology, informatics etc., see e.g. [1–7]. These realistic decision-making tasks are often difficult to solve by limited available input information items (III).

Moreover, realistic industrial and natural science decision-making applications are based on vague data sets. For example, chemical, petrochemical and food processes take into decision making processes inaccurate measurements, e.g. gross systematic errors, chaotic changes of concentrations, flow rates, etc., see e.g. [8]. Bankruptcies, Loans and Investments are examples from economics. Management of many common resources, e.g. environment and safety

Responsibility: Doc. RNDr. Bedrich Puze, CSc. Additional support was provided by the Internal Grant Agency at Brno University of Technology Grant number: FP-S-13-2052 'Microeconomic and macroeconomic principles and their impact on corporate behaviour Grant Recipient: Institute of Economics, Faculty of Business and Management Responsibility: prof. Ing. Oldřich Rejnus, CSc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

problems, is implemented and solved via group interactions, which is best described by public goods games [9].

Most of decision-making tasks can be represented by single root trees, i.e. decision trees, and sets of available III, such as probabilities, penalties, plausibility, etc., see e.g. [8]. In real-world problems, the complete III set is usually not available; there is lack and uncertainty input information, see e.g. [10, 11].

If uncertainties cannot be quantified using a simple probabilistic way then the topic of possibility decision theory is often a natural one to consider, see e.g. [12, 13]. In many studies the problem of information uncertainty is handled using fuzzy sets, e.g. fuzzy numbers, see e.g. [14–17].

Problems related to under-specification / lack information are usually solved by metaheuristics, see e.g. [18, 19]. If all information items, e.g. probabilities, penalties, required by a chosen decision making algorithm are not available then the decision task is solved under the total ignorance. It means that the decision tree topology is the only input information available. The traditional algorithms cannot take into consideration the individual heterogeneity / features and the topology of complex systems / networks [20, 21]. This is partially correct for simple structures as trees as well. EBAs (Entropy-based algorithms) can be applied under III conditions [22]. They cover a broad spectrum of task of different natures, see e.g. [23, 24].

If isolated information items are available then the decision task is solved under the so-called partial ignorance. Problems under partial ignorance are usually solved by Bayes' Theorem (Bayesian), see e.g. [25, 26]. However well-established decision making algorithms, e.g. Bayesian [25, 27] cannot feasibly absorb these isolated additional information. This is particularly true if additional information items are represented by vague heuristics based on common sense reasoning.

Common sense heuristics are often the only available generators of III if the problem under total ignorance is solved. Such heuristic generates ARII (all required information items). However, the partial ignorance problem incorporates some AII (additional information items). ARII and AII are not consistent. A feasible reconciliation is inevitable. Common sense reasoning indicates that AII are more trustworthy. This fact must be taken into consideration.

Reconciliation algorithms are based on different formal tools, see e.g. [28, 29]. There are several incomplete decision making problems and the relevant algorithms how to solve them [4, 6, 30–32].

Decision making algorithms will be accepted by users from industry, business etc. if they are simple. A decision making under total ignorance is simple if relevant heuristics are simple. The bellow proposed algorithm is a simple variant how to solve decision problems under partial ignorance. The advantage of this approach lies in the fact that is based on easy to understand heuristics.

Heuristics

Decisions are often based on heuristics, see e.g. [33, 34]. This paper is based on an assumption that decision makers are ready to accept some general heuristics based on common sense reasoning. There are many possible heuristics, which can be mutually contradictory.

An example of a pair of mutually exclusive heuristics is

H1: A longer decision tree sub-path is less probable,

H2: A longer decision tree sub-path is more probable.

Both heuristics reflect some features of common sense reasoning. The heuristic H1 reflects, e.g. reliability and maintenance aspects [35]. A reliability interpretation of the heuristic H1 is:

Simpler Systems are More Reliable.

The heuristic H2 is usually much more specific. For example—If credit related tasks are studied then this heuristic becomes meaningful e.g.:

If a credit applicant satisfies more requirements then the final probability of the credit granting is higher.

This paper is based on the heuristics H1, however, it is a simple task to modify the below given theory using the heuristics H2.

The algorithm studied in this paper is based on a strong analogy between a water flow through a one root tree system of pipes and the decision tree of the same topology. Let us suppose that one litre of water is pumped into the root node of the decision tree each second and there is no accumulation of water inside the tree. The consequence is that one litre of water must leave the tree through its terminal nodes each second; see e.g. [36]. Flows through all branches of the tree under study must be balanced. The relevant balance equation for a node with *one* flow in, *k* flows out is written [14]:

$$IN_i = \sum_{j=1}^k OUT_j \tag{1}$$

where *IN* is flow into *i*th node and *OUT* are flows out of *i*th node. For example, the flow into node no. 1 is the sum of flows into nodes no. 4 and 5, see Fig 1. Flow balancing equations of a decision tree define a linear system of equations, see [14]:

$$A \cdot y^T = B \tag{2}$$

where *y* is a vector of flows.

This paper deals with the second case where the available information is not reliable and precise. There is a relationship between a flowrate through node and a probability of this node. Then a reinterpretation of the heuristic H1 is

- H3: The flowrate of water through a node is equal to its probability.

The following classical axiomatic definition of probability can be found in a number of standard texts [37]. Let *A* be an event, and Ω is a sample space. A probability $p(A)$ of event *A*, must satisfy three axioms:

$$\text{for all } A \in \Omega, p(A) \geq 0 \tag{3}$$

$$p(\Omega) = 1 \tag{4}$$

$$p(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n p(A_i) \tag{5}$$

The relevant water flows through a node satisfies the axioms (3)–(5).

Topological Resistance

A decision tree has one root node *r*, see e.g. Fig 1, where circles / nodes represent either lotteries or decisions [38].

The following definitions are used below:

T Set of terminals, see e.g. nodes 4, 6, 7 and 8, Fig 1.

N Set of all nodes.

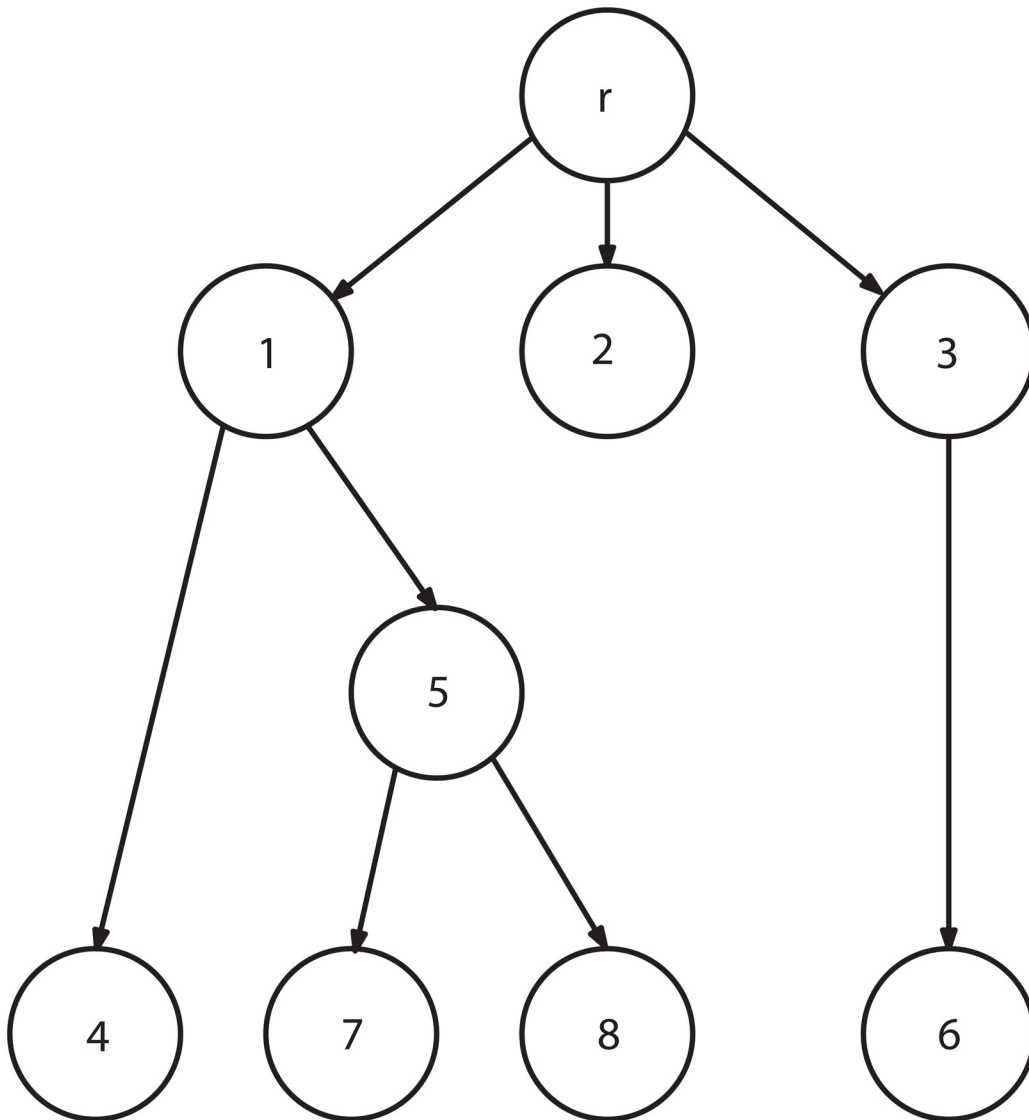


Fig 1. A decision tree.

doi:10.1371/journal.pone.0131590.g001

s_{ij} Number of edges of the sub-tree where i is the root of the sub-tree and j is the node next to the sub-root, see e.g. $s_{15} = 3$ namely the edges 1–5, 5–7, 5–8.

S_i Number of all edges of the sub-tree where i is the sub-root (water resistance of i th node).

$$S_i = \sum_j s_{ij}, \text{ see e.g. } S_1 = s_{14} + s_{15} = 1 + 3 = 4 \quad (6)$$

where j represents nearest downstream node of the sub-tree next to the i th node.

$$s_{ij}^* = S_i - s_{ij}, \text{ for all } i, j \in N - T, s_{15}^* = S_1 - s_{15} = 4 - 3 \quad (7)$$

S_i^* Number of edges of the i th sub-tree:

$$S_i^* = \sum_j s_{ij}^*, \text{ see e.g. } S_1^* = s_{15}^* + s_{14}^* = 4 \tag{8}$$

where j represents nearest downstream node of the sub-tree next to the i th node.

To each node an appropriate splitting fraction is assigned [39]. This paper is based on the following definition of the splitting fraction $\alpha_{i,j}$. Splitting ratio from i th node to j th node:

$$\alpha_{i,j} = s_{ij}^* / S_i^*, \text{ for all } j \in N - T, \text{ see e.g. } \alpha_{1,5} = s_{15}^* / S_1^* = 1/4 \tag{9}$$

p_j of j th terminal for $j \in N$ is a flowrate of water through j th node. The value P_r of a root node always equals one.

$$P_r = 1 \tag{10}$$

Non-root node probability is, see Fig 2.

$$P_j = P_i \cdot \alpha_{i,j}, j \in (N - T), \tag{11}$$

where i represents nearest upstream node (the sub-root node of the sub-tree). The Eq (11) is based on the balance Eq (1).

The set of $N-T$ linear Eq (11) (the set of balance equations) where the set P is a vector of unknown variables and the splitting ratios α (9) are numerical constants can be easily solved.

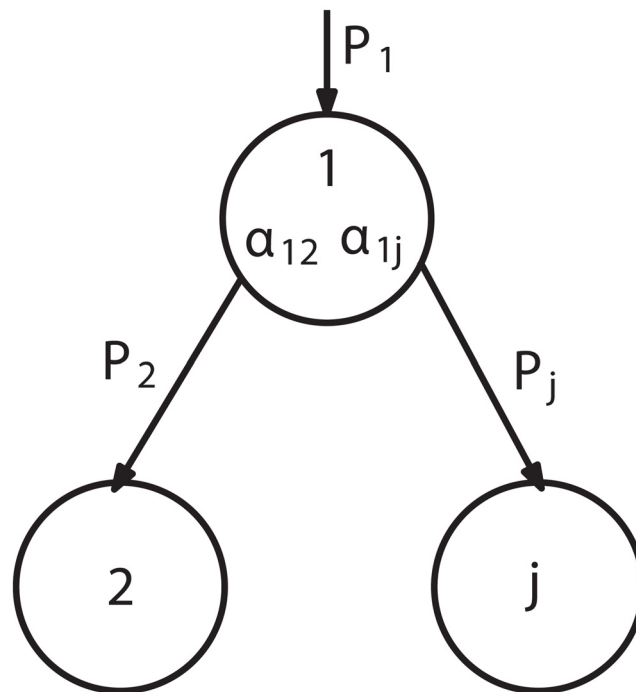


Fig 2. A flowrate of water through j th node.

doi:10.1371/journal.pone.0131590.g002

Example The heuristic H3 can be used to evaluate resistances $S(6)$, see Fig 1:

i th	node	s_{ij}	S_i
r		$s_{r1} = 5, s_{r2} = 1, s_{r3} = 2$	8
1		$s_{14} = 1, s_{15} = 3$	4
2			—
3		$s_{36} = 1$	1
4			—
5		$s_{57} = 1, s_{58} = 1$	2
6			—
7			—
8			—

where symbol “-” in third column indicates the i th node is a terminal.

The splitting ratios (9) are:

$$\begin{aligned}
 \alpha_{r,1} &= 3/16 = 0.187 \\
 \alpha_{r,2} &= 7/16 = 0.438 \\
 \alpha_{r,3} &= 6/16 = 0.375 \\
 \alpha_{r,1} + \alpha_{r,2} + \alpha_{r,3} &= 1 \\
 \alpha_{1,4} &= 3/4 = 0.750 \\
 \alpha_{1,5} &= 1/4 = 0.250 \\
 \alpha_{1,4} + \alpha_{1,5} &= 1 \\
 \alpha_{3,6} &= 1 \\
 \alpha_{5,7} &= 1/2 = 0.500 \\
 \alpha_{5,8} &= 1/2 = 0.500 \\
 \alpha_{5,7} + \alpha_{5,8} &= 1
 \end{aligned} \tag{12}$$

The system of linear Eq (11) with splitting ratios (12) is as follows, see Fig 1:

$$\begin{aligned}
 1 &= P_r \\
 0.187 \cdot P_r &= P_1 \\
 0.438 \cdot P_r &= P_2 \\
 0.375 \cdot P_r &= P_3 \\
 0.750 \cdot P_1 &= P_4 \\
 0.250 \cdot P_1 &= P_5 \\
 1 \cdot P_3 &= P_6 \\
 0.500 \cdot P_5 &= P_7 \\
 0.500 \cdot P_5 &= P_8
 \end{aligned} \tag{13}$$

Solving system of Eq (13) gives the following probabilities:

$$\begin{aligned}
 P_r &= 1.000 \\
 P_1 &= 0.187 \\
 P_2 &= 0.438 \\
 P_3 &= 0.375 \\
 P_4 &= 0.140 \\
 P_5 &= 0.047 \\
 P_6 &= 0.375 \\
 P_7 &= 0.024 \\
 P_8 &= 0.024
 \end{aligned}
 \tag{14}$$

Partial Ignorance

A typical feature of all realistic decision tasks is a shortage of information. However, it is not true that nothing is known. The concept of the total ignorance represented by e.g. the heuristic H1 helps to incorporate a set of isolated specific information items within a general framework of the heuristics H3.

The set of additional information can be expressed by an additional set of probabilities for some nodes:

$$\mathbf{R} \equiv (R_1, R_2, \dots, R_h)
 \tag{15}$$

has h elements, $h \leq N$. For example the probability of node 2, see Fig 1 can be expressed by verbal value ‘small’:

$$R_2 = \textit{small}
 \tag{16}$$

The additional value(s) see Eqs (15) and (16) are reconciled with the heuristic H3. The first reconciliation step is to prove that the reconciliation is inevitable. The additional data set \mathbf{R} Eq (15) is ignored, i.e. the total ignorance probabilities P_{TI} is solved. It means that the set of linear Eq (11) is solved and the set \mathbf{P}_{TI} is quantified, see e.g. Eq (13). If

$$\mathbf{P}_{TI} \textit{sufficiently equal} \mathbf{R}
 \tag{17}$$

then no reconciliation is required.

The concept sufficiently equal can be based on different soft calculus, e.g. fuzzy sets [40], rough sets. It is not studied in this paper.

Fuzzy Reconciliation of Balancing Task

There are different types of reconciliation [41–43]. Data reconciliation consists in measured or estimated quantities (in our paper probabilities) in order to balance the flows in a given decision tree. The vector of flows y can be subdivided into vectors x and u , i.e. known probabilities x and totally unknown probabilities u , see [14]. Let us denote x^* the vector of available probabilities. In general, the system $A(xu)^t = B$ has no solution. The solved problem is to modify x , while remaining as close as possible to x^* , such that the balance Eq (2), with $y = (xu)$, are satisfied [14, 44]. There are several approaches to solve reconciliation problem.

The traditional, i.e. least-squares, approach to the reconciliation considers measured data sets. Measurement errors follow a normal distribution with zero average and a diagonal covariance matrix [14]. The accuracy of each measurement x_i^* , denoted as a mean value, is characterized by its standard deviation σ_i . Data reconciliation becomes a problem of optimization under linear constraints. The solution of this problem is known by e.g. weighted least-squares, see [14, 45]. The main limitation of this approach is the fulfilment of the Gaussian hypothesis (a random variable must be unbounded). The other limitations are shown in [14].

Fuzzy interval reconciliation, see [14], is the approach based on estimated quantities (in our paper probabilities). Fuzzy interval reconciliation does not provide the same result as the least-square method. This approach provides only intervals instead of precise values. More or less possible values of each probability are limited by a fuzzy interval. Constraints of the mass balance Eq (2) are satisfied to a certain degree.

The problem of searching for a possible solution is based on a seeking an optimal position within all the (fuzzy) intervals of possible values, see [14]. This optimization problem can be solved by the max-min method, resolution methods such as α -cuts using fuzzy linear programming, see [8, 14, 46, 47]. The main limitation of this approach is the inability to detect a meaningful value of reconciliation (trade-off).

The reconciliation studied in this paper is based on fuzzy linear programming by [48] which provides a possibility to detect a meaningful value of reconciliation (trade-off). Its result is an optimal compromise:

Results of water probabilities (11) versus Additional probabilities (15).

The mathematical aspects of reconciliation are very important [49]. A fuzzy reconciliation is a solution of an over-specified set of equations:

$$\mathbf{A} \cdot \mathbf{P} = \mathbf{B} \cup \mathbf{P} = \mathbf{R} \tag{18}$$

where \mathbf{A} is the matrix of the splitting ratios (9) and the equations $\mathbf{A} \cdot \mathbf{P} = \mathbf{B}$ is given in (11), $\mathbf{P} = \mathbf{R}$ is the set of additional probabilities.

The set of Eq (18) has $n + h$ equations and n variables \mathbf{P} . The reconciliation is solved by a fuzzy linear programming [50–54].

Eq (18) are differently important / reliable and therefore their violations are differently accepted by decision makers. Rather often certain equations cannot be violated at all. For example probability of a terminal event, see e. g. node 4 in Fig 1, is known very accurately. Therefore the corresponding probability P_4 must be kept unchanged irrespective of any reconciliation algorithm.

Our experience indicates that the fuzzy set R_i is meaningfully characterized by triangular grades of membership [55], see e.g. Fig 3.

The triangular grades of membership can be expressed by a triplet (a, b, d) , where

$$a < b < d \tag{19}$$

Then the additional probability from (15) can be written as follows

$$P_i = R_i = (a, b, d) \tag{20}$$

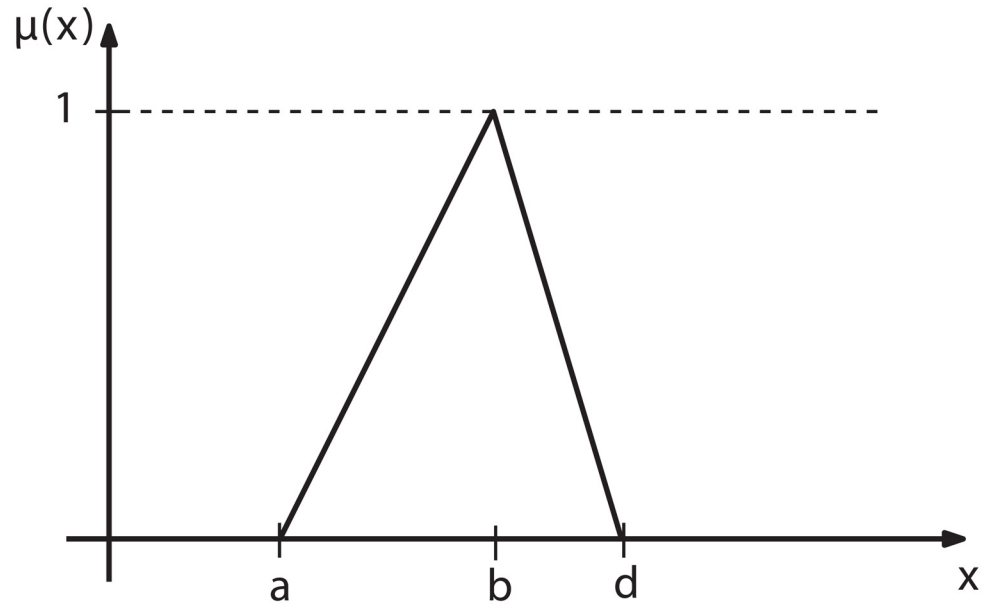


Fig 3. Triangular grades of membership.

doi:10.1371/journal.pone.0131590.g003

The *i*-th Eq (20) can be transformed into four linear inequalities [47] by introducing two vectors of slack variables S_{uj} and S_{lj} .

$$\sum_{i=1}^h (\alpha_{ij} \cdot P_i) + S_{uj} \geq b_j \tag{21}$$

$$\sum_{i=1}^h (\alpha_{ij} \cdot P_i) - S_{lj} \leq b_j \tag{22}$$

$$0 \leq S_{uj} \leq b_j - a_j \tag{23}$$

$$0 \leq S_{lj} \leq d_j - b_j \tag{24}$$

where

b_j, a_j, d_j - see Eq (19),

S_{uj} - is the *j*-th upper slack variable,

S_{lj} - is the *j*-th lower slack variable.

The set of inequalities (21)–(24) represents fuzzy description of *j*-th linear Eq (11).

The total ignorance can generate rather misleading decision making results. Therefore it is highly desirable to over specify the problem by using additional information items (15). This set of additional information allows us to cross check not only the heuristics but the additional information items among themselves as well.

One possible objective function which represents a meaningful reconciliation (trade-off) is

$$Q = \min_{S_u, S_l} \left(\sum_{j=1}^m (S_{uj}/(b_j - a_j) + S_{lj}/(d_j - b_j)) \right) \tag{25}$$

The minimization problem (25) can be finally solved as a conventional linear programming task [53].

For example the probability of node 2, see e.g. Fig 1, is defined as ‘small’ (16). This verbal value ‘small’ is expressed by a fuzzy set. The fuzzy set is characterized by its triangular grade of membership (19)

$$a = 0.1; b = c = 0.2; d = 0.25 \tag{26}$$

and can be written as $R_2 = (0.1;0.2;0.25)$. This equation can be transformed by (21)–(24) into four linear inequalities by introducing two slack variables S_u and S_l .

$$\begin{aligned} 0.1 \cdot P_2 + S_u &\geq 0.2 \\ 0.1 \cdot P_2 - S_l &\leq 0.2 \\ 0 \leq S_u &\leq 0.1 \\ 0 \leq S_l &\leq 0.05 \end{aligned} \tag{27}$$

The inequalities (27) are added to the balance Eq (11). The objective function, see Eq (25) is

$$Q = \min_{S_u, S_l} (S_u/0.1 + S_l/0.05) \tag{28}$$

The classical linear programming task (27), (28) is solved using available software e.g. GAMS [54].

The above described algorithm has the following advantages:

- It is simple and therefore transparent and easy to understand. Majority of well-established decision making and consequently reconciliation algorithms have not taken into consideration limits / extent of formal education (mathematics,

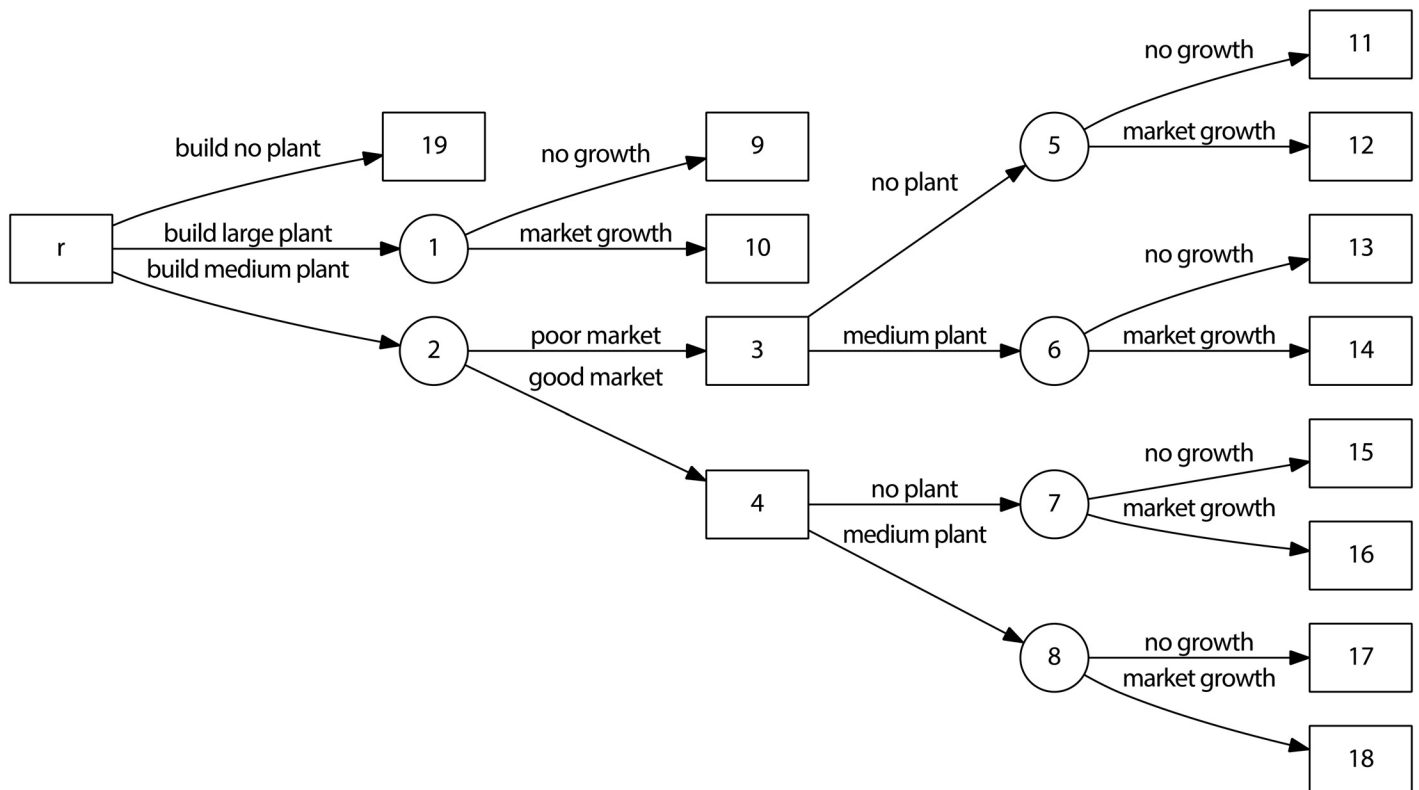


Fig 4. Decision Tree (Source: edited by [57]).

doi:10.1371/journal.pone.0131590.g004

Table 1. Profit and additional probabilities of terminal, see Fig 4.

Node	Probability	Profit (£)	Node	Probability	Profit (£)
1	-	-	11	-	-30,000
2	-	-	12	-	10,000
3	-	-	13	-	-70,000
4	-	-	14	-	-10,000
5	-	-	15	-	30,000
6	-	-	16	-	70,000
7	-	-	17	-	80,000
8	-	-	18	-	120,000
9	-	20,000	19	-	0
10	-	100,000			

doi:10.1371/journal.pone.0131590.t001

computer science) of decision makers in e.g. industry, investment and public sector [56]. However, just simple and therefore transparent algorithms will be used extensively in industry, etc.

- The heuristics can be easily modified / changed to fit into common sense reasoning of decision makers.

Table 2. Splitting ratios for all non-terminal nodes, see Fig 4.

Node	Splitting ratios	Node	Splitting ratios
r	$\alpha_{r,1} = 0.4211, \alpha_{r,2} = 0.1052, \alpha_{r,19} = 0.4736$	5	$\alpha_{5,11} = 0.5, \alpha_{5,12} = 0.5$
1	$\alpha_{1,9} = 0.5, \alpha_{1,10} = 0.5$	6	$\alpha_{6,13} = 0.5, \alpha_{6,14} = 0.5,$
2	$\alpha_{2,3} = 0.5, \alpha_{2,4} = 0.5$	7	$\alpha_{7,15} = 0.5, \alpha_{7,16} = 0.5,$
3	$\alpha_{3,5} = 0.5, \alpha_{3,6} = 0.5$	8	$\alpha_{8,17} = 0.5, \alpha_{8,18} = 0.5,$
4	$\alpha_{4,7} = 0.5, \alpha_{4,8} = 0.5$		

doi:10.1371/journal.pone.0131590.t002

Table 3. Node probabilities.

Node	Probability	Node	Probability
r	1	10	0.2106
1	0.4211	11	0.01315
2	0.1052	12	0.01315
3	0.0526	13	0.01315
4	0.0526	14	0.01315
5	0.0263	15	0.01315
6	0.0263	16	0.01315
7	0.0263	17	0.01315
8	0.0263	18	0.01315
9	0.2106	19	0.4736

doi:10.1371/journal.pone.0131590.t003

Table 4. Fuzzy Probabilities I, II, III for Good Market (R_d).

	a	b = c	D
I	0.05	0.07	0.08
II	0.04	0.09	0.10
III	0.11	0.16	0.17

doi:10.1371/journal.pone.0131590.t004

Table 5. Fuzzy approach—calculated probabilities.

Node	r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Probab. (I)	1	0.4211	0.1052	0.0352	0.07	0.0176	0.0176	0.035	0.035	0.2106	0.2106	0.0088	0.0088	0.0088	0.0088	0.0175	0.0175	0.0175	0.0175	0.0175	0.4736
Probab. (II)	1	0.4211	0.1052	0.05	0.055	0.025	0.025	0.0276	0.0276	0.2106	0.2106	0.0125	0.0125	0.0125	0.0125	0.0138	0.0138	0.0138	0.0138	0.0138	0.4736
Probab. (III)		No solution																			

doi:10.1371/journal.pone.0131590.t005

Table 6. Violation details.

	Equation	S_u	S_l	Value of membership function	Value of objective function Q, see Eq (25)
I	21 (P_4)	0.00	0.00	1.00	0.00
II	21 (P_4)	0.035	0.00	0.30	0.70
III	21 (P_4)	The corresponding linear programming task has <i>No Solution</i>			

doi:10.1371/journal.pone.0131590.t006

A set of potentially useful heuristics can be ad hoc generated for different branches of human activities, e.g. for safety studies—a complex system is less reliable. Such ad hoc heuristics can replace the general heuristics H1, H2 mentioned above.

- Complex decision trees can be easily treated.
If the problems under complete ignorance are solved then the linear programming problem is the only limit. Therefore it is possible to solve very large trees.
- Incorporation of additional constraints is simple.
Introduction of slack variables for each / some additional probability(ies) (20) allows us to have flexible control of the newly generated linear programming and perform, e.g. different sensitivity analyses.

Illustrative Example

A modified version of the original decision tree [57], see Fig 4 and Table 1, is self-explanatory. The numbers of arcs given in Fig 4 are used to identify the corresponding probabilities. The decision tree under study contains many nodes and edges, where circles mean lottery nodes and squares mean decision nodes [57]. The tree has 20 nodes and 19 edges. The nodes 9–19 are terminal nodes. Additional probabilities R (15) are not given. It means that the total ignorance problem is solved, see Table 1.

The heuristic H1 is used to assess the probabilities. The resulting splitting ratios (9) are given in Table 2.

The resulting probabilities are given in Table 3.

The problem under partial ignorance is represented by the total ignorance problem; see Table 1 and Fig 4, and one additional probability R. The only additional probability R is the probability of the *Good Market*, see Fig 4. The above described procedure (18)–(25) is used to solve three cases of *Good Market* probabilities, see Table 4.

There are 20 nodes in Fig 4, therefore 20 variables (P_1 – P_{19}) are used and 20 linear Eq (9) are compiled. The number of additional probabilities is 1. Therefore the total number of equations is $20 + 1 = 21$. The first fuzzy probability I, see Table 4, gives the following equation:

$$P_4 = R_4 = (0.05; 0.07; 0.08)$$

This equation is transformed, using Eqs (21)–(24), into the following four linear inequalities by introducing four slack variables S_{u1}, S_{l1} .

$$P_4 + S_{u1} \geq 0.07$$

$$P_4 - S_{l1} \leq 0.07$$

$$0 \leq S_{u1} \leq 0.02$$

$$0 \leq S_{l1} \leq 0.01$$

Table 7. Least-squares approach—calculated probabilities.

Node	r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Probab. (I)	1	0.420	0.0875	0.0293	0.058	0.0147	0.0147	0.0291	0.0291	0.2102	0.2102	0.0098	0.0098	0.0098	0.0098	0.0098	0.0170	0.0170	0.0170	0.0170	0.4727
Probab. (II)	1	0.421	0.0976	0.0463	0.051	0.0232	0.0232	0.0256	0.0256	0.2104	0.2104	0.0127	0.0127	0.0127	0.0127	0.0127	0.0139	0.0139	0.0139	0.0139	0.4732
Probab. (III)		No solution																			

doi:10.1371/journal.pone.0131590.t007

The system of 24 constraints (equations) is solved together with the objective function (25) using a linear programming algorithm. Table 5 gives the resulting probabilities based on the splitting ratios (9), balance Eq (11) and partial ignorance value for each fuzzy probability from Table 4.

The details of the calculations are listed in the Table 6.

The deviations S_u and S_l are zero for the first case I. It means that the corresponding equations (21)–(24) are satisfied perfectly. In the second case II, the corresponding equations are not satisfied perfectly. It means that it is a compromise solution (with value 0.7, see Table 6).

The evaluated probabilities, see Table 5, allow us to solve the decision making problem using some classical methods, e.g. the expected value criterion [57].

The first case I (Table 4) the expected value of profit (the expected value criterion) is 63 150 £. The company should build the medium plant in the first step. If the market is good then the company should build the other medium plant. If the market is poor then the company should build no plant.

The second case II (Table 4) the expected value of profit is 60 000 £. The company should build the large plant in the first step.

Table 7 gives the resulting probabilities based on the least-squares method for each initial probability, i.e. 0.07, 0.09 and 0.16, from Table 4. All calculations, see Table 7, were performed using GNU Octave, see e.g. [58].

Tables 5 and 7 show that the difference between the additional probabilities (15), see Table 4, and the resulting probabilities (11) is smaller if the fuzzy linear programming is used than in the case of the least-squares method.

Conclusion

Evaluations of realistic complex decision trees attract attention. Decision makers / field experts in their desperation to satisfy increasingly demanding laws and regulations (e.g. safety and environmental engineering) and / or pressure of competition (e.g. exchange rates hedging) are ready to believe that there is a theoretical answer to their needs. However, the only solution is to increase data / knowledge inputs into decision making processes. It means that no available information item may be ignored. Therefore known isolated fuzzy probabilities must be meaningfully incorporated into the decision making tasks.

The key reconciliation problem is the choice of the probabilities generation heuristic. If this heuristic is not accepted by a decision maker then some modifications of this heuristic are inevitable to cover specific requirements of the decision making problem under study. This is, however, an ad hoc procedure.

The heuristic used in this paper is based on a strong analogy between a water flow through a one root tree system of pipes and the decision tree of the same topology. The heuristic solves decision problems under total ignorance, i.e. the decision tree topology is the only information available. However, isolated information items e.g. some vaguely known probabilities (e.g. fuzzy probabilities) are usually available. It means that a realistic problem is analysed under partial ignorance. The present paper shows how a fuzzy linear programming is used to reconcile the probabilities generated by the water heuristic and the available set of fuzzy probabilities.

The presented algorithm is a simple variant how to solve decision problems under partial ignorance. The advantage of this approach lies in the fact that the objective function and the equations of the individual constrain are linear, and therefore easily solvable using commonly known simplex method. However, it is an easy task to incorporate not just fuzzy probabilities but e.g. fuzzy penalties and profits into trees evaluations / reconciliations. In mathematical

terms it means to use more complex fuzzy optimization algorithms [17, 48]. Moreover different heuristics can be used to generate missing values. The negative consequence is that multidimensional nonlinear optimization problems must be solved.

Decision trees can be generated using different algorithms. For example a genetic algorithm can induce decision trees [59]. It means that not just missing values but some trees themselves can be generated.

Limitation of proposed approach is the need to build decision tree (it can be large for extensive projects). Also it should be noted that the proposed heuristics H1 is not the only one possible and it may not suit all decision-making problems. The choice of suitable heuristics depends on the type of solved decision-making problem. It means it depends on the decision maker requirements. Water heuristic H1 can be replaced by more general entropy based heuristics using e.g. powers degrees [60].

The proposed approach have broad spectrum of applications, e.g. investment, biotechnology, chemical engineering and medicine. Moreover it is possible to take into consideration tasks where the probabilities are given vaguely, e.g. using fuzzy numbers, or they are specified from different sources, e.g. from different members of project teams, experts of project management, etc.

Author Contributions

Wrote the paper: KD MD.

References

1. Dondi R, El-Mabrouk N, Swenson K. Gene tree correction for reconciliation and species tree inference: Complexity and algorithms. *J Discrete Algorithms*. 2013; 25: 51–65 doi: [10.1016/j.jda.2013.06.001](https://doi.org/10.1016/j.jda.2013.06.001)
2. Vats S, Vats G, Vaish R, Kumar V. Selection of optimal electronic toll collection system for India: A subjective-fuzzy decision making approach. *Appl Soft Comput*. 2014; 21: 444–452. doi: [10.1016/j.asoc.2014.04.006](https://doi.org/10.1016/j.asoc.2014.04.006)
3. Daniel K, Hirshleifer D, Teoh S. Investor Psychology in Capital Markets: Evidence and Policy Implications. *J Monet Econ*. 2011; 48: 139–209. doi: [10.1016/S0304-3932\(01\)00091-5](https://doi.org/10.1016/S0304-3932(01)00091-5)
4. Hodouin D. Methods for automatic control, observation, and optimization in mineral processing plants. *J Process Control*. 2011; 21: 211–225. doi: [10.1016/j.procont.2010.10.016](https://doi.org/10.1016/j.procont.2010.10.016)
5. Maroney J, McGarry C, Ó hÓgartaigh C. Familiarity, home bias and investors' reactions to 20-F reconciliation gains and losses and perceptions of the quality of accounting principles. *Br Account Rev*. 2008; 40: 103–122. doi: [10.1016/j.bar.2008.01.001](https://doi.org/10.1016/j.bar.2008.01.001)
6. Vasebi A, Poulin É, Hodouin D. Dynamic data reconciliation in mineral and metallurgical plants. *Annu Rev Control*. 2012; 36: 235–243. doi: [10.1016/j.arcontrol.2012.09.005](https://doi.org/10.1016/j.arcontrol.2012.09.005)
7. Stiglic G, Kocbek S, Pernek I, Kokol P. Comprehensive decision tree models in bioinformatics. *PLoS ONE*. 2012; 7: s.e33812. doi: [10.1371/journal.pone.0033812](https://doi.org/10.1371/journal.pone.0033812)
8. Thipwivatpotjana P, Lodwick W. Pessimistic, optimistic, and minimax regret approaches for linear programs under uncertainty. *Fuzzy Optim Decis Mak*. 2014; 13: 151–171. doi: [10.1007/s10700-013-9171-z](https://doi.org/10.1007/s10700-013-9171-z)
9. Yuan WJ, Xia CY. Role of Investment Heterogeneity in the Cooperation on Spatial Public Goods Game. *PLoS One*. 2014; 9: e91012. doi: [10.1371/journal.pone.0091012](https://doi.org/10.1371/journal.pone.0091012) PMID: [24632779](https://pubmed.ncbi.nlm.nih.gov/24632779/)
10. Kuzilek J, Kremen V, Soucek F, Lhotska L. Independent Component Analysis and Decision Trees for ECG Holter Recording De-Noising. *PLoS ONE*. 2014; 9: e98450. doi: [10.1371/journal.pone.0098450](https://doi.org/10.1371/journal.pone.0098450) PMID: [24905359](https://pubmed.ncbi.nlm.nih.gov/24905359/)
11. Qin G, Luo L, Lv L, Xiao Y, Tu J, Tao L, et al. Decision tree analysis of traditional risk factors of carotid atherosclerosis and a cutpoint-based prevention strategy. *PLoS ONE*. 2014; 9: e111769. doi: [10.1371/journal.pone.0111769](https://doi.org/10.1371/journal.pone.0111769) PMID: [25398126](https://pubmed.ncbi.nlm.nih.gov/25398126/)
12. Dubois D. On various ways of tackling incomplete information in statistics. *Int J Approx Reason*. 2014; 55: 1570–1574.
13. Fargier H, Amor N, Guezguez W. On the Complexity of Decision Making in Possibilistic Decision Trees; 2012. Available: [arXiv:1202.3718](https://arxiv.org/abs/1202.3718). Accessed 17 January 2015.

14. Dubois D, Fargier H, Guyonnet D. Data Reconciliation under Fuzzy Constraints in Material Flow Analysis. Proceedings of the 8th conference of the European Society for Fuzzy Logic and Technology. Atlantis Press, 2013.
15. Lee Y, Li M, Yen T, Huang T. Analysis of fuzzy Decision Making Trial and Evaluation Laboratory on technology acceptance model. *Expert Syst Appl*. 2011; 38: 14407–14416. doi: [10.1016/j.eswa.2011.04.088](https://doi.org/10.1016/j.eswa.2011.04.088)
16. Pedrycz W, Al-Hmouz R, Morfeq A, Balamash A. Building granular fuzzy decision support systems. *Knowl Based Syst*. 2014; 58: 3–10. doi: [10.1016/j.knosys.2013.07.022](https://doi.org/10.1016/j.knosys.2013.07.022)
17. Zeinalkhani M, Eftekhari M. Fuzzy partitioning of continuous attributes through discretization methods to construct fuzzy decision tree classifiers. *Info Sci*. 2014; 278: 715–735. doi: [10.1016/j.ins.2014.03.087](https://doi.org/10.1016/j.ins.2014.03.087)
18. Jegadeesh N, Kim J, Krische S, Lee C. Analyzing the analysts: When do recommendations add value?. *J Finance*. 2004; 59: 1083–1124. doi: [10.1111/j.1540-6261.2004.00657.x](https://doi.org/10.1111/j.1540-6261.2004.00657.x)
19. Bradshaw M. How Do Analysts Use Their Earnings Forecasts in Generating Stock Recommendations?. SSRN Working Paper Series. 2004. doi: [10.2139/ssrn.256438](https://doi.org/10.2139/ssrn.256438)
20. Xia CY, Wang L, Sun S, Wang J. An SIR model with infection delay and propagation vector in complex networks. *Nonlinear Dynamics*. 2012; 69: 927–934. doi: [10.1007/s11071-011-0313-y](https://doi.org/10.1007/s11071-011-0313-y)
21. Xia CY, Sun SW, Liu ZX, Chen ZQ, Yuan ZZ. Epidemics of sirs model with nonuniform transmission on scale-free networks. *Int J Mod Phys B*. 2009; 23: 2203–2213. doi: [10.1142/S021797920905211X](https://doi.org/10.1142/S021797920905211X)
22. Dehmer M, Kraus V. On extremal properties of graph entropies. *Match-Commun Math Cmput Chem*. 2012; 68: 889–912.
23. Dehmer M, Grabner M. The Discrimination Power of Molecular Identification Numbers Revisited. *Match-Commun Math Cmput Chem*. 2013; 69: 785–794.
24. Kraus V, Dehmer M, Schutte M. On Sphere-Regular Graphs and the Extremality of Information-Theoretic Network Measures. *Match-Commun Math Cmput Chem*. 2013; 70: 885–900.
25. Kyburg H. Jr. Bayesian and non-bayesian evidential updating. *Artif Intell*. 1987; 31: 271–293.
26. Kim M, Cheeti A, Yoo C, Choo M, Paick J, Oh S. Non-invasive clinical parameters for the prediction of urodynamic bladder outlet obstruction: analysis using causal bayesian networks. *PLoS ONE*. 2014; 9: e113131. doi: [10.1371/journal.pone.0113131](https://doi.org/10.1371/journal.pone.0113131) PMID: [25397903](https://pubmed.ncbi.nlm.nih.gov/25397903/)
27. Wang P. The limitation of Bayesianism. *Artif Intel*. 2004; 158: 97–106. doi: [10.1016/j.artint.2003.09.003](https://doi.org/10.1016/j.artint.2003.09.003)
28. Sais F, Thomopoulos R. Ontology-aware prediction from rules: A reconciliation-based approach. *Knowl Based Syst*. 2014; 67: 117–130. doi: [10.1016/j.knosys.2014.05.023](https://doi.org/10.1016/j.knosys.2014.05.023)
29. Jung J. Computational reputation model based on selecting consensus choices: An empirical study on semantic wiki platform. *Expert Syst Appl*. 2012; 39: 9002–9007. doi: [10.1016/j.eswa.2012.02.035](https://doi.org/10.1016/j.eswa.2012.02.035)
30. Schallehn E, Sattler K, Saake G. Efficient similarity-based operations for data integration. *Data Knowl Eng*. 2004; 48: 361–387. doi: [10.1016/j.datak.2003.08.004](https://doi.org/10.1016/j.datak.2003.08.004)
31. Kashkoush M, ElMaraghy H. Matching Bills of Materials Using Tree Reconciliation. *Procedia CIRP*. 2013; 7: 169–174. doi: [10.1016/j.procir.2013.05.029](https://doi.org/10.1016/j.procir.2013.05.029)
32. Alhaj-Dibo M, Maquin D, Ragot J. Data reconciliation: A robust approach using a contaminated distribution. *Control Eng Pract*. 2008; 16: 159–170. doi: [10.1016/j.conengprac.2007.01.003](https://doi.org/10.1016/j.conengprac.2007.01.003)
33. Mousavi S, Gigerenzer G. Risk, uncertainty, and heuristics. *J Bus Res*. 2014; 67: 1671–1678. doi: [10.1016/j.jbusres.2014.02.013](https://doi.org/10.1016/j.jbusres.2014.02.013)
34. Sundar S, Singh A. New heuristic approaches for the dominating tree problem. *Appl Soft Comput*. 2013; 13: 4695–4703. doi: [10.1016/j.asoc.2013.07.014](https://doi.org/10.1016/j.asoc.2013.07.014)
35. Dohnal M, Vykydal J, Kvapilík M, Bures P. Practical uncertainty assessment of reasoning paths (fault trees) under total uncertainty ignorance. *J Loss Prev Process Ind*. 1992; 5: 125–131. doi: [10.1016/0950-4230\(92\)80009-W](https://doi.org/10.1016/0950-4230(92)80009-W)
36. Fitch EC, Hong I. Hydraulic component design and selection. Stillwater: BarDyne; 2004.
37. Dimitri P B, Tsitsiklis J N. Introduction to Probability. Belmont, Mass.: Athena Scientific; 2008.
38. Rokach L, Maimon O. Data Mining with Decision Trees: Theory and Applications. Singapore: World Scientific; 2008.
39. Rutkowski L, Jaworski M, Pietruczuk L, Duda P. The CART decision tree for mining data streams. *Information Sciences*. 2014; 266: 1–15. doi: [10.1016/j.ins.2013.12.060](https://doi.org/10.1016/j.ins.2013.12.060)
40. Parsons S, Dohnal M. The qualitative and semiquantitative analysis of environmental problems. *Environ Softw*. 1995; 10: 75–85. doi: [10.1016/0266-9838\(95\)00008-9](https://doi.org/10.1016/0266-9838(95)00008-9)

41. Janssens D, Wets G, Brijs T, Vanhoof K, Arentze T, Timmermans H. Integrating Bayesian networks and decision trees in a sequential rule-based transportation model. *Eur J Oper Res.* 2006; 175: 16–34. doi: [10.1016/j.ejor.2005.03.022](https://doi.org/10.1016/j.ejor.2005.03.022)
42. Vandaele N, Landrieux B, Vandaele C, Landrieux C, Decouttere C. The sustainable reconciliation between technological, financial and people aspects in R&D portfolio assessment. Proceedings of the 6th International Conference on Management of Innovation and Technology, ICMIT 2012. doi: [10.1109/ICMIT.2012.6225779](https://doi.org/10.1109/ICMIT.2012.6225779)
43. Narasimhan S, Jordache C. *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data.* 1st ed. Houston: Gulf Professional Publishing; 1999.
44. Monekosso D, Remagnino P. Data reconciliation in a smart home sensor network. *Expert Syst Appl.* 2013; 40: 3248–3255. doi: [10.1016/j.eswa.2012.12.037](https://doi.org/10.1016/j.eswa.2012.12.037)
45. Zhang Z, Shao Z, Chen X, Wang K, Qian J. Quasi-weighted least squares estimator for data reconciliation. *Comput Chem Eng.* 2010; 34: 154–162. doi: [10.1016/j.compchemeng.2009.09.007](https://doi.org/10.1016/j.compchemeng.2009.09.007)
46. Ubando A, Culaba A, Aviso K, Ng D, Tan R. Fuzzy mixed-integer linear programming model for optimizing a multi-functional bioenergy system with biochar production for negative carbon emissions. *Clean Technol Environ Policy.* 2014; 16: 1537–1549. doi: [10.1007/s10098-014-0721-z](https://doi.org/10.1007/s10098-014-0721-z)
47. Tan R, Briones L, Culaba A. Fuzzy data reconciliation in reacting and non-reacting process data for life cycle inventory analysis. *J Clean Prod.* 2007; 15: 944–949. doi: [10.1016/j.jclepro.2005.09.001](https://doi.org/10.1016/j.jclepro.2005.09.001)
48. Tanaka H, Asai K. Fuzzy linear programming problems with fuzzy numbers. *Fuzzy Sets Syst.* 1984; 13: 1–10. doi: [10.1016/0165-0114\(84\)90022-8](https://doi.org/10.1016/0165-0114(84)90022-8)
49. Watson S. The meaning of probability in probabilistic safety analysis. *Reliab Eng Syst Saf.* 1994; 45: 261–269. doi: [10.1016/0951-8320\(94\)90142-2](https://doi.org/10.1016/0951-8320(94)90142-2)
50. Fedrizzi M, Kacprzyk J, Verdegay JL. A survey of fuzzy optimization and mathematical programming. In: Fedrizzi M, Kacprzyk J, Roubens M, editors. *Interactive Fuzzy Optimization.* Lecture Notes in Economics and Mathematical Systems. Berlin: Springer; 1991. pp. 15–28.
51. Lai YJ, Hwang CI. *Fuzzy mathematical programming: Methods and applications.* Berlin: Springer-Verlag; 1992.
52. Guua S, Wu Y. Two-phase approach for solving the fuzzy linear programming problems. *Fuzzy Sets Syst.* 1999; 107: 191–195. doi: [10.1016/S0165-0114\(97\)00304-7](https://doi.org/10.1016/S0165-0114(97)00304-7)
53. Huang G, Moore RD. Grey linear programming, its solving approach, and its application. *Int J Syst Sci.* 1993; 24: 159–172. doi: [10.1080/00207729308949477](https://doi.org/10.1080/00207729308949477)
54. Brooke A, Kendrick D, Meeraus A. *GAMS: User's Guide.* Redwood City, California: The Scientific Press; 1988.
55. Kikuchi S. A method to defuzzify the fuzzy number: Transportation problem application. *Fuzzy Sets Syst.* 2000; 116: 3–9. doi: [10.1016/S0165-0114\(99\)00033-0](https://doi.org/10.1016/S0165-0114(99)00033-0)
56. Nie G, Zhang L, Liu Y, Zheng X, Shi Y. Decision analysis of data mining project based on Bayesian risk. *Expert Syst Appl.* 2009; 36: 4589–4594. doi: [10.1016/j.eswa.2008.05.014](https://doi.org/10.1016/j.eswa.2008.05.014)
57. Rose LM. *Engineering investment decisions: planning under uncertainty.* Amsterdam: Elsevier; 1976.
58. Eaton JW, Bateman D, Hauberg S. *Gnu Octave Version 3.0.1 Manual: A High-Level Interactive Language for Numerical Computations.* 1st ed. CreateSpace Independent Publishing Platform; 2009.
59. Basgalupp MP, Barros RC, De Carvalho A, Freitas AA. Evolving decision trees with beam search-based initialization and lexicographic multi-objective evaluation. *Inf Sci.* 2014; 258: 160–181. doi: [10.1016/j.ins.2013.07.025](https://doi.org/10.1016/j.ins.2013.07.025)
60. Cao S, Dehmer M, Shi Y. Extremality of degree-based graph entropies. *Inf Sci.* 2014; 278: 22–33. doi: [10.1016/j.ins.2014.03.133](https://doi.org/10.1016/j.ins.2014.03.133)