Technological University Dublin
ARROW@TU Dublin

Articles | School of Computer Sciences

2023

Comparing and Extending the Use of Defeasible Argumentation with Quantitative Data in Real-World Contexts

Lucas Rizzo
Technological University Dublin, lucas.rizzo@tudublin.ie

Luca Longo
Technological University Dublin, luca.longo@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/scschcomart

Part of the Computer Engineering Commons

Recommended Citation

Rizzo, Lucas and Longo, Luca, "Comparing and Extending the Use of Defeasible Argumentation with Quantitative Data in Real-World Contexts" (2023). *Articles*. 185.
https://arrow.tudublin.ie/scschcomart/185

This Article is brought to you for free and open access by the School of Computer Sciences at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, gerard.connolly@tudublin.ie, vera.kilshaw@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

Information Fusion 89 (2023) 537–566

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Comparing and extending the use of defeasible argumentation with quantitative data in real-world contexts

Lucas Rizzo *, Luca Longo

School of Computer Science, Technological University Dublin, Dublin, Ireland

ARTICLE INFO

Keywords:
Defeasible argumentation
Knowledge-based systems
Non-monotonic reasoning
Argumentation theory
Fuzzy logic
Expert systems
Computational trust

ABSTRACT

Dealing with uncertain, contradicting, and ambiguous information is still a central issue in Artificial Intelligence (AI). As a result, many formalisms have been proposed or adapted so as to consider non-monotonicity. A non-monotonic formalism is one that allows the retraction of previous conclusions or claims, from premises, in light of new evidence, offering some desirable flexibility when dealing with uncertainty. Among possible options, knowledge-base, non-monotonic reasoning approaches have seen their use being increased in practice. Nonetheless, only a limited number of works and researchers have performed any sort of comparison among them. This research article focuses on evaluating the inferential capacity of defeasible argumentation, a formalism particularly envisioned for modelling non-monotonic reasoning. In addition to this, fuzzy reasoning and expert systems, extended for handling non-monotonicity of reasoning, are selected and employed as baselines, due to their vast and accepted use within the AI community. Computational trust was selected as the domain of application of such models. Trust is an ill-defined construct, hence, reasoning applied to the inference of trust can be seen as non-monotonic. Inference models were designed to assign trust scalars to editors of the Wikipedia project. Scalars assigned to recognised trustworthy editors provided the basis for the analysis of the models' inferential capacity according to evaluation metrics from the domain of computational trust. In particular, argument-based models demonstrated more robustness than those built upon the baselines despite the knowledge bases or datasets employed. This study contributes to the body of knowledge through the exploitation of defeasible argumentation and its comparison to similar approaches. It provides publicly implementations for the designed models of inference, which might be a useful aid to scholars interested in performing non-monotonic reasoning activities. It adds to previous works, empirically enhancing the generalisability of defeasible argumentation as a compelling approach to reason with quantitative data and uncertain knowledge.

1. Introduction

Representing and manipulating knowledge with computers is still one of the main challenges in AI. This knowledge includes the ability to perform common sense reasoning, which is often non-monotonic [1]. Non-monotonic reasoning allows additional information to invalidate old claims or conclusions [1–6]. The classic non-monotonic reasoning example is given by '*birds fly*'. It is reasonable to assume that a particular bird, Tweety, flies, *unless* it is an exceptional bird: ostrich, duck, penguin, and so on [2]. This type of reasoning can be modelled in AI by several non-monotonic formalisms [1], such as inheritance networks with exception [7], semantic networks using Dempster's rule [8], non-monotonic logics [4,9,10] and knowledge-based systems [11]. Still, to the best of the authors' knowledge, there is an absence in the literature of empirical comparisons among some of these formalisms. This research article focuses on the comparison of knowledge-based, non-monotonic systems, which make use of rules or arguments supporting or contradicting certain conclusions to formalise non-monotonic reasoning. For instance, fuzzy reasoning [12] and expert systems [13] with the addition of non-monotonic layers, and computational argumentation, also referred to as defeasible argumentation [14,15]. All these approaches have led to the development of non-monotonic reasoning models usually based upon knowledge bases provided by human experts. Therefore, their performance depends on the amount and quality of knowledge available. However, they also allow such knowledge, possibly fragmented, vague, and non-algorithmic, to be represented in a natural and structured way [16]. Intuitively, this provides a higher degree of interpretability and transparency to the reasoning process. The reason for that is because these models attempt to use human language and to follow the way humans reason. This

* Corresponding author.
E-mail addresses: lucas.rizzo@tudublin.ie (L. Rizzo), luca.longo@tudublin.ie (L. Longo).

https://doi.org/10.1016/j.inffus.2022.08.025
Received 21 February 2022; Received in revised form 16 August 2022; Accepted 26 August 2022
Available online 3 September 2022
1566-2535/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

attempt potentially increases their explainability, which is essential for their adoption and usage. Nonetheless, these advantages have not been sufficient to increase the use of defeasible argumentation technology when performing quantitative reasoning under uncertainty. In this case, quantitative reasoning is understood as reasoning built with domain knowledge and performed on quantitative data, thus being able to provide numerical inferences. Bench-Capon and Dunne [17] identified a set of challenges that need to be overcome for achieving this goal, including the lack of a strong link between argumentation and other formalisms and the lack of engineering solutions for the application of argumentation. Another problem arises from the early stage of research trying to add quantitative approaches to argumentation [18,19]. Often, quantitative approaches in AI are deemed as limited for their inability to provide justifiable conclusions [20]. Hence, this study attempts to empirically evaluate the inferential capacity of defeasible argumentation models against other similar reasoning approaches. In addition to defeasible argumentation, non-monotonic fuzzy reasoning and expert systems are selected and employed as baselines, due to their vast and accepted use within the AI community.

To perform this comparison, the problem of representing the construct of computational trust has been chosen. Trust is a crucial human construct investigated by several disciplines, such as psychology, sociology, and philosophy [21]. It is an ill-defined construct, whose application lies in the domain of knowledge representation and reasoning. In this research article, the modelling of reasoning applied to the inference of computational trust is proposed in the context of the Wikipedia project. The goal is to design non-monotonic, knowledge-based models capable of assigning a trust value in the range $[0, 1] \subset \mathbb{R}$ to Wikipedia editors on a case-by-case basis. One means complete trust should be assigned to an editor, while 0 means an absence of trust assigned to the editor. These models are built upon domain knowledge and instantiated by quantitative data, thus can provide numerical inferences. Moreover, the domain knowledge employed contains pieces of evidence and arguments that can be withdrawn in light of new information, allowing the proposed assignment of trust to be seen as a form of defeasible reasoning activity.

It is expected that the comparison of non-monotonic reasoning approaches applied in the domain of computational trust will improve the perception of defeasible argumentation in relation to other similar alternatives. Moreover, such an experiment will also add to previous works that have made similar empirical comparisons in different domains of application, including human mental workload modelling [22,23] and mortality occurrence modelling [24,25]. These applications are connected by their presumptively non-monotonic nature. In other words, they all present incomplete, ambiguous and retractable pieces of evidence. Hence, reasoning applied to them is likely suitable for being modelled by non-monotonic reasoning systems. The experimentation with new applications is necessary to enhance the generalisability of defeasible argumentation as a compelling approach to reason with quantitative data and uncertain knowledge. In turn, this enhancement could possibly enable different applications and experiments, likely to be defeasibly modelled, to be carried out.

The remainder of this paper is organised as follows: Section 2 provides the related work on expert systems and fuzzy reasoning, including their options for handling non-monotonicity, followed then by defeasible argumentation and a short description of computational trust. Section 3 presents the design of the empirical experiment proposed to allow the envisioned comparison. Section 4 describes the results, performs its analysis and presents the respective discussion. Lastly, Section 5 concludes with a summary of the study, limitations, findings and recommendations of future research.

2. Literature and related work

In order to enhance the understanding of non-monotonic reasoning approaches, in particular defeasible argumentation, this section provides the reader with a precise description of the studied knowledge-based systems. Lastly, the section concludes with a short review of some works that have attempted to compare non-monotonic formalisms, followed by an introduction to computational trust, the chosen application domain for comparison purposes from a real-world context.

IF (antecedent) **THEN** (consequent)

Fig. 1. Typical form of rules employed in rule-based expert systems.

2.1. Expert systems

Succinctly, expert systems are defined as systems that try to transfer a vast body of specific knowledge from a human to a computer. They attempt to emulate such a human in a given field [13] and are aimed at accomplishing tasks that require human expertise or at playing the role of an assistant [26]. Their structure is usually composed of two internal components: a *knowledge base* and an *inference engine* [13] The former is provided by a human expert and generally translated into a set of logical rules. The latter is aimed at eliciting, firing and aggregating such rules towards a conclusive inference. Rules are used to define what to do and what to conclude in different scenarios. Usually, they follow the form depicted in Fig. 1.

These rules, in conjunction with a set of facts (for instance data) and an interpreter that decides the application of the rules, are what constitute a rule-based system. They can model a large range of problems, once the domain knowledge is represented as IF-THEN rules. It is important for the number of rules not to be too large so as to make the interpreter inefficient [27]. In this case, the system can exploit users' inputs and pieces of information stored in the knowledge base to reason with.

2.1.1. Non-monotonicity in expert systems

The use of non-monotonic logics in expert systems has been studied for several decades [28]. Nonetheless, the general use of non-monotonic reasoning in industry has not been extensive [29–31]. A few examples of expert systems that deal with non-monotonicity are proposed through the use of inheritance with exceptions in semantic networks [29], through the use of defeasible logic [32], through the use of default reasoning [33], and through the use of probabilistic reasoning [34]. In this research article, knowledge in expert systems is represented by rules, and the respective reasoning is performed in a single step. In other words, data is imported, and all rules are fired at once. Thus, to retract a rule, the notion of 'contradictions' or 'exceptions' are employed. These are defined by domain experts and describe special cases in which a rule is no longer valid. Once a special case is triggered, a backtrack search is employed to remove affected rules [31, chap. 9]. This might require excessive efforts, depending on the amount of data to be managed and the number of reasoning steps firing the backtrack search. However, note that even though this is a simplistic procedure, it can still be effectively implemented in a problem of single reasoning step, with a reasonable number of rules. That is the case with computational trust, the application chosen here for comparison purposes. Because reasoning applied to model computational trust can be performed in a single step with a reasonable number of rules, expert systems designed in this study do not follow a usual multi-step reasoning process. Further information on the domain of application and design methodologies will be detailed afterwards.

2.2. Fuzzy reasoning

Fuzzy reasoning models are the product of knowledge-based systems that incorporate fuzzy logic and/or fuzzy sets [12] into their reasoning and knowledge representation techniques [35]. Fuzzy logic calculus coupled with the notion of fuzzy membership functions allows for an

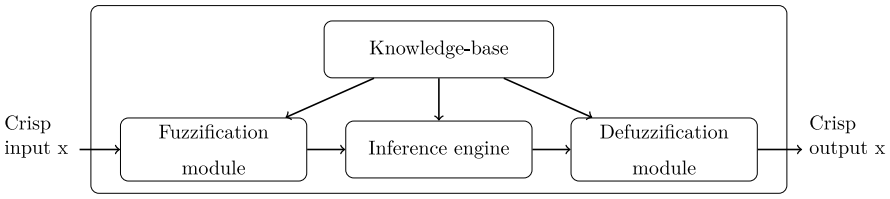


Fig. 2. General structure of a fuzzy control system.
Source: Adapted from [37].

effective knowledge representation of imprecise and uncertain pieces of information. Such notions can be exploited by IF-THEN rules of similar structure to those depicted in Fig. 1. In this case, a fuzzy rule-based system arises, which is most useful when modelling systems that make use of linguistic variables as their antecedents and consequents. In turn, rules can be employed for the construction of a fuzzy control system. Usually, such a system is composed of a set of crisp inputs, a knowledge base, a fuzzification module, an inference engine and a defuzzification module [36], as depicted by the diagram in Fig. 2.

This structure begins by the fuzzification module assessing the membership grades of crisp inputs associated with fuzzy sets. Subsequently, a fuzzy inference method must be applied, in order to produce an inference. Many of such inference methods are available [38–40]. This research article employs the 'Mamdani' fuzzy inference method [40], which is often used in practice [41]. Finally, defuzzification methods [42], such as centroid or mean max membership, must be applied to convert a fuzzy set into a crisp output in different fashions, contrarily to fuzzification methods that convert crisp inputs to fuzzy sets.

2.2.1. Non-monotonicity in Fuzzy reasoning

Some supplementary extensions of fuzzy inference systems have been suggested, in order to tackle the use of non-monotonic rules. Unfortunately, these extensions are few and not well established. For example, in [43], conflicting rules have their conclusions aggregated by a possible averaging function. However, an inference from a non-monotonic fuzzy rule cannot be propagated since the theory does not allow circularity. Another type of non-monotonicity in fuzzy systems is investigated in [44]. In this instance, non-monotonicity arises when identical consequents are inferred by distinct permutations of variables in the antecedents. The goal of this approach is to remove redundant rules, while preserving the crisp values returned by the defuzzification module of a Mamdani fuzzy system. A third approach is given by Siler and Buckley [45, chap. 8], whereby possibility theory [46] is included into the fuzzy reasoning system to tackle conflicting instructions. Siler and Buckley [45] make the assumption that 'truth values represent *necessity*, the extent to which the data support a proposition'. At the same time, they also treat 'truth values that represent *possibility*, the extent to which a truth value represents the extent to which the data fails to refute a proposition'. The possibility of proposition A is denoted $Pos(A)$, while its necessity is denoted $Nec(A)$. Both are values between $[0, 1] \subset \mathbb{R}$. Necessity is also assumed to represent the traditional truth values reviewed on the previous subsections, while truth values that represent possibility need to be added to the system. Moreover, it is assumed that adding supporting evidence can affect the necessity but not the possibility of a proposition, and adding contradicting evidence can never increase possibilities. In other words, $Nec(A) \leq Pos(A)$, for any proposition A. In this case it is guaranteed that propositions are defeasible. For example, if $Nec(a) = 1$ and $Pos(a) = 0$, it would not be possible to refute a. Under these circumstances, the effect on the necessity of a proposition a by a set of propositions $\{Q_1, \ldots, Q_k\}$ contradicting a, and a set of propositions $\{P_1, \ldots, P_j\}$ supporting a, is derivable as:

$$Nec(a) = (Nec(a) \bigcup_{j} Nec(P_j)) \bigcap_{k} (\neg Nec(Q_k)) \tag{1}$$

where $\neg Nec(Q_i) = 1 - Nec(Q_i)$, the union is implemented by the 'max' operator and the intersection by the 'min' operator. A few axioms used to develop conventional possibility theory are not considered in this approach, due to their incompatibility with other fuzzy logics. However, according to [45] the advantage provided is a functional theory, when incorporated into fuzzy reasoning with rule-based systems. For instance, suppose a proposition a whose necessity is 0.3 and possibility is 1.0; if P (necessity 0.4) supports a and Q (necessity 0.2) refutes a, then $Nec(a) = min(max(0.3, 0.4), 1 - 0.2)$. Thus, the new extent to which a support its truth is 0.4, because of the support of P and the failed attempt of refutation from Q, due to its low necessity. Unlike other reviewed implementations of non-monotonicity, note that this approach does not restrict the type of membership functions, methods of fuzzy inference, methods of defuzzification or propagation of inferences generated by non-monotonic rules. However, it does require that possibility values be defined. Simple approaches might be to assume possibility 1 for propositions that can be refuted by any other piece of information, and possibility 0 for propositions that cannot be refuted by any other piece of information.

As for applications of non-monotonic fuzzy systems, to the best of our knowledge, no paper has adopted these reviewed approaches in real-world contexts. Instead, they have been evaluated under simulated environments or by hypothetical examples. In particular, Siler and Buckley [45] exemplify its proposal of adding possibility theory into a fuzzy reasoning system with a simplified application in the medical field for demonstration purposes. The goal was to determine the anatomical significance of regions in an echocardiogram composed by ultrasound images of a beating heart. Rules were automatically created from a classification database. Contradictions arose from noisy images inferring mutually exclusive conclusions and were resolved with the aid of Eq. (1).

2.3. Defeasible argumentation

Argumentation deals with the study of assertion and definition of arguments usually emerged from divergent opinions. In AI, unlike expert systems and fuzzy reasoning, computational argumentation theory was introduced as a formalism for modelling non-monotonic reasoning. Thus, it does not require an implementation of non-monotonicity as described in the previous reasoning approaches. Argumentation, when aimed at developing computational models of arguments, might also be referred to as *defeasible argumentation* [15] — a paradigm that has become increasingly significant [17] and widely employed for modelling non-monotonic reasoning [47].

Computational argumentation systems are usually structured around layers specialising in the definition of internal structure of arguments, the definition of arguments interactions, the resolution of conflicts between arguments and the possible resolution strategies for reaching a justifiable conclusion [15]. However, as the boundaries of such layers might not be precisely defined, a few layered structures have been proposed [48,49] for the development of computational models of argument. Another example of multi-layered structure can be found in [18] and is depicted in Fig. 3.

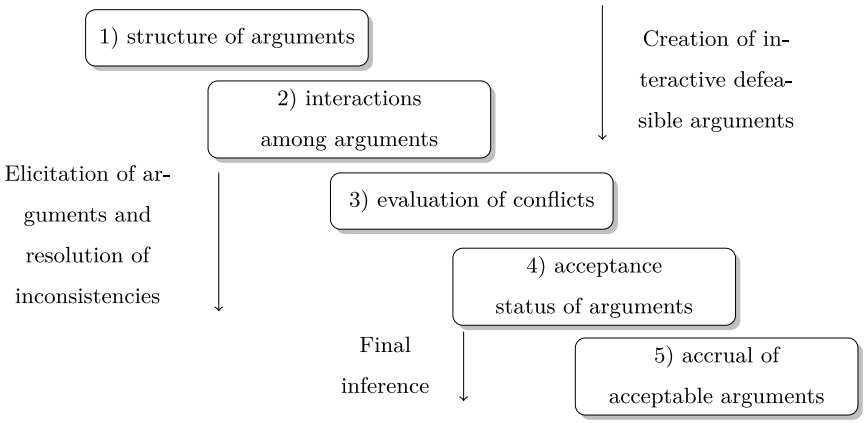


Fig. 3. Five layers structure [18] for the creation of argument-based models of inference.

The present research article adopts this structure, due to the nature of the experiment and application selected for comparison purposes (computational trust), which relies on the knowledge of a single expert, in order to achieve numerical inferences. In the literature of defeasible argumentation, many works are focused on each of these layers.

For instance, in Layer 1, Toulmin [50] was one of the first to introduce a conceptual model of argument with better structured arguments. Other structures are also possible [51,52] and they help to clarify possible ways in which arguments can be represented. Nonetheless, argumentation systems might still be constructed with simpler arguments, for instance when these are represented by a pair of premises and a conclusion.

In Layer 2, several works attempt to model the relationship between arguments and the management of their interactions, which means the actual arguing process. The classification proposed by Prakken [53] exemplifies three different classes: *undercutting* and *rebutting* attack, first formalised in [54], and *undermining* attack, introduced in [55]. An undermining attack refers to an argument being attacked on one of its premises, thus it is the only possible class of attack for deductive inferences. In contrast, the classes of undercutting and rebutting attack target respectively the inference link and the conclusion of an argument — structures which can be denied only in a defeasible argument. Another type of interaction is given by supporting relations between arguments, or *bipolarity* [56]. This concept is not employed in this research article, but the interested reader is referred to [57–60].

Subsequently in Layer 3, is possible to find works focused on characterising the success of an attack (or often referred to as *defeat*). Commonly, attacks have a form of binary relation. However, in order to determine a defeat, two other trends have been observed in the literature of argumentation: the preferentiality/strength of arguments and the preferentiality/strength of attacks. Strength of arguments is recognised as a valid source of information for those deciding on a collection of acceptable arguments [61], and can be employed in a variety of ways [62–64]. Dunne et al. [61] proposed, instead, the use of strength of attack relations. Their approach is justified by the fact that it is not only the strength of arguments that is important, but also the strength of the attack that one argument makes on another.

Once attacks have been evaluated the acceptance status of arguments can be defined through acceptability semantics in Layer 4. Conflicts by themselves do not demonstrate which arguments should be ultimately accepted. To do so, it is necessary to evaluate the overall interaction of arguments across the conflicting set. Most frequently, this evaluation relies on the abstract argumentation theory proposed by Dung [65] and later extended [66–68]. Hence, acceptability semantics return *extensions*, or subsets of arguments that can be mutually acceptable, according to a specific rationale of the semantics. Therefore, a notion of *scepticism* is usually employed in the informal discussion behind the behaviour of semantics [69]. For example, the grounded semantics is considered more *sceptical* for taking fewer committed choices and always providing a single *extension*. By contrast, the preferred semantics is seen as a more *credulous* approach for being more audacious when accepting arguments and has consequently been able to provide more than one extension. These notions are formally presented in Appendix A. Still, Dung's semantics and its variations are not the only class of semantics employed in abstract argumentation theory. Another well-known class of semantic is the ranking-based one [70–75]. In ranking-based semantics, the goal is not to provide an extension, but to rank arguments to define the most important one(s). It is a more flexible class of semantics, in the sense that arguments are not strictly rejected or accepted, but instead a graded assessment of arguments is provided, based on the topology of the argumentation framework. Appendix A covers the formal definitions of the semantics employed in this research article.

Finally, in Layer 5, the assessment of the statements supported by acceptable arguments is performed. When performing defeasible inferences, for practical purposes, usually a single decision needs to be made or a single action performed. However, multiple acceptable arguments may be computed in the previous steps. Usually, these coincide with possible consistent points of view that can be simultaneously considered for describing the knowledge being modelled. In the case of extension-based semantics, extensions might contain multiple arguments, multiple extensions might be computed, or both. In the case of ranking-based semantics, multiple arguments might be ranked at the top – a situation that can easily occur when multiple arguments are not attacked. Despite the suggestion in [76] of ranking the arguments by the number of extensions they belong to, this is not a problem usually addressed in the literature. Thus, in the present research, premises and conclusions of employed arguments are linked to categorical or numerical datasets. This allows for a simplified quantification of their values and aggregation in different fashions, namely average, sum or median. In turn, this aggregation produces a final inference (number), which is only possible due to the use of arguments built upon associated numerical datasets.

Some works make use of all the reviewed multi-layer structures (Fig. 3) in their systems [77–79], whereas others do not [80–82]. Table 1 lists several argument-based systems and their approaches for each layer in some prominent areas of argumentation. Some of these approaches may not have been previously reviewed, since they are beyond the scope of this study.

Table 1
Examples of argument-based systems and their internal configurations across layers.

Ref.	Application	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
[78]	Treatment outcomes	Inference rule	Rebuttals	Preference list	Extens. based	Utility theory
[79]	Healthcare	Natural language	Binary relation	Preference list	Preferred with pref.	User preference
[81]	Care planning	Toulmin	n/a	n/a	n/a	n/a
[83]	Mental workload	Inference rule + fuzzy logic	Undercutting + rebutting	Strength of att. and arg.	Grounded + Preferred	Extens. cardinality + fuzzy logic
[25]	Healthcare	Standard logic	Undercutting + rebutting	Strength of arg.	Grounded, Preferred, Categoriser	Extens. cardinality + average
[84]	Computational trust	Argument schemes	n/a	n/a	n/a	Sum aggregation function

2.4. Comparison of non-monotonic formalisms

A comprehensive comparison of different approaches to non-monotonic reasoning is a subject not sufficiently developed in the literature. In particular, to the best of the authors' knowledge, there is a very limited number of empirical works that incorporates the diversity of reasoning methodologies as done in this research article. Still, some works have proposed different comparisons among non-monotonic formalisms and are important to mention. For instance, Brewka et al. [85] provide a good overview in order to categorise non-monotonic logics by modal-preference logics, fixed-point logics and abductive methods. The recent work in [86] presents guidelines for selection of non-monotonic logics resulting in 17 different types of logics. Konolige [87] also studies the relation between non-monotonic logics, specifically between default and autoepistemic logic.

Delladio et al. [88] investigate the relations between a normal default logic and a variant of defeasible logic programming. The former is a special case in which justifications and consequents of default rules are the same, or in terms of default logic: $p(x) : c(x)/c(x)$. The latter is a formalism that combines logic programming and defeasible argumentation, to allow the representation of defeasible and non-defeasible rules. The authors show an equivalence between the consequents from the normal default logic and different answers given by the defeasible logic programming. Still, it investigates a theoretical relationship, limited by certain cases of the selected formalisms.

Dutilh Novaes and Veluwenkamp [89] make an empirical test on the accuracy of two formal non-monotonic reasoning models: preferential logic, a non-monotonic logic extended from a monotonic one; and screened belief revision, a particular version of belief revision theories [90]. The experiment attempts to demonstrate which of the two formalisms can better predict belief bias, a form of human reasoning. This examination is different from the investigation proposed in this research article, which instead assesses the inferential capacity of non-monotonic, knowledge-based reasoning approaches for quantitative inferences in real-world contexts.

More recently, Arieli et al. [91] perform a comparative study among logic-based approaches to formal argumentation and a theoretical discussion about the relations of these and other non-monotonic reasoning formalisms. For instance, it describes how certain extensions of autoepistemic logics [9] and default logic [10] could be translated to an extension of assumption-based argumentation [92], another formalism designed to capture and generalise non-monotonic reasoning.

Yang et al. [93] compare first-order predicate logic, fuzzy logic and non-monotonic logic implemented through negation as failure. The methods were contrasted using a simulation approach in which experiment facts were considered as random numbers. In turn, a set of algorithms is provided for their investigated transformations and modifications. This work possesses a very similar motivation to the present research article when comparing different reasoning formalisms, even if some are monotonic. It attempts to evaluate the capacity of inference and the complexity of these distinct approaches. Yet, despite proposing an interesting mechanism for experimentation, the study does not evaluate the subject in real-world domains. Instead, the simulation approach seems to elucidate the performance of such methods only in a computationally-controlled environment, for instance when different types of transformation must be applied to numerical datasets.

2.5. Computational trust modelling

Computational trust modelling, a knowledge-representation and reasoning problem, has been selected in order to allow the comparison of defeasible argumentation with other similar reasoning techniques. It has been investigated by several disciplines from different perspectives, such as psychology, sociology, and philosophy [21]. Many definitions of trust can be found in the literature [94]. Briefly, it can be described as a prediction that a trusted entity will bring to completion the expectations of a trustier in some specific context. The first computational model of trust was proposed in [95]. Its goal was to enable artificial agents to make trust-based decisions in the domain of Distributed Artificial Intelligence. The modelling of computational trust has also several applications in digital systems, for instance: reputation management [96,97]; social search and collective intelligence [98,99]; user behaviour modelling [100]; and self-adaptive recommendations [101,102].

Several works have examined the relation between argumentation and computational trust. For instance, Matt et al. [103] review how argumentation can help agents make decisions. It also discusses how arguments can improve the assessment of the trustworthiness of certain agents by supporting predictions on these agents' future behaviours. In turn, Parsons et al. [104] present a set of argument schemes for reasoning about trust. It is aimed at providing a computational mechanism for establishing arguments about trustworthiness. These schemes are also followed by a set of critical questions that can rule out their use. Other works have also focused on defining argument-based approaches for reasoning about trust [105,106].

In this research article, the context under evaluation comes from the Wikipedia project. Collaborative, user-generated content is of essential importance in the web. Hence, sites such as Wikipedia, TripAdvisor and Flickr leverage the interest and contribution of people all over the world. The drawback comes from the discrepant origin and quality of such contributions, leading to the complication of visitors and content moderators assessing their reliability. Wikipedia itself is under continual change from different sources, ranging from domain experts and to casual contributors, to vandals and committed editors. Therefore, many works have investigated the problem of computing the trust of Wikipedia editors or articles. For instance, Adler and de Alfaro [107] present a content-driven reputation system for Wikipedia editors, assuming that the reputation of editors can be used as a rough guide to the trust assigned to articles edited by them. In turn, reputation is assigned according to the longevity of the text inserted, and the longevity of the text edited by each editor. In a subsequent work, Adler et al. [108] compute the trust of a word in a Wikipedia article according to the reputation of the original editor of the word, as well as the reputation of editors who edited content in the vicinity of the word. The

study demonstrates that text labelled as high trust has a significantly lower chance of being edited in the future. Similarly, Zeng et al. [109] explore the revision history of an article to assess the trustworthiness of the article through a dynamic Bayesian network. In short, other works evaluate the trust of Wikipedia's contributors through a multi-agent trust model [110] and the Wikipedia editor reputation through the stability of content inserted [111].

To conclude, let us point out that the proposed use of defeasible argumentation and other reasoning approaches in this research article is *not* aimed at enhancing the assessment of computational trust. Hence, the performed experiments are not compared with the aforementioned works. Nonetheless, to the best of the author's knowledge, the use of non-monotonic reasoning, instantiated by quantitative information, for the inference of trust of Wikipedia editors has only been attempted in previous works [112,113]. These employed different sets of data and/or reasoning approaches. Thus, the investigation proposed in this research article is a more comprehensive one, extending these and other previous works [22,24,25] that have compared knowledge-based, non-monotonic reasoning approaches applied in different domains of application. Therefore, the main goal of this research article is to enhance the generalisability of defeasible argumentation as an effective approach to reason with quantitative, uncertain and conflicting information in real-world contexts. The next section provides a precise description of the research problem, the formulated hypothesis and the methods applied to test it.

3. Design and methodology

The research problem being addressed is how defeasible argumentation compares to similar reasoning approaches, such as expert systems and fuzzy reasoning, when used for the formalisation of non-monotonic reasoning models of inference. Moreover, this paper focuses on the case in which such models can be instantiated by quantitative data from real-world domains. Most recently, there seems to be an increase in the use of defeasible argumentation as the basis of current models employed in practice. Therefore, the assumption is that this approach could also be more suitable for modelling non-monotonic reasoning and producing non-monotonic reasoning models of inference in the domain of computational trust. The confirmation of this assumption would reinforce the applicability and generalisability of defeasible argumentation with quantitative data in real-world contexts. Consequently, it could also aid other scholars to adopt the proposed approach in other domains of applications. To investigate this, an *inductive type of research* is proposed — that is, one which attempts to propose broader generalisations from specific observations. An observation comes from the recent increased use of argument-based models in fields such as health care, knowledge-representation and reasoning, and multi-agent systems. This leads to the following hypothesis:

Hypothesis. If computational trust is modelled with defeasible argumentation, then the inferential capacity of its models will be superior than that achieved by non-monotonic fuzzy reasoning and expert systems models according to a predefined set of evaluation metrics from the domain of application.

To test this hypothesis, non-monotonic reasoning models of inference are designed and built with the pre-existing theoretical knowledge of the investigated reasoning approaches, as reviewed in Section 2. The goal is to design non-monotonic reasoning models capable of assigning a trust value in the range $[0, 1] \subset \mathbb{R}$ to Wikipedia editors. One means complete trust should be assigned to an editor, while 0 means an absence of trust assigned to the editor. In turn, such models are instantiated with real-world data, allowing a statistical comparison of produced inferences. Fig. 4 depicts the designed experiment with the evaluation phases incorporated into the flow. First, knowledge bases structured around natural language terms are employed by the non-monotonic reasoning approaches for the design of inferential models. These same models are later instantiated by real-world datasets from the domain of computational trust, producing three sets of inferences, one for each reasoning approach. These sets are subsequently analysed for the comparison of the inferential capacity of the reasoning approaches. This comparison is done by assessing the values assigned to Wikipedia Barnstar editors. A Barnstar[1] represents an award used by Wikipedia to recognise valuable editors. It is a non-automatic award bestowed from a Wikipedia editor to another Wikipedia editor. Therefore, it is not a ground truth for trust. Instead, it is used as a proxy measure to identify trustworthy editors and to allow the selection of evaluation metrics later detailed in this section.

3.1. Datasets employed and knowledge bases design

Wikipedia makes all its data available for download through HTML or XML dumps,[2] including articles, articles' history and complete text data. Hundreds of different language editions are available for download. Since no natural language information is analysed, but only quantitative data related to editors, the XML dump of the Portuguese-language[3] edition and the XML dump of the Italian-language[4] edition of the Wikipedia were selected for examination and downloaded on 8 January 2020. These were selected mainly due their respective sizes that were more appropriate to the computational resources available. Moreover, it was expected that two sets of data would be able to reinforce findings and confirm possible observed differences in the inferences produced by designed non-monotonic reasoning models. According to the Wikimedia Foundation's Analytics,[5] the Portuguese file contained 999,696 pages created (excluding pages being redirects), 1,947,023 editors, and 133 Barnstar editors, while the Italian file contained 1,576,621 articles, 2,804,142 editors, and 106 Barnstar editors, both up to January 2020. Each dumped Wikipedia page is identified by its title and it has a number of associated revisions containing: (i) its own ID; (ii) a time stamp; (iii) a contributor (editor) identified by a user name or IP address if anonymous; (iv) an optional commentary left by the editor; (v) the current number of bytes of the page on current revision; (vi) an optional tag indicating whether the revision is minor or major and should be reviewed by other editors. Fig. A.14 depicts an example of the XML structure. This data was extracted for the definition of features listed in Table 2 for each editor, resulting in two datasets.[6] Temporal factors such as presence, regularity and frequency factor were first proposed in [114]. A time window of 30 days was selected for evaluation of the frequency and regularity factors, similarly to the statistical analyses performed by the Wikimedia Foundation.

The set of extracted features was employed for the construction of two knowledge bases with the author's knowledge and intuition in this domain. Appendix B lists the information contained in them. The set of IF-THEN rules constructed was the same for both knowledge bases. They were defined intuitively and with the aid of external sources. For instance, the numerical ranges associated with natural language terms employed to describe `activity factor` and `bytes` were defined with the aid of the Wikimedia Foundation's Analytics. According to the reports of this foundation, an editor is considered a contributor if he/she has made more than 10 editions in his/her life cycle. Hence an `activity factor` ≥ 10 was used to infer *medium high* trust, while `activity factor` ≥ 20 was used to infer *high* trust. Similarly, the last report on the mean size of articles in the Portuguese Wikipedia showed a mean of 2388 bytes per article, while 90% of articles had more than

[1] https://en.wikipedia.org/wiki/Wikipedia:Barnstars.
[2] https://dumps.wikimedia.org/.
[3] File `ptwiki-20200820-stub-meta-history.xml`.
[4] File `itwiki-20200801-stub-meta-history.xml`.
[5] https://stats.wikimedia.org.
[6] https://doi.org/10.6084/m9.figshare.14939319.v1.

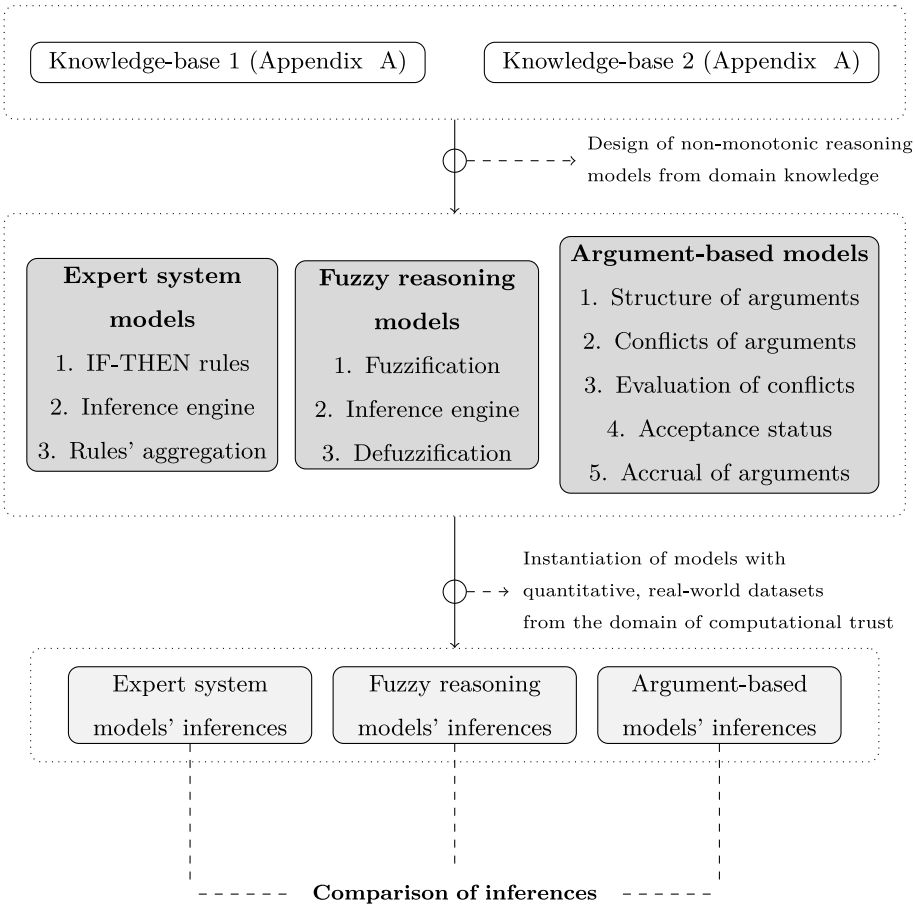


Fig. 4. Design of a comparative empirical research study and its evaluation schema aimed at evaluating the inferential capacity of defeasible argumentation.

Table 2
Summary of features employed by a human reasoner for trust assessment.

Feature	Description
Pages	Number of unique pages ($\mathbb{N}_{\geq 0}$) edited by the user.
Activity	Number of edits ($\mathbb{N}_{\geq 0}$) performed by the user.
Anonymity	Categorical value (Yes [1], No [0]) indicating whether the user is anonymous or not. Anonymous users are identified by their IP.
Not Minor	Ratio [0, 1] $\subset \mathbb{R}$ of edits flagged by the own editor for revision. 1 (0) means all (No) edits of the editor flagged by him or herself as not minor.
Comments	Ratio [0, 1] $\subset \mathbb{R}$ of edits in which a comment was included. One comment allowed per edition.
Presence	Ratio [0, 1] $\subset \mathbb{R}$ between the registration date of the user and the date of the beginning of the system (January 2001).
Frequency	Frequency ratio [0, 1] $\subset \mathbb{R}$ of edits per time window of 30 days in the editor's life cycle. Maximum value limited at 1.
Regularity	Regularity ratio [0, 1] $\subset \mathbb{R}$ per time window of 30 days. 1 means at least one interaction every 30 days in the editor's life cycle.
Bytes	Overall integer number of bytes edited by the user. Insertions/ deletions respectively increase/decrease the amount of bytes.

512 bytes. This information was used to intuitively infer *high* (*medium high*) if an editor had contributed at least 2388 bytes (512 bytes) throughout all his/her editions. Other features were normalised in the range [0, 1] $\subset \mathbb{R}$, in order to provide a more standard reasoning process. For instance, the feature `not minor` was divided by the `activity factor`, hence providing the percentage of editions flagged as major for all editions of the editor. Features normalised in the range [0, 1] $\subset \mathbb{R}$ were then described by four natural language terms: *low*, *medium low*, *medium high* and *high*. Table A.9 lists the feature transformations employed, the associated natural language terms and their respective numerical ranges. It is important to highlight that the constructed IF-THEN rules are limited by the type of data employed. While Wikipedia provides the text history of all its articles' editions, no natural language data is exploited in this study. A knowledge base also formed by features that take advantage of natural language data would likely contain stronger information for the inference of the editors' trust.

Following this, a set of contradictions among IF-THEN rules was defined. This process was made in two different ways, resulting in two different knowledge bases. The first was made by means of an intuitive manner, trying to establish evident relationships, such as low `frequency factor` contradicting the use of high `presence factor` to infer high trust. In other words, high `presence factor` was interpreted as an indication of high trust, unless the same editor also had a low `frequency factor`. A set of premises was also considered for the definition of agents whose trust should be low, such as a vandal or a bot. For example, an editor who is anonymous, has a low number of `comments`, a very low number of `not minor` edits, a high number of `pages` edited, and a high number of `bytes` inserted was considered a bot. In other words, this set of characteristics was considered sufficient to assume that an editor was a bot. In turn, this

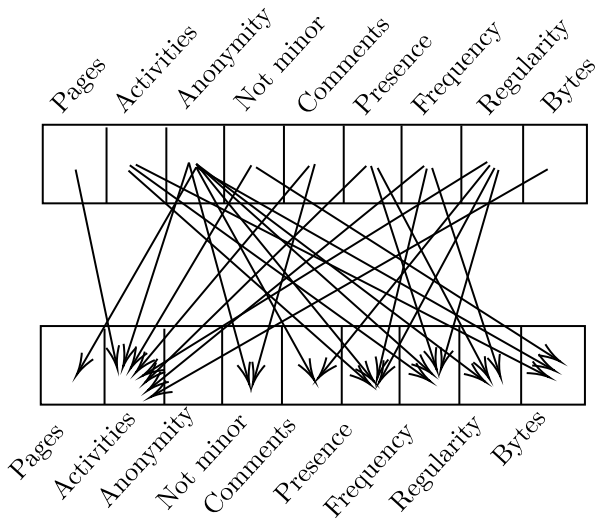


Fig. 5. Visual aid in the form of a relationship matrix employed by the knowledge base designer. Each arrow represents a possible conflict that could exist intuitively when employing two features to infer some level of computational trust.

set of premises was used to contradict several IF-THEN rules inferring high trust. This full knowledge base is reported in Appendix B, with the resulting graphical representation depicted in Fig. A.15.

In the second knowledge base designed, the contradictions among IF-THEN rules were identified with the visual aid of a relationship matrix, Fig. 5. It depicted the name of the features in two equal rows in order to help the author to identify pairs of features that could possibly have conflicts according to his/her knowledge. Many more contradictions were identified in this case, since it was easier to visualise all the possible pairs of features that could have some conflicting set of beliefs. Also note that it would be likely easier to identify the conflicts by adding the name of the conflicting feature in each box instead of the arrows. For example, the arrow from `pages` to `activity factor` was drawn as a reminder that while a high `activity factor` could be an indicator of high computational trust, this should not be the case if the editor has only modified a low number of pages. The idea is that a trustworthy editor would, based on the author's belief and intuition, collaborate on a high number of pages when performing a high number of edits. In addition, IF-THEN rules that did not contradict each other, but which inferred different trust levels, were considered this time to contradict each other, resulting in a much larger number of contradictions. The rationale for this was to assume that only one trust level should be accepted, demanding more from the conflict resolution strategy of each non-monotonic reasoning approach and ideally performing fewer calculations for the aggregation of rules/arguments before a final inference. Fig. A.16 depicts the resulting graphical representation of this knowledge base.

Fuzzy membership functions (FMF) were also designed by the author to model natural language terms such as high, low, very low, and others (Fig. A.17). These are necessary for the implementation of fuzzy reasoning models. Theoretically, such functions can provide higher precision for the modelling of natural language terms, or 'fuzzy' concepts. Natural language terms related to the same feature, for instance, low/medium low `frequency factor`, were designed in such a way that some intersection was possible between the terms. However, note that defining FMF is also a fuzzy process. Hence, two types of FMF were attempted, a linear and a gaussian one, which are often employed by fuzzy systems. The reason for having different types was to investigate their impact on the inferential capacity of fuzzy reasoning models. Other types could have been defined and further research can be made in this aspect.

In conclusion, note that these knowledge bases were built by a single agent, without the collaboration of other domain experts. Still, it is important to highlight that, despite such knowledge bases not being subject to a formal process of validation, for instance by being inspected by other experts, it is reasonable to assume that they are credible. This assumption comes from the fact that the author has more than 10 years of heavy internet usage, competent qualifications in computer science, and is experienced in a multitude of digital collaborative environments. A different approach would be to collaborate with a larger group of experts, producing greater and more sophisticated knowledge bases. Nonetheless, this approach could also lead to other issues, for example due to an expert's difficulty in understanding the knowledge used by computational models, an expert's capacity for verbalising knowledge, or an expert's capacity for understanding the amount of detail required [115]. In summary, the process of knowledge acquisition is a familiar problem and a frequent bottleneck of knowledge-driven approaches in general. It is not the goal of this research article to propose a solution for such issues. Rather, this research creates trustworthy, credible knowledge bases. In turn, it employs the same knowledge bases to perform the envisioned comparison among non-monotonic reasoning approaches.

3.2. Design of non-monotonic reasoning models employing expert systems

This section provides a step-by-step description of the inferential process of possible expert system models when applied for modelling a non-monotonic reasoning process in the domain of computational trust. A running example is depicted in Fig. 6. This example is referred to throughout this section and is aimed at providing a complete overview of the designed expert system inferential process.

3.2.1. IF-THEN rules

The first step of a rule-based expert system is to model a knowledge base usually gathered from an expert with rules in the form *"IF (antecedent) THEN (consequent)"*. In this research article, the antecedent is a set of premises associated with several quantitative features that are believed by the expert to influence the consequent being inferred (computational trust). The consequent might have different levels and is assumed to be derived from the premises by a domain expert. Therefore, no single deductive system is applied. The same premise/s might be used by different domain experts, but leading to different conclusions. Each level of a premise in the antecedent, as well as each level of the consequent, is mapped to a numerical range also by the domain expert. In this way, features associated to a certain level, such as low and high, can be evaluated true also according to continuous values. To formalise the generic case, IF-THEN rules are precisely defined. This definition follows the logic structure in which antecedents can have multiple premises joined with AND/OR boolean operators, while the consequent is a single statement [13, chap. 3].

Definition 1 (*Generic IF-THEN Rule*)**.** A generic IF-THEN rule is defined, without loss of generalisability for OR and AND operators, as:

$$\textbf{IF}\ \ (i_1 \in [l_1, u_1]\ \ \text{AND}\ \ i_2 \in [l_2, u_2]\ )\ \ \text{OR}\ \ (i_3 \in [l_3, u_3]\ \ \text{AND}\ \ i_4 \in [l_4, u_4])$$
$$\textbf{THEN}\ \ \textit{consequent}\ \ \in [l_c, u_c]$$

where $i_n \in \mathbb{R}$ is the input value of the feature n with numerical range $[l_n \in \mathbb{R}, u_n \in \mathbb{R}]$; the range $[l_c \in \mathbb{R}, u_c \in \mathbb{R}]$ is the numerical range for the consequent level being inferred with; and AND and OR are boolean logical operators.

3.2.2. Inference engine and non-monotonic extension

The inference engine starts with the activation of IF-THEN rules by input data. This data is used to evaluate antecedents of rules, activating a subset whose evaluation returns true. This evaluation is based on the numerical ranges provided by the knowledge base designer. For instance, in rule C4, *high* `comments` means that a `comments` input

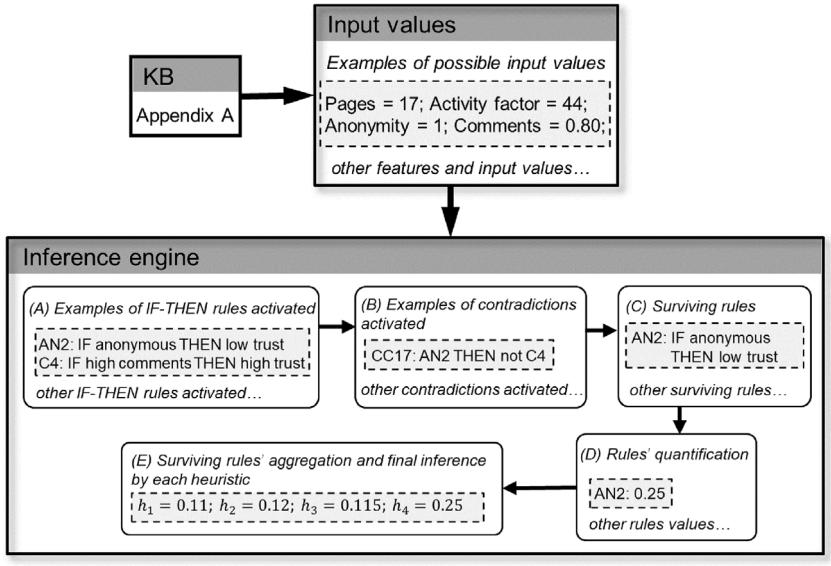


Fig. 6. An illustration of a reasoning process modelled by an expert system for a single Wikipedia editor.

value between [0.75, 1] $\subset \mathbb{R}$ provided by a user will activate the rule. A rule can also be contradicted by other rules that intend to bring forward and support contradictory information. Formally, these can also be seen as *meta-rules* [13, chap. 3], or rules that describe how other rules should be used, as in the contradictions in the running example. These are activated in the same fashion as IF-THEN rules. The main difference comes from the fact that the consequent of these meta-rules, or contradictions, might impact other meta-rules or other IF-THEN rules, while the consequent of IF-THEN rules is being employed only for the inference of computational trust. If both an IF-THEN rule and at least one contradiction challenging the rule have been activated, then the inference engine discards the rule. This mechanism will eventually form a set of surviving rules. The running example depicts some input values, two activated rules and one surviving rule.

In this research article, the rules in the set of surviving rules will always be inferring the same consequent, but most likely at different levels. Since the goal is to aggregate them and to extract a unique scalar, with the most representative of the consequents being inferred, an aggregation strategy is needed. In this situation, a usual expert system would have a typical set of choices for selection of rules [13, 116], for example, deciding a priority for each rule, returning multiple outcomes or choosing the first rule activated. However, none of these strategies are applicable in this research study. The constructed knowledge bases do not give explicit preferences among rules, order of activation or possibility to compute more than one output. Because of this, IF-THEN rules must be quantified and aggregated to infer a single scalar. In the present study, this value is defined according to the numerical range of the consequent of the rule, the numerical range of its premises and the input values provided for the rule activation. In the basic scenario of an IF-THEN rule with only one premise, it is quantified as the minimum (respective maximum) value of the numerical range of its consequent if its premise is activated with its minimum (respective maximum) value. For instance, consider rule C4 rewritten with illustrative numerical ranges:

– C4 rewritten: **IF** `comments` $\subset$ [0.75, 1] **THEN** trust $\subset$ [0.75, 1]

In this case, if the input value for `comments` is 0.75, then C4 value will be 0.75. Analogously, if the input value for `comments` is 1, then C4 value will be 1. Activation values greater than 0.75 and less than 1 are evaluated according to a linear relationship. This is defined by a function f as proposed in [113]:

Definition 2 (*Generic Rule Value*)**.** The value of a generic IF-THEN rule r is given by the function:

$$f(r) = \frac{|u_c - l_c|}{R_{max} - R_{min}} \times (v - R_{max}) + u_c,$$

where

$$v = min[max(i_1, i_2), max(i_3, i_4)],$$

$$R_{max} = min[max(u_1, u_2), max(u_3, u_4)],$$

$$R_{min} = min[max(l_1, l_2), max(l_3, l_4)]$$

The value of a rule will always lie between the numerical range [l_c, u_c] of its consequent. Moreover, the boundaries in this range will define the type of relationship between premises and consequent:

- If $l_c < u_c$, then Definition 2 will model a linear relationship. The higher the value of the premise/s, the higher the value of the conclusion.
- If $l_c > u_c$, then Definition 2 will model a contrary linear relationship. The higher the value of the premise/s, the lower the value of the conclusion.
- If $l_c = u_c$, then Definition 2 will model a constant function whose every input results in the same output (u_c). This might be useful to model consequents with categorical levels.

Briefly, Definition 2 provides an evaluation formula for rules that employ logical operators AND/OR, replacing them for *max* and *min* operators. Different operators could have been employed if required by the domain of application or human reasoner. Moreover, if there is no reason to select one operator over another, they could also be a parameter when designing expert system models. However, having in mind the rules contained in the knowledge bases employed in this study, the adoption of other operators would likely not have a significant impact on the results. The antecedents of these rules are often formed by a single premise, thus, their evaluation would follow a simple linear relationship regardless of the aggregation strategy adopted for multiple premises.

Finally, four heuristics defined in previous works [117] are employed for the aggregation of values assigned to surviving IF-THEN rules. The strategies are defined, in order to extract different points of view from remaining rules and to accommodate the use of rule weights when weights are provided. Weights among rules are provided in the

employed knowledge bases. They are the result of a pairwise comparison procedure between the 9 employed features (Table A.9) performed by the knowledge base designer. Hence, they will be numbers in the range [0, 8] $\subset \mathbb{N}$, being 0 if a feature is considered less important than any other feature for the inference of computational trust, and 8 if it is considered more important than any other feature. The weight of a feature will also represent the weight of the rule employing this feature. The aim is to investigate the impact of adding this extra information on the inferential capacity of the expert systems models. Thus, the heuristics are:

- h_1: definition of the sets of surviving rules grouped by their consequent level. Extraction of the largest set. Average of the values of the rules in the largest set. In case two or more of the largest sets exist, the above process is repeated for each, and their average is returned. The idea is to give importance to the largest set of surviving rules supporting the same consequent level.
- h_2: same as h_1 but applying the weighted average instead of the average. The goal here is to allow the possibility of defining weights to specific rules.
- h_3: average value of all surviving IF-THEN rules. This is to give equal importance to all surviving IF-THEN rules, regardless of which level of the consequent they were supporting.
- h_4: same as h_3 but applying the weighted average instead of the average. Again, the goal is to allow the use of weights attributed to specific rules.

The running example depicts an illustrative output for the four heuristics in the last step of the inference engine.

3.3. Design of non-monotonic reasoning models employing Fuzzy reasoning

Fuzzy reasoning provides a robust representation of linguistic information by using fuzzy membership functions. In this research article, the structure of a Mamdani fuzzy control system and the use of possibility theory as defined in [45, chap. 8], and reviewed in Section 2.2, are employed for the definition of fuzzy reasoning models of inference. As with expert systems, a running example of a single inference is depicted in Fig. 7 and referred throughout this subsection.

3.3.1. Fuzzification module

The first step, the fuzzification module, starts with the definition of *fuzzy* IF-THEN rules and *fuzzy* contradictions. Their structure is the same as that presented for expert systems, but they are computed in a different fashion. Each linguistic term associated with a feature level or consequent level, such as *high* or *low*, is then described by a FMF, which is also provided by the knowledge base designer. In the running example, some FMF employed are depicted in the fuzzification module.

3.3.2. Inference engine and a non-monotonic extension

Once the fuzzification step has been completed and the knowledge base of the expert translated into fuzzy IF-THEN rules and fuzzy contradictions, the next step is to evaluate the initial truth values of the fuzzy IF-THEN rules. To do so, each membership grade on the antecedent of these rules needs to be evaluated according to the input data. For instance, consider rule U2 in the running example. If `pages` = 17, then the membership grade of the linguist term *medium high* is 0.75, according to the FMF for `pages`. In this case, the initial truth value of U2, before solving contradictions, is also 0.75. U2 is a simplified example, but more than one feature can be contained in each rule's antecedents. Hence, a t-conorm and a t-norm are necessary to implement the notions of union and intersection. In the literature of fuzzy logic, t-norms (fuzzy intersection) and t-conorms (fuzzy union) [118], are a function $F : [0,1]^2 \rightarrow [0,1]$ such that the axioms of commutativity, associativity, monotonicity, and boundary condition are satisfied. Most commonly, the Zadeh, the Product and the Łukasiewicz operators are employed and are those selected for investigation in this research study, as listed in Table 3. Other operators can be seen in [118].

Table 3
T-norms and t-conorms usually employed by fuzzy systems. $x, y \in [0,1]$.

Fuzzy operator	T-Norm	T-Conorm
Zadeh	$T_m(x,y) = \min(x,\ y)$	$C_m(x,y) = \max(x,\ y)$
Łukasiewicz	$T_L(x,y) = \max(x + y - 1,\ 0)$	$C_L(x,y) = \min(x + y,\ 1)$
Product	$T_p(x,y) = x \cdot y$	$C_p(x,y) = x + y - x \cdot y$

Following the calculation of the fuzzy IF-THEN rules' initial truth values, possibility theory is adopted for the resolution of contradictions. According to the approach proposed in [45], truth values are used to represent *possibility* (Pos) and *necessity* (Nec) as defined in Section 2.2.1. In this study, necessity is represented by the membership grade of a proposition and possibility is always 1 for all propositions. This means that all propositions (or rules) are open to be retracted. Since there is no addition of supporting information, but only attempts to contradict or refute information, it is possible to employ Eq. (1) to deal with the contradictions in the knowledge bases of this study. For instance, the new necessity of fuzzy rule U2, if it is contradicted only by the fuzzy contradiction CC14, is given by:

- Nec(U2) = min (Nec(*medium high* `pages`), 1 - Nec(`anonymous`))

Nec(*medium high* `pages`) is the membership grade of the linguistic variable *medium high* `pages`. Three situations might arise in this case:

- If Nec(`anonymous`) = 0, then CC14 has no impact on the necessity of U2.
- If Nec(`anonymous`) = 1, then U2 is refuted and assumes new necessity 0.
- If 0 < Nec(`anonymous`) < 1, then U2 can only maintain the same necessity or have it decreased to a value greater than 0, indicating a partial refutation.

The new necessity of the fuzzy rule U2 represents the truth value of *medium high* `trust` in this rule. However, it is important to highlight that the approach developed in [45] has been inspired by a multi-step forward-chaining reasoning system. In this research study, reasoning is done in a single step, in the sense that data is imported, and all rules are fired at once. Nonetheless, it is possible to define a precedence order of fuzzy contradictions. More precisely, it is possible to define a tree structure in which the consequent of a fuzzy contradiction is the antecedent of the next fuzzy contradiction. In this way, Eq. (1) can be applied from the root or roots to the leaves. In case of cyclic contradictions, or contradictions whose consequents impact each other's premises, they are solved simultaneously. For that, the truth value of all fuzzy rules is stored before solving any cyclic fuzzy contradictions. In turn, the final truth value of fuzzy rules is calculated according to these stored values.

Having the fuzzy contradictions solved by the proposed mechanism, rule weights (if defined) can be applied to the current truth values of fuzzy IF-THEN rules. In this study, the approach proposed in [119] is selected. In this case, rule weights are normalised in the range $[0,1] \subset \mathbb{R}$ and multiplied by the current truth value of each rule. In this example, it is also assumed that the weight of a feature represents the weight of the rule that contains this feature, as implemented by expert system models.

Eventually, a disjunctive approach is employed for computing the truth values of the consequent levels. Hence, each consequent level is given by the maximum value of the truth values of the fuzzy IF-THEN rules that infer the same consequent level. If a conjunctive approach was selected (using the minimum value instead), the set of rules would be jointly satisfied, representing a stricter proposal. Here, the disjunctive approach is selected for being a more flexible proposal that guarantees that at least one rule is satisfied. The membership grade of updated fuzzy IF-THEN rules will then propagate to their consequents, producing a set of truncated membership functions associated

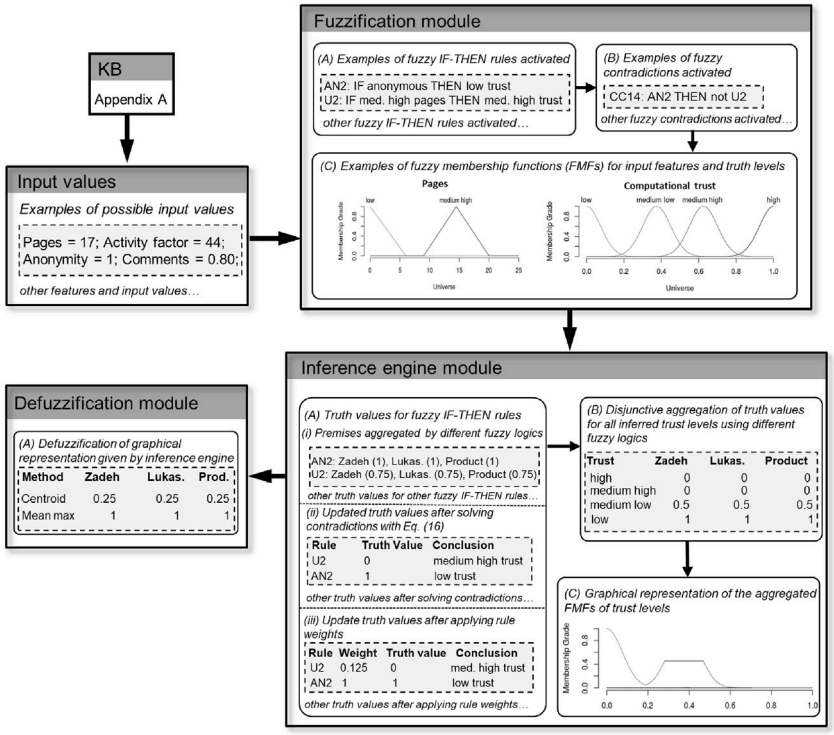


Fig. 7. An illustration of a reasoning process modelled by fuzzy reasoning for a single Wikipedia editor.

with their consequents. The inference engine in the running example depicts the truth values of fuzzy rules, their updated values after solving contractions, and, finally, after applying rule weights. This is followed by the disjunctive aggregation of trust levels and the definition of the respective graphical representation.

3.3.3. Defuzzification module

The output of the inference engine is a graphic representation of the aggregation of these truncated membership functions. Several methods can be used for calculating a single defuzzified scalar from this graphic representation [42]. Two are commonly employed and selected here: *mean of max* and *centroid*. The first returns the average of all elements (consequent levels) with maximal membership grade. The second returns the coordinates (x, y) of the centre of gravity of the graphic representation. The defuzzified scalar is represented then by the x coordinate of the centroid, as per the defuzzification module in the running example.

3.4. Design of non-monotonic reasoning models employing defeasible argumentation

The definition of argument based-models follows the five-layer modelling approach proposed in [18] and depicted in Fig. 3. It starts with the definition of the internal structure of arguments, continues with the definition of their conflicts, the computation of their acceptance status, and ends with the aggregation of accepted arguments. A running example is depicted in Fig. 8 and referred throughout this subsection. Note that all arguments and their interactions come from the knowledge bases employed in this research article and described in Appendix B.

3.4.1. Layer 1: Definition of the internal structure of arguments

Most commonly, an argument is composed of one or more premises and a conclusion derivable by applying an inference rule. Hence, the first step of an argumentation process is to define its *forecast arguments*:

$$Forecast\ argument : premises \rightarrow conclusion$$

This structure includes a set of premises (believed to influence the conclusion being inferred) and a conclusion derivable by applying an inference rule $\rightarrow$. It is an uncertain implication, which is used to represent a defeasible argument. As with the rules of expert systems, premises and conclusions are strictly bounded in numerical ranges associated with natural language terms (for instance *low* and *high*). Forecast arguments are activated if their premises evaluate as true, according to the input data provided. Boolean logical operators AND and OR can be applied for the use of multiple premises, similar to the rules employed by expert system models.

3.4.2. Layer 2: Definition of the interactions between arguments

In order to evaluate inconsistencies, the notion of *mitigating argument* [103] is introduced. These are arguments that attack other forecast arguments or other mitigating arguments. Both forecast and mitigating arguments are special *defeasible rules*, as defined in [53]. Informally, if their premises hold then *presumably* (defeasibly) their conclusions also hold. Different types of attacks and consequently mitigating arguments exist in the literature, as reviewed in Section 2.3. For instance, in this research article, an undermining attack is represented by a forecast argument and an inference $\Rightarrow$ to a negated argument B (forecast or mitigating):

$$Undermining\ attack : forecast\ argument \Rightarrow \neg B$$

Note that these are undermining attacks because the conflict arises from the premises in each forecast argument. For instance, in the running example, B3 is attacking AF3 by stating that `activity factor` should not be *high*, since `bytes` is *low*. An undercutting attack would be defined if for some reason the inference being performed by an argument was being contested. For example, contradictions "OnlyAge" described in Table A.12. Here, an undercutting attack is represented by a set of premises and an inference to a negated argument B (forecast or mitigating):

$$Undercutting\ attack : premises \Rightarrow \neg B$$

Lastly, a rebuttal attack would be created if, for some reason, it was believed that the conclusion of an argument was false. For instance,

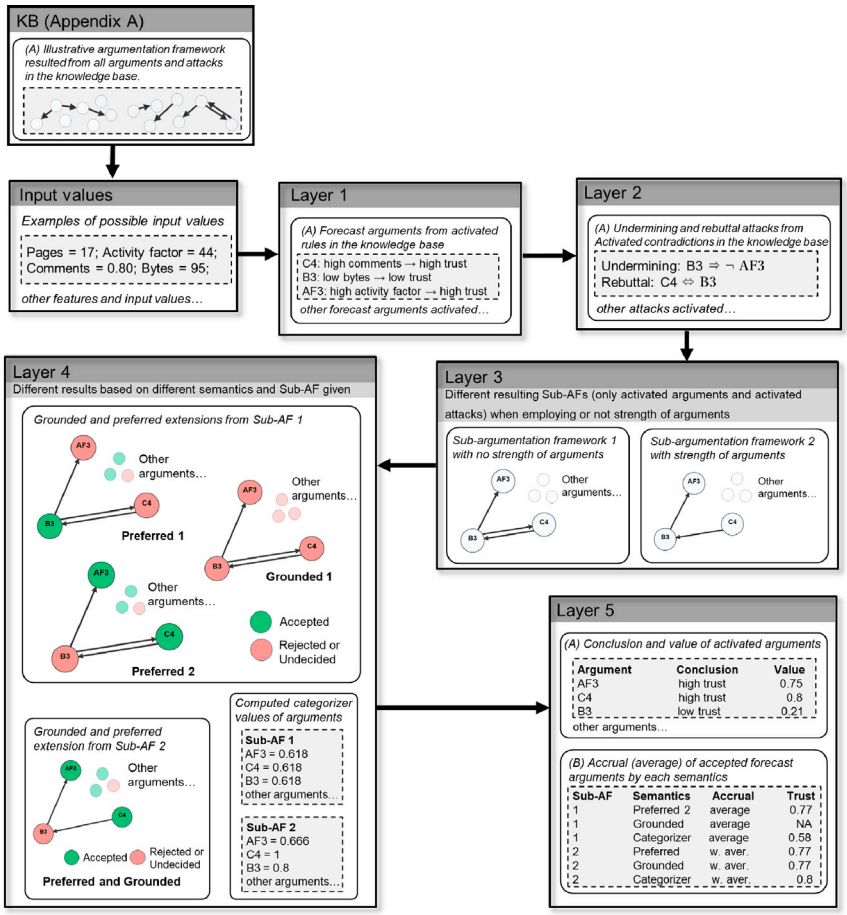


Fig. 8. An illustration of a reasoning process modelled by defeasible argumentation for a single Wikipedia editor.

some domain expert could define an attack targeted at C4 by saying that there is evidence to infer another level of trust instead of *high*. Note that in the context of computational trust, different consequents (trust levels) might coexist or not according to the expert's reasoning, hence not all arguments with different conclusions should necessarily lead to rebuttal attacks. Moreover, since all arguments in the constructed knowledge bases are defeasible (or not strict) rebuttals would be mutual (both arguments would attack each other). Rebuttal attacks (⇔) occur in this research article when two forecast arguments support mutually exclusive conclusions according to a domain expert, hence, are represented as:

$$Rebuttal\ attack: forecast\ argument \Leftrightarrow forecast\ argument$$

Let us point out that different types of attacks can enhance the explainability of argument-based models and aid in the process of creating knowledge bases. However, they do not impact on the computation of the acceptability status of arguments and final numerical scalar being produced by such models in the next layers. This computation is performed via abstract argumentation theory as proposed in [65]. In this case, all attacks are seeing as a binary relation as described in Section 2.3.

3.4.3. Layer 3: Evaluation of the conflicts of arguments

At this stage, forecast and mitigating arguments can be seen as an argumentation framework, which can be elicited with data. Arguments will then be activated or discarded, based on whether their premises evaluate as true or false. For example, consider the argument *C4: high comments → high trust*. If the input for `comments` is lower than 0.75, then the premise of the argument will be false (comments is not high), and the argument will be discarded (not considered in the next steps of the reasoning process). Attacks between activated arguments will be evaluated before being activated as well. These usually have a form of a binary relation. In a binary relation, a successful (activated) attack occurs whenever both of its source (argument attacking) and its target (argument being attacked) are activated. This study also makes use of the notion of strength of arguments as presented in [63]. In this case, an attack is considered successful only if the strength of its source is equal to, or greater than, the strength of its target. In the running example, feature weights are employed for defining the strength of arguments. As with the definition of rule weights in expert systems and fuzzy reasoning, the weight of a feature will also represent the strength of the argument employing this feature. The running example depicts two sub-AFs from activated arguments and attacks, with and without strength or arguments. These sub-AFs are argumentation frameworks contained in the original ones created in Layer 2, but only considering activated arguments and attacks.

3.4.4. Layer 4: Definition of the acceptance status of arguments

Given a sub-AF, all its arguments are considered abstract as proposed in [65]. In other words, only the topology of the sub-AFs built in Layer 3 (and not the internal structure of their arguments) is relevant to decide whether an argument should be accepted or not before proceeding to Layer 5. In turn, acceptability semantics [65–67,120,121] are applied to compute the acceptance status of each argument, or its acceptability. As previously defined, acceptability semantics evaluate the overall interaction of arguments across the AF (or sub-AF in this argumentation process), in order to select the arguments that should ultimately be accepted. In the running example, two well-known extension-based semantics, preferred and grounded [65], and one ranking-based semantics, categoriser [121], are illustrated. For the sake of simplicity, their formal definitions are presented in Appendix A.

Table 4
Non-monotonic models built upon expert systems.

Model	Knowledge base (Appendix B)	Heuristic
E1 (E5)	KB1 (KB2)	h_1
E2 (E6)	KB1 (KB2)	h_2
E3 (E7)	KB1 (KB2)	h_3
E4 (E8)	KB1 (KB2)	h_4

3.4.5. Layer 5: Accrual of acceptable arguments

Eventually, in the last step of the reasoning process, a final inference must be produced. In the case of extension-base semantics if multiple extensions are computed, one might be preferred over the others. In this study, the cardinality of an extension (number of accepted arguments) is used as a mechanism for the quantification of its credibility. Intuitively, a larger extension of arguments, that by definition are also conflict-free, might be seen as more credible than smaller extensions. The idea comes from the fact that more evidence could be seen as a justification to make some decision instead of another one. It is a limited, simple way to solve the issue of multiple extensions. This is also solved by the models employing the strength of arguments. In this case the remaining sub-AFs have no rebuttals, hence, there are no cycles and no multiple extensions when using the preferred semantics with the designed knowledge bases. However, note that models employing the strength of arguments might present worse inferential capacity, leaving no ideal solution given the knowledge bases employed. If the computed extensions have all the same cardinality, these are all brought forward in the reasoning process. After the selection of the larger extension/s or best-ranked argument/s, a single scalar is produced through the accrual of the values assigned to its/their forecast arguments (arguments that infer some trust level). Mitigating arguments have already completed their role by contributing to the resolution of conflicting information (previous layer) and thus, are not considered in this layer. The values of forecast arguments follow from the same formula described in Definition 2. An assumption is made here that forecast arguments have a similar structure to the IF-THEN rules defined for expert systems. Premises are associated with numerical ranges and concatenated by boolean operators AND and OR, and the conclusion has a numerical range as the consequents of the IF-THEN rules. Having their values assigned, the accrual of forecast arguments can be made in different ways, for instance considering measures of central tendency. Here, the average is accounted for models that use a binary relation of attacks, while the weighted average is accounted for models that use the notion of strengths of arguments. Note that in the case of two or more preferred extensions with the same number of accepted forecast arguments, the outcome of the preferred semantics is the mean of all its extensions.

3.5. Summary of models and comparative metrics

The list of models built with different reasoning approaches can be seen in Tables 4, 5, and 6. Overall, 68 models were implemented, with different configuration parameters selected for evaluation, as described in the previous sections. To compare the inferences produced by them, three evaluation metrics for computational trust are employed: *rank of Barnstars*, *spread*, and *percentage of NAs* (not assigned). Table 7 lists the calculation method of each metric.

The rationale behind the Rank of Barnstars metric, is that when sorting editors in descending order by their assigned trust values, it is assumed that the ranking of the best models will result in Barnstar editors being placed at the highest positions. Non-Barnstar editors may also be highly trustworthy. Nonetheless, Barnstar editors still should, presumably, be ranked at the highest positions. Moreover, since trust is not a binary concept, it is expected that the distribution of the trust values assigned by these same models to Barnstar editors should have a positive, continuous spread. Spread is measured by the standard deviation of the values assigned to Barnstar editors. Finally, models that are capable of producing a final inference in more cases are considered better, thus a higher percentage of NAs, or cases without an assigned inference, is deemed as a disadvantage. Note that no metric was defined for computing an overall difference between trust values assigned to Barnstar editors and trust values assigned to non-Barnstar editors. The reason for this derives from the lack of knowledge about the non-Barnstar editors.

Table 5
Non-monotonic models built upon fuzzy reasoning. Fuzzy operators selected are Zadeh (Z), Product (P) and Łukasiewicz (Ł). Knowledge bases (KB) and fuzzy membership functions (FMF) are detailed in Appendix B.

Model	Oper.	Def.method	Ruleweight	KB + FMF (Appendix B)
$FL1$ ($FL13$)	Z	Centroid	No	KB1 (KB2) + Triangular
$FL2$ ($FL14$)	Z	Mean of max	No	KB1 (KB2) + Triangular
$FL3$ ($FL15$)	P	Centroid	No	KB1 (KB2) + Triangular
$FL4$ ($FL16$)	P	Mean of max	No	KB1 (KB2) + Triangular
$FL5$ ($FL17$)	Ł	Centroid	No	KB1 (KB2) + Triangular
$FL6$ ($FL18$)	Ł	Mean of max	No	KB1 (KB2) + Triangular
$FL7$ ($FL19$)	Z	Centroid	Yes	KB1 (KB2) + Triangular
$FL8$ ($FL20$)	Z	Mean of max	Yes	KB1 (KB2) + Triangular
$FL9$ ($FL21$)	P	Centroid	Yes	KB1 (KB2) + Triangular
$FL10$ ($FL22$)	P	Mean of max	Yes	KB1 (KB2) + Triangular
$FL11$ ($FL23$)	Ł	Centroid	Yes	KB1 (KB2) + Triangular
$FL12$ ($FL24$)	Ł	Mean of max	Yes	KB1 (KB2) + Triangular
$FC1$ ($FC13$)	Z	Centroid	No	KB1 (KB2) + Gaussian
$FC2$ ($FC14$)	Z	Mean of max	No	KB1 (KB2) + Gaussian
$FC3$ ($FC15$)	P	Centroid	No	KB1 (KB2) + Gaussian
$FC4$ ($FC16$)	P	Mean of max	No	KB1 (KB2) + Gaussian
$FC5$ ($FC17$)	Ł	Centroid	No	KB1 (KB2) + Gaussian
$FC6$ ($FC18$)	Ł	Mean of max	No	KB1 (KB2) + Gaussian
$FC7$ ($FC19$)	Z	Centroid	Yes	KB1 (KB2) + Gaussian
$FC8$ ($FC20$)	Z	Mean of max	Yes	KB1 (KB2) + Gaussian
$FC9$ ($FC21$)	P	Centroid	Yes	KB1 (KB2) + Gaussian
$FC10$ ($FC22$)	P	Mean of max	Yes	KB1 (KB2) + Gaussian
$FC11$ ($FC23$)	Ł	Centroid	Yes	KB1 (KB2) + Gaussian
$FC12$ ($FC24$)	Ł	Mean of max	Yes	KB1 (KB2) + Gaussian

4. Results and discussion

This section presents the results of the instantiation of the designed non-monotonic reasoning models using two sets of data, each built from a Wikipedia XML dump, first from the Italian-language and then from the Portuguese-language edition. Defeasible argumentation and expert system models were implemented through an online framework using entirely custom code in PHP and JavaScript [122]. Differently, fuzzy reasoning models were implemented using the C++ programming language [123]. The presentation of results is structured around each evaluation metric, a summary, and a final discussion.

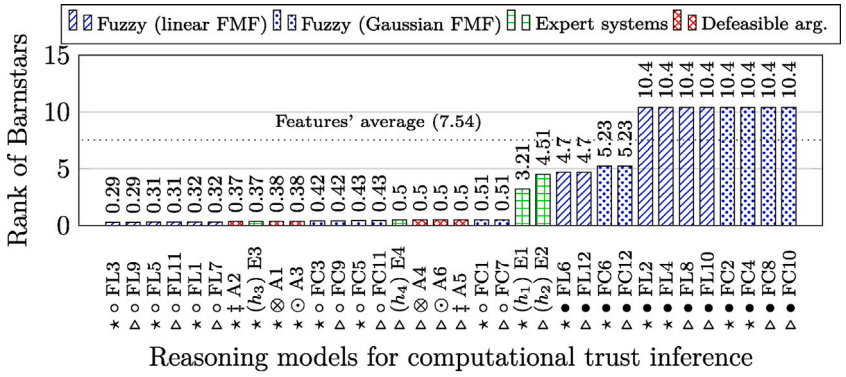
4.1. Rank of barnstars

This metric is targeted at evaluating whether the investigate non-monotonic reasoning models are capable of ranking the Barnstar editors at the highest positions. Figs. 9 and 10 depicts the normalised sum of Barnstar ranks in the range $[0, 100] \subset \mathbb{R}$ for the models built with knowledge base 1 (KB1) and knowledge base 2 (KB2) respectively. Each figure depicts first results when models are instantiated by the Italian-language dataset and then by the Portuguese-language dataset. Instances without an assigned trust scalar were removed from this analysis. A baseline was computed by the average of the normalised features reported by each editor, also resulting in inferences in the range $[0, 1] \subset \mathbb{R}$. This baseline was defined only to indicate whether the non-monotonic reasoning process performed by the implemented models was effective in comparison to a non-deductive and simplified inference.

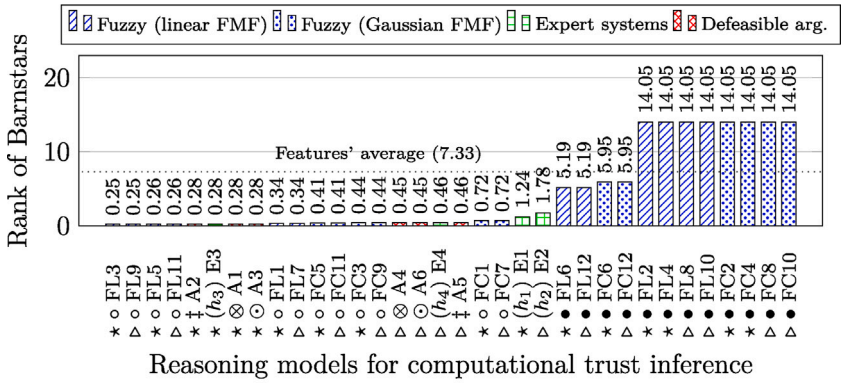
From both figures, it is possible to observe that the computed ranks were effective, ranging from 0 (perfect rank) to 14.05 across

Table 6
Non-monotonic models built upon defeasible argumentation. Knowledge bases employed in Layer 1 and 2 are detailed in Appendix B.

Model	Layer 1 and Layer 2	Layer 3	Layer 4	Layer 5
	Arguments and Conflicts	Attack relation	Semantics	Accrual
A1 (A7)	KB1 (KB2)	Binary	Preferred	card. + average
A2 (A8)	KB1 (KB2)	Binary	Categoriser	average
A3 (A9)	KB1 (KB2)	Binary	Grounded	average
A4 (A10)	KB1 (KB2)	Strength of arg.	Preferred	card. + w. average
A5 (A11)	KB1 (KB2)	Strength of arg.	Categoriser	w. average
A6 (A12)	KB1 (KB2)	Strength of arg.	Grounded	w. average



(a) Models instantiated with the Italian-language dataset.



(b) Models instantiated with the Portuguese-language dataset.

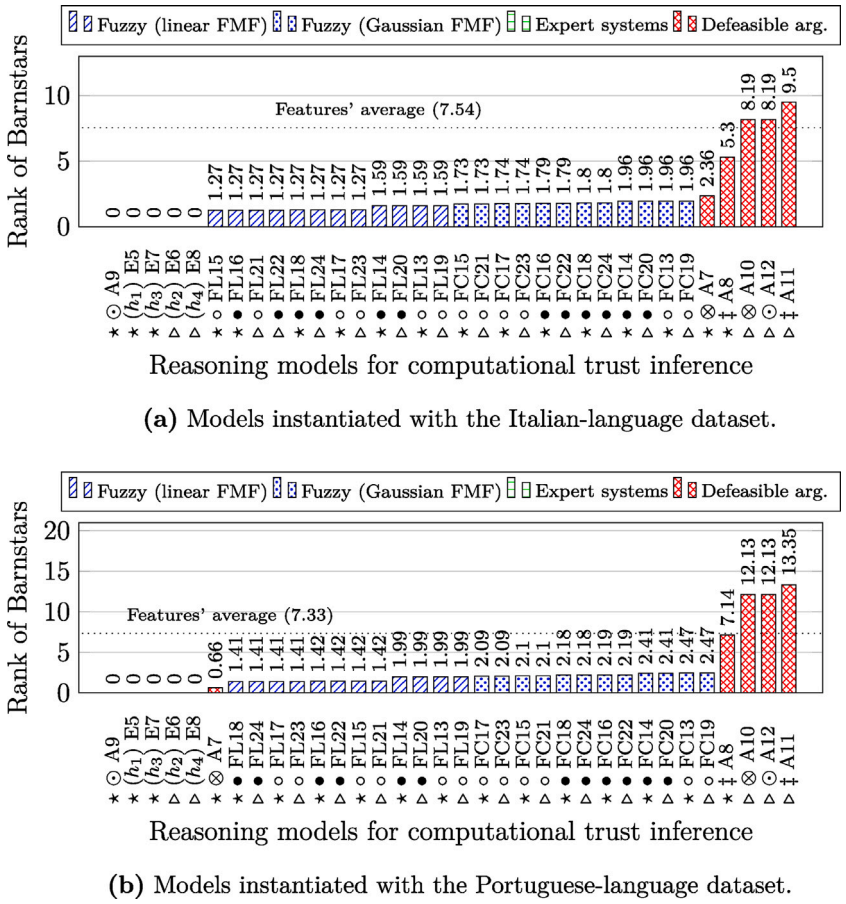
Fig. 9. Normalised sum of the rank of Barnstars users for models built with knowledge base 1 Appendix B. Inferior symbols are used to represent: centroid (∘) and mean of max (•) defuzzification approach; heuristics h_1 to h_4; grounded (⊙), preferred (⊗), and categoriser semantics (‡); and use (respectively no use) of the rule weights/arguments strength (▷, respectively ⋆).

Table 7
Calculation of evaluation metrics employed to assess the trust inferences performed by non-monotonic reasoning models in the Wikipedia Project.

Metric	Calculation
Rank of Barnstars	Remove editors with No trust value assigned from the dataset. Sort all other editors by their trust values in descending order. Non-Barnstars tied with Barnstars are ranked above. Sum the ranks of the Barnstar editors and normalise the result in the range [0, 100] ⊂ ℝ. 0 means all Barnstars with an assigned trust value are ranked above any non-Barnstar, while 100 means they are ranked below any non-Barnstar.
Spread	Standard deviation of the trust values assigned to Barnstars.
Percentage of NAs	Percentage of editors that had No assigned trust value.

all models and input datasets. This suggests that the non-monotonic reasoning approaches were all capable of capturing, to some degree, the notions of the ill-defined construct of trust, even when presenting worse values compared to the features' average. Moreover, the use of different datasets does not seem to have a significant impact on the produced rank of Barnstars. It highlights the stability of models when using different sets of data. However, some differences can still be noted among the models built with different knowledge bases. Different results were expected especially due to the contrasting topologies of these two knowledge bases. Figs. A.15 and A.16 depict their graphical representations. Particularly, while certain parameters appear to be effective for models built with one knowledge base, the same parameters can be less effective when the other knowledge base is applied instead. For instance:

1. Expert system models provided ranks between 0.28–4.51 when built with KB1 and perfect ranks when built with KB2. The



(a) Models instantiated with the Italian-language dataset.

(b) Models instantiated with the Portuguese-language dataset.

Fig. 10. Normalised sum of the rank of Barnstars users for models built with knowledge base 2 (Appendix B). Inferior symbols are used to represent: centroid (∘) and mean of max (•) defuzzification approach; heuristics h_1 to h_4; grounded (⊙), preferred (⊗), and categoriser semantics (‡); and use (respectively no use) of the rule weights/arguments strength (▷, respectively ⋆).

variance when employing KB1 was observed due to the filtering of surviving rules proposed by heuristics h_1 and h_2, which seems to diminish the quality of the ranks produced ($E1$, $E2 \times E3$, $E4$). Noticeably, it does not seem to affect the models built with the same heuristics but using KB2;

2. Fuzzy reasoning models presented higher variance when instantiated with KB1 (ranks between 0.25–14.05) than with KB2 (ranks between 1.27–2.47). The use of linear or Gaussian FMF did not contribute to this variance or had a significant impact in the quality of the ranks produced. Instead, when built with KB1, fuzzy models using the centroid defuzzification approach (labelled with ∘) seem to be much more effective than those using the max defuzzification approach (labelled with •). When built with KB2, no particular parameter seems to provide significantly better ranks, possibly due to the difficulty of these models in dealing with a higher number of contradictions contained in this knowledge base;
3. Argument-based models demonstrated great stability when instantiated with KB1 (ranks between 0.28–0.5), but higher variance when instantiated with KB2 (ranks between 0–13.35). It shows that acceptability semantics and the use (or not) of strength or arguments did not impact results when using KB1. Contrarily, for the argument-based models built with KB2, the use of grounded and preferred semantics (labelled with ⊙ and ⊗) resulted in better ranks. Lastly, models built with KB2 and no strength of arguments (labelled with ⋆) also performed better than their counterparts (labelled with ▷), suggesting that the strengths of arguments defined by the author did not improve the rank of Barnstar editors.

In summary, the first knowledge base presents a simple topology. It seems to work more effectively with argument-based models, a subset of expert system models, and a subset of fuzzy reasoning models. Contrarily, the second knowledge base is built with many more attacks, resulting in a more complex topology. Despite such complexity, certain models did achieve a perfect rank of Barnstar editors. However, as it will be evaluated in the next subsection, a higher topological complexity might also hamper the capacity of these same models of producing inferences. Thus, the next subsection evaluates the percentage of NAs, or the percentage of instances without a trust value assigned by each model. It further investigates a probable trade-off between these two knowledge bases: while one is more simplified and allows inferences to be produced for all instances, the other is more complex and precise, but prevents models from reaching a conclusion in a number of cases.

4.2. Percentage of NAs

The capacity for assigning trust values under conflicting information is assumed to be a favourable property. Thus, it is evaluated through the percentage of NAs. It is known that certain acceptability semantics employed by argument-based models, such as the categoriser, will always return a non-empty extension. In addition, the preferred semantics is less likely to return an empty extension compared to the grounded semantics. In fact, due to the topology of the knowledge

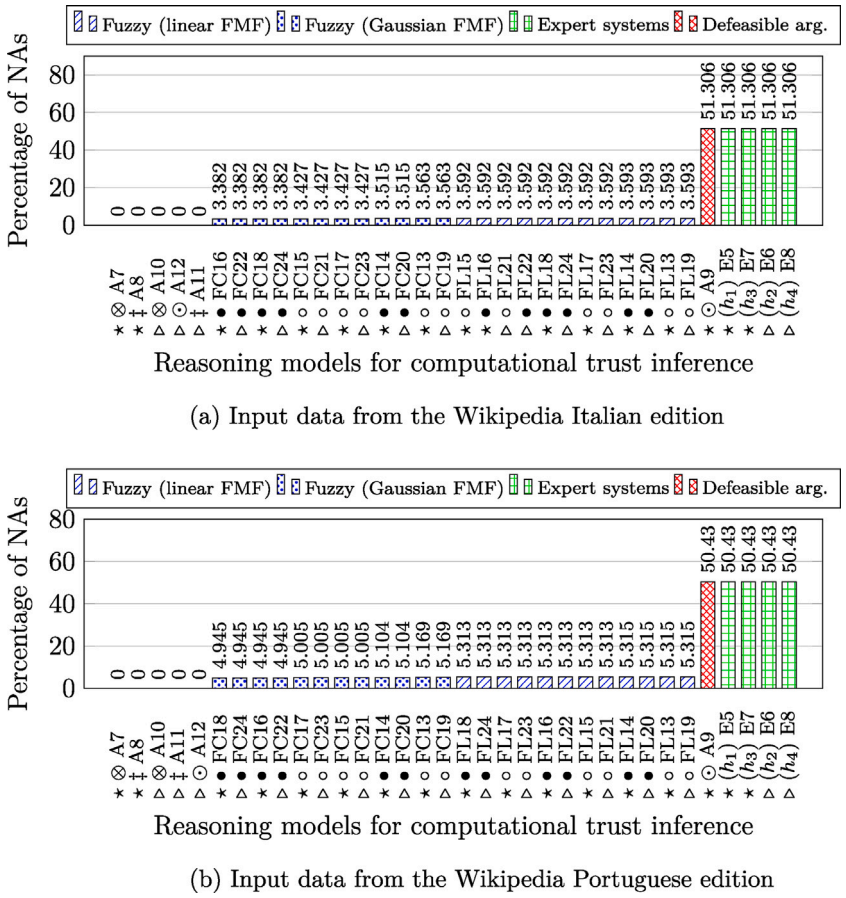


Fig. 11. Percentage of instances without a computational trust scalar assigned by models built with knowledge base 2 (Appendix B). Inferior symbols are used to represent: centroid (∘) and mean of max (•) defuzzification approach; heuristics h_1 to h_4; grounded (⊙), preferred (⊗), and categoriser (‡) semantics; and use (respectively no use) of the rule weights/arguments strength (▷, respectively ⋆).

bases employed in this research article, the preferred semantics always returns a non-empty extension. However, the models implemented with other semantics and other reasoning approaches might not reach a final inference for certain cases. Hence, it is important to evaluate the extent to which this can impact the quality of the designed reasoning models. Fig. 11 depicts the percentage of NAs for models instantiated with KB2. Due to the simplified topology of the KB1, NAs were not reported for any reasoning model built with it. Similarly to the previous evaluation metric, the instantiation of models with different datasets did not result in different trends.

As for expert system models, it seems clear that the simplistic conflict resolution strategy employed by them might work well when applied to knowledge bases of simplified topology (as per Fig. 10). However, once a higher number of conflicts is presented several instances might not have an assigned inference. When instantiated with KB2, 51.3% (Italian-language dataset) and 50.43% (Portuguese-language dataset) NAs were reported for all expert system models. Thus, limiting the applicability of the reasoning approach with KB2.

With respect to fuzzy reasoning models, similar percentages of NAs were observed, indicating an equivalent capability of resolving the conflicts provided in KB2. Thus, note that the use of possibility theory for the resolution of large amounts of conflicts does not appear to be impacted by other configuration parameters of fuzzy reasoning models in this knowledge base.

Finally, in relation to argument-based models, it is possible to observe that $A12$ and $A9$, both built with the grounded semantics, presented very different results. While $A12$ makes use of strength of arguments and presented 0% of NAs, $A9$ makes no use of strength of arguments and presented 51.3% (Italian-language dataset) and 50.43% (Portuguese-language dataset) of NAs. Therefore, the use of strength of arguments can assist in the issue of empty extensions when employing the grounded semantics. However, the quality of the inferences is not maintained when using such strengths. $A9$ had a perfect rank, while $A12$ had a rank value of 12.13 as depicted in Fig. 10. This is an indication that the strengths defined by the author helped to solve the excessive amount of conflicts in KB2, but did not seem to enhance the rank of Barnstar editors. This is likely due to the way rule weights and strength of arguments were defined, feature by feature. Weights/strengths could have been defined for each rule and arguments directly, in a more time-consuming manner and requiring more domain knowledge, but likely better capable of quantifying their importance. The only argument-based models that could achieve a strong rank of Barnstars while not reporting NAs were $A7$ (preferred semantics) and $A8$ (categoriser semantics), which were built with no strength of arguments. It reinforces the suitability of the preferred semantics and the categoriser semantics (despite worse rank of Barnstars editors) for the inference of computational trust.

4.3. Spread

Another metric selected for evaluating the quality of the inferences produced by the non-monotonic reasoning models was the spread of the trust scalars assigned to Barnstar editors. This was measured through the standard deviation (σ) of these scalars. As previously mentioned,

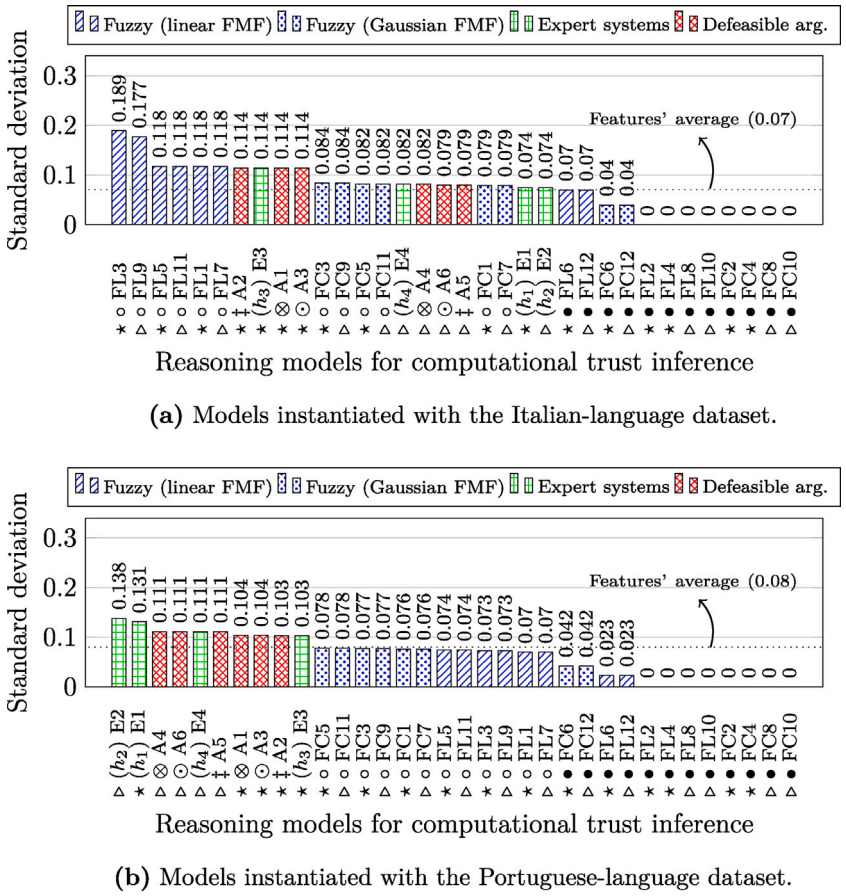


Fig. 12. Spread (calculated with the standard deviation) of computational trust scalars inferred to Barnstar users with models built with knowledge base 1 (Appendix B). Inferior symbols are used to represent: centroid (∘) and mean of max defuzzification approach (•); heuristics h_1 to h_4; grounded (⊙), preferred (⊗), and categoriser (‡) semantics; and use (respectively no use) of the rule weights/arguments strength (▷, respectively ⋆).

trust is not a binary concept. Thus, if we assume that Barnstar editors are trustworthy, we should also expect that they will have different trust levels. Figs. 12 and 13 depicts the results for the models built with KB1 and KB2. The baseline instrument is depicted again. It produces trust scalars through the average of the normalised values of the features reported by each editor. It is employed as an attempt to show whether the non-monotonic reasoning processes implemented could outperform, to some degree, a non-deductive and simplified inference.

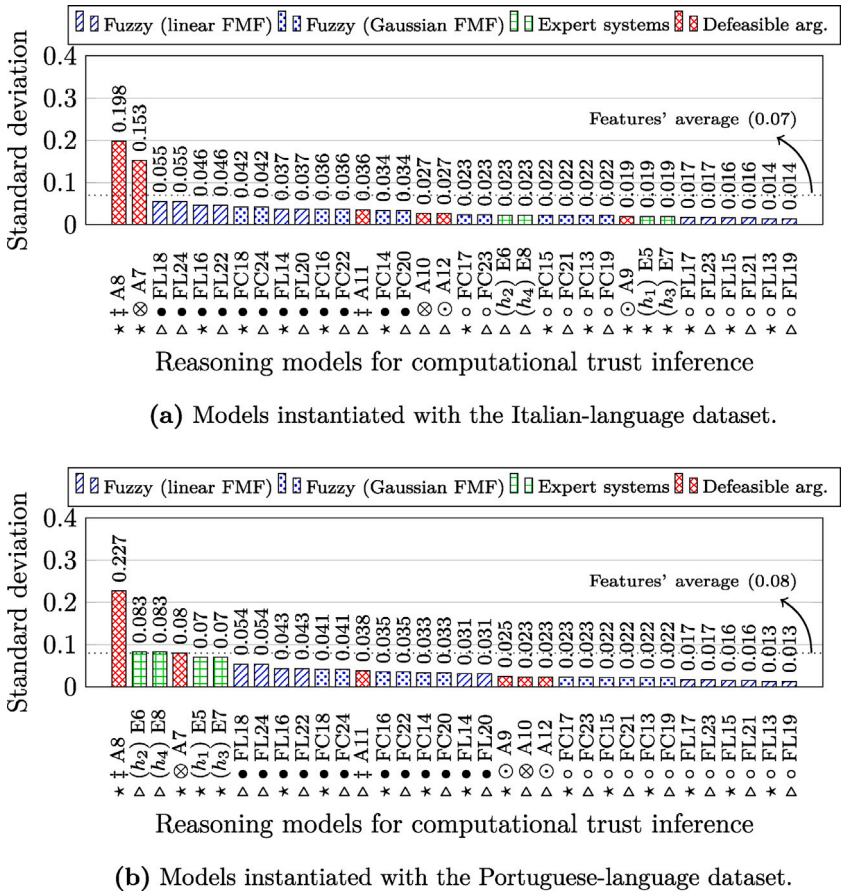
As depicted in Fig. 12, most models built with KB1 achieved a σ between 0.07–0.18, close to or higher than the results reported by the features' average. Hence, the capacity to differentiate trust levels by most designed models is similar and can be considered appealing. Low σ values were expected due to the required difference in trust scalars assigned to Barnstars and non-Barnstar editors. In other words, the higher the σ the higher the chance of overlapping trust values assigned to Barnstar and non-Barnstar editors and, consequently, the worse the rank of Barnstar editors.

In addition, contrary to the previous analysed evaluation metrics, the models' σ varies greatly when instantiated with the Italian-language and the Portuguese-language dataset. This result indicates that no single reasoning approach is better suited to design inference models able to achieve higher values of σ. An exception was noted for fuzzy reasoning models built with KB1 and employing the mean max defuzzification approach (labelled with •). These reported the lowest σ (between 0–0.07), regardless of the dataset employed and below the features' average. One possibility is that the lower number of inputs required by the mean max approach (only the maximum membership grades) resulted in inferences of less variability when compared to the centroid defuzzification approach, which considers the whole aggregated fuzzy set given by the models' fuzzification module.

As for results depicted in Fig. 13, note that most models built with KB2 achieved even lower σ (between 0.01–0.08), and were below the features' average. Therefore, their capacity to differentiate trust levels is similar, but lower when compared to the counterpart models built with KB1. No single expert system or fuzzy reasoning model reported a high σ when built with KB2. This is an indication that expert system and fuzzy reasoning models are likely not capable of inferring robust trust scalars when built with knowledge bases containing a large number of conflicting information. Particularly, they are presumably not able to follow the assumption of trust not being a binary concept. Some few exceptions were observed for argument-based models. In particular, models $A7$ and $A8$ achieved 0.19 and 0.15 σ (Italian-language dataset), and 0.22 and 0.08 σ (Portuguese-language dataset). These seem to be the only models built with KB2 able to achieve a favourable spread, while maintaining 0% of NAs and a robust rank of Barnstar editors (between 0.66–2.36 for $A7$, and between 5.3–7.14 for $A8$). This balance illustrates the strong capacity for modelling non-monotonic reasoning by argument-based models defined with the preferred and categoriser semantics while built with KB2. This capacity was not observed in any other investigated reasoning approach.

4.4. Summary and discussion

The inferential capacity of the employed non-monotonic reasoning approaches is examined in terms of them being able to produce models



(a) Models instantiated with the Italian-language dataset.

(b) Models instantiated with the Portuguese-language dataset.

Fig. 13. Spread (calculated with the standard deviation) of computational trust scalars inferred to Barnstar users with models built with knowledge base 2 (Appendix B). Inferior symbols are used to represent: centroid (∘) and mean of max defuzzification approach (•); heuristics h_1 to h_4; grounded (⊙), preferred (⊗), and categoriser (‡) semantics; and use (respectively no use) of the rule weights/arguments strength (⊳, respectively ⋆).

Table 8
Status of the reasoning approaches with respect to the successfully modelling of computational trust by models implemented with them. In case of mixed results, it is reported the number of models built with the reasoning approach that were successful over all the ones implemented. Reasons for failing are detailed in parenthesis.

Dataset/KB	Expert systems	Fuzzy reasoning	Defeasible argumentation
Italian/KB1	All	14/24 (low σ)	All
Italian/KB2	None (high % of NAs and low σ)	None (low σ)	**2/6** (high % of NAs and low σ)
Portuguese/KB1	All	12/24 successful with borderline σ	All
Portuguese/KB2	None (high % of NAs)	None (low σ)	**2/6** (high % of NAs and low σ)

whose inferences could be considered valid in the domain of computational trust. Reasons for the inferences of a model being considered invalid include spread (σ) lower than the baseline (features' average) or very high percentage of NAs. In terms of rank of Barnstar all models were considered effective. Table 8 summarises the results.

As it can be observed, expert systems models presented appealing results when built with KB1. This demonstrates that the reasoning approach can be used effectively when instantiated with knowledge bases with a lower number of conflicts, supporting the vast use of expert systems in an ample range of domains present in the literature. Thus, it is important to highlight strengths such as clear reasoning process, capacity to keep the language of the domain, and capacity to add and retract rules. By contrast, when built with KB2, all expert system models were considered unsuccessful when modelling reasoning applied to computational trust. This was mostly due to the high percentage of NAs regardless of the dataset employed. As anticipated, limitations were expected in line with the performed literature review. The lack of a built-in non-monotonicity layer or options for implementing it reinforce the disadvantage of expert systems in dealing with large amounts of uncertain, vague and contradictory information.

Fuzzy reasoning models presented the most divergent reasoning process, when compared to expert system and defeasible argumentation models. While expert systems and defeasible argumentation share similarities such as quantification of rules/arguments and their aggregation through measures of central tendency, fuzzy reasoning adopts the notions of fuzzy sets and FMF, thus providing a disparate inferential process. This difference offers advantages such as higher precision for the modelling of natural language terms and capacity to handle fuzzy concepts. Certainly, such advantages are of great importance when non-monotonic reasoning is being performed. However, this higher precision comes with a number of disadvantages, including the definition of membership functions that differ from the way in which humans reason. Moreover, in order to manipulate these functions and define an inferential process with them, the definition of a number of configuration parameters is necessary. For example, a fuzzy logic operator, a fuzzy inference method, and a defuzzification approach are all needed. Some mathematical reasoning is required to select the most appropriate parameters, limiting the applicability of the reasoning approach by domain experts who are not familiar with fuzzy parameters and their

interpretations. In addition, the available options for implementation of non-monotonicity are not well developed. In this research article, the use of possibility theory was selected for being intuitively the only approach that allowed the retraction of rules in a fuzzy sense. In other words, partial retractions were allowed, according to the truth value of the propositions being evaluated. However, this approach was limited by its order of application, not being commutative. The requirement of all these configurations, and the variability in the inferences produced, seem to place fuzzy reasoning in between expert systems and defeasible argumentation, in terms of inferential capacity when performing non-monotonic reasoning.

Lastly, defeasible argumentation presented the most robust results, being able to produce successful models despite the knowledge base and dataset employed. Similar to expert systems, all models were considered successful when built with KB1, in addition to producing better ranks of Barnstar. When built with KB2, argument-based models were deemed successful when employing the preferred and categoriser semantics without strength of arguments. This shows that some knowledge for the selection of configuration parameters is also required when domain experts decide to use defeasible argumentation. However, it is argued here that the amount of knowledge is lesser than that required by fuzzy reasoning. Moreover, the notion of scepticism behind the behaviour of semantics has been widely discussed in the literature of defeasible argumentation, making its adoption by experts in other fields more accessible. Lastly, when built with KB2, only argument-based models were able to produce instances in all cases while reporting appealing spread and rank of Barnstars.

To sum up, defeasible argumentation proved to be the most balanced reasoning approach, with models capable of maintaining strong results, despite the complexity of the knowledge base. Nonetheless, it is important to mention other limitations that could be noted when employing it. For instance, as with the other reasoning approaches, knowledge acquisition was a boundary when developing inferential models. This limitation could be observed, for example, in the accrual of arguments performed in the last stage of the reasoning process. The accrual based on the cardinality of extensions and measures of central tendency, such as average and weighted average of accepted arguments, are simplistic approaches. They could be seen as a surrogate for conflicts not known or not established between arguments in the knowledge base. Ideally, if the goal is to reach a single conclusion, another iteration of knowledge acquisition should be performed, instead of choosing extensions of higher cardinality (when necessary) and averaging accepted arguments of distinct consequents. This could likely improve the percentage of NAs and increase the number of models deemed successful. Still, defeasible argumentation was seemingly the most suitable to model non-monotonic reasoning applied in the domain of computational trust, allowing the creation of models that were constantly among the top-performing ones. Hence, in conclusion, the acceptance status of the hypothesis formulated in Section 3 is described below.

- **Hypothesis**: If computational trust is modelled with defeasible argumentation, then the inferential capacity of its models will be superior than that achieved by the selected baselines, non-monotonic fuzzy reasoning and expert systems models, according to a predefined set of evaluation metrics from the domain of application.
- **Acceptance status**: *accepted*. Reasoning models built with defeasible argumentation were superior for maintaining a balance between the metrics of evaluation, while being able to provide inferences for all instances in the datasets. Expert systems and fuzzy reasoning models achieved appropriated inferences with less-complex knowledge bases, but did not guarantee the production of inferences in all cases or appealing models regardless of the knowledge base.

5. Conclusion and future work

This research article focused on reviewing possible implementations of non-monotonic reasoning by knowledge-based approaches and conducting an empirical comparison between them. Non-monotonic reasoning allows the retraction of previous conclusions in light of new information. It is a compelling approach to model reasoning applied to domains of uncertain information. The implementation of non-monotonic reasoning required the use of credible domain knowledge bases, which in this study were designed by the author. Their process of creation was exemplified, resulting in two variations: one simplified, with less conflicts; and another with more conflicts and higher topological complexity. Both contained sets of rules, contradictions, natural language terms, and fuzzy membership functions as reported in Appendix B. At last, these knowledge bases were exploited for the design of inference models built upon three knowledge-based, reasoning approaches able to perform non-monotonic reasoning: expert systems, fuzzy reasoning, and defeasible argumentation. Eventually, these same models were instantiated with real-world, quantitative datasets extracted from two Wikipedia dumps of different language editions. Two dumps were selected to reinforce findings and extend the generalisability of the study in terms of datasets employed. After being instantiated, the designed models were used to infer a trust scalar in the $[0, 1] \subset \mathbb{R}$ to each editor of the encyclopaedia, where 1 means complete trust, and 0 means complete absence of trust should be assigned. Still, the proposed comparison was *not* aimed at enhancing the assessment of computational trust. Instead, it attempted to situate defeasible argumentation, between two other well-known reasoning techniques: expert systems and fuzzy reasoning. In particular, the inferential capacity of reasoning models built with such approaches were examined. Three evaluation metrics were selected for this analysis: (1) the sum of the ranks by assigned scalars to recognised trustworthy individuals; (2) the spread of assigned scalars to these same trustworthy editors; and (3) the percentage of instances without an assignment.

This research included a number of limitations. These are important and can inform future work for extending research on further domains of application and for considering additional non-monotonic reasoning approaches. In terms of its design, the domain-specific metrics selected for the evaluation of reasoning models increased the complexity of the proposed comparison due to the lack of ground truths. The scope of the design was also bounded by an empirical comparison between knowledge-based, non-monotonic reasoning approaches using quantitative data in a real-world context. Alternatively stated, the method employed (empirical study) to perform the envisioned comparison, was restricted. The employed models of inference were designed upon different existing building blocks in the literature of the reasoning approaches they were grounded on. This approach was necessary in order to achieve working models in the domain of computational trust with the available data and knowledge bases. Therefore, the understanding of the differences between such reasoning approaches in theoretical scenarios, or scenarios unlikely to happen in the context of modelling reasoning applied to computational trust, is also limited. In addition, the knowledge bases employed by the non-monotonic reasoning models were limited by their process of formation. This process was highly dependent on domain experts and on the time required for manual acquisition of information and formation/validation of structured knowledge. Moreover, each knowledge base was the result of the information extracted from a single human reasoner and not from multiple reasoners. The selection of multiple reasoners might require additional financial resources not available in this research work.

Despite such limitations, findings indicated how all the employed reasoning approaches allowed for the construction of models that were capable of assigning satisfactory trust scalars in certain scenarios. In the scenario of the knowledge base with less conflicts, all expert systems and defeasible argumentation models were deemed successful despite the dataset selected. Contrarily, fuzzy reasoning models led to

mixed results, due to its higher number of configuration parameters for design of models. These parameters provided greater flexibility but their complexity and abundance likely limited the applicability and use of fuzzy reasoning by domain experts. When faced with the knowledge base of higher topological complexity, models built with expert systems reported a significant number of instances that remained unsolved, suggesting that a lower applicability in this kind of scenario is possible despite the dataset chosen. Fuzzy reasoning models reported less unsolved instances, but the spread of assigned scalars to trustworthy editors was too low for any model to be considered effective, also regardless of the dataset chosen. In contrast, some argument-based models were able to solve all the instances while demonstrating a better capacity of inferring robust trust scalars, or scalars able to show an effective balance among the selected evaluation metrics.

These results could lead to a possible interpretation of defeasible argumentation being better suited to capture the underlying reasoning of the knowledge bases employed in this study. Still, some differences could be observed by argument-based models employing different configuration parameters. For instance, models built with a more credulous rationale (preferred semantics), achieved better inferences. The reason for such performance may be due to a high uncertainty of the domain of computational trust. This higher uncertainty could have originated from: the fact that no other experts were consulted for validation or collaboration during the process of creation of knowledge bases; or by the fact that computational trust is not a well defined construct. Therefore, knowledge bases could likely be further improved in order to be used by sceptical argument-based models.

In conclusion, the originality of this research lies in the extensive comparison performed among defeasible argumentation and two other approaches capable of performing non-monotonic reasoning in the domain of computational trust. Previous works [113] have attempted to employ defeasible argumentation and to perform similar comparisons also in other domains, such as mental workload modelling [22,23] and mortality occurrence modelling [24,25]. In this research article different sets of data and/or models of inference were employed, extending the use of defeasible argumentation in real-world, quantitative contexts. Moreover, this research article is the first to employ a large amount of data for eliciting reasoning models. Large data sets reduce the risk of bias that could occur when having knowledge bases designed by single agents. Since millions of records are being evaluated (compared to hundreds from previous works), it is not possible to manually inspect all of them. Hence, a broad study has been performed, empirically enhancing the generalisability of defeasible argumentation as a possible approach to reason with quantitative data and conflicting/uncertain knowledge. In addition, a review of the investigated reasoning approaches was carried out, including their options for adding a non-monotonic layer. The practical use of such approaches coupled with a modular design that facilitates similar experiments was exemplified and their respective implementations made public [122,123]. Moreover, the addition and use of a non-monotonic layer in the inferential processes of models built with expert systems and fuzzy reasoning was also exemplified. Such use is seldom reported in the field of non-monotonic reasoning. It might be a useful aid to scholars familiar with these reasoning approaches and also interested in performing non-monotonic reasoning activities. Overall, this study attempts to serve as a beginning point for other scholars, who could use it to replicate and/or improve the proposed approach in other domains of application. Consequently, this could contribute to the long-term goal of demonstrating the applicability and generalisability of defeasible argumentation with quantitative data in real-world contexts.

Different avenues can be pursued for future work. For example, a comparison of broader scope can be performed by adopting different structures and configurations of reasoning models. In particular, the issue of multiple extensions could be addressed by a mix of extension-based and ranking-based semantics as employed in [124], or using attack weights for selecting extensions [125]. Another improvement to make the applicability of defeasible argumentation more generalisable could come from the analysis of other argumentation systems, such as fuzzy argumentation [126,127], possibilistic abstract dialectical frameworks [128], probabilistic argumentation [129], or bipolar argumentation [56,130]. Hybrid reasoning techniques, such as neuro-fuzzy systems [131], genetic fuzzy systems [132] and fuzzy argumentation [126] are also recommended. An investigation of different knowledge bases, both in the domain of computational trust and in new ones, will result in new findings. For example, in other contexts of digital collaborative environments, such as blogs, forums and social networks. Another interesting technique for the construction of knowledge bases could be the use of multiple reasoners for knowledge acquisition. Contradictions are generally hard to be formalised, and more reasoners might argue among themselves, leading to the creation of more conflicting rules/arguments. A less time-consuming method to produce new knowledge bases could also be attempted through the development of human-in-the-loop solutions, partially automating the construction of arguments and attacks. For instance, several works have proposed different techniques for rule extraction from machine learning models [133–136]. Finally, a higher explanatory capacity might lead to higher levels of adoption and deployment of non-monotonic reasoning. Explainability is a multifaceted concept [137–141]. Thus, an in-depth investigation in this aspect is also suggested.

CRediT authorship contribution statement

Lucas Rizzo: Conceptualization, Methodology, Software, Investigation, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Data curation. **Luca Longo:** Conceptualization, Methodology, Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data/code at the manuscript.

Appendix A. Definitions of acceptability semantics

In this research article, it is important to define the most common Dung semantics [65], such as grounded, preferred, and stable, as well as other important notions such as reinstatement and conflict-freeness.

Definition 3 (*Shorthand Notations [67]*)**.** Let $\langle Ar, att\rangle$ be an Argumentation Framework (AF), $A, B \in Ar$ and $Args \subseteq Ar$.

- A^+ as $\{B \mid (A, B) \in att\}$.
- $Args^+$ as $\{B \mid (A, B) \in att$ for some $A \in Args\}$.
- A^- as $\{B \mid (B, A) \in att\}$.
- $Args^-$ as $\{B \mid (B, A) \in att$ for some $A \in Args\}$.

A^+ indicates the arguments attacked by A, while A^- indicates the arguments attacking A. $Args^+$ indicates the set of arguments attacked by $Args^+$, while $Args^-$ indicates the set of arguments attacking $Args^-$.

Definition 4 (*Conflict-free [67]*)**.** Let $\langle Ar, att\rangle$ be an AF and $Args \subseteq Ar$. $Args$ is *conflict-free* iff $Args \cap Args^+ = \emptyset$.

Definition 5 (*Defence [67]*)**.** Let $\langle Ar, att\rangle$ be an AF, $A \in Ar$ and $Args \subseteq Ar$. $Args$ defends an argument A iff $A^- \subseteq Args^+$.

Definition 6 (*Reinstatement Labelling [68]*)**.** Let $\langle Arg, att\rangle$ be an AF and $Lab : Arg \rightarrow \{in, out, undec\}$ be a labelling function. Lab is a *reinstatement labelling* iff it satisfies:

- $\forall A \in Ar : (Lab(A) = out \equiv \exists B \in Ar : (B \text{ defends } A \wedge Lab(B) = in))$ and
- $\forall A \in Ar : (Lab(A) = in \equiv \forall B \in Ar : (B \text{ defends } A \supset Lab(B) = out))$

Definition 7 (*Dung's Acceptability Semantics [65], as defined in [67]*)**.** Let $Args$ be a conflict-free set of arguments, $F : 2^{Args} \rightarrow 2^{Args}$ a function such that $F(Args) = \{A \mid A \text{ is defended by } Args\}$ and $Lab : Args \rightarrow \{in, out, undec\}$ a reinstatement labelling function. Also consider $in(Lab)$ short for $\{A \in Args \mid Lab(A) = in\}$, $out(Lab)$ short for $\{A \in Args \mid Lab(A) = out\}$ and $undec(Lab)$ short for $\{A \in Args \mid Lab(A) = undec\}$.

- $Args$ is *admissible* if $Args \subseteq F(Args)$.
- $Args$ is a *complete* extension if $Args = F(Args)$.
- $in(Lab)$ is a *grounded* extension if $undec(Lab)$ is maximal, or $in(Lab)$ is minimal, or $out(Lab)$ is minimal.
- $in(Lab)$ is a *preferred* extension if $in(Lab)$ is maximal or $out(Lab)$ is maximal.
- $in(Lab)$ is a *stable* extension if $undec(Lab) = \emptyset$.

The categoriser ranking-based semantics [121] is also employed in this research article. It assigns a value to each argument based on its number of attackers. To do so, a categoriser function and a categoriser semantics are defined as follows:

Definition 8 (*Categoriser Function [121]*)**.** Let $\langle Args, att \rangle$ be an argumentation framework. Then, $Cat : Args \rightarrow (0, 1]$ is the categoriser function defined as:

$$Cat(a) = \begin{cases} 1 & \text{if } a^- = \emptyset \\ \frac{1}{1+\sum_{c \in a^-} Cat(c)} & \text{otherwise} \end{cases}$$

Definition 9 (*Categoriser Semantics [121]*)**.** Given an argumentation framework $\langle Args, att \rangle$ and a categoriser function $Cat : Args \rightarrow (0, 1]$, a ranking-based categoriser semantics associates a ranking $\succeq_{AF}^{Cat}$ on $Args$ such that $\forall a, b \in Args, a \succeq_{AF}^{Cat} b$ iff $Cat(a) \geq Cat(b)$.

Appendix B. Knowledge bases

In this section, two knowledge bases are defined for the inference of computational trust. Their features were extract from the files provided by Wikipedia dumps (Fig. A.14). Nine quantitative features were selected and are detailed next. Rules and contradictions are defined by the author, who is qualified in computer science and has appropriate experience in a multitude of digital collaborative environments. Both knowledge bases are consisted of:

- A set of features employed for the modelling and assessment of computational trust (Table A.9).
- A set of natural language terms associated with numerical ranges used for reasoning with such features, for instance *low* and *high* (Table A.9).
- A set of inferential rules in the form:
  - IF B feature A THEN C trust. Where B is a level of feature A and C is a trust level. For instance "**IF** *high* `bytes` **THEN** *high* `trust`". Boolean operators AND/OR might also be used to add other premises.
- A set of contradictions or meta-rules in the form:
  - IF B feature A THEN not Rule B
  - IF Rule A THEN not Rule B
- A graphical representation of rules and contradictions.
- A set of fuzzy membership functions associated with the natural language terms.

```
...
<page>
  <title>AccessibleComputing</title>
  <ns>0</ns>
  <id>10</id>
  <redirect title="Computer accessibility" />
  <revision>
    <id>233192</id>
    <timestamp>2001-01-21T02:12:21Z</timestamp>
    <contributor>
      <username>RoseParks</username>
      <id>99</id>
    </contributor>
    <comment>*</comment>
    <model>wikitext</model>
    <format>text/x-wiki</format>
    <text id="233192" bytes="124" />
    <sha1>8kul9tlwjm9oxgvqzbwuegt9b2830vw</sha1>
  </revision>
  <revision>
    ...
  </revision>
</page>
...
```

Fig. A.14. XML file structure of Wikipedia.

Features and natural language terms of knowledge bases 1 and 2

See Table A.9.

IF-THEN rules employed by knowledge bases 1 and 2

See Table A.10.

Contradictions employed by knowledge bases 1 and graphical representation

See Table A.11 and Fig. A.15.

Contradictions employed by knowledge bases 2 and graphical representation

See Table A.12 and Fig. A.16.

Fuzzy membership functions

Fig. A.17 depicts the possible fuzzy membership functions employed for modelling the natural language terms listed in Table A.9. Some terms present only triangular membership functions because they were modelled with absolute values extracted from the Wikimedia Foundation's Analytics.

Appendix C. List of results by model

See Table B.13.

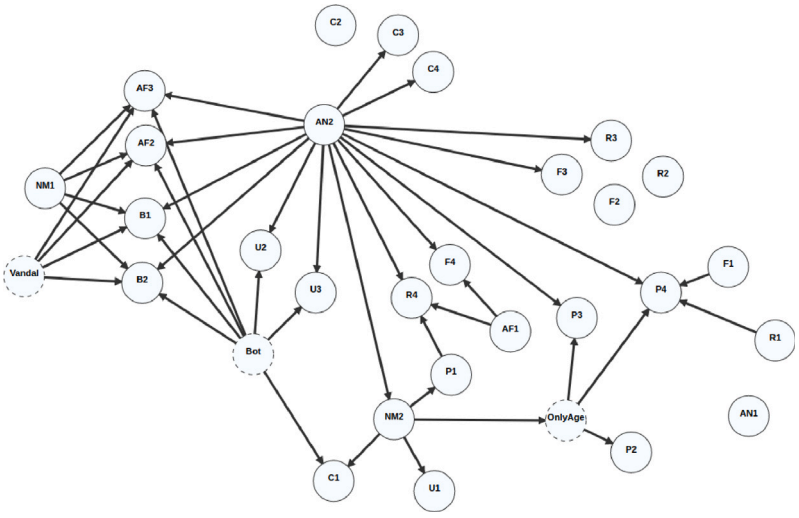


Fig. A.15. Graphical representation of knowledge base 1. Nodes might represent (fuzzy) IF-THEN rules (continuous circles) or premises of contradictions (dashed circles). Arrows represent contradictions between two rules.

Table A.9
List of features employed by the author for reasoning and inference of trust (as described in Table 2), followed by transformations applied to each of them, possible values found in the dataset and natural language terms with respective numerical range associated. Weights were also defined by the author through a pairwise comparison process.

Feature	Transformation	Values	Weight	Natural lang. terms
Pages	None	[1, 1,576,621] $\in \mathbb{N}$	1	– *low* $\in [0, 5]$ – *medium high* $\in [10, 19]$ – *high* $\in [20, 20+]$
Activity factor	None	[1, 2,850,913] $\in \mathbb{N}$	3	– *low* $\in [0, 5]$ – *medium high* $\in [10, 19]$ – *high* $\in [20, 20+]$
Anonymous	None	1 or 0	8	– *Yes* $= 1$ – *No* $= 0$
Not minor	$\frac{\text{Not minor}}{\text{Activity fac.}}$	[0, 1] $\in \mathbb{R}$	7	– *very low* $\in [0, 0.05]$ – *medium to high* $\in [0.25, 1)$
Comments	$\frac{\text{Comments}}{\text{Activity fac.}}$	[0, 1] $\in \mathbb{R}$	5	– *low* $\in [0, 0.25)$ – *medium low* $\in [0.25, 0.5)$ – *medium high* $\in [0.5, 0.75)$ – *high* $\in [0.75, 1]$
Presence factor	$\frac{\text{Pres. factor}}{\text{Wiki life time}}$	[0, 1] $\in \mathbb{R}$	3	– *low* $\in [0, 0.25)$ – *medium low* $\in [0.25, 0.5)$ – *medium high* $\in [0.5, 0.75)$ – *high* $\in [0.75, 1]$
Frequency factor	Capped at 1	[0, 1] $\in \mathbb{R}$	5	– *low* $\in [0, 0.25)$ – *medium low* $\in [0.25, 0.5)$ – *medium high* $\in [0.5, 0.75)$ – *high* $\in [0.75, 1]$
Regularity factor	Capped at 1	[0, 1] $\in \mathbb{R}$	3	– *low* $\in [0, 0.25)$ – *medium low* $\in [0.25, 0.5)$ – *medium high* $\in [0.5, 0.75)$ – *high* $\in [0.75, 1]$
Bytes	None	$[-1 \times 10^8, 8 \times 10^8] \in \mathbb{N}$	1	– *low* $\in [0, 110]$ – *medium high* $\in [512, 2387]$ – *high* $\in [2388, 2388+]$

Table A.10
(fuzzy) IF-THEN rules employed in knowledge bases 1 and 2 designed by the author for inference of computational trust of Wikipedia editors.

Label	Internal structure
B1	**IF** *medium high* bytes **THEN** *medium high* trust
B2	**IF** *high* bytes **THEN** *high* trust
B3	**IF** *low* bytes **THEN** *low* trust
AF1	**IF** *low* activity factor **THEN** *low* trust
AF2	**IF** *medium high* activity factor **THEN** *medium high* trust
AF3	**IF** *high* activity factor **THEN** *high* trust
AN1	**IF** *No* anonymous **THEN** *high* trust
AN2	**IF** *Yes* anonymous **THEN** *low* trust
U1	**IF** *low* pages **THEN** *low* trust
U2	**IF** *medium high* pages **THEN** *medium low* trust
U3	**IF** *high* pages **THEN** *medium high* trust
C1	**IF** *low* comments **THEN** *low* trust
C2	**IF** *medium low* comments **THEN** *medium low* trust
C3	**IF** *medium high* comments **THEN** *medium high* trust
C4	**IF** *high* comments **THEN** *high* trust
P1	**IF** *low* presence factor **THEN** *low* trust
P2	**IF** *medium low* presence factor **THEN** *medium low* trust
P3	**IF** *medium high* presence factor **THEN** *medium high* trust
P4	**IF** *high* presence factor **THEN** *high* trust
F1	**IF** *low* frequency factor **THEN** *low* trust
F2	**IF** *medium low* frequency factor **THEN** *medium low* trust
F3	**IF** *medium high* frequency factor **THEN** *medium high* trust
F4	**IF** *high* frequency factor **THEN** *high* trust
R1	**IF** *low* regularity factor **THEN** *low* trust
R2	**IF** *medium low* regularity factor **THEN** *medium low* trust
R3	**IF** *medium high* regularity factor **THEN** *medium high* trust
R4	**IF** *high* regularity factor **THEN** *high* trust
NM1	**IF** *very low* not minor **THEN** *low* trust
NM2	**IF** *medium to high* not minor **THEN** *high* trust

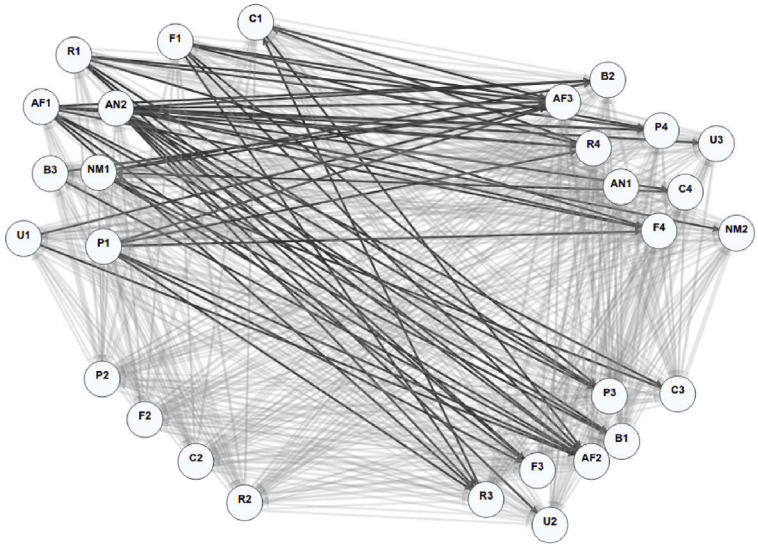


Fig. A.16. Graphical representation of knowledge base 2. Nodes represent (fuzzy) IF-THEN rules. Dark arrows represent contradictions between two rules due to some believed inconsistency by the author grounded in his domain knowledge. Light arrows represent contradictions between rules due to different consequents being inferred by each rule.

Table A.11
(fuzzy) Contradictions for knowledge base 1 designed by the author for inference of computational trust of Wikipedia editors.

Label	Internal structure
CC1	**IF** NM1 **THEN not** B1
CC2	**IF** NM1 **THEN not** B2
CC3	**IF** NM2 **THEN not** OnlyAge
CC4	**IF** P1 **THEN not** R4
CC5	**IF** AF1 **THEN not** R4
CC6	**IF** AF1 **THEN not** F4
CC7	**IF** R1 **THEN not** P4
CC8	**IF** F1 **THEN not** P4
CC9	**IF** NM1 **THEN not** AF2
CC10	**IF** NM1 **THEN not** AF3
CC11	**IF** NM2 **THEN not** U1
CC12	**IF** NM2 **THEN not** C1
CC13	**IF** NM2 **THEN not** P1
CC14	**IF** AN2 **THEN not** U2
CC15	**IF** AN2 **THEN not** U3
CC16	**IF** AN2 **THEN not** C3
CC17	**IF** AN2 **THEN not** C4
CC18	**IF** AN2 **THEN not** AF2
CC19	**IF** AN2 **THEN not** AF3
CC20	**IF** AN2 **THEN not** R4
CC21	**IF** AN2 **THEN not** F4
CC22	**IF** AN2 **THEN not** F3
CC23	**IF** AN2 **THEN not** R3
CC24	**IF** AN2 **THEN not** P3
CC25	**IF** AN2 **THEN not** P4
CC26	**IF** AN2 **THEN not** B2
CC27	**IF** AN2 **THEN not** B1
CC28	**IF** AN2 **THEN not** NM2
Bot.a	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** U4
Bot.b	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** U3
Bot.c	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** U2
Bot.d	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** C1
Bot.e	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** B2
Bot.f	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** B1
Bot.g	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** AF2
Bot.h	**IF** *Yes* anonymous *AND low* comments *AND* (*medium high* bytes OR *high* bytes) *AND very low* not minor *AND* (*high* unique pages OR *medium high* unique pages) **THEN not** AF3
Vandal.a	**IF** (*low* presence factor OR *medium low* presence factor) *AND low* regularity factor *AND low* comments *AND low* unique pages **THEN not** AF2
Vandal.b	**IF** (*low* presence factor OR *medium low* presence factor) *AND low* regularity factor *AND low* comments *AND low* unique pages **THEN not** AF3
Vandal.c	**IF** (*low* presence factor OR *medium low* presence factor) *AND low* regularity factor *AND low* comments *AND low* unique pages **THEN not** B1
Vandal.d	**IF** (*low* presence factor OR *medium low* presence factor) *AND low* regularity factor *AND low* comments *AND low* unique pages **THEN not** B2
OnlyAge.a	**IF** *low* frequency factor *AND low* regularity factor *AND low* activity factor **THEN not** P4
OnlyAge.b	**IF** *low* frequency factor *AND low* regularity factor *AND low* activity factor **THEN not** P3
OnlyAge.c	**IF** *low* frequency factor *AND low* regularity factor *AND low* activity factor **THEN not** P2

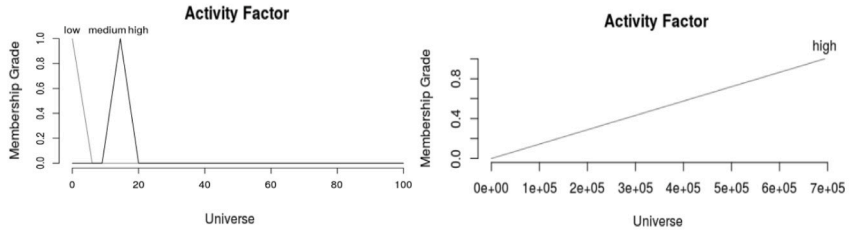
Table A.12
(fuzzy) Contradictions for knowledge base 2 designed by the author for inference of computational trust of Wikipedia editors. "Rule A $\nrightarrow$ Rule B" reads as "If Rule A then not Rule B". "Rule A $\nleftrightarrow$ Rule B" reads as two rules: "If Rule A then not Rule B", and "If Rule B then not Rule A".

Contradic. rules	Contradict. rules	Contradict. rules	Contradict. rules
AN2 $\nrightarrow$ AF3	AN2 $\nrightarrow$ AF2	C1 $\nrightarrow$ AF3	C1 $\nrightarrow$ AF2
F1 $\nrightarrow$ AF3	F1 $\nrightarrow$ AF2	NM1 $\nrightarrow$ AF3	NM1 $\nrightarrow$ AF2
R1 $\nrightarrow$ AF2	R1 $\nrightarrow$ AF3	P1 $\nrightarrow$ AF3	P1 $\nrightarrow$ AF2
U1 $\nrightarrow$ AF2	U1 $\nrightarrow$ AF3	AN2 $\nrightarrow$ B1	AN2 $\nrightarrow$ B2
AN2 $\nrightarrow$ C3	AN2 $\nrightarrow$ C4	AN2 $\nrightarrow$ F4	AN2 $\nrightarrow$ F3
AN2 $\nrightarrow$ NM2	AN2 $\nrightarrow$ R4	AN2 $\nrightarrow$ R3	AN2 $\nrightarrow$ P3
AN2 $\nrightarrow$ P4	AN2 $\nrightarrow$ U2	AN2 $\nrightarrow$ U3	AF1 $\nrightarrow$ B2
AF1 $\nrightarrow$ B1	B3 $\nrightarrow$ AF2	B3 $\nrightarrow$ AF3	NM1 $\nrightarrow$ B1
NM1 $\nrightarrow$ B2	R3 $\nrightarrow$ C1	R4 $\nrightarrow$ C1	AF1 $\nrightarrow$ F4
AF1 $\nrightarrow$ F3	F1 $\nrightarrow$ R3	F1 $\nrightarrow$ R4	R1 $\nrightarrow$ F4
R1 $\nrightarrow$ F3	F1 $\nrightarrow$ P4	F1 $\nrightarrow$ P3	P1 $\nrightarrow$ F4
P1 $\nrightarrow$ F3	C4 $\nrightarrow$ NM1	C3 $\nrightarrow$ NM1	AF1 $\nrightarrow$ R3
AF1 $\nrightarrow$ R4	R1 $\nrightarrow$ P4	R1 $\nrightarrow$ P3	P1 $\nrightarrow$ R3
P1 $\nrightarrow$ R4	AF1 $\nrightarrow$ P4	AF1 $\nrightarrow$ P3	AF2 $\nleftrightarrow$ AF1
AF3 $\nleftrightarrow$ AF1	AN1 $\nleftrightarrow$ AF1	C2 $\nleftrightarrow$ AF1	C3 $\nleftrightarrow$ AF1
C4 $\nleftrightarrow$ AF1	F2 $\nleftrightarrow$ AF1	NM2 $\nleftrightarrow$ AF1	R2 $\nleftrightarrow$ AF1
P2 $\nleftrightarrow$ AF1	U2 $\nleftrightarrow$ AF1	U3 $\nleftrightarrow$ AF1	AF3 $\nleftrightarrow$ AF2
AN1 $\nleftrightarrow$ AF2	B2 $\nleftrightarrow$ AF2	C2 $\nleftrightarrow$ AF2	C4 $\nleftrightarrow$ AF2
F2 $\nleftrightarrow$ AF2	F4 $\nleftrightarrow$ AF2	NM2 $\nleftrightarrow$ AF2	R2 $\nleftrightarrow$ AF2
R4 $\nleftrightarrow$ AF2	P2 $\nleftrightarrow$ AF2	P4 $\nleftrightarrow$ AF2	U3 $\nleftrightarrow$ AF2
B1 $\nleftrightarrow$ AF3	C2 $\nleftrightarrow$ AF3	C3 $\nleftrightarrow$ AF3	F2 $\nleftrightarrow$ AF3
F3 $\nleftrightarrow$ AF3	R2 $\nleftrightarrow$ AF3	R3 $\nleftrightarrow$ AF3	P2 $\nleftrightarrow$ AF3
P3 $\nleftrightarrow$ AF3	U2 $\nleftrightarrow$ AF3	AN2 $\nleftrightarrow$ AN1	B1 $\nleftrightarrow$ AN1
C1 $\nleftrightarrow$ AN1	C2 $\nleftrightarrow$ AN1	C3 $\nleftrightarrow$ AN1	F1 $\nleftrightarrow$ AN1
F2 $\nleftrightarrow$ AN1	F3 $\nleftrightarrow$ AN1	NM1 $\nleftrightarrow$ AN1	R1 $\nleftrightarrow$ AN1
R2 $\nleftrightarrow$ AN1	R3 $\nleftrightarrow$ AN1	P1 $\nleftrightarrow$ AN1	P2 $\nleftrightarrow$ AN1
P3 $\nleftrightarrow$ AN1	U1 $\nleftrightarrow$ AN1	U2 $\nleftrightarrow$ AN1	B3 $\nleftrightarrow$ AN1
C2 $\nleftrightarrow$ AN2	F2 $\nleftrightarrow$ AN2	R2 $\nleftrightarrow$ AN2	P2 $\nleftrightarrow$ AN2
B2 $\nleftrightarrow$ B1	C1 $\nleftrightarrow$ B1	C2 $\nleftrightarrow$ B1	C4 $\nleftrightarrow$ B1
F1 $\nleftrightarrow$ B1	F2 $\nleftrightarrow$ B1	F4 $\nleftrightarrow$ B1	NM2 $\nleftrightarrow$ B1
R1 $\nleftrightarrow$ B1	R2 $\nleftrightarrow$ B1	R4 $\nleftrightarrow$ B1	P1 $\nleftrightarrow$ B1
P2 $\nleftrightarrow$ B1	P4 $\nleftrightarrow$ B1	U1 $\nleftrightarrow$ B1	U3 $\nleftrightarrow$ B1
B3 $\nleftrightarrow$ B1	C1 $\nleftrightarrow$ B2	C2 $\nleftrightarrow$ B2	C3 $\nleftrightarrow$ B2
F1 $\nleftrightarrow$ B2	F2 $\nleftrightarrow$ B2	F3 $\nleftrightarrow$ B2	R1 $\nleftrightarrow$ B2
R2 $\nleftrightarrow$ B2	R3 $\nleftrightarrow$ B2	P1 $\nleftrightarrow$ B2	P2 $\nleftrightarrow$ B2
P3 $\nleftrightarrow$ B2	U1 $\nleftrightarrow$ B2	U2 $\nleftrightarrow$ B2	B3 $\nleftrightarrow$ B2
C2 $\nleftrightarrow$ C1	C3 $\nleftrightarrow$ C1	C4 $\nleftrightarrow$ C1	F2 $\nleftrightarrow$ C1
F3 $\nleftrightarrow$ C1	F4 $\nleftrightarrow$ C1	NM2 $\nleftrightarrow$ C1	R2 $\nleftrightarrow$ C1
P2 $\nleftrightarrow$ C1	P3 $\nleftrightarrow$ C1	P4 $\nleftrightarrow$ C1	U2 $\nleftrightarrow$ C1
U3 $\nleftrightarrow$ C1	C3 $\nleftrightarrow$ C2	C4 $\nleftrightarrow$ C2	F1 $\nleftrightarrow$ C2
F3 $\nleftrightarrow$ C2	F4 $\nleftrightarrow$ C2	NM1 $\nleftrightarrow$ C2	NM2 $\nleftrightarrow$ C2
R1 $\nleftrightarrow$ C2	R3 $\nleftrightarrow$ C2	R4 $\nleftrightarrow$ C2	P1 $\nleftrightarrow$ C2
P3 $\nleftrightarrow$ C2	P4 $\nleftrightarrow$ C2	U1 $\nleftrightarrow$ C2	U2 $\nleftrightarrow$ C2
U3 $\nleftrightarrow$ C2	B3 $\nleftrightarrow$ C2	C4 $\nleftrightarrow$ C3	F1 $\nleftrightarrow$ C3
F2 $\nleftrightarrow$ C3	F4 $\nleftrightarrow$ C3	NM2 $\nleftrightarrow$ C3	R1 $\nleftrightarrow$ C3
R2 $\nleftrightarrow$ C3	R4 $\nleftrightarrow$ C3	P1 $\nleftrightarrow$ C3	P2 $\nleftrightarrow$ C3
P4 $\nleftrightarrow$ C3	U1 $\nleftrightarrow$ C3	U3 $\nleftrightarrow$ C3	B3 $\nleftrightarrow$ C3
F1 $\nleftrightarrow$ C4	F2 $\nleftrightarrow$ C4	F3 $\nleftrightarrow$ C4	R1 $\nleftrightarrow$ C4
R2 $\nleftrightarrow$ C4	R3 $\nleftrightarrow$ C4	P1 $\nleftrightarrow$ C4	P2 $\nleftrightarrow$ C4
P3 $\nleftrightarrow$ C4	U1 $\nleftrightarrow$ C4	U2 $\nleftrightarrow$ C4	B3 $\nleftrightarrow$ C4
F2 $\nleftrightarrow$ F1	F3 $\nleftrightarrow$ F1	F4 $\nleftrightarrow$ F1	NM2 $\nleftrightarrow$ F1
R2 $\nleftrightarrow$ F1	P2 $\nleftrightarrow$ F1	U2 $\nleftrightarrow$ F1	U3 $\nleftrightarrow$ F1
F3 $\nleftrightarrow$ F2	F4 $\nleftrightarrow$ F2	NM1 $\nleftrightarrow$ F2	NM2 $\nleftrightarrow$ F2
R1 $\nleftrightarrow$ F2	R3 $\nleftrightarrow$ F2	R4 $\nleftrightarrow$ F2	P1 $\nleftrightarrow$ F2
P3 $\nleftrightarrow$ F2	P4 $\nleftrightarrow$ F2	U1 $\nleftrightarrow$ F2	U2 $\nleftrightarrow$ F2
U3 $\nleftrightarrow$ F2	B3 $\nleftrightarrow$ F2	F4 $\nleftrightarrow$ F3	NM1 $\nleftrightarrow$ F3
NM2 $\nleftrightarrow$ F3	R2 $\nleftrightarrow$ F3	R4 $\nleftrightarrow$ F3	P2 $\nleftrightarrow$ F3
P4 $\nleftrightarrow$ F3	U1 $\nleftrightarrow$ F3	U3 $\nleftrightarrow$ F3	B3 $\nleftrightarrow$ F3
NM1 $\nleftrightarrow$ F4	R2 $\nleftrightarrow$ F4	R3 $\nleftrightarrow$ F4	P2 $\nleftrightarrow$ F4
P3 $\nleftrightarrow$ F4	U1 $\nleftrightarrow$ F4	U2 $\nleftrightarrow$ F4	B3 $\nleftrightarrow$ F4
NM2 $\nleftrightarrow$ NM1	R2 $\nleftrightarrow$ NM1	R3 $\nleftrightarrow$ NM1	R4 $\nleftrightarrow$ NM1
P2 $\nleftrightarrow$ NM1	P3 $\nleftrightarrow$ NM1	P4 $\nleftrightarrow$ NM1	U2 $\nleftrightarrow$ NM1
U3 $\nleftrightarrow$ NM1	R1 $\nleftrightarrow$ NM2	R2 $\nleftrightarrow$ NM2	R3 $\nleftrightarrow$ NM2
P1 $\nleftrightarrow$ NM2	P2 $\nleftrightarrow$ NM2	P3 $\nleftrightarrow$ NM2	U1 $\nleftrightarrow$ NM2
U2 $\nleftrightarrow$ NM2	B3 $\nleftrightarrow$ NM2	R2 $\nleftrightarrow$ R1	R3 $\nleftrightarrow$ R1
R4 $\nleftrightarrow$ R1	P2 $\nleftrightarrow$ R1	U2 $\nleftrightarrow$ R1	U3 $\nleftrightarrow$ R1
R3 $\nleftrightarrow$ R2	R4 $\nleftrightarrow$ R2	P1 $\nleftrightarrow$ R2	P3 $\nleftrightarrow$ R2
P4 $\nleftrightarrow$ R2	U1 $\nleftrightarrow$ R2	U2 $\nleftrightarrow$ R2	U3 $\nleftrightarrow$ R2
B3 $\nleftrightarrow$ R2	R4 $\nleftrightarrow$ R3	P2 $\nleftrightarrow$ R3	P4 $\nleftrightarrow$ R3

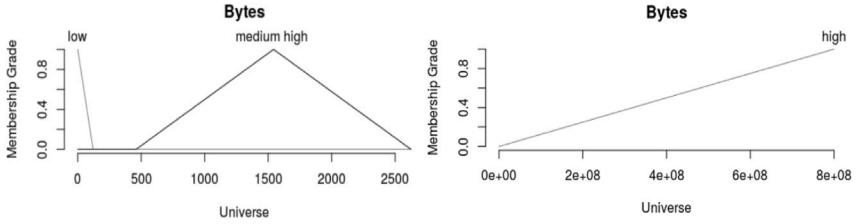
(*continued on next page*)

Table A.12 (*continued*).

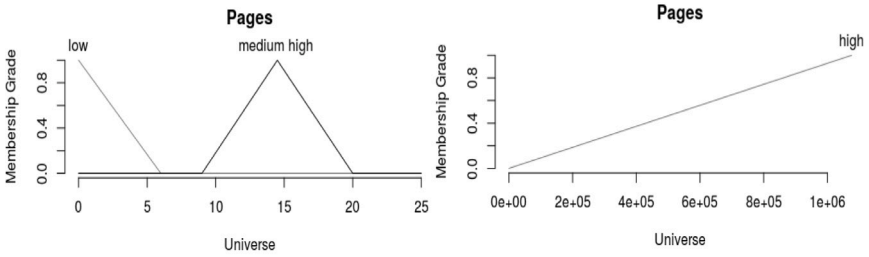
Contradic. rules	Contradict. rules	Contradict. rules	Contradict. rules
U1 ⇎ R3	U3 ⇎ R3	B3 ⇎ R3	P2 ⇎ R4
P3 ⇎ R4	U1 ⇎ R4	U2 ⇎ R4	B3 ⇎ R4
P2 ⇎ P1	P3 ⇎ P1	P4 ⇎ P1	U2 ⇎ P1
U3 ⇎ P1	P3 ⇎ P2	P4 ⇎ P2	U1 ⇎ P2
U2 ⇎ P2	U3 ⇎ P2	B3 ⇎ P2	P4 ⇎ P3
U1 ⇎ P3	U3 ⇎ P3	B3 ⇎ P3	U1 ⇎ P4
U2 ⇎ P4	B3 ⇎ P4	U2 ⇎ U1	U3 ⇎ U1
U3 ⇎ U2	B3 ⇎ U2	B3 ⇎ U3	



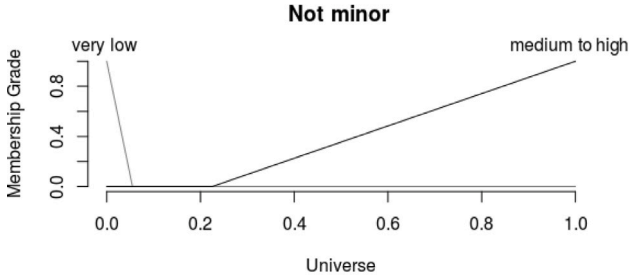
(a) Activity factor levels



(b) Bytes levels

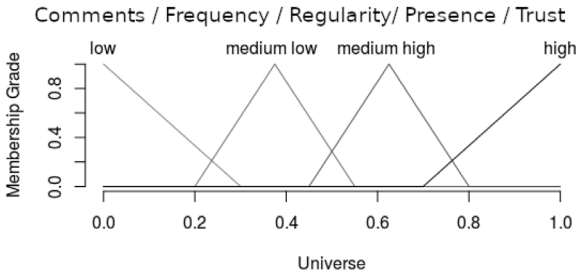


(c) Pages levels

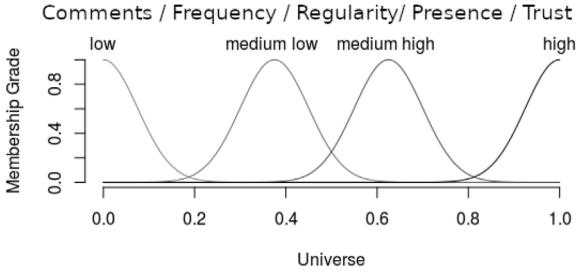


(d) Not minor levels

Fig. A.17. Employed fuzzy membership functions for different levels related to computational trust itself and its selected features.



(e) Triangular functions for comments, frequency factor, regularity factor, presence factor, and computational trust



(f) Gaussian functions for comments, frequency factor, regularity factor, and presence factor, and computational trust

Fig. A.17. (*continued*).

Table B.13
List or results achieved by each model for each evaluated metric when instantiated by the Wikipedia Italian edition and by the Wikipedia Portuguese edition. The full list of models and their configurations are given in Tables 4, 5, and 6. The sum of the absolute difference between results achieved for each dataset is given in the last column. The closer to 0 this difference gets, the more stable a model behaved when instantiated with two different datasets.

Model	Italian dataset			Portuguese dataset			Sum abs. diff.
	Rank	Spread	NAs	Rank	Spread	NAs	
*A*01	0.38	0.11	0	0.28	0.1	0	0.11
*A*02	0.37	0.11	0	0.28	0.1	0	0.11
*A*03	0.38	0.11	0	0.28	0.1	0	0.11
*A*04	0.42	0.08	0	0.45	0.11	0	0.06
*A*05	0.29	0.08	0	0.46	0.11	0	0.21
*A*06	0.29	0.08	0	0.45	0.11	0	0.2
*A*07	2.36	0.15	0	0.66	0.08	0	1.77
*A*08	5.3	0.2	0	7.14	0.23	0	1.87
*A*09	0	0.02	51.31	0	0.02	50.43	0.88
*A*10	8.19	0.03	0	12.13	0.02	0	3.94
*A*11	9.5	0.04	0	13.35	0.04	0	3.86
*A*12	8.19	0.03	0	12.13	0.02	0	3.94
*E*01	0.32	0.07	0	1.24	0.13	0	0.98
*E*02	0.32	0.07	0	1.78	0.14	0	1.51
*E*03	0.37	0.11	0	0.28	0.1	0	0.11
*E*04	0.42	0.08	0	0.46	0.11	0	0.07
*E*05	0	0.02	51.31	0	0.07	50.43	0.93
*E*06	0	0.02	51.31	0	0.08	50.43	0.94
*E*07	0	0.02	51.31	0	0.07	50.43	0.93
*E*08	0	0.02	51.31	0	0.08	50.43	0.94
*FC*01	0.51	0.08	0	0.72	0.08	0	0.21
*FC*02	10.4	0	0	14.05	0	0	3.66
*FC*03	0.43	0.08	0	0.44	0.08	0	0.01
*FC*04	10.4	0	0	14.05	0	0	3.66
*FC*05	0.31	0.08	0	0.41	0.08	0	0.11
*FC*06	4.7	0.04	0	5.95	0.04	0	1.26
*FC*07	0.51	0.08	0	0.72	0.08	0	0.21
*FC*08	10.4	0	0	14.05	0	0	3.66
*FC*09	0.43	0.08	0	0.44	0.08	0	0.01
*FC*10	10.4	0	0	14.05	0	0	3.66
*FC*11	0.31	0.08	0	0.41	0.08	0	0.11

(*continued on next page*)

Table B.13 (*continued*).

Model	Italian dataset			Portuguese dataset			Sum abs. diff.
	Rank	Spread	NAs	Rank	Spread	NAs	
*FC*12	4.7	0.04	0	5.95	0.04	0	1.26
*FC*13	1.96	0.02	3.56	2.47	0.02	5.17	2.12
*FC*14	1.96	0.03	3.52	2.41	0.03	5.1	2.04
*FC*15	1.73	0.02	3.43	2.1	0.02	5.01	1.94
*FC*16	1.79	0.04	3.38	2.19	0.04	4.94	1.96
*FC*17	1.74	0.02	3.43	2.09	0.02	5.01	1.93
*FC*18	1.8	0.04	3.38	2.18	0.04	4.94	1.95
*FC*19	1.96	0.02	3.56	2.47	0.02	5.17	2.12
*FC*20	1.96	0.03	3.52	2.41	0.03	5.1	2.04
*FC*21	1.73	0.02	3.43	2.1	0.02	5.01	1.94
*FC*22	1.79	0.04	3.38	2.19	0.04	4.94	1.96
*FC*23	1.74	0.02	3.43	2.09	0.02	5.01	1.93
*FC*24	1.8	0.04	3.38	2.18	0.04	4.94	1.95
*FL*01	0.5	0.12	0	0.34	0.07	0	0.2
*FL*02	10.4	0	0	14.05	0	0	3.66
*FL*03	4.51	0.19	0	0.25	0.07	0	4.38
*FL*04	10.4	0	0	14.05	0	0	3.66
*FL*05	0.5	0.12	0	0.26	0.07	0	0.28
*FL*06	5.23	0.07	0	5.19	0.02	0	0.09
*FL*07	0.5	0.12	0	0.34	0.07	0	0.21
*FL*08	10.4	0	0	14.05	0	0	3.66
*FL*09	3.21	0.18	0	0.25	0.07	0	3.06
*FL*10	10.4	0	0	14.05	0	0	3.66
*FL*11	0.5	0.12	0	0.26	0.07	0	0.28
*FL*12	5.23	0.07	0	5.19	0.02	0	0.09
*FL*13	1.59	0.01	3.59	1.99	0.01	5.32	2.12
*FL*14	1.59	0.04	3.59	1.99	0.03	5.32	2.12
*FL*15	1.27	0.02	3.59	1.42	0.02	5.31	1.88
*FL*16	1.27	0.05	3.59	1.42	0.04	5.31	1.88
*FL*17	1.27	0.02	3.59	1.41	0.02	5.31	1.86
*FL*18	1.27	0.05	3.59	1.41	0.05	5.31	1.86
*FL*19	1.59	0.01	3.59	1.99	0.01	5.32	2.12
*FL*20	1.59	0.04	3.59	1.99	0.03	5.32	2.12
*FL*21	1.27	0.02	3.59	1.42	0.02	5.31	1.88
*FL*22	1.27	0.05	3.59	1.42	0.04	5.31	1.88
*FL*23	1.27	0.02	3.59	1.41	0.02	5.31	1.86
*FL*24	1.27	0.05	3.59	1.41	0.05	5.31	1.86

References

[1] G. Brewka, Nonmonotonic Reasoning: Logical Foundations of Commonsense, Vol. 12, Cambridge University Press, 1991.

[2] R. Reiter, Nonmonotonic reasoning, in: Exploring Artificial Intelligence, Elsevier, 1988, pp. 439–481.

[3] J.L. Pollock, Knowledge and Justification, Princeton University Press, Princeton, 1974.

[4] J. McCarthy, Circumscription—a form of non-monotonic reasoning, Artificial Intelligence 13 (1–2) (1980) 27–39.

[5] R.A. Kowalski, F. Sadri, Logic programs with exceptions, New Gener. Comput. (ISSN: 1882-7055) 9 (3) (1991) 387–400, http://dx.doi.org/10.1007/BF03037170.

[6] L. Longo, A defeasible reasoning framework for human mental workload representation and assessment, Behav. Inf. Technol. 34 (8) (2015) 758–786.

[7] J.F. Horty, R.H. Thomason, D.S. Touretzky, A skeptical theory of inheritance in nonmonotonic semantic networks, Artificial Intelligence 42 (2–3) (1990) 311–348.

[8] M.L. Ginsberg, Non-monotonic reasoning using Dempster's rule, in: AAAI, Vol. 84, 1984, pp. 112–119.

[9] R.C. Moore, Semantical considerations on nonmonotonic logic, Artificial Intelligence 25 (1) (1985) 75–94.

[10] R. Reiter, A logic for default reasoning, Artificial Intelligence 13 (1–2) (1980) 81–132.

[11] R. Akerkar, P. Sajja, Knowledge-Based Systems, Jones & Bartlett Publishers, 2009.

[12] L.A. Zadeh, et al., Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.

[13] J. Durkin, J. Durkin, Expert Systems: Design and Development, Prentice Hall PTR, 1998.

[14] D. Bryant, P. Krause, A review of current defeasible reasoning implementations, Knowl. Eng. Rev. 23 (3) (2008) 227–260.

[15] H. Prakken, G. Vreeswijk, Logics for defeasible argumentation, in: Handbook of Philosophical Logic, Springer, 2001, pp. 219–318.

[16] G. Guida, C. Tasso, Design and Development of Knowledge-Based Systems: From Life Cycle to Methodology, John Wiley & Sons, Inc., 1995.

[17] T.J. Bench-Capon, P.E. Dunne, Argumentation in artificial intelligence, Artificial Intelligence 171 (10–15) (2007) 619–641.

[18] L. Longo, Argumentation for knowledge representation, conflict resolution, defeasible inference and its integration with machine learning, in: Machine Learning for Health Informatics, Springer, 2016, pp. 183–208.

[19] S.A. Gomez, C.I. Chesñevar, Integrating defeasible argumentation and machine learning techniques, in: Technical Report, Universidad Nacional del Sur, 2004.

[20] C. Chesñevar, A.G. Maguitman, M.P. González, Empowering recommendation technologies through argumentation, in: G. Simari, I. Rahwan (Eds.), Argumentation in Artificial Intelligence, Springer US, Boston, MA, ISBN: 978-0-387-98197-0, 2009, pp. 403–422.

[21] J. Sabater, C. Sierra, Review on computational trust and reputation models, Artif. Intell. Rev. (ISSN: 1573-7462) 24 (1) (2005) 33–60.

[22] L. Rizzo, L. Longo, An empirical evaluation of the inferential capacity of defeasible argumentation, non-monotonic fuzzy reasoning and expert systems, Expert Syst. Appl. (ISSN: 0957-4174) 147 (2020) 113220, http://dx.doi.org/10.1016/j.eswa.2020.113220.

[23] L. Longo, L. Rizzo, P. Dondio, Examining the modelling capabilities of defeasible argumentation and non-monotonic fuzzy reasoning, Knowl.-Based Syst. (ISSN: 0950-7051) 211 (2021) 106514, http://dx.doi.org/10.1016/j.knosys.2020.106514.

[24] L. Rizzo, L. Majnaric, P. Dondio, L. Longo, An investigation of argumentation theory for the prediction of survival in elderly using biomarkers, in: L. Iliadis, I. Maglogiannis, V. Plagianakos (Eds.), Artificial Intelligence Applications and Innovations, Springer International Publishing, Cham, 2018, pp. 385–397.

[25] L. Rizzo, L. Majnaric, L. Longo, A comparative study of defeasible argumentation and non-monotonic fuzzy reasoning for elderly survival prediction using biomarkers, in: C. Ghidini, B. Magnini, A. Passerini, P. Traverso (Eds.), AI*IA 2018 – Advances in Artificial Intelligence, Springer International Publishing, Cham, 2018, pp. 197–209.

[26] P. Jackson, P. Jackson, Introduction to Expert Systems, Vol. 2, Addison-Wesley, Reading, MA, 1990.

[27] C. Grosan, A. Abraham, Rule-based expert systems, in: Intelligent Systems, Springer, 2011, pp. 149–185.

[28] D.M. Gabbay, Theoretical foundations for non-monotonic reasoning in expert systems, in: Logics and Models of Concurrent Systems, Springer, 1985, pp. 439–457.

[29] L. Morgenstern, M. Singh, An expert system using nonmonotonic techniques for benefits inquiry in the insurance industry, in: IJCAI (1), 1997, pp. 655–661.

[30] L. Morgenstern, Inheritance comes of age: Applying nonmonotonic techniques to problems in industry, Artificial Intelligence 103 (1–2) (1998) 237–271.

[31] F. Puppe, Systematic Introduction to Expert Systems: Knowledge Representation and Problem-Solving Methods, Springer-Verlag, Berlin, 1993.

[32] D. Nute, R.I. Mann, B.F. Brewer, Controlling expert system recommendations with defeasible logic, Decis. Support Syst. 6 (2) (1990) 153–164.

[33] E.-S. El-Azhary, A. Edrees, A. Rafea, Diagnostic expert system using non-monotonic reasoning, Expert Syst. Appl. 23 (2) (2002) 137–144.

[34] R. Lewis, C. Roberts, Using non-monotonic reasoning to manage uncertainty in railway asset diagnostics, Expert Syst. Appl. (ISSN: 0957-4174) 37 (5) (2010) 3616–3623, http://dx.doi.org/10.1016/j.eswa.2009.10.030.

[35] A. Kandel, Fuzzy Expert Systems, CRC Press, 1991.

[36] K.M. Passino, S. Yurkovich, M. Reinfrank, Fuzzy Control, Vol. 42, Citeseer, 1998.

[37] O. Cordón, A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems, Internat. J. Approx. Reason. 52 (6) (2011) 894–913.

[38] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, in: Readings in Fuzzy Sets for Intelligent Systems, Elsevier, 1993, pp. 387–403.

[39] Y. Tsukamoto, An approach to fuzzy reasoning method, in: M.M. Gupta, R.K. Ragade, R.R. Yager (Eds.), Advances in Fuzzy Set Theory and Applications, NorthHolland, Amsterdam, 1979.

[40] E.H. Mamdani, Application of fuzzy algorithms for control of simple dynamic plant, Proc. Inst. Electr. Eng. (ISSN: 0020-3270) 121 (12) (1974) 1585–1588, http://dx.doi.org/10.1049/piee.1974.0328.

[41] T.J. Ross, Fuzzy Logic with Engineering Applications, McGraw-Hill, New York, 1995.

[42] H. Hellendoorn, C. Thomas, Defuzzification in fuzzy controllers, J. Intell. Fuzzy Systems 1 (2) (1993) 109–123.

[43] J.L. Castro, E. Trillas, J.M. Zurita, Non-monotonic fuzzy reasoning, Fuzzy Sets and Systems 94 (2) (1998) 217–225.

[44] A. Gegov, N. Gobalakrishnan, D. Sanders, Rule base compression in fuzzy systems by filtration of non-monotonic rules, J. Intell. Fuzzy Systems 27 (4) (2014) 2029–2043.

[45] W. Siler, J.J. Buckley, Fuzzy Expert Systems and Fuzzy Reasoning, John Wiley & Sons, 2005.

[46] D. Dubois, H. Prade, Possibility theory: qualitative and quantitative aspects, in: Quantified Representation of Uncertainty and Imprecision, Springer, 1998, pp. 169–226.

[47] C.I. Chesñevar, A.G. Maguitman, R.P. Loui, Logical models of argument, ACM Comput. Surv. 32 (4) (2000) 337–383.

[48] H. Prakken, G. Sartor, The role of logic in computational models of legal argument: A critical survey, in: A.C. Kakas, F. Sadri (Eds.), Computational Logic: Logic Programming and beyond: Essays in Honour of Robert a. Kowalski Part II, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN: 978-3-540-45632-2, 2002, pp. 342–381.

[49] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. Simari, M. Thimm, S. Villata, Towards artificial argumentation, AI Mag. 38 (3) (2017) 25–36.

[50] S. Toulmin, The Uses of Argument, Cambridge University Press, 1958.

[51] J. Bentahar, B. Moulin, M. Bélanger, A taxonomy of argumentation models used for knowledge representation, Artif. Intell. Rev. 33 (3) (2010) 211–259.

[52] D. Walton, Argumentation Schemes for Presumptive Reasoning, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[53] H. Prakken, An abstract framework for argumentation with structured arguments, Argum. Comput. 1 (2) (2010) 93–124.

[54] J.L. Pollock, Defeasible reasoning, Cogn. Sci. 11 (4) (1987) 481–518.

[55] G. Vreeswijk, Studies in Defeasible Argumentation (Ph.D. thesis), Vrije University Amsterdam, 1993.

[56] C. Cayrol, M.-C. Lagasquie-Schiex, Bipolar abstract argumentation systems, in: Argumentation in Artificial Intelligence, Springer US, Boston, MA, ISBN: 978-0-387-98197-0, 2009, pp. 65–84.

[57] S. Villata, G. Boella, D.M. Gabbay, L. van der Torre, Modelling defeasible and prioritized support in bipolar argumentation, Ann. Math. Artif. Intell. (ISSN: 1573-7470) 66 (1) (2012) 163–197.

[58] C. Cayrol, M.-C. Lagasquie-Schiex, Coalitions of arguments: A tool for handling bipolar argumentation frameworks, Int. J. Intell. Syst. 25 (1) (2010) 83–109.

[59] L. Amgoud, C. Cayrol, M.C. Lagasquie-Schiex, P. Livet, On bipolarity in argumentation frameworks, Int. J. Intell. Syst. 23 (10) (2008) 1062–1093.

[60] F. Nouioua, V. Risch, Bipolar argumentation frameworks with specialized supports, in: 2010 22nd IEEE International Conference on Tools with Artificial Intelligence, Vol. 1, 2010, pp. 215–218.

[61] P.E. Dunne, A. Hunter, P. McBurney, S. Parsons, M. Wooldridge, Weighted argument systems: Basic definitions, algorithms, and complexity results, Artificial Intelligence 175 (2) (2011) 457–486.

[62] H. Prakken, G. Sartor, Argument-based extended logic programming with defeasible priorities, J. Appl. Non-Class. Log. 7 (1–2) (1997) 25–75.

[63] J.L. Pollock, Cognitive Carpentry: A Blueprint for how to Build a Person, Mit Press, 1995.

[64] L. Amgoud, S. Vesic, Rich preference-based argumentation frameworks, Internat. J. Approx. Reason. (ISSN: 0888-613X) 55 (2) (2014) 585–606.

[65] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (2) (1995) 321–358.

[66] Y. Wu, M. Caminada, M. Podlaszewski, A labelling-based justification status of arguments, Studia Logica 3 (4) (2010) 12–29.

[67] M.W. Caminada, D.M. Gabbay, A logical account of formal argumentation, Studia Logica 93 (2–3) (2009) 109.

[68] M. Caminada, On the issue of reinstatement in argumentation, in: European Workshop on Logics in Artificial Intelligence, Springer, 2006, pp. 111–123.

[69] P. Baroni, M. Giacomin, Semantics of abstract argument systems, in: Argumentation in Artificial Intelligence, Springer, 2009, pp. 25–44.

[70] L. Amgoud, J. Ben-Naim, Ranking-based semantics for argumentation frameworks, in: International Conference on Scalable Uncertainty Management, Springer, 2013, pp. 134–147.

[71] L. Amgoud, J. Ben-Naim, D. Doder, S. Vesic, Ranking arguments with compensation-based semantics, in: Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning, 2016, pp. 12–21.

[72] C. Cayrol, M.-C. Lagasquie-Schiex, Graduality in argumentation, J. Artificial Intelligence Res. 23 (2005) 245–297.

[73] P.-A. Matt, F. Toni, A game-theoretic measure of argument strength for abstract argumentation, in: European Workshop on Logics in Artificial Intelligence, Springer, 2008, pp. 285–297.

[74] F. Pu, J. Luo, Y. Zhang, G. Luo, Argument ranking with categoriser function, in: R. Buchmann, C.V. Kifor, J. Yu (Eds.), Knowledge Science, Engineering and Management, Springer International Publishing, Cham, ISBN: 978-3-319-12096-6, 2014, pp. 290–301.

[75] P. Dondio, Ranking semantics based on subgraphs analysis, in: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2018, pp. 1132–1140, URL http://dl.acm.org/citation.cfm?id=3237383.3237864.

[76] S. Konieczny, P. Marquis, S. Vesic, On supported inference and extension selection in abstract argumentation frameworks, in: S. Destercke, T. Denoeux (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Springer International Publishing, Cham, 2015, pp. 49–59.

[77] C.F. Chang, A. Miller, A. Ghose, Mixed-initiative argumentation: Group decision support in medicine, in: P. Kostkova (Ed.), Electronic Healthcare, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN: 978-3-642-11745-9, 2010, pp. 43–50.

[78] A. Hunter, M. Williams, Argumentation for aggregating clinical evidence, in: Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on, Vol. 1, IEEE, 2010, pp. 361–368.

[79] R. Craven, F. Toni, C. Cadar, A. Hadad, M. Williams, Efficient argumentation for medical decision-making, in: Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning, 2012, pp. 598–602.

[80] V. Patkar, C. Hurt, R. Steele, S. Love, A. Purushotham, M. Williams, R. Thomson, J. Fox, Evidence-based guidelines and decision support services: a discussion and evaluation in triple assessment of suspected breast cancer, Br. J. Cancer 95 (11) (2006) 1490.

[81] D. Glasspool, J. Fox, A. Oettinger, J. Smith-Spark, Argumentation in decision support for medical care planning for patients and clinicians, in: AAAI Spring Symposium: Argumentation for Consumers of Healthcare, 2006, pp. 58–63.

[82] M.A. Grando, L. Moss, D. Sleeman, J. Kinsella, Argumentation-logic for creating and explaining medical hypotheses, Artif. Intell. Med. 58 (1) (2013) 1–13.

[83] L. Longo, Formalising Human Mental Workload as a Defeasible Computational Concept (Ph.D. thesis), Trinity College Dublin, 2014.

[84] P. Dondio, S. Barrett, Presumptive selection of trust evidence, in: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS '07, ACM, New York, NY, USA, ISBN: 978-81-904262-7-5, 2007, pp. 166:1–166:8.

[85] G. Brewka, J. Dix, K. Konolige, Nonmonotonic Reasoning: An Overview, Vol. 73, CSLI publications Stanford, 1997.

[86] U. Hlobil, Choosing your nonmonotonic logic: A shopper' guide, in: P. Arazim, T. Lávika (Eds.), The Logica Yearbook 2017, College Publications, London, 2018, pp. 109–123.

[87] K. Konolige, On the relation between default and autoepistemic logic, Artificial Intelligence (ISSN: 0004-3702) 35 (3) (1988) 343–382, http://dx.doi.org/10.1016/0004-3702(88)90021-5, URL https://www.sciencedirect.com/science/article/pii/0004370288900215.

[88] T. Delladio, N.D. Rotstein, G.R. Simari, A comparison between non-monotonic formalisms, in: VIII Workshop de Investigadores en Ciencias de la Computación, 2006.

[89] C. Dutilh Novaes, H. Veluwenkamp, Reasoning biases, non-monotonic logics and belief revision, Theoria 83 (1) (2017) 29–52.

[90] P. Gärdenfors, Belief Revision, Vol. 29, Cambridge Tracts in Theoretical Computer Science, 1992.

[91] O. Arieli, A. Borg, J. Heyninck, C. Straßer, Logic-based approaches to formal argumentation, J. Appl. Log. 2631 (6) (2021) 1793.

[92] A. Bondarenko, P.M. Dung, R.A. Kowalski, F. Toni, An abstract, argumentation-theoretic approach to default reasoning, Artificial Intelligence 93 (1–2) (1997) 63–101.
[93] K.H. Yang, D. Olson, J. Kim, Comparison of first order predicate logic, fuzzy logic and non-monotonic logic as knowledge representation methodology, Expert Syst. Appl. 27 (4) (2004) 501–519.
[94] S. Parsons, P. McBurney, E. Sklar, Reasoning about trust using argumentation: A position paper, in: Workshop on Argumentation in Multi-Agent Systems, 2010, pp. 159–170.
[95] S.P. Marsh, Formalizing Trust as a Computational Concept (Ph.D. thesis), University of Stirling, Department of Computer Science and Mathematics, 1994.
[96] E. Yashkina, A. Pinigin, J. Lee, M. Mazzara, A.S. Adekotujo, A. Zubair, L. Longo, Expressing trust with temporal frequency of user interaction in online communities, in: International Conference on Advanced Information Networking and Applications, Springer, 2019, pp. 1133–1146.
[97] A. Melnikov, J. Lee, V. Rivera, M. Mazzara, L. Longo, Towards dynamic interaction-based reputation models, in: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications, AINA, IEEE, 2018, pp. 422–428.
[98] L. Longo, S. Barrett, P. Dondio, Information foraging theory as a form of collective intelligence for social search, in: International Conference on Computational Collective Intelligence, Springer, 2009, pp. 63–74.
[99] L. Longo, P. Dondio, S. Barrett, Enhancing social search: a computational collective intelligence model of behavioural traits, trust and time, in: Transactions on Computational Collective Intelligence II, Springer, 2010, pp. 46–69.
[100] L. Longo, S. Barrett, P. Dondio, Toward social search - from explicit to implicit collaboration to predict users' interests, in: WEBIST 2009 - Proceedings of the Fifth International Conference on Web Information Systems and Technologies, Lisbon, Portugal, March 23-26, 2009, 2009, pp. 693–696.
[101] L. Longo, P. Dondio, R. Bresciani, S. Barrett, A. Butterfield, Enabling adaptation in trust computations, in: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009. COMPUTATIONWORLD'09. Computation World, IEEE, 2009, pp. 701–706.
[102] P. Dondio, L. Longo, S. Barrett, A translation mechanism for recommendations, in: IFIP International Conference on Trust Management, Springer, 2008, pp. 87–102.
[103] P.-A. Matt, M. Morgem, F. Toni, Combining statistics and arguments to compute trust, in: 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, Canada, Vol. 1, ACM, 2010, pp. 209–216.
[104] S. Parsons, K. Atkinson, Z. Li, P. McBurney, E. Sklar, M. Singh, K. Haigh, K. Levitt, J. Rowe, Argument schemes for reasoning about trust, Argum. Comput. 5 (2–3) (2014) 160–190.
[105] L. Amgoud, R. Demolombe, An argumentation-based approach for reasoning about trust in information sources, Argum. Comput. 5 (2–3) (2014) 191–215.
[106] Y. Tang, K. Cai, P. McBurney, E. Sklar, S. Parsons, Using argumentation to reason about trust and belief, J. Log. Comput. 22 (5) (2012) 979–1018.
[107] B.T. Adler, L. de Alfaro, A content-driven reputation system for the wikipedia, in: Proceedings of the 16th International Conference on World Wide Web, WWW '07, ACM, New York, NY, USA, ISBN: 978-1-59593-654-7, 2007, pp. 261–270.
[108] B.T. Adler, K. Chatterjee, L. de Alfaro, M. Faella, I. Pye, V. Raman, Assigning trust to wikipedia content, in: Proceedings of the 4th International Symposium on Wikis, WikiSym '08, ACM, New York, NY, USA, ISBN: 978-1-60558-128-6, 2008, pp. 26:1–26:12.
[109] H. Zeng, M.A. Alhossaini, L. Ding, R. Fikes, D.L. McGuinness, Computing trust from revision history, in: Intl. Conf. on Privacy, Security and Trust, 2006.
[110] Y. Krupa, L. Vercouter, J.F. Hübner, A. Herzig, Trust based evaluation of wikipedia's contributors, in: H. Aldewereld, V. Dignum, G. Picard (Eds.), Engineering Societies in the Agents World X, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 148–161.
[111] S. Javanmardi, C. Lopes, P. Baldi, Modeling user reputation in wikis, Stat. Anal. Data Min.: ASA Data Sci. J. 3 (2) (2010) 126–139.
[112] L. Rizzo, P. Dondio, L. Longo, Exploring the potential of defeasible argumentation for quantitative inferences in real-world contexts: An assessment of computational trust, in: L. Longo, L. Rizzo, E. Hunter, A. Pakrashi (Eds.), Proceedings of the 28th Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Republic of Ireland, December 7-8, 2020, in: CEUR Workshop Proceedings, vol. 2771, CEUR-WS.org, 2020, pp. 37–48.
[113] L. Rizzo, Evaluating the Impact of Defeasible Argumentation as a Modelling Technique for Reasoning under Uncertainty (Ph.D. thesis), Technological University Dublin, 2020.
[114] L. Longo, P. Dondio, S. Barrett, Temporal factors to evaluate trustworthiness of virtual identities, in: Third International Conference on Security and Privacy in Communications Networks and the Workshops-SecureComm, IEEE, 2007, pp. 11–19.
[115] N.R. Milton, Knowledge Acquisition in Practice: A Step-by-Step Guide, Springer, London, 2007.
[116] J. McDermott, C. Forgy, Production system conflict resolution strategies, in: Pattern-Directed Inference Systems, Elsevier, 1978, pp. 177–199.
[117] L. Rizzo, P. Dondio, S.J. Delany, L. Longo, Modeling mental workload via rule-based expert system: A comparison with NASA-TLX and workload profile, in: L. Iliadis, I. Maglogiannis (Eds.), Artificial Intelligence Applications and Innovations: 12th IFIP WG 12.5 International Conference and Workshops, AIAI 2016, Thessaloniki, Greece, September 16-18, 2016, Proceedings, Springer International Publishing, Cham, 2016, pp. 215–229.
[118] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms. Dordrecht, Kluwer Academic Publishers, Dordrecht, 2002.
[119] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 9 (4) (2001) 506–515.
[120] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A comparative study of ranking-based semantics for abstract argumentation, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[121] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artificial Intelligence 128 (1–2) (2001) 203–235.
[122] L. Rizzo, Argumentation framework, 2021, https://doi.org/10.5281/zenodo.6980180. (accessed 10 August 2022).
[123] L. Rizzo, Non-monotonic fuzzy reasoning implementation, 2021, https://doi.org/10.5281/zenodo.5088015. (accessed 10 August 2022).
[124] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, Combining extension-based semantics and ranking-based semantics for abstract argumentation, in: Sixteenth International Conference on Principles of Knowledge Representation and Reasoning, 2018, pp. 118–127.
[125] S. Coste-Marquis, S. Konieczny, P. Marquis, M.A. Ouali, Selecting extensions in weighted argumentation frameworks, in: COMMA, 2012, pp. 342–349.
[126] P. Dondio, Propagating degrees of truth on an argumentation framework: an abstract account of fuzzy argumentation, in: Proceedings of the Symposium on Applied Computing, ACM, 2017, pp. 995–1002.
[127] J. Janssen, M. De Cock, D. Vermeir, Fuzzy argumentation frameworks, in: Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2008, pp. 513–520.
[128] J. Heyninck, M. Thimm, G. Kern-Isberner, T. Rienstra, K. Skiba, On the relation between possibilistic logic and abstract dialectical frameworks, in: Proceedings of the 19th International Workshop on Non-Monotonic Reasoning (NMR'21), 2021.
[129] H. Li, N. Oren, T.J. Norman, Probabilistic argumentation frameworks, in: S. Modgil, N. Oren, F. Toni (Eds.), Theorie and Applications of Formal Argumentation, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 1–16.
[130] C. Cayrol, M.C. Lagasquie-Schiex, On the acceptability of arguments in bipolar argumentation frameworks, in: L. Godo (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN: 978-3-540-31888-0, 2005, pp. 378–389.
[131] D. Nauck, F. Klawonn, R. Kruse, Foundations of Neuro-Fuzzy Systems, John Wiley & Sons, Inc., 1997.
[132] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends, Fuzzy Sets and Systems 141 (1) (2004) 5–31.
[133] M.G. Augasta, T. Kathirvalavakumar, Reverse engineering the neural networks for rule extraction in classification problems, Neural Process. Lett. (ISSN: 1573-773X) 35 (2) (2012) 131–150.
[134] N. Barakat, A.P. Bradley, Rule extraction from support vector machines: A review, Neurocomputing (ISSN: 0925-2312) 74 (1) (2010) 178–190, Artificial Brains.
[135] Y. Chen, N.R. Pal, I. Chung, An integrated mechanism for feature selection and fuzzy rule extraction for classification, IEEE Trans. Fuzzy Syst. (ISSN: 1941-0034) 20 (4) (2012) 683–698.
[136] G. Vilone, L. Rizzo, L. Longo, A comparative analysis of rule-based, model-agnostic methods for explainable artificial intelligence, in: L. Longo, L. Rizzo, E. Hunter, A. Pakrashi (Eds.), Proceedings of the 28th Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Republic of Ireland, December 7-8, 2020, in: CEUR Workshop Proceedings, vol. 2771, CEUR-WS.org, 2020, pp. 85–96.
[137] G. Vilone, L. Longo, Notions of explainability and evaluation approaches for explainable artificial intelligence, Inf. Fusion (ISSN: 1566-2535) 76 (2021) 89–106, http://dx.doi.org/10.1016/j.inffus.2021.05.009.
[138] A. Abdul, J. Vermeulen, D. Wang, B.Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, ACM, New York, NY, USA, ISBN: 978-1-4503-5620-6, 2018, pp. 582:1–582:18.
[139] A. Preece, Asking 'Why' in AI: Explainability of intelligent systems – perspectives and challenges, Intell. Syst. Account. Finance Manag. 25 (2) (2018) 63–72.
[140] Z.C. Lipton, The mythos of model interpretability, Queue 16 (3) (2018) 30:31–30:57.
[141] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38.