Константин Фролов[1]

О МОДЕЛИ БЕНЧ-КАПОНА ДЛЯ СТРУКТУРЫ АРГУМЕНТАЦИИ, ОСНОВАННОЙ НА ЦЕННОСТЯХ[2]

Аннотация. В статье рассматриваются выразительные возможности модели Бенч-Капона для структуры аргументации, основанной на ценностях. Основным отличием этой модели от классического подхода Дунга является то обстоятельство, что, помимо отношения атаки, определённого на упорядоченных парах аргументов, в этой модели выделяется особый подкласс данного отношения — отношение *отмены*, то есть успешной атаки с точки зрения определённой аудитории с учетом её ценностных приоритетов. Эта особенность расширяет выразительные возможности данной модели, а также позволяет иметь *единственное* непустое предпочтительное расширение для данной аудитории для всякой абстрактной структуры аргументации, в которой отсутствуют циклы, состоящие из аргументов, поддерживающих одну и ту же ценность.
Ключевые слова: аргументация, абстрактная структура аргументации, ценности, предпочтительное расширение.

Konstantin Frolov

ON THE BENCH-CAPON VALUE-BASED ARGUMENTATION FRAMEWORK

Abstract. In this paper I examine the expressive power of the Bench-Capon value-based argumentation framework. In frameworks of this type we have two types of relations between arguments: the attack relation and the defeat relation, which is a class of successful attacks from the point of view of a certain audience, taking into account its value priorities. The main advantage of this approach is that every audience-specific value-based argumentation framework with no single-valued cycles has a unique, nonempty preferred extension.
Keywords: argumentation, argumentation framework, values, preferred extension.

Для цитирования: *Фролов К. Г.* О модели Бенч-Капона для структуры аргументации, основанной на ценностях // Логико-философские штудии. 2022. Т. 20, № 4. С. 447–455. DOI: 10.52119/LPHS.2022.33.41.006.

---

[1] *Фролов Константин Геннадьевич* — научный сотрудник Международной лаборатории логики, лингвистики и формальной философии НИУ ВШЭ, Москва; старший научный сотрудник Института философии СПбГУ, Санкт-Петербург.
Konstantin Frolov, International Laboratory for Logic, Linguistics and Formal Philosophy, HSE University, Moscow.
kgfrolov@hse.ru

[2] Статья подготовлена при финансовой поддержке Российского научного фонда, проект 20-18-00158 «Формальная философия аргументации и комплексная методология поиска и отбора решений спора».

Основные идеи моделирования абстрактной структуры аргументации сформулированы П. Дунгом в (Dung 1995). Согласно его подходу, который к настоящему моменту стал уже классическим, структура аргументации может быть представлена в виде ориентированного графа. Вершины такого графа репрезентируют отдельные доводы и аргументы, которые принимаются в качестве неких атомов аргументации, внутренняя структура которых нас в данном случае не интересует, поскольку целью такого моделирования является анализ отношений между аргументами. Соответственно, направления релевантной атаки, которые имеются между аргументами, отображаются направленными дугами данного графа.

Идеи подхода Дунга наследует модель Т. Бенч-Капона для структуры аргументации, основанной на ценностях (*value-based argumentation framework, VAF-модель*; Bench-Capon 2003), в рамках которой становится возможным провести различие между отношением атаки, которое определено на упорядоченных парах аргументов, и подвидом данного отношения — отношением *отмены*, то есть успешной атаки с точки зрения определённой аудитории с учетом её ценностей и приоритетов (Bench-Capon, Dunne 2002).

В первой части статьи мы приведём ряд строгих определений, которые существенны для данной модели, а затем проиллюстрируем её выразительные возможности на наглядном примере.

Определение 1. *Структура аргументации, основанной на ценностях*, представляет собой кортеж $\langle AR, attacks, V, val, P\rangle$, где

- AR — множество аргументов;
- $attacks$ — бинарное отношение атаки, представляющее собой подмножество декартова произведения AR на само себя, $attacks \subseteq AR \times AR$;
- V — множество ценностей;
- val — функция, которая ставит в соответствие каждому аргументу из множества AR ценность из множества V, которой этот аргумент соответствует в наибольшей степени;
- P — множество индексов для потенциальных аудиторий спора, характеристической чертой которых являются различия в ценностных приоритетах, проявляющиеся в виде перестановок ценностей из множества V в упорядоченном списке, начинающемся с наиболее приоритетной ценности и заканчивающемся ценностью, наименее приоритетной для данной аудитории.

Определение 2. *Чувствительная к аудитории структура аргументации, основанной на ценностях*, представляет собой кортеж $\langle AR, attacks, V, val, Vpref_a\rangle$, где

- $AR, attacks, V, val$ интерпретируются так же, как и в предыдущем случае;
- $Vpref_a$ представляет собой транзитивное и асимметричное отношение предпочтения со стороны аудитории a между ценностями из множества V, т. е. $Vpref_a \subseteq V \times V$.

Асимметричность отношения предпочтения $Vpref_a$ влечёт его антирефлексивность, что в сочетании с транзитивностью гарантирует в рамках данной модели отсутствие циклов в ценностных предпочтениях аудитории, то есть невозможность наличия в множестве V такого подмножества ценностей $\{v_1, \dots, v_n\}$, что $Vpref_a(v_1, v_2), \dots, Vpref_a(v_{n-1}, v_n)$, $Vpref_a(v_n, v_1)$.

Определение 3. Аргумент $A \in AR$ *отменяет* аргумент $B \in AR$ для аудитории a, если и только если A атакует B и при этом ценность, соответствующая атакуемому аргументу B, не является для аудитории a более значимой, чем ценность, соответствующая аргументу B. Формально: $attacks(A, B)$, и не имеет места $Vpref_a(val(B), val(A))$.

Определение 4. Аргумент $A \in AR$ является *приемлемым* для аудитории a на множестве аргументов S, если для любого аргумента B из множества AR, отменяющего для аудитории a аргумент A, найдётся аргумент C из множества S, отменяющий для аудитории a аргумент B. В формальной записи:

$$(\forall x)((x \in AR \wedge defeats_a(x, A)) \rightarrow (\exists y)(y \in S \wedge defeats_a(y, x)))$$

Определение 5. Множество аргументов S является *бесконфликтным* для аудитории a, если ни один из аргументов в S не отменяет для аудитории a ни один из аргументов в S. В формальном виде:

$$\neg(\exists x)(\exists y)(x \in S \wedge y \in S \wedge defeats_a(x, y)),$$

что эквивалентно

$$(\forall x)(\forall y)((x \in S \wedge y \in S) \rightarrow (\neg attacks_a(x, y) \vee Vpref(val(y), val(x)) \in Vpref_a))$$

Определение 6. Множество аргументов S, бесконфликтное для аудитории a, называется *допустимым для аудитории* a, если оно состоит исключительно из аргументов, приемлемых для аудитории a.

Определение 7. Множество аргументов S в рамках структуры ценностно обоснованной аргументации VAF называется *предпочтительным расширением для аудитории* a (обозначается $Pref_a$), если оно является максимальным допустимым для аудитории a множеством аргументов, то есть если в множество S не может быть добавлен ни один не входящий в него аргумент из AR без того, чтобы данное множество не перестало быть допустимым для аудитории a.

Определение 8. Аргумент $A \in AR$ является *объективно приемлемым*, если и только если он является элементом предпочтительного расширения для каждой аудитории из множества P. В формальной записи: $(\forall p \in P)(A \in Pref_p)$.

Определение 9. Аргумент $A \in AR$ является *субъективно приемлемым*, если и только если он является элементом предпочтительного расширения для какой-нибудь аудитории из множества P. В формальной записи: $(\exists p \in P)(A \in Pref_p)$.

Рассмотрим теперь иллюстративный спор, на примере которого в дальнейшем будут продемонстрированы выразительные возможности VAF-модели.

Пусть у нас имеются агенты K и D, которые обмениваются между собой следующими аргументами.

K: (1) Тебе надо сделать прививку. (2) Это поможет достичь коллективного иммунитета, что является необходимым условием для полного снятия многочисленных ограничений.

D: (3) Я не согласен делать прививку. Абсолютно безопасных прививок не существует. При таких процедурах можно заразиться даже ВИЧ, не говоря уже о непредсказуемых индивидуальных аллергических реакциях. Не вижу смысла рисковать.

K: (4) Этими рисками следует поступиться ради здоровья и безопасности других людей. Негативные последствия в результате прививок носят единичный характер, тогда как позитивные последствия в виде выработки иммунитета к данной инфекции носят массовый характер.

D: (5) Даже если я сделал бы прививку, это не гарантировало бы здоровья и безопасности для других людей. Одной моей прививки было бы явно недостаточно для достижения коллективного иммунитета и снятия ограничений. Необоснованно утверждать, что от моего отказа или согласия сделать прививку вообще зависит судьба коллективного иммунитета и карантинных ограничений.

K: (6) Никто не говорит, что от тебя зависит судьба всей страны или человечества, но твоя прививка определённо могла бы способствовать достижению коллективного иммунитета и скорейшему снятию ограничений. Именно поэтому её и стоит сделать.

D: (7) Для меня в этом нет никакого смысла. Если другие, следуя твоим аргументам, сделают прививку, то коллективный иммунитет будет достигнут, ограничения сняты, а я при этом избегу всех рисков, связанных с прививкой. Для меня это оптимально.

K: (8) Но если другие, следуя твоим аргументам, не сделают прививку, то коллективный иммунитет будет достигнут еще очень нескоро, в результате чего от этой инфекции умрёт множество людей, которые могли бы жить дальше. В их числе можешь оказаться ты сам.

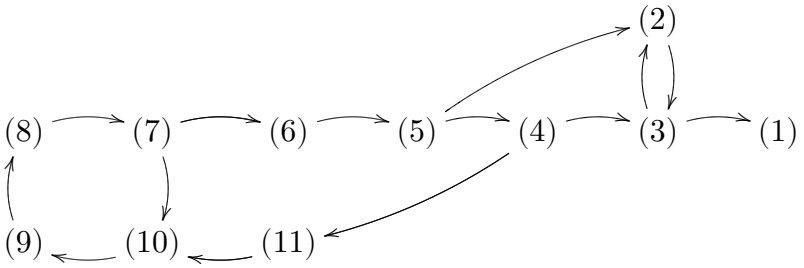
D: (9) Уверяю тебя, в этом случае я не стал бы, в отличие от тебя, винить в произошедшем окружающих, которые отказались прививаться для того,

чтобы меня спасти. Не следует возлагать ответственность за события, имеющие естественные причины, на других людей.

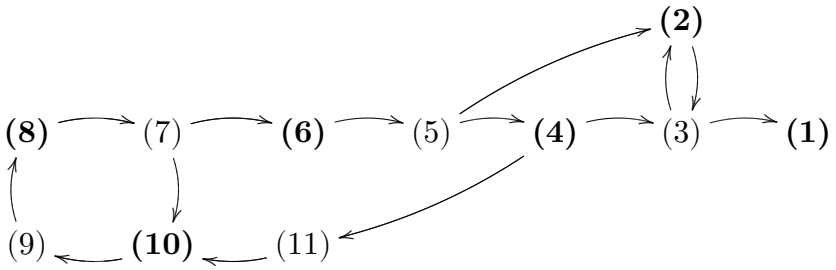
K: (10) Но если пандемия продолжится из-за недостаточных темпов вакцинации, то у этого будут не только естественные причины! Это произойдёт в том числе и из-за дефицита солидарности и неготовности людей принять на себя незначительные риски ради общего блага.

D: (11) Я уже говорил, что принимать такие риски, на мой взгляд, нерационально.

Абстрактная структура аргументации AF для данного множества аргументов, представленная в виде ориентированного графа, выглядит следующим образом:



На основании данного графа можно заметить, что для абстрактной структуры AF характерно наличие двух различных предпочтительных расширений: $Pref_1 = \{(1), (2), (4), (6), (8), (10)\}$ и $Pref_2 = \{(3), (5), (7), (9), (11)\}$. Чтобы убедиться в этом, рассмотрим для примера множество $Pref_1$:



Во-первых, отметим, что оно бесконфликтно, то есть в нем не найдётся таких двух аргументов, чтобы один из них атаковал другой. Во-вторых, оно состоит исключительно из приемлемых аргументов, то есть для любого аргумента X, атакующего любой аргумент из $Pref_1$, в множестве $Pref_1$ найдётся аргумент, атакующий данный X. Например, для аргумента (3), атакующего аргумент (1), в $Pref_1$ найдётся аргумент (2), атакующий (3). Для защиты от атаки на аргумент (4) со стороны аргумента (5) в $Pref_1$ найдется аргумент (6). И так далее. И в-третьих,

$Pref_1$ является максимальным расширением, поскольку к нему нельзя добавить ни одного элемента из AR без того, чтобы оно перестало быть бесконфликтным.

Аналогичным образом ситуация обстоит и для множества $Pref_2$.

Заметим, что $Pref_1$ состоит исключительно из аргументов, приводимых агентом K, тогда как $Pref_2$ содержит исключительно аргументы агента D. В этом нет ничего удивительного. Это лишь косвенно свидетельствует о том, что оба агента в достаточной мере коммуникативно рациональны. При этом AF-модель оставляет нас как наблюдателей данного спора в ситуации неопределённости относительно его исхода. В рамках AF-подхода мы не можем отдать предпочтение в данном случае ни одной из сторон, но можем лишь констатировать, что позиция каждой из сторон является структурно приемлемой в данном споре.

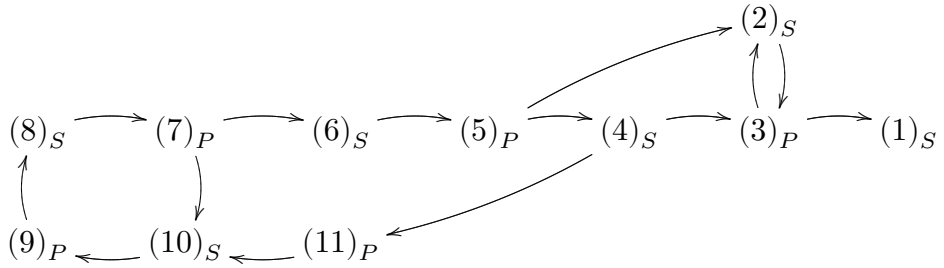
Существенно иначе ситуация обстоит в рамках VAF-модели, где мы, будучи наблюдателями данного спора, имеем возможность зачислить себя в состав той или иной аудитории, характеризующейся набором ценностных приоритетов (Perelman, Olbrechts-Tyteca 1969), которые позволят нам однозначно определиться с тем, позиция какой из сторон является для нас убедительной, а какой — нет.

Для демонстрации таких возможностей VAF-модели добавим к имеющейся структуре, помимо множества аргументов, предъявленных в споре, и отношений атаки между этими аргументами, также множество ценностей $V = \{V_S, V_P\}$, где V_S — это ценность общественной (social) безопасности и благополучия; V_P — ценность личной (personal) безопасности и благополучия. Следующим элементом, который нам понадобится для построения VAF-модели, является функция val, которая каждому элементу из множества AR ставит в соответствие элемент из множества V, характеризуя тем самым все аргументы на основании того, чьим интересам они соответствуют в наибольшей степени.

Ясно, что вопросы эпистемологии ценностей и методы верификации и фальсификации утверждений о соответствии тех или иных аргументов тем или иным ценностям связаны с целым рядом сложных философских проблем, которые мы в данном случае вынужденно обойдём вниманием, задав функцию val простым перечислением упорядоченных пар. Пусть наша функция val для рассматриваемого примера выглядит следующим образом:

$$val(1) = V_S; val(2) = V_S; val(3) = V_P; val(4) = V_S; val(5) = V_P; val(6) = V_S;$$
$$val(7) = V_P; val(8) = V_S; val(9) = V_P; val(10) = V_S; val(11) = V_P.$$

Внесём эту информацию в наш граф в качестве индексов для каждого аргумента:



На основании этого графа видно, что агент D в данном споре приводит только те аргументы, которые соответствуют интересам и безопасности отдельной личности, тогда как агент K приводит те аргументы, которые в наибольшей степени соответствуют поддержанию общественной безопасности и благополучия.

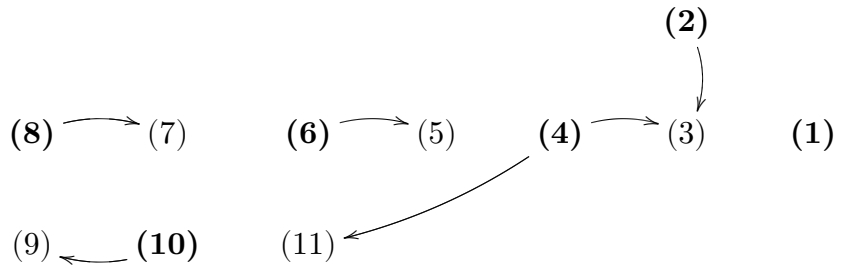
Поскольку множество ценностей в нашей модели состоит из двух элементов, множество потенциальных аудиторий для данного спора состоит из двух элементов — по числу возможных перестановок в списке приоритетных ценностей. Таким образом, множество P включает в себя:

$$a1, \text{ для которой } Vpref_{a1} = \{\langle V_S, V_P\rangle\};$$
$$a2, \text{ для которой } Vpref_{a2} = \{\langle V_P, V_S\rangle\}.$$

Допустим теперь, что агент D, будучи убеждённым индивидуалистом, сам принадлежит к аудитории $a2$, то есть ставит свои интересы выше общественных. Допустим также, что агент K, будучи альтруистом, сам принадлежит к аудитории $a1$, то есть ставит интересы общества выше своих личных интересов.

В соответствии с идеями VAF-подхода теперь абстрактная структура аргументации может быть трансформирована в набор из двух различных основанных на ценностях аргументационных структур, чувствительных к ценностным приоритетам каждой конкретной аудитории.

Ценностно обоснованная аргументационная структура $AVAF_{a1}$ для аудитории $a1$ в таком случае будет выглядеть в виде графа следующим образом:

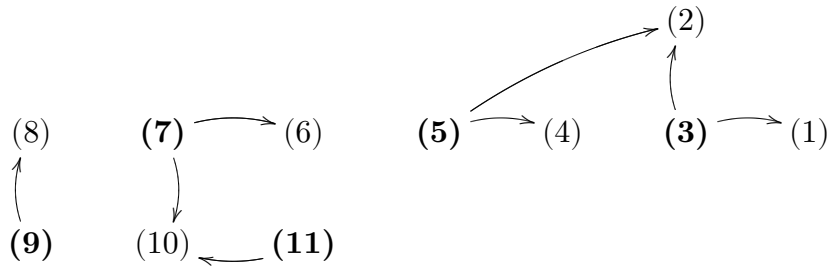


Эта структура получена из исходного графа простым удалением всех атак со стороны аргументов, поддерживающих менее приоритетные для данной аудитории ценности (ценности личной безопасности и благополучия), направленных на

аргументы, поддерживающие более приоритетные для данной аудитории ценности (ценности общественной безопасности и благополучия). При этом стрелками на данном графе обозначены теперь не отношения атаки между аргументами, а отношения *успешной атаки*, то есть отношения *отмены*.

Соответственно, предпочтительным расширением для аудитории $a1$ будет множество $Pref_{a1} = \{(1), (2), (4), (6), (8), (10)\}$.

В свою очередь, основанная на ценностях аргументационная структура $AVAF_{a2}$ для аудитории $a2$ для того же спора выглядит следующим образом:



Так же, как и в предыдущем случае, этот граф получен простым удалением из исходного графа всех атак со стороны аргументов, поддерживающих менее приоритетную для данной аудитории ценность общественной безопасности, направленных на аргументы, поддерживающие более приоритетную для данной аудитории ценность личной безопасности. Предпочтительным расширением для аудитории $a2$ будет множество $Pref_{a2} = \{(3), (5), (7), (9), (11)\}$.

Нетрудно заметить, что множество $Pref_{a1}$ полностью совпадает по своему составу с $Pref_1$ в рамках AF-модели, тогда как $Pref_{a2}$ совпадает с $Pref_2$. В чем же тогда состоит существенное различие между AF-подходом и VAF-подходом к моделированию структуры аргументации для практических споров?

Наиболее важной особенностью VAF-подхода является наличие теоремы Бенч-Капона, которая гласит, что если чувствительная к аудитории ценностно обоснованная аргументационная структура $AVAF_p$ не имеет циклов, состоящих из аргументов, поддерживающих одну и ту же ценность v, то такая структура имеет единственное непустое предпочтительное расширение для данной аудитории p.

Это означает, что для VAF-подхода класс споров, для которых мы как наблюдатели не в состоянии определить победителя, значительно меньше по объему, чем для AF-подхода. Так, при оценке результатов нашего иллюстративного спора каждый воспринимающий его агент, действуя в рамках VAF-подхода, имеет возможность на первом шаге причислить себя к одной из двух возможных аудиторий, после чего на втором шаге для него станет возможным выявить *единственного* победителя в данном споре (с точки зрения соответствующей аудитории; Bench-Capon 2002: 236).

Таким образом, выразительные возможности VAF-подхода значительно превосходят возможности AF-подхода при моделировании практических споров по

поводу действий, в рамках которых ценностные приоритеты заинтересованных сторон зачастую имеют определяющее воздействие на исход спора.

Литература

Bench-Capon 2002 — *Bench-Capon T. J. M.* Agreeing to Differ: Modelling Persuasive Dialogue Between Parties With Different Values // Informal Logic. 2002. Vol. 22, no. 3. P. 231–245.

Bench-Capon 2003 — *Bench-Capon T. J. M.* Persuasion in Practical Argument Using Value-based Argumentation Frameworks // Journal of Logic and Computation. 2003. Vol. 13, no. 3. P. 429–448.

Bench-Capon, Dunne 2002 — *Bench-Capon T. J. M., Dunne P. E.* Value Based Argumentation Frameworks. Research Report ULCS-02-001, Department of Computer Science, University of Liverpool, 2002. URL: `https://intranet.csc.liv.ac.uk/research/techreports/tr2002/ulcs-02-001.pdf` (accessed: 13.11.2022).

Dung 1995 — *Dung P. H.* On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games // Artificial Intelligence. 1995. Vol. 77, iss. 2. P. 321–357.

Perelman, Olbrechts-Tyteca 1969 — *Perelman C., Olbrechts-Tyteca L.* The New Rhetoric: A Treatise on Argumentation. Notre Dame: University of Notre Dame Press, 1969.