Метод решения обратных задач Смаллиана

А. С. Полушин

Аннотация. В данной статье рассматриваются обратные задачи Смаллиана о рыцарях и лжецах и предлагается метод их формализации и решения. DOI: 10.52119/LPHS.2024.79.55.005.

Ключевые слова: Р. Смаллиан, рыцари и лжецы, логические задачи, обратные задачи, логика высказываний, формализация, нормальные формы.

A Method for Solving Reversed Smullyan's Puzzles

A. S. Polushin

Abstract. This article considers reversed Smullyan's logical puzzles of Knights and Knaves; a method of their formalization and solution is proposed.

Keywords: Smullyan, Knights and Knaves, logical puzzles, reversed puzzles, propositional logic, formalization problem, normal forms.

Задачи Раймонда Смаллиана [1–3] не только увлекательны сами по себе, но и чрезвычайно интересны с точки зрения логики. Дело в том, что эти задачи формулируются в рамках содержательно очень простых и наглядных моделей (которые могут являться прообразом моделей куда более сложных), но при этом они уже предъявляют определенные требования к аппарату формальной логики и ставят вопрос о существовании и эффективности методов их решения — методов, которые могут быть полезны и в других моделях.

Как нам представляется, в контексте задач Смаллиана есть три основных проблемы: формализации, типологии и методов решения задач. Не касаясь вопросов формализации, мы сейчас разберем на примерах два типа задач — *прямые и обратные* — и методы их решения.

В качестве примера *прямой задачи* рассмотрим задачу № 33 из книги Смаллиана «Как же называется эта книга?» [2]:

> На некоем острове обитают только рыцари, которые всегда говорят правду, и лжецы, которые всегда лгут. Нам встретились два жителя этого острова (обозначим их А и В), и А заявил: «Я лжец, а В — рыцарь». Спрашивается: кем (рыцарем или лжецом) является А, и кем (рыцарем или лжецом) является В?

Способ формализации и метод решения подобного рода задач предложил В. П. Мухачев [4]. Для формализации данной задачи нам достаточно двух высказываний:

A : «А является рыцарем»

B : «В является рыцарем»

Все условие задачи при этом укладывается в формулу $A \leftrightarrow (\neg A \wedge B)$. Как показал В. П. Мухачев, для ответа на вопрос задачи достаточно привести эту формулу к виду СДНФ:

1. $A \leftrightarrow (\neg A \wedge B)$
2. $(A \to (\neg A \wedge B)) \wedge ((\neg A \wedge B) \to A)$
3. $(\neg A \vee (\neg A \wedge B)) \wedge (\neg(\neg A \wedge B) \vee A)$
4. $(\neg A \vee (\neg A \wedge B)) \wedge ((A \vee \neg B) \vee A)$
5. $(\neg A \vee (\neg A \wedge B)) \wedge (A \vee \neg B)$

6. $(\neg A \wedge A) \vee (\neg A \wedge \neg B) \vee (\neg A \wedge B \wedge A) \vee (\neg A \wedge B \wedge \neg B)$
7. $\neg A \wedge \neg B$, что и является ответом задачи (A и B являются лжецами).

По сути, мы доказали формулу $(A \leftrightarrow (\neg A \wedge B)) \leftrightarrow (\neg A \wedge \neg B)$, то есть эквивалентность условия и ответа задачи.

В качестве примера *обратной задачи* используем для наглядности ту же задачу № 33, но изменим соответствующим образом ее условие и вопрос:

> На некоем острове обитают только рыцари, которые всегда говорят правду, и лжецы, которые всегда лгут. Нам встретились два жителя этого острова (обозначим их A и B), и A сделал заявление, из которого можно заключить, что A и B являются лжецами. Спрашивается: какое именно заявление сделал A?

Для формализации этой задачи нам достаточно трех высказываний:

A : «A является рыцарем»

B : «B является рыцарем»

F : неизвестное высказывание, сделанное персонажем A

Условие задачи, таким образом, принимает вид формулы $(A \leftrightarrow F) \leftrightarrow (\neg A \wedge \neg B)$, *содержащей неизвестное высказывание, и в этом мы видим главное отличие обратных задач от прямых*. Для решения задачи нужно привести выражение $(A \leftrightarrow F) \leftrightarrow (\neg A \wedge \neg B)$ к форме эквивалентности, левой подформулой которой является в точности само высказывание F, а правая подформула F не содержит.

В случае данной задачи мы можем сразу воспользоваться равносильностью $((P \leftrightarrow Q) \leftrightarrow R) \leftrightarrow (P \leftrightarrow (Q \leftrightarrow R))$ (а затем привести правую подформулу к стандартному виду СДНФ, хотя это и необязательно):

1. $(A \leftrightarrow F) \leftrightarrow (\neg A \wedge \neg B)$
2. $F \leftrightarrow (A \leftrightarrow (\neg A \wedge \neg B))$
3. $F \leftrightarrow ((A \to (\neg A \wedge \neg B)) \wedge ((\neg A \wedge \neg B) \to A))$
4. $F \leftrightarrow ((\neg A \vee (\neg A \wedge \neg B)) \wedge (\neg(\neg A \wedge \neg B) \vee A))$
5. $F \leftrightarrow ((\neg A \vee (\neg A \wedge \neg B)) \wedge ((A \vee B) \vee A))$
6. $F \leftrightarrow ((\neg A \vee (\neg A \wedge \neg B)) \wedge (A \vee B))$
7. $F \leftrightarrow ((\neg A \wedge A) \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B \wedge A) \vee (\neg A \wedge \neg B \wedge B))$
8. $F \leftrightarrow (\neg A \wedge B)$

Однако в общем случае мы не можем рассчитывать на то, что наша формула имеет вид такой двойной эквивалентности, а F входит в нее только один раз и именно таким образом. Поэтому мы предлагаем другой, общий метод для решения обратных задач. Сначала приведем нашу формулу к виду СДНФ:

1. $(A \leftrightarrow F) \leftrightarrow (\neg A \wedge \neg B)$
2. $((A \leftrightarrow F) \to (\neg A \wedge \neg B)) \wedge ((\neg A \wedge \neg B) \to (A \leftrightarrow F))$
3. $(((A \to F) \wedge (F \to A)) \to (\neg A \wedge \neg B)) \wedge ((\neg A \wedge \neg B) \to ((A \to F) \wedge (F \to A)))$
4. $(((\neg A \vee F) \wedge (\neg F \vee A)) \to (\neg A \wedge \neg B)) \wedge ((\neg A \wedge \neg B) \to ((\neg A \vee F) \wedge (\neg F \vee A)))$
5. $(\neg((\neg A \vee F) \wedge (\neg F \vee A)) \vee (\neg A \wedge \neg B)) \wedge (\neg(\neg A \wedge \neg B) \vee ((\neg A \vee F) \wedge (\neg F \vee A)))$
6. $((\neg(\neg A \vee F) \vee \neg(\neg F \vee A)) \vee (\neg A \wedge \neg B)) \wedge (A \vee B) \vee ((\neg A \vee F) \wedge (\neg F \vee A)))$
7. $(((A \wedge \neg F) \vee (F \wedge \neg A)) \vee (\neg A \wedge \neg B)) \wedge (A \vee B) \vee (\neg A \wedge \neg F) \vee (F \wedge A))$

8. $(A \wedge \neg F \wedge A) \vee (A \wedge \neg F \wedge B) \vee (A \wedge \neg F \wedge \neg A \wedge \neg F) \vee (A \wedge \neg F \wedge A \wedge F) \vee (F \wedge \neg A \wedge A) \vee (F \wedge \neg A \wedge B) \vee (F \wedge \neg A \wedge \neg A \wedge \neg F) \vee (F \wedge \neg A \wedge F \wedge A) \vee (\neg A \wedge \neg B \wedge A) \vee (\neg A \wedge \neg B \wedge B) \vee (\neg A \wedge \neg B \wedge \neg A \wedge \neg F) \vee (\neg A \wedge \neg B \wedge F \wedge A)$
9. $(A \wedge \neg F) \vee (A \wedge \neg F \wedge B) \vee (F \wedge \neg A \wedge B) \vee (\neg A \wedge \neg B \wedge \neg F)$
10. $(A \wedge \neg F \wedge \neg B) \vee (A \wedge \neg F \wedge B) \vee (F \wedge \neg A \wedge B) \vee (\neg A \wedge \neg B \wedge \neg F)$
11. $(\neg F \wedge A \wedge \neg B) \vee (\neg F \wedge A \wedge B) \vee (F \wedge \neg A \wedge B) \vee (\neg F \wedge \neg A \wedge \neg B)$

Теперь, пользуясь равносильностью $((P \wedge Q) \vee (P \wedge R)) \leftrightarrow (P \wedge (Q \vee R))$, сводим все дизъюнкты к двум, один из которых содержит F, другой — $\neg F$:

12. $(\neg F \wedge ((A \wedge \neg B) \vee (A \wedge B) \vee (\neg A \wedge \neg B))) \vee (F \wedge (\neg A \wedge B))$

По закону де Моргана приписываем отрицание конъюнкту, стоящему в паре с $\neg F$:

13. $(\neg F \wedge \neg(\neg(A \wedge \neg B) \wedge \neg(A \wedge B) \wedge \neg(\neg A \wedge \neg B))) \vee (F \wedge (\neg A \wedge B))$

Наша формула теперь имеет вид $(\neg F \wedge \neg P) \vee (F \wedge Q)$, то есть близка к форме эквивалентности. Если P эквивалентно Q, то, исходя из равносильности $((\neg F \wedge \neg P) \vee (F \wedge P)) \leftrightarrow (F \leftrightarrow P)$, имеет место $F \leftrightarrow P$. Проверить эквивалентность P и Q можно, например, приведя их к СДНФ (для эквивалентности необходимо и достаточно, чтобы СДНФ содержали одинаковые наборы дизъюнктов; в противном случае задача не имеет решения). Соответствующие преобразования мы произведем прямо в теле формулы:

14. $(\neg F \wedge \neg((\neg A \vee B) \wedge (\neg A \vee \neg B) \wedge (A \vee B))) \vee (F \wedge (\neg A \wedge B))$
15. $(\neg F \wedge \neg((\neg A \vee B) \wedge ((\neg A \wedge B) \vee (\neg B \wedge A)))) \vee (F \wedge (\neg A \wedge B))$
16. $(\neg F \wedge \neg((\neg A \wedge \neg A \wedge B) \vee (\neg A \wedge \neg B \wedge A) \vee (B \wedge \neg A \wedge B) \vee (B \wedge \neg B \wedge A))) \vee (F \wedge (\neg A \wedge B))$
17. $(\neg F \wedge \neg(\neg A \wedge B)) \vee (F \wedge (\neg A \wedge B))$

СДНФ для P и Q совпадают и имеют вид $\neg A \wedge B$. Следовательно, мы получили форму эквивалентности

18. $F \leftrightarrow (\neg A \wedge B)$, что и является ответом задачи (то есть А сделал заявление «Я лжец, а В — рыцарь» или любое другое, эквивалентное ему).

По сути, мы доказали формулу $((A \leftrightarrow F) \leftrightarrow (\neg A \wedge \neg B)) \leftrightarrow (F \leftrightarrow (\neg A \wedge B))$, то есть эквивалентность условия и ответа задачи.

Заметим, что рассмотренные нами методы решения прямых и обратных задач можно заменить нестандартным построением таблиц истинности, а именно: исходя из эквивалентности условия и ответа *прямой задачи* $(A \leftrightarrow P(A, B)) \leftrightarrow Q(A, B)$ мы находим истинностные значения функции Q через данные нам истинностные значения функции P, а в случае обратной задачи мы находим значения P через данные нам значения Q.

Литература

1. Смаллиан Р. *Алиса в стране смекалки.* М.: Мир, 1987.
2. Смаллиан Р. *Как же называется эта книга?* М.: Мир, 1981.
3. Смаллиан Р. *Принцесса или тигр?* М.: Мир, 1985.
4. Мухачев В. П. Формализация задач Смаллиана. *Первые Смирновские чтения по логике. Материалы международной научной конференции. 18–20 марта 1997 года.* М.: ИФ РАН, 1997.