Крипке+ДеГроот: эпистемико-доксатическая модель социального влияния

В. В. Долгоруков[1]

Предлагается эпистемико-доксатическая модель с социальным влиянием, позволяющая расширить классическую модель ДеГроота за счет более богатого описания мнения агентов.

Ключевые слова: модель ДеГроота, социальное влияние, эпистемическая логика

1. Введение

Американским статистиком Моррисом ДеГроотом [1] была предложена модель, описывающая динамику изменения мнений агентов, подверженных влиянию друг друга.

Модель ДеГроота состоит из двух компонентов: 1) описание мнения агентов и 2) описание влияния агентов друг на друга. Мнение агента в момент времени t представляет собой действительное число: $x_i(t) \in [0, 1]$. Влияние агентов друг на друга описывается стохастической матрицей T, в которой $Tji \in [0, 1]$ интуитивно соответствует пропорции, в которой i-агент учитывает мнение j-го агента.

Обновление мнений описывается как умножение матрицы T на вектор мнений $x(t) = \langle x_1(t) ... x_n(t)\rangle$ на предыдущем шаге: $x(t+1) = T \cdot x(t)$.

Мы бы хотели предложить модель, которая бы сохраняла описание влияния агентов друг на друга в виде стохастической матрицы, однако более подробным образом описывала бы мнения агентов. В качестве модели-носителя мнения агентов мы будем использоваться эпистемико-доксатическую модель Крипке.

2. Эпистемико-доксатическая модель Крипке

Будем называть эпистемико-доксатической моделью Крипке (см. [2]) следующую структуру $M = (W, (\sim_i)_{i \in Ag}, (\preceq_i)_{i \in Ag}, V)$, где W — непустое

[1] *Долгоруков Виталий Владимирович* — зам.заведующего, Международная лаборатория логики, лингвистики и формальной философии НИУ ВШЭ, e-mail: vdolgorukov@hse.ru

Dolgorukov Vitaliy Vladimirovich — deputy head, HSE University, International Laboratory for Logic, Linguistics and Formal Philosophy.

множество возможных миров; $\sim_i$ — отношение эквивалентности на множестве W (интуитивно понимается как отношение неразличения миров агентом $i \in Ag$; $\preceq_i$ — рефлексивное и транзитивное отношение на множестве (интуитивно понимается как субъективное восприятие i-м агентом одного мира как не менее правдоподобного, чем другой); V — функция оценки, которая каждой пропозициональной переменной сопоставляет множество возможных миров. Предполагается, что отношения $\preceq_i$ и $\sim_i$ связаны следующим образом: $\forall w' \forall w''((w' \preceq_i w'' \vee w'' \preceq_i w') \equiv w' \sim_i w'')$.

Введем следующие обозначения: $[\varphi]_M := \{w \in W \mid M\}, w \models \varphi\}$; $[w]_i := \{w' \in W \mid w \sim_i w'\}$; $max_{\preceq_i}(X) := \{w \in X \mid \forall w' \in X : w' \preceq_i w\}$, где $X \subseteq W$.

Язык эпистемико-доксатической модели Крипке определяется следующей грамматикой $\varphi, \psi ::= p \mid \ \neg\varphi \mid (\varphi \wedge \psi) \mid K_i\varphi \mid B_i\varphi$, где p - пропозициональная переменная, $i \in Ag$. «$K_i\varphi$» читается как «агент i знает, что φ», а «$B_i\varphi$» читается как «агент i верит, что φ». Опишем условия истинности формулы в отмеченной модели (M, w).

- $M, w \models p$ е.т.е. $w \in V(p)$
- $M, w \models \neg\varphi$ е.т.е. $M, w \not\models \varphi$
- $M, w \models \varphi \wedge \psi$ е.т.е. $M, w \models \varphi$ и $M, w \models \psi$
- $M, w \models K_i\varphi$ е.т.е. $\forall w'(w \sim_i w' \Rightarrow M, w' \models \varphi)$
- $M, w \models B_i\varphi$ е.т.е. $\forall w' \in max_{\preceq_i}([w]_i) : M, w' \models \varphi$

3. Крипке+ДеГроот: эпистемико-доксатическая модель с социальным влиянием

Модель $M = (W, (\sim_i)_{i \in Ag}, (\preceq_i)_{i \in Ag}, V, Infl, \theta)$ будем называть эпистемико-доксатической моделью с социальными влиянием. Данная модель добавляет к модели Крипке функцию социального влияния $Infl : Ag \times Ag \mapsto [0, 1]$ и пороговое значение $\theta \in [0, 1]$. Функция социального влияния соответствует стохастический матрице в модели ДеГроота, то есть, предполагается, что $\sum_{j \in Ag} Infl(j, i) = 1$. Величина θ определяет порог, при котором агент готов изменить свою точку зрения.

Введем оператор «$[I\varphi]\psi$», который читается следующим образом: «после того как агенты обменяются мнениями по поводу φ имеет место ψ».

Таким образом, язык эпистемико-доксатической модели с социальными влиянием определяется следующей грамматикой:

$$\varphi, \psi ::= p \mid \ \neg\varphi \mid (\varphi \wedge \psi) \mid K_i\varphi \mid B_i\varphi \mid [I\varphi]\psi$$

Для того, чтобы описать семантику оператора $[I\varphi]$ введем несколько обозначений и опишем механизм обновления модели.

Пусть $Ag_a^{M,w}(\varphi) := \{i \in Ag \mid M, w \models B_a B_i \varphi\}$, то есть, множество агентов, которые верят в φ, с точки зрения агента a, в отмеченной модели (M, w). В обновленной модели $M^{I\varphi}$ меняется только отношение $\preceq_i$ по следующему принципу:

$$x \preceq_i^{I\varphi} y = \begin{cases} x \preceq_i^{\uparrow\varphi} y, \text{ если} \sum_{j \in Ag_a^{M,w}(\varphi)} Infl(j,i) \geq \theta \\ x \preceq_i y, \text{ иначе} \end{cases}$$

Здесь $\preceq_i^{\uparrow\varphi}$ обозначает консервативное обновление отношения $\preceq_i$, то есть, минимальное отношение, которое удовлетворяет следующим условиям (см. [2]):

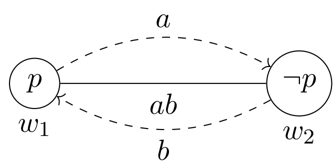
- $\forall x \forall y((x \in best_i(\varphi, w) \wedge y \in [w]_i) \to y \prec_i^{\uparrow\varphi} x)$
- $\forall x \forall y((x, y \in [w]_i - best_i(\varphi, w) \wedge x \preceq_i y) \to x \preceq_i^{\uparrow\varphi} y)$

Здесь $best_i(\varphi, w) := max_{\preceq_i}([w]_i \cap [\varphi]_M)$.

Теперь мы можем дать определение оператору социального обновления:

$$M, w \models [I\varphi]\psi \text{ е.т.е. } M^{I\varphi}, w \models \psi$$

Рассмотрим простейший пример.



Пусть в модели M_1 агенты a и b обладают противоположными мнениями по вопросу p, а функция социального влияния выглядит следующим образом:

$$\begin{pmatrix} 0,2 & 0,8 \\ 0 & 1 \end{pmatrix}$$

Тогда, в отмеченной модели (M_1, w_1) верно, что $B_a\neg p \wedge B_b p \wedge [Ip](B_b p \wedge B_a p)$.

Можно рассмотреть более сложные случаи: агенты могут обладать истинным мнением, но заблуждаться относительно мнений друг друга: $p \wedge B_a p \wedge B_b p \wedge B_a B_b \neg p \wedge B_b B_a \neg p \wedge [Ip](B_a \neg p \wedge B_b \neg p)$.

Отметим некоторые свойства данного оператора: (a) $[I\varphi](\psi \to \chi) \to ([I\varphi]\psi \to [I\varphi]\chi)$; (b) $\neg[I\varphi]\bot$; (c) $\models \psi \Rightarrow \models [I\varphi]\psi$; (d) $[I\varphi]p \equiv p$

Предполагается, что возможно сформулировать аксиомы редукции для логики с оператором «$[I\varphi]\psi$» за счет обогащения статического фрагмента конструкцией «$Infl(G, a)$» – суммарное влияние агентов из группы G на агента превышает пороговое значение.

4. Заключение

Предложенная эпистемико-доксатическая модель с социальным влиянием позволяет моделировать такие феномены информационные каскады, плюралистическое неведение, поляризация мнений и др.

Также возможно расширение модели за счет возможности динамического изменения функции влияния. В частности, агент a может перестать доверять мнению агента b, если a узнает, что b заблуждается относительно достоверной информации, которая есть у a.

Список литературы

[1] DeGroot Morris, "Reaching a Consensus", *Journal of the American Statistical Association*, **69**:345 (1974), 118–121.

[2] Pacuit Eric, "Dynamic Epistemic Logic II: Logics of Information Change", *Philosophy Compass*, **8**:9 (2013), 815–833.

Kripke+DeGroot: An Epistemic-Doxastic Model for Social Influence

Dolgorukov V.V.

We propose an extension of DeGroot's learning model with an epistemic-doxstic modal logic. It is argued that this framework can be applied for the analysis of informational cascades, pluralistic ignorance and other phenomenon.

Keywords: DeGroot, social influence, epistemic logic

References

[1] DeGroot Morris, "Reaching a Consensus", *Journal of the American Statistical Association*, **69**:345 (1974), 118–121.

[2] Pacuit Eric, "Dynamic Epistemic Logic II: Logics of Information Change", *Philosophy Compass*, **8**:9 (2013), 815–833.