Социальные вычисления и моральный номогенез

И. Ф. Михайлов (Институт философии РАН)

Современное состояние когнитивной науки характеризуется пониманием когнитивных процессов как распределённых статистических вычислений. Мозг как естественная нейросеть и общество как сфера обмена информацией могут быть поняты как естественные распределённые вычислительные устройства, оперирующие вероятностными предсказаниями и нацеленные на достижение максимальной энергетической эффективности. Мозг характеризуется наличием жёстких связей между нейронами, тогда как социальная сеть не предполагает устойчивых физических связей. В начале социальной эволюции общества пытаются эмулировать жёсткие связи через системы рабства, сословий и т. п. Однако с возрастанием скорости информационных обменов необходимость в этом постепенно ослабевает.

Распределённые вычисления осуществляются в системе взаимосвязанных процессоров, каждый из которых выполняет несложные алгоритмы с обратной связью, но в результате система в целом демонстрирует сложное адаптивное (интеллектуальное) поведение [*Wolfram*, 2002]. В сфере искусственного интеллекта известны как минимум две принципиальные распределённые модели: нейросеть и мультиагентная система [*Cederman*, 2005. *Михайлов*, 2018]. Нейросети (НС) состоят из множества взаимосвязанных нейронов, каждый из которых умеет рассчитывать величину входящего сигнала относительно заданного порога и взвешивать свои связи с другими нейронами. Мультиагентные системы (МАС) состоят из программных или реальных агентов, каждый из которых отслеживает действия других и реагирует на них в соответствии с заложенными в него правилами. НС очень приблизительно эмулируют деятельность мозга, МАС — поведение животных или человеческих сообществ. Обе архитектуры, по сути, представляют собой статистические машины, способные на ходу создавать, реализовывать и изменять вероятностные алгоритмы адаптации к изменчивому окружению.

Если представить себе МАС, агенты которой управляются обучаемыми нейросетями, то правила их взаимодействия будут вырабатываться эволюционно, стремясь к минимизации энергетических потерь системы в целом [*Macy; Willer*, 2002]. Процесс подбора таких правил остановится, когда система по его результатам достигнет энергетического оптимума.

Моральные нормы в этом контексте могут быть поняты как статистические алгоритмы распространения социальных взаимодействий, реализуемые на уровне отдельных процессоров (агентов, индивидов). Их функциональная роль — настройка системы в целом в направлении энергетического оптимума. Именно поэтому отдельным агентом они воспринимаются как внутренние регуляторы загадочного происхождения («нравственный закон внутри нас»). Конкретное содержание подразумеваемых ими запретов («заповедей») исторически случайно — значение имеет только их статистическое влияние на показатели системы в целом. Но запреты действуют как механизм сокращения числа возможных состояний системы, что равнозначно нарастанию степени её структурности [*Анохин*, 1978] и уменьшению энтропии [*Gao et al.*, 2013]. Обучаемость агентов, в свою очередь, формирует эволюционный алгоритм достижения системой энергетического оптимума.

Поскольку реальные биологические агенты (например, люди) сами представляют собой крайне сложные адаптивно-статистические компьютеры, состоящие из многих иерархически вложенных подсистем, естественная социальная система, в отличие от искусственных МАС, может обеспечить их подчинение правилам только с некоторой долей вероятности. Отсюда феномен «моральной свободы» — принципиальная возможность не подчиниться правилу. Однако, как показывают нейрофизиологические исследования [*Mehta; Josephs*, 2011], механизм общественного одобрения/осуждения действует напрямую через гормональную сферу мозга, и поэтому реакция социального «организма»

на поведение агента и в целом его социальная (не)успешность переживается им непосредственно на эмоциональном уровне. Это обстоятельство, как представляется, объясняет и субъективное ощущение «долга».
Вычислительный подход к этике преодолевает индуктивизм натуралистических и утилитаристских теорий, научно объясняя «трансцендентальность» нравственного закона, но в то же время позволяет обойтись без идеалистической метафизики деонтизма кантовского толка.

Литература:

Cederman, 2005 — Cederman L. E. Computational models of social forms: Advancing generative process theory . *American Journal of Sociology. Vol. 110. No. 4*. P. 864–893.
Gao et al., 2013 — Gao J. et al. Information Entropy As a Basic Building Block of Complexity Theory . *Entropy. Vol. 15. No. 12*. P. 3396–3418.
Macy; Willer, 2002 — Macy M. W., Willer R. From factors to actors: Computational sociology and agent-based modeling . *Annual Review of Sociology. Vol. 28*. P. 143–166.
Mehta; Josephs, 2011 — Mehta P. H., Josephs R. A. Social endocrinology: Hormones and social motivation. // Social motivation Frontiers of social psychology. New York, NY, US: Psychology Press. C. 171–189.
Wolfram, 2002 — Wolfram S. A New Kind of Science. : Wolfram Media Inc. 1192 p.
Анохин, 1978 — Анохин П. К. Принципиальные вопросы общей теории функциональных систем // Философские аспекты теории функциональной системы. Москва: Издательство “Наука.” С. 49–106.
Михайлов, 2018 — Михайлов И. Ф. Философские проблемы моделирования мультиагентных систем . *Философские науки. No. 12*. P. 56–74.