

## К РАЗНООБРАЗИЮ ПРЕДСТАВЛЕНИЯ МЕНТАЛЬНЫХ СВОЙСТВ АГЕНТОВ В ЭПИСТЕМИЧЕСКОЙ ЛОГИКЕ

Правильность и своевременность принятия решений являются важными моментами во многих сферах целенаправленной человеческой деятельности, в которых возникает проблема выбора между альтернативными вариантами. Окончательно о правильности сделанного выбора можно судить лишь по соответствию его результатов заданным целям. Поскольку часто в момент выбора, вследствие сложности ситуации или недостаточной информированности, мы не можем однозначно определить результат или оценить точное влияние его последствий на движение к заданной цели, критерием оценки правильности выбора становится его обоснованность. В каждой конкретной ситуации обоснованность выбора может иметь различные доводы. В одних случаях они зависят от степени информированности об альтернативах. В других — от представления о выполняемой деятельности в целом, от знания о взаимосвязях между внутренними процессами этой деятельности, от знания методов принятия решений, от учета внешних факторов. Во многих случаях на процесс выбора влияет ограниченность во времени или других ресурсах.

Благодаря развитию кибернетики появилась возможность в некоторых случаях автоматизировать процесс выработки и обоснования вариантов решения с целью их дальнейшего принятия человеком или автономно функционирующей робототехникой. В работах по искусственному интеллекту процесс выработки, обоснования и принятия решений связывается с деятельностью некоторого обобщенного и идеализированного объекта, именуемого *интеллектуальным агентом* или просто *агентом*. В каждом частном случае агент действует относительно некоторой *области знаний*, позволяющей обосновать его выбор. Во многих случаях агент действует не изолированно, а в сообществе других агентов, которые могут решать ту же или другие задачи. При этом области знаний агентов могут не совпадать. Каждый из агентов может строить гипотезы о причинах поведения других агентов: почему они сделали данный

выбор, был ли у них выбор и т. п. В результате анализа ситуации или поведения других агентов и при наличии определенных ментальных предпосылок агент может представить модель мира другого агента — его знания и сомнения.

В некоторых случаях для выработки правильного решения достаточно иметь адекватное и релевантное описание (представление или формализацию) окружающей действительности. Это характерно для задач, в которых агент принимает решения, относительно некоторой области знаний, изолированно от решений и действий других агентов. Однако если области знаний агентов пересекаются, то для большей обоснованности выбора следует учитывать еще и способ рассуждения других агентов, их намерения, цели, поведение и многое другое. В качестве вспомогательного инструмента можно использовать эпистемическую логику. Сейчас она является одной из наиболее интересных и перспективных формальных систем, позволяющих моделировать процесс выработки решений на основании знания о ситуации в целом и на основании представления о знаниях сообщества агентов об этой ситуации.

Способ представления ментальных свойств агентов в моделях базовой эпистемической логики был унаследован большинством ее расширений. Особенности семантики Хинтикки, лежащей в основе эпистемической логики, а также особенности способа ее интерпретации накладывают ограничения на то, какие ментальные свойства агентов могут быть выражены в классе эпистемических логик.

Во-первых, семейство эпистемических логик позволяет формализовать знания и сомнения агента относительно некоторой *онтологической реальности*, которую мы будем связывать с предметной реальностью, воспринимаемой некоторым социумом (множеством агентов), и противопоставлять множеству *эпистемических реальностей*, каждую из которых мы будем связывать со знаниями некоторого индивидуума (агента).

В зависимости от того, каким социумом и в какой ситуации воспринимается онтологическая реальность, при ее описании могут быть использованы различные *аспекты*, выражающие либо знание этого социума о наличии или отсутствии тех или иных черт онтологической реальности, либо неопределенность в знании о наличии или отсутствии этих черт. Таким образом, при описании онтологической реальности берутся не только те ее аспекты, наличие или отсутствие которых очевидно для некоторого социума, но и те, в наличии или отсутствии которых он сомневается, но на которых

все же сосредоточивается. Очевидно, что социум может обратить свое внимание лишь на небольшую часть аспектов многогранной онтологической реальности. Применительно к процессу принятия решений в рамках какой-либо задачи часто оказывается, что на некоторые из ее аспектов обращается внимание лишь в ходе ее решения. Причем для ряда задач предварительное внимание к таким аспектам означало бы другой ход их решения.

В традиционной интерпретации семантика Хинтикки требует, чтобы все аспекты, на которые обращается внимание на каком-либо шаге процесса познания, были описаны на всех предшествующих шагах. С точки зрения модели в целом это требование можно связать со свойством информационной полноты, с точки зрения ее динамики — со свойством монотонности. Такой подход позволяет формализовать только те задачи, в которых фаза выделения всех аспектов онтологической реальности полностью предшествует фазе мыслительных операций, т. е. он не позволяет описывать ряд открытых информационных систем.

Кроме того, традиционная семантика Хинтикки подразумевает, что все аспекты онтологической реальности, на которые обращает внимание один агент, находятся в области внимания всех других агентов, пусть и с различной степенью уверенности относительно их присутствия. Благодаря этому ограничению, мы не сможем формализовать множество задач, например со скрытой мотивацией.

Во-вторых, семейство эпистемических логик позволяет формализовать знания и сомнения агентов относительно эпистемических реальностей, т. е. относительно своих знаний и сомнений, а также относительно знаний и сомнений других агентов. Однако и здесь при формализации мы сталкиваемся с рядом ограничений.

Семантика базовой эпистемической логики предъявляет к моделям некоторые требования, зачастую отсутствующие в реальных когнитивных ситуациях. Это связано с тем, что помимо четко установленных знаний агентов и выявленных неопределенностей существуют невыявленные неопределенности. Эта область может поставлять существенную информацию, влияющую на структуру выявленных знаний. Поэтому отсутствие области невыявленной неопределенности в моделях эпистемической логики может значительно сузить область ее применения.

Одно из предъявляемых требований состоит в том, что в исходной модели требуется перечислить все возможные варианты когнитивных состояний. В ослабленном варианте это требование сво-

дится к тому, что частные знания и сомнения каждого из агентов должны быть объединены в одной общей модели. Таким образом, знания становятся в некоторой степени общедоступными. В определенном смысле это соответствует знаниям стороннего наблюдателя. В результате такие модели не позволяют описывать ситуации, в которых знания агентов о знаниях визави попадают в область невыявленных неопределенностей. Кроме того, они не дают возможности описывать также ситуации, в которых обновление эпистемических состояний связано с адаптацией вновь полученной, но несовместимой информации.

В качестве одного из наиболее простых способов изменения исходной модели предлагается различать индивидуальные представления каждого из агентов о состоянии знаний визави. Для этого будет использовано Декартово произведение исходных отношений достижимости. Дальнейшее изменение предлагается связать с определением индивидуальных особенностей структуры модели и даже функции означивания для каждого из агентов. Учесть эти особенности можно, перейдя к использованию неполных исходных моделей. В результате для сохранения возможности построения совместимых заключений потребуется задать операцию, позволяющую перестраивать структуру модели. Этого можно достичь посредством определения публичного анонсирования, порождающего новые миры и связи между ними. Похожая идея используется при задании игр с неполной информацией.

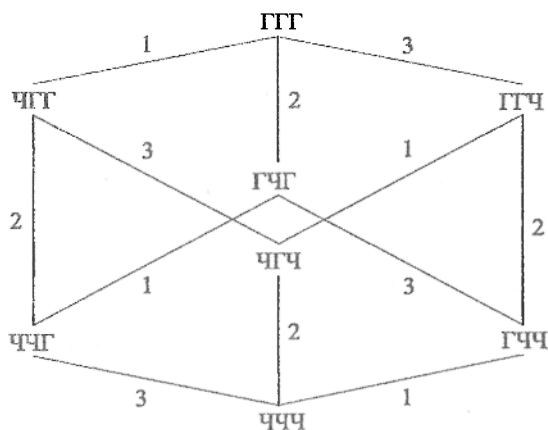
Естественно предположить, что, для того чтобы агент согласился скорректировать структуру своей модели, уровень доверия к полученной информации должен быть выше, чем к той, которой он располагает. Для этого определяются отправитель, получатель информации и иерархия уровней доверия к ней. Первое приводит к тому, что действия, как и знания, сопоставляются с агентами.

Для иллюстрации возможных когнитивных состояний обратимся к известной эпистемической задаче «Испачкавшиеся дети». Рассмотрим ее исходный вариант для анализа проблемы представления прямого и косвенного знания, а также соотношения частичного и исчерпывающего представления косвенного знания.

Трое детей играли в саду, и двое из них испачкали лица. Своих лиц дети не видят, но видят лица друг друга. Им сказали, что, по крайней мере, один из них испачкал свое лицо. После этого их спросили, может ли каждый из них сказать о себе,

является ли его (ее) лицо грязным. Все дети, при одновременном ответе, скажут: «Нет». Затем их можно еще раз спросить, могут ли они теперь сказать о себе, кто из них грязный. При этом дети, испачкавшие свои лица, имеют возможность распознать это, используя полученную ранее информацию. И лишь при третьем вопросе ребенок с чистым лицом получит возможность вывести заключение об этом.

Зададим начальный граф этой задачи (см. рис.), описывающий когнитивные состояния детей. Каждое из когнитивных состояний выражает неопределенность детей относительно того, какая из онтологических ситуаций (возможностей) является реальной. Для каждого из детей его когнитивное состояние включает не только выражение собственной неопределенности относительно реальной онтологической ситуации, но и представление о неопределенности других детей относительно реальной онтологической ситуации.



Начальный граф задачи «Испачкавшиеся дети»

Поскольку в соответствии с постановкой задачи дети сосредоточены на выяснении того, кто из них испачкал лицо, а кто нет, абстрагируемся от остальных факторов, описывающих реальную онтологическую ситуацию, и остановимся лишь на том, является ли лицо данного ребенка чистым либо оно испачкано. В узлах графа размещены упорядоченные тройки литер, представляющие все возможные комбинации онтологических ситуаций. Упорядочив детей,

можно связать каждую из литер с конкретным ребенком. Литерой «Г» обозначим то, что ребенок ипачкал лицо, а литерой «Ч» -- что оно чистое. В свою очередь, помеченные дуги графа также связаны с конкретными детьми (метка дуги соответствует порядковому номеру определенного ребенка). При этом помеченная дуга графа, связывающая пару онтологических ситуаций, выражает их эпистемическую неразличимость для данного ребенка. Согласно условию задачи, в начальной эпистемической ситуации каждый из детей не сможет различать онтологические ситуации, в которых описание двух других детей остается неизменным.

Согласно предложенному условию задачи, реальной онтологической ситуации может соответствовать только один из узлов ГГЧ, ЧГГ или ГЧГ. Допустим, таким узлом является ГГЧ. Рассмотрим вопрос: зачем нужно указывать дуги и узлы помимо тех, которые непосредственно связаны с узлом ГГЧ? Действительно, казалось бы, на начальном этапе опроса неопределенность первого ребенка связана лишь с двумя узлами. Ведь он видит, что второй ребенок грязный, а третий — чистый. Это будет означать, что он не смог бы различить лишь онтологические ситуации ГГЧ и ЧГЧ. Аналогично второй ребенок не смог бы различить ГГЧ и ГЧЧ, а третий — ГГЧ и ГГГ. В результате правомерно выдвинуть предположение о том, что этих четырех узлов достаточно для исчерпывающего описания когнитивной ситуации.

Однако данный анализ когнитивной ситуации содержит только непосредственное знание, не содержащее гипотез о знании визави. В результате граф понимается как некое хранилище, полностью доступное для анализа внешнему эксперту, в то время как детям информация о знаниях друг друга недоступна. Непосредственным знанием можно ограничиться, только оставив дуги и состояния, непосредственно примыкающие к действительному состоянию. Однако такая интерпретация не представляется интересной.

Более того, нетрудно заметить, что как добавление новых дуг и узлов потребует привлечения интерпретации, учитывающей косвенное знание, содержащее гипотезы о знаниях визави, так и, наоборот, переход к интерпретации, учитывающей косвенное знание, потребует добавления новых дуг и узлов на начальном графе задачи.

Действительно, косвенность позволяет неограниченную итерацию, поэтому ее допущение немедленно приведет к добавлению дополнительных дуг и состояний. Вместе с тем новые дуги и состоя-

пия не будут соответствовать неопределенностям детей относительно собственного представления о реальной онтологической ситуации. Однако появится возможность включить в описание представления детей относительно неопределенности своих визави.

Мы выделили две возможные интерпретации представленного графа. В первом случае, несмотря на то что на одном графе представлена эпистемическая информация, касающаяся всех детей, ни один из детей не имеет доступа к информации о визави. Во втором случае вся когнитивная ситуация полностью доступна всем детям.

Однако многие когнитивные ситуации не соответствуют ни тому ни другому случаю. В самом деле, вполне обычно, что нам могут быть неизвестны абсолютно все знания и сомнения наших визави, однако известно хоть что-то об этом. Другими словами, мы хотим исключить лишь часть косвенных прочтений, но не все. Более того, наши представления о сомнениях наших визави (об их эпистемическом состоянии) могут не совпадать с действительной картиной их сомнений. Значит, семантику эпистемической логики следует модифицировать так, чтобы наличие некоторой дуги или состояния на графе не означало автоматически их доступность всем агентам.

Внесем изменение в задачу «Испачкавшиеся дети». Допустим, что один из испачкавшихся детей видит свое отражение в зеркале, но об этом кроме него никому не известно. В результате окажется, что ребенок, наблюдающий свое отражение, будет иметь знание, отличное от представлений о его знании у других детей. Таким образом, мы не можем интерпретировать одну и ту же дугу графа и как выражение неопределенности самого ребенка, и как выражение представления других детей о его неопределенности. В качестве решения можно изменить определение отношения достижимости в модели эпистемической логики.

Помимо того что агенты могут иметь неверное представление о знаниях визави, в их знаниях может отсутствовать информация о некоторых когнитивных альтернативах. С точки зрения представления в виде графа это выразится в отсутствии соответствующих этим альтернативам состояний и дуг, с ними связанных. Причем в некоторых случаях различие между ситуациями, в которых агент не знает о возможности существования какой-либо из альтернатив, и ситуациями, в которых агент не может выделить эту альтернативу среди других, является существенным.

Так, если в модифицированную задачу «Испачкавшиеся дети» добавить знание о существовании зеркала в качестве возможной

альтернативы другим детям, то это может полностью изменить не только их представление о ситуации, но и ход их дальнейших рассуждений.

Допустим, дети не подозревают о возможности существования зеркала. Тогда на первый вопрос, может ли каждый из них сказать о себе, является ли его (ее) лицо грязным, ребенок, стоящий напротив зеркала, скажет, что его лицо грязное. В результате при повторении этого вопроса оставшиеся дети скажут, что они чистые, предположив, что первый ребенок получил информацию о своем лице, увидев два чистых лица напротив. В данном случае эти дети опирались бы на недостоверные причины умозаключения первого ребенка, и один из них в результате пришел бы к ложному заключению.

Напротив, если бы дети догадывались о возможности существования зеркала, они, возможно, не стали бы делать поспешного вывода, не убедившись в отсутствии подвоха. Как следствие, результат их рассуждений мог бы оказаться и иным.

Для того чтобы учитывать подобные ситуации, мы предлагаем, как и в случае с отношением достижимости, для каждого из агентов задать свою отдельную систему миров, непосредственно недоступную для других агентов. При этом, конечно, каждая из таких отдельных систем миров будет связана своей группой отношений достижимости и будет определяться индивидуальной функцией означивания. Доступ к структуре знаний друг друга агенты могут осуществлять путем публичного анонсирования.

В итоге мы получили, что в семействе эпистемических логик, основанном на традиционном понимании семантики Хинтикки, оказывается выраженным только узкий спектр ментальных свойств агентов. В настоящее время в моделях эпистемической логики никак не учитывается то, что в реальности как способности агентов к переработке информации могут различаться, так и их области знания могут иметь различные очертания. Более реалистичные модели представления агентов в эпистемической логике только начинают разрабатываться. Они появляются в единичных экземплярах и поэтому сами по себе пока не позволяют получить общей картины о возможностях представления ментальных свойств агентов в эпистемической логике.