

ВЫРАЖЕНИЕ МОДЕЛИРУЕМЫХ ЭФФЕКТОВ ЭМОЦИЙ У ИСКУССТВЕННЫХ АГЕНТОВ КАК ВИЗУАЛЬНЫЙ ЯЗЫК

А. В. Шиллер

Московский государственный университет им. М. В. Ломоносова, Россия
shiller.a@gmail.com

В течение последних десятилетий существенно возросло число исследований эмоций, появились новые теории и модели эмоциональной системы. Исследователям стало ясно, что феномен эмоций может быть понят, использован и смоделирован с помощью двух фундаментальных процессов: генерации эмоций и эффектов эмоций, а также связанных с ними задач моделирования. Эти задачи служат строительными блоками для аффективных моделей и включают для обоих процессов следующие составляющие: определение набора областей; определение соотношений между этими областями (от триггеров эмоций к генерации эмоций и от эмоций к их эффектам); вычисление интенсивности и магнитуды для расчета интенсивностей эмоций в процессе генерации и магнитуд эффектов в процессе возникновения этих эффектов; определение функций, которые соединяют и интегрируют сложные эмоции.

Интерес к проблеме эмоций возрастает вместе с развитием исследований искусственного интеллекта. Все чаще возникает потребность в человекоподобном поведении, воспроизвести которое очень сложно без попыток моделирования эмоционального аппарата. Особенно важным становится моделирование эмоций в контексте создания агентов, функциональность которых связана с коммуникацией с человеком. Для многих практических задач и проблем (например, распознавание эмоций, реализация эффектов или последствий эмоций) перспективным кажется развитие технологий машинного обучения. Однако решение отдельных задач (таких как распознавание эмоций по фото/тексту и т. д.) пока не привело к качественно прорыву в моделировании эмоциональных систем. Более того, исследователи все чаще отказываются от создания отдельной эмоциональной системы, апеллируя к тому, что эффекты эмоций реализуются в поведении агента автоматически, если они были включены в данные, использованные для обучения агента. Примером практической реализации задачи является разработанный компанией Microsoft чатбот Тэй в Твиттере, который быстро научился писать эмоциональные тексты, но при этом очевидно, что такое поведение не является следствием работы его собственной эмоциональной системы.

Но многие исследователи в области робототехники, ИИ, интерфейсов человек-компьютер и когнитивных наук все-таки создают вычислительные модели с учетом разработанных ранее теорий и природы эмоций. Целью появления таких моделей является создание более достоверных и аффективных искусственных персонажей (в том числе NPC – non-personal

character) и роботов и, кроме того, повышение качества взаимодействия человек-компьютер.

В статье проанализированы методологические трудности моделирования эффектов эмоций. Этот анализ представляет собой шаг в направлении формализации моделирования эмоций и предлагает основания для разработки более систематического, общего подхода к моделированию, а также отдельных подходов для создания моделей эффектов эмоций и эмоций в целом. В качестве результата анализа выделен ряд принципов, которые необходимо учитывать при моделировании эмоций. Данные принципы, положенные в основу моделируемой эмоциональной системы агента, могут помочь исследователям продвинуться на пути создания человекоподобного ИИ, использующего эмоциональную систему в качестве визуального языка при коммуникации.

Ключевые слова: моделирование эмоций, визуальный язык, эффекты эмоций, архитектура искусственного агента, ИИ – искусственный интеллект.

EXPRESSION OF THE SIMULATED EFFECTS OF EMOTIONS IN ARTIFICIAL AGENTS AS A VISUAL LANGUAGE

Alexandra Shiller

Lomonosov State University, Moscow, Russian Federation
shiller.a@gmail.com

The number of studies on emotions has increased significantly during the past decades; new theories and models of the emotional system have appeared. It became clear to researchers that the phenomenon of emotions can be understood, used and modeled using two fundamental processes: the generation of emotions and the effects of emotions, as well as the associated modeling tasks. These tasks serve as building blocks for affective models and, for both processes, include the following components: identification of the set of areas; determination of the relationship between these areas (from emotion triggers to emotion generation, and from emotions to their effects); calculation of the intensity and magnitude to calculate the intensity of emotions in the process of the generation and the magnitudes of the effects in their occurrence; determination of functions that connect and integrate complex emotions. Interest in the problem of emotions also increases with the development of a new research direction: artificial intelligence research. Increasingly, there is a requirement for demonstration and reproduction of human-like behavior, which is very difficult to achieve without trying to model the emotional apparatus. Especially important is the modeling of

emotions in the context of creating agents whose functionality is associated with communication with a human. For many practical tasks and problems (for example, recognition of emotions, realization of emotion effects or consequences), the development of machine learning technologies seems promising. However, the solution of particular tasks (such as the recognition of emotions by photo, text, etc.) has not yet led to a qualitative breakthrough in the modeling of emotional systems. Moreover, researchers are increasingly refusing to create a separate emotional system, appealing to the fact that the effects of emotions are realized in the behavior of the agent automatically if they were included in the datasets used to train the agent. An example of the practical implementation is a Microsoft chatter bot Tay released via Twitter. It quickly learned to write emotional texts, but it is clear that its behavior is not the result of the work of its own emotional system. But many researchers in the field of robotics, AI, man-computer interfaces, and cognitive sciences still create computational models based on the previously developed theories and the nature of emotions. The purpose of the emergence of such models is to create more reliable, humanlike and effective artificial characters (including NPCs, non-personal characters) and robots, and also to improve the quality of human-computer interaction. The article presents an analysis of the methodological difficulties of modeling the effects of emotions. This analysis represents a step towards formalizing the modeling of emotions and suggests a basis for developing a more systematic, general approach to modeling, as well as particular approaches to create models of the effects of emotions and generation of emotions. As a result of the analysis, a number of modeling principles are revealed, which must be taken into account. These principles, which form the basis of an agent's modeled emotional system, can help researchers move towards the creation of a humanlike AI, which uses the emotional system as a visual language in communication.

Keywords: visual language, effects of emotions, modeling of emotions, architecture of artificial agent, AI – artificial intelligence.

DOI 10.23951/2312-7899-2019-4-223-243

Вопреки прогрессу в исследованиях эмоций, наблюдающемуся в последние 20 лет, эмоции остаются неуловимым феноменом. В то время как были открыты физиологические процессы, лежащие в основе некоторых эмоций (например, роль амигдалы в обработке угрожающего стимула, роль орбитофронтальной коры в регуляции эмоций), многое о механизмах эмоциональной регуляции остается неизвестным. Мультимодальная природа эмоций, их сложность и наличие межмодальных взаимодействий отражают сложность

этого феномена и наталкивают на мысль, что само существование «фундаментальных процессов», лежащих в основе эмоций, находится в настоящее время под большим вопросом. Тем не менее не стоит оставлять попыток найти и создать такую фундаментальную теорию эмоций.

Для целей символического моделирования эмоций и моделирования эмоций в рамках агентной архитектуры наиболее перспективным способом исследования является разделение проблемного поля моделирования эмоций на две широкие категории процессов. На те, которые ответственны за *возникновение эмоций*, и на те, которые сопровождают когнитивные *эффекты активированных эмоций*, экспрессивное поведение (например, лицевые выражения, речь) и выбор действий.

Выражение эмоций с помощью разных экспрессивных каналов является тем процессом, который лежит «на поверхности», виден взаимодействующим с агентом пользователям или другим агентам, и поэтому может рассматриваться как особый визуальный язык.

Целью данной статьи является анализ методологических трудностей моделирования эффектов эмоций. В ней рассматриваются некоторые важные шаги для разработки модели эмоций у искусственных агентов: проблема поиска теоретических оснований моделирования эффектов эмоций, различия между внутренними и внешними эффектами эмоций, роль формы телесного (материального) воплощения агента для моделирования эффектов эмоций, взаимоотношения триггеров эмоций и эффектов эмоций, процесс моделирования динамики эффектов эмоций.

Рассмотрение этих шагов позволяет понять, что необходимо учитывать при моделировании эффектов эмоций, чтобы выражение эмоций у искусственных агентов не только отражало заложенные теоретические конструкции эмоций, но и стало полноценным способом коммуникации, позволяющим назвать искусственного агента человекоподобным ИИ или ИИ третьей волны.

Поиск общей теории моделирования эффектов эмоций

Прежде всего, необходимо отметить особую важность моделирования двух основных процессов (генерации эмоций и эффектов эмоций) с учетом имеющегося теоретического и эмпирического бэкграунда. Это требование означает, что аффективные модели

должны использовать авторитетные теоретические подходы к моделированию эмоций (например, дискретный, пространственный, компонентный) и там, где это возможно, использовать устоявшиеся теории в качестве основы моделирования определенных процессов (например, теории когнитивной оценки ОСС для генерации эмоций). ОСС – несколько вариантов теорий, предложенных Ортони и коллегами, теперь обобщенно упоминающиеся как ОСС теория [Ortony et al. 1988; Frijda 1986]. ОСС теория является наиболее часто используемой теорией когнитивной оценки в вычислительных моделях эмоций [Andre et al. 2000; Aylett 2004; Bates et al. 1992; De Rosis et al. 2003; El-Nasr et al. 2000; Elliot 1992; Gratch, Marsella 2004; Loyall 1997; Prendinger et al. 2004; W. S. R. Reilly 1996; Staller, Petta 1998].

Теоретические подходы к моделированию различаются по предлагаемым ими семантическим примитивам для представления эмоций (*базовые эмоции* в дискретных/категориальных подходах, *удовольствие-возбуждение* (*Pleasure-Arousal – PA*) или *удовольствие-возбуждение-доминантность* (*Pleasure-Arousal-Dominance – PAD*) в пространственном подходе и *оценочные переменные* в компонентном подходе). Теории также различаются по набору областей, в которых происходит моделирование и отражаются отношения между процессами генерации эмоций и возникновения эффектов эмоций. Принято выделять несколько таких областей (см. таблицу 1).

Таблица 1

Необходимые области моделирования

Набор областей в этой таблице представляет максимально возможный набор, требуемый для моделирования эмоций. Для конкретной модели и конкретной теории могут быть необходимы только некоторые из представленных областей.

Название области	Описание	Примеры элементов области
Объектная	Элементы внешнего мира (физические, социальные), представленные сигналами (перцептивный вход агента)	Другие агенты, события, физические объекты
Когнитивная	Внутренние ментальные понятия, необходимые для генерации эмоций или выражения их влияния на когниции	Когнитивные подсказки, ситуации, цели, ожидания, убеждения, нормы, конгруэнтность цели...

Продолжение таблицы 1

Название области	Описание	Примеры элементов области
Абстрактная	Зависимые от теории, например, векторы или пространства, оценочные переменные, ОСС оценочные критерии	Удовольствие, возбуждение, доминантность, релевантность цели, конгруэнтность цели, безусловность ...
Аффективная	Аффективные состояния (эмоции, настроения) и личностные черты	Радость, грусть, страх, гнев, гордость, зависть, ревность, экстраверсия
Физиологическая	Симулируемые физиологические особенности	Уровень энергии
Экспрессивные каналы	Внутренние каналы, с помощью которых могут быть выражены эмоции агента: лицевые выражения, позы, жесты, движения глаз и головы, движения тела, речь	Лицевые выражения (улыбка, хмурое выражение), речь (грустная, возбужденная), позы (спокойные, неловкие), движения (быстрые, медленные) – выраженные через примитивы, присущие каждому каналу, например, набор выражений лица)
Поведенческая	Поведенческий репертуар агента в его физическом и социальном окружении	Походка, бег, спокойное стояние на месте, схватывание объекта, жест – здороваться за руку применительно к другому агенту

Существуют данные, которые подтверждают роль прямого вклада отдельной характеристики события в формируемые эффекты эмоции для поведенческой или экспрессивной области, например, новизна вызывает приподнимание бровей и век, нарушения голоса или появление вздохов, выпрямление позы тела [Scherer 1992].

Для многих других характеристик события или контекста такие данные недоступны. Поэтому в некоторых случаях одинаковый набор абстрактных переменных служит одновременно для сопровождения генерации эмоций и эффектов эмоций. Теоретические подходы также различаются по числу эмоций, для которых они применимы, т. е. пространства, определяемые PA или PAD, меньше, чем пространство, определяемое с помощью оценочных переменных. Наконец, подходы различаются по тому, насколько существующие в них семантические примитивы подходят для выражения эмо-

ций по экспрессивным каналам и для комбинации сложных эмоций. Например, пространства в пространственном подходе хорошо подходят для описания продолжительных и протекающих постепенно изменений лицевых выражений, которые необходимы для реалистичного выражения эмоций на лице. С другой стороны, использование дискретного/категориального подхода для сопровождения лицевых выражений потребует дополнительных вычислений для обеспечения реалистичной аффективной динамики.

Описанные выше различия позволяют определить, какой теоретический подход подходит лучше всего для конкретного агента. Например, если у агента только несколько базовых эмоций и ему не нужно одновременно представлять (переживать) страх и злость (или любую другую пару эмоции, имеющую одинаковые показатели по характеристикам «возбуждение» и «валентность»), тогда пространственный подход всего с двумя пространствами (возбуждение и валентность/удовольствие) приемлем. Если требуются и страх, и злость, тогда пространственный подход может все еще быть приемлемым, но необходимо добавить третье измерение (доминантность), чтобы иметь возможность различить страх и злость, характеризующиеся негативной валентностью и высоким возбуждением. Если конкретный агент должен обладать большим числом эмоций, включая сложные и социальные эмоции, то более подходящим будет компонентный подход, поскольку он позволяет описать большой набор аффективных состояний. Дискретный/категориальный подход может быть использован как для описанных выше случаев, так и может служить эксплицитным дополнением для специфических эмоциональных процессов для каждой представленной модели эмоций.

Прежде чем переходить к описанию теорий возникновения эффектов эмоций, необходимо подробно остановиться на описании теорий генерации эмоций, поскольку оба процесса – генерация эмоций и эффекты эмоций – качественно взаимосвязаны, и для понимания методологических сложностей поиска теории-основания для процесса возникновения эффектов эмоций необходимо прояснить возможности существования и требований к такой теории для генерации эмоций.

Наиболее исследованной теорией *генерации эмоций* является когнитивная оценка, и это самая часто встречающаяся основа для создания моделей генерации эмоций в моделировании аффективной системы агента. Доминирующей теорией для описания когнитивной оценки является теория ОСС, которая предлагает

разработанную таксономию триггеров эмоций, а также независимые от области применения оценочные критерии, используемые для генерации различных эмоций. У теории ОСС есть преимущество, которое заключается в том, что ее применяли уже много раз, а существующие модели генерации эмоций могут служить конкретными успешными примерами для разработки аналогичных моделей для новых агентов. Недостатком теории ОСС является абстрактность ее оценочных критериев, которые не напрямую соотносятся с эффектами эмоций. Если теория ОСС выбирается в качестве основы для генерации эмоций агента, то потребуется выбрать другую теорию (или подход) в качестве основы для моделирования эффектов эмоций.

В качестве другой теории-основания можно привести компонентную теорию. Оценочные переменные компонентной теории имеют ряд преимуществ, например, могут служить для моделирования и генерации эмоций, и эффектов эмоций, поскольку множество значений оценочных переменных напрямую соотносятся с элементами внутриэкспрессивной области. Не так давно были сделаны попытки определить соотношения оценочных переменных и когнитивной области (то есть насколько хорошо определенные оценки коррелируют с процессом использования эвристических, аналитическими процессами, решением проблем).

Учитывая мультимодальную природу эффектов эмоций, в отличие от генерации эмоций, для поддержки *моделирования эффектов эмоций* не существует наиболее признанной теории, которая могла бы выступать аналогом теории ОСС для генерации эмоций. Следствием отсутствия всеобъемлющей теории для моделирования кроссмодальных эффектов эмоций является то, что для разных модальностей должны использоваться разные теоретические подходы. Кроме того, важно провести различие между внутренними эффектами эмоций – когнитивными эффектами, то есть влиянием эмоций на протекающие психические процессы, и внешними выражениями эмоций с помощью экспрессивных каналов и поведения.

Рассматривая влияние эмоций на когнитивную сферу (учитывается вся сфера когнитивных процессов, представленная у агента), основанные на параметрах теории представляют собой хорошую базу для моделирования разнообразных эффектов эмоций по отношению к когнитивной сфере. В таких теориях определен набор контрольных параметров для отдельных частей процесса внутри аффективной модели и архитектуры агента. Например, параметры могут контролировать скорость и объем внимания, оценку ситуа-

ции, управление целями, планирование и даже сама по себе генерацию эмоций. Когда эмоция генерируется у агента, она транслируется через набор специальных значений параметров, которые запускают изменения в архитектурных модулях. Результаты изменений когнитивных процессов в итоге отражаются в принимаемых агентом решениях и в выборе специфических действий.

Относительно эффектов эмоций, выражаемых через экспрессивные каналы, разнообразие процессов-медиаторов различных эффектов внутри отдельного канала (например, лицевые выражения, поза, жесты, речь) и отсутствие «общей теории» для эффектов эмоций приводят к тому, что эффекты эмоций лучше всего передаются с помощью набора прямых соотношений между экспрессивными каналами, доступными той телесной форме, в которую помещен агент.

По причине отсутствия «общей теории» эффектов эмоций наличие глубоко проработанных представлений об информационном процессе (что не обязательно означает создание процессуально-уровневой модели) является преимуществом при моделировании архитектуры агента. Эти представления способствуют координации сложных модальностей, внутри которых выражаются эмоции, и представляют собой более крепкую и гибкую основу для моделирования, нежели набор прямых соотношений эмоций и эффектов. Эта точка зрения в настоящее время поддерживается большинством создателей моделей и архитектур для правдоподобных материально воплощенных агентов [Aylett 2004; De Rosi et al. 2003; Prendinger et al. 2002].

Определение формы телесного (материального) воплощения агента

Форма воплощения агента зависит от набора факторов, включающих общие характеристики (каким будет агент – например, игровой персонаж, виртуальный тренер, робот, используемый для реабилитации, робот-компаньон для детей и т. д.), необходимые функции эмоций у агента (например, автономность или аффективный реализм), общий визуальный реализм контекста использования агента (например, приложение для мобильного устройства с маленьким экраном или полноразмерное изображение агента в виртуальной реальности/агент в форме робота); и от доступных для разработки формы воплощения агента ресурсов (например, уровень

детализации выражения лица, речи); а также от набора эмоций, который должен уметь выражать агент. В конечном итоге форма воплощения агента определит и его коммуникативные возможности – то, сможет ли он использовать моделируемые эффекты эмоций в качестве визуального языка.

Форма воплощения агента определит набор доступных экспрессивных каналов для выражения эмоций. Экспрессивными каналами, с помощью которых можно выразить эмоции, являются: лицевые выражения, взгляд, движения головы, жесты, позы тела, особенности движения персонажа в целом, речь (как вербальные, так и ее невербальные характеристики, а также просодические компоненты речи – манера или подача, темп и объем). Разработчик должен определить, какие из этих экспрессивных каналов необходимы для целей агента и какие реальны для воплощения, учитывая возможные технические и теоретические трудности (soft&hard). Разработчику также необходимо решить, какие экспрессивные каналы подходят лучше всего для выражения включенных в модель эмоций, и учесть возможные ограничения для воплощения, как, например, размер экрана. Обычно лицевые выражения являются наилучшим выбором для выражения эмоций. Однако в ситуациях, когда размер экрана мал и его разрешение низко, более приемлемыми могут быть другие каналы (например, жесты и речь).

Разработчик должен также принимать во внимание возможности генерации аффективных выражений для того канала, который планирует использовать. Например, для отображения аффективных выражений на лице наиболее часто используются элементы из FACS [Ekman, Friesen 1978] в качестве семантических примитивов, в то время как аффективно окрашенная синтетическая речь в настоящий момент не так хорошо развита.

Определение формы воплощения агентов подразумевает выделение индивидуальных каналов внутри *экспрессивной области* модели и устанавливает присутствие им семантические примитивы.

Другим компонентом воплощения агента является симулируемая физиология. Конечно, не для всех агентов требуется эксплицитная модель, содержащая физиологическую модальность. Предполагая овеществленную природу большинства агентов, заложенное представление и симуляция некоторых физиологических характеристик сделают персонажей более реалистичными, если симулируемые физиологические состояния смогут оказывать влияние на поведение агента. Это отдельная задача моделирования – определение вклада этих характеристик в генерацию эмоций (на-

пример, уставший или голодный агент с большей вероятностью будет вести себя раздраженно или зло), а также в эффекты эмоций (например, NPC (non-player character), долго испытывающий страх, может быстрее начать уставать, возрастающий уровень усталости может сделать NPC более раздражительным). В той степени, в которой физиологические характеристики моделируются у аффективных агентов, они очень абстрактны и просты, обычно состоят из нескольких измерений/уровней (например, возбуждение, энергия) и доступности физических ресурсов (например, степень голода, усталость или «здоровье»). Согласно данным разработчиков, добавленная характеристика «возбуждение» для агента Max [Becker et al. 2005] позволяет ему выражать изменения ритма дыхания и моргания, что делает его более правдоподобным. Эти характеристики затем составляют элементы *физиологической области* модели.

Определение специфичных триггеров и эффектов эмоций, представляемых в модели

Также одним из важных шагов при моделировании эмоций у искусственных агентов является определение соотношений генерации и эффектов эмоций. Это означает, что сначала должны быть определены соотношения, которые связывают вызывающие эмоции триггеры с конкретными эмоциями. Потом должны быть определены соотношения, которые связывают различные эмоции и настроения с их выражениями в модальностях, доступных для агента. В зависимости от выбранного теоретического основания модели и особенностей связанной с ним *абстрактной области*, если она присутствует, некоторые соотношения могут также включаться в абстрактную область. Например, если выбран пространственный теоретический подход, триггеры могут быть соотнесены с пространствами, которые соотносятся с конкретными эмоциями. С другой стороны, пространства могут быть соотнесены напрямую с отдельными эффектами (возбуждение будет соотнесено с более быстрыми движениями и с симулируемыми физиологическими характеристиками, например с возрастанием частоты дыхания).

Определение этих соотношений состоит из двух этапов. *Вопервых*, уточняются входящие в модель области (см. выше таблицу 1). Для модели, использующей *дискретный/категориальный подход*, где отдельные эмоции являются семантическими примитивами, соотношения, сопровождающие генерацию эмоций, будут

отображать связи элементов из объектной области (О) напрямую с элементами аффективной области (А). В модели, использующей *компонентный подход*, сначала определяются соотношения элементов объектной области с оценочными переменными внутри абстрактной области (Аб), которые затем соотносятся с аффективной областью (А). Во-вторых, содержание соотношений должно быть специфицировано, что выразится в генерацию конкретных эмоций или настроений, а также их эффектов. Как минимум это означает выражение эффектов в поведении, но обычно также включает аффективные выражения и еще реже влияние эффектов эмоций на когнитивные процессы. Поскольку некоторые агенты также обладают симуляциями «физиологических» характеристик, например уровнем энергии, усталости и т. п., то и эффекты эмоций, влияющие на эти характеристики, и эффекты этих характеристик, влияющие на генерацию эмоций, также должны быть представлены для усиления аффективного реализма.

Как минимум для аффективной модели агента требуются такие области и связанные с ними элементы, как объектная область (О), аффективная область (А) и поведенческая область (П). Обычно агенты выражают генерированные эмоции с использованием определенных экспрессивной области (Э) и ее элементов. Для более сложных агентов может быть необходима также когнитивная область (К), чтобы агент мог выражать более интересное и аффективно реалистичное поведение как результат генерированных эмоций. Для многих агентов также определяют физиологическую область (Ф) в понятиях характеристик, отражающих «физические» состояния агента (например, усталость, энергия). Дополнительно может быть определена абстрактная область (Аб), которая отражает конкретные теоретические подходы и/или отдельные теории и чьи элементы представляют собой дополнительные переменные, сопровождающие эффекты эмоций в различных модальностях, например использование параметров, отражающих эффекты эмоций для разных когнитивных процессов.

Специфическое содержание этих соотношений должно быть основано на эмпирических данных, относительно типичных триггеров конкретных эмоций и типичных эффектов эмоций в разных модальностях и представленных в модели каналах, включая любые дополнительные переменные из абстрактной области, возникающие под действием особого контекста мира агента и его функциональности. Например, тот факт, что физическая угроза безопасности робота будет являться триггером страха должна подтверждаться

контекстом, в котором находится робот, с помощью идентификации специфических угроз для его безопасности (например, падение с платформы, падение в щель, разрушение от удара об скалу). Эмпирические данные о том, что страх вызывает концентрацию внимания на стимуле и смещение интерпретации стимулов к оценке нейтральных ситуаций как опасных, должны подтверждаться в контексте «мира» робота путем существования различия между опасными и безопасными стимулами (например, быстро приближающийся большой объект является опасным стимулом), и определения альтернативных интерпретаций ситуаций согласно спектру опасности, чтобы у наделенного страхом агента была возможность выбрать наиболее опасные ситуации и действовать согласно им.

Результатом этого шага при моделировании эмоций является зафиксированный набор соотношений между представленными у агента областями, который затем служит основой для определения необходимых вычислительных задач для воплощения генерации и эффектов эмоций в аффективной модели агента. Конкретный набор необходимых соотношений зависит от представленных у агента экспрессивных модальностей, которые определены формой его воплощения, и от абстрактных областей, представленных в модели, которые определяются теоретической основой модели.

Нужно отметить, что в данной статье описывается агент без возможности обучения, поэтому специфическое содержание соотношений между аффективной и поведенческой областью должно быть предзадано разработчиком модели. Это ограничение, скорее всего, исчезнет в будущем, как только машинное обучение будет повсеместно внедрено в игровую и научную индустрии.

Определение аффективной динамики

Аффективная динамика в моделировании эмоций отражается в вычислении интенсивности эмоций во время генерации эмоций и в определении магнитуд эффектов эмоций во время моделирования эффектов, а также в определении зарождения и распада как интенсивностей, так и магнитуд эффектов эмоций. Определение аффективной динамики конкретной модели также включает в себя выбор того, как соединить (интегрировать) сложные эмоции (в процессе генерации эмоций) и сложные эффекты эмоций (в процессе моделирования эффектов) в одновременном процессе создания

сложных аффективных состояний или при попытках соединить вновь созданные состояния с уже существующими.

Реалистичное представление аффективной динамики является главной задачей аффективного моделирования с очень малым числом идей и вариантов решения, доступных на сегодня, а также малочисленными данными, за исключением утверждений и качественных отношений (например, тревога смещает внимание в сторону определения угрожающего стимула; возбуждение запускает более высокий тембр и темп речи). Во многих существующих аффективных моделях определение количества доступных качественных данных происходит *ad hoc*, с некоторыми значениями, определяемыми эмпирически, и основано на производительности модели. Вычисления интенсивности эмоций обычно включают в себя простые формулы (например, интенсивность эмоции = желательность какого-то события \times его схожесть) или линейные комбинации дифференциально взвешенных содействующих факторов. Эти формулы могут требовать значительной тонкой настройки для регулирования производительности модели.

Важный этап в разработке модели аффективной динамики это принятие решения о том, как интенсивность эмоций должна изменяться во времени, это означает, что необходимо продумать темпы появления и угасания эмоций. Простейшим решением здесь будет исключение этих компонентов и использование простой ступенчатой функции, где эмоция просто появляется на пике своей интенсивности, длится на протяжении какого-то временного интервала, а потом возвращается к значению 0 или базовому уровню для данного агента. Разные временные интервалы определяются для разных эмоций, и для разных личностей (например, в целом счастливый агент (высокая экстраверсия, высокая степень согласия, низкий нейротизм в терминах Большой Пятерки) будет поддерживать позитивный настрой и эмоции продолжительный период времени; вспыльчивый агент будет быстро становиться злым и испытывать злость долгое время).

С другой стороны, темпы появления и распада могут быть выражены монотонно возрастающей и убывающей (соответственно) функцией для моделирования аффективной динамики в более реалистичной манере. В моделях генерации эмоций используется множество разных функций, включая линейные, экспоненциальные, сигмоидные и логарифмические [Reilly 2006].

Наконец, разработчик должен решить, как будут интегрироваться сложные эмоции. Они могут генерироваться с помощью про-

цесса оценки, в котором уже существующие эмоции будут соединяться с новыми, только что появившимися эмоциями. Этот аспект аффективной динамики гораздо меньше изучен как в существующих психологических теориях и концептуальных моделях, так и в вычислительных моделях. Обычно для решения этой сложной проблемы используются относительно простые подходы, которые ограничивают реализм результирующей модели в любых ситуациях, кроме самых простых.

Рейлли проанализировал несколько существующих подходов для соединения простых эмоций (позитивные с позитивными, негативные с негативными) и выделил их недостатки и преимущества. Простое добавление интенсивности может привести к слишком большому усилению интенсивности (например, некоторые «низко интенсивные» эмоции будут приводить к «высоко интенсивной» реакции). Усреднение интенсивностей может приводить к финальной интенсивности, которая ниже, чем составляющие ее интенсивности, а это очень нежелательная ситуация для биологических агентов. Подход максимализма (где победитель забирает все «the winner takes it all») игнорирует кумулятивный эффект сложных эмоций. Ограничения этих простых функций мотивируют разработчиков использовать другие подходы, включая использование логарифмической и сигмоидной функции [Reilly 2006; Picard 1997]. Часто используются зависимые от области весовые коэффициенты, чтобы конкретная эмоция генерировалась в первую очередь в качестве функции личности персонажа.

Более проблематичная ситуация возникает, когда противоположные или совершенно различные эмоции возникают одновременно (например, некоторая ситуация вызывает и радость, и грусть сразу). Ни существующие теории, ни эмпирические данные в настоящее время не могут предоставить основу для разработки решения этой проблемы. Должны ли противоположные эмоции отменять друг друга? (Будем ли мы испытывать себя спокойными и нейтральными, если наш дом горит, но при этом мы только что выиграли в лотерею?) Правильно ли в принципе думать об эмоциях как о парах противоположностей? Можем ли мы заключить, что более сильная эмоция является более подходящей [Hudlicka 2007]? На какой стадии эмоции соединяются и противоречие разрешается? Должны ли конфликты эмоций разрешаться на стадии оценки, чтобы избежать проблемы заранее? Или на стадии когнитивных эффектов, например, при выборе цели? Или на стадии выбора поведения? Последнее является наиболее проблематичным, так как

довольно очевидно, что такой феномен наблюдается у биологических агентов, поскольку ошибка при разрешении противоречия конфликтующих поведенческих тенденций может быть фатальной для агента.

Вычислительные решения для комбинаций неконгруэнтных эмоций в целом зависят от задачи или области и принимаются *ad hoc*. Существующие подходы можно разделить на четыре категории: *основанные на интенсивности* (выбор эмоций или настроений с наивысшей интенсивностью); *основанные на правиле старшинства* (определение старшинства между представленными у агента эмоциями и разрешение любого конфликта между сложными состояниями с помощью этого правила (отношения старшинства часто устанавливаются на основе личности агента); *основанные на конгруэнтности настроению подходы* (выбор эмоции, конгруэнтной текущему настроению); и определение нового состояния *на основе абстрактных представлений* (комбинация семантических примитивов, внутри которых представлена эмоция или настроение (например, пространства PAD (Pleasure-Arousal-Dominance), оценочные переменные) и генерация нового аффективного состояния, которое соединяет сложные состояния).

После того как эмоция возникла, должна моделироваться динамика ее эффектов. Этот процесс является более сложным по причине мультимодальной природы эффектов эмоций, которую мы обсуждали выше, что приводит к тому, что доступны и применяются разные подходы для разных модальностей или каналов.

Моделирование разных магнитуд эффектов в поведенческой области часто включает в себя использование сложных порогов интенсивности для выражения разных эмоций по отдельным каналам (например, разные по интенсивности пороги могут требоваться для триггеров эффектов в когнитивной сфере/лицевых выражениях/поведении) [Sonnemans, Frijda 1994]. Разнообразные пороги также часто используются для обозначения соотношения разных интенсивностей внутри отдельных действий. Например, различные интенсивности радости могут выражаться с помощью смеха, хлопков или подпрыгивания.

Для некоторых каналов *экспрессивной области* интенсивность эмоции прямо пропорциональна магнитуде связанных с ней эффектов, что выражается с помощью связанных с конкретным экспрессивным каналом семантических примитивов. Например, для лицевого выражения, наивысшей интенсивностью обладает счастье, наибольшее движение скуловых мышц приведет к наиболее широ-

кой улыбке. Сходным образом эффекты возрастающего уровня злости, выражаемые в просодических характеристиках речи, будут иметь высшие значения некоторых просодических функций, например объема и темпа речи. Нужно, однако, отметить, что существуют значимые индивидуальные различия в экспрессивном выражении эмоций, например культурные особенности. Важно, чтобы разработчик модели принимал их во внимание для создания правдоподобных NPC (non-playing characters). Например, в интерактивном приложении агент А может выражать свою злость через очень громкий крик, в то время как агент Б может внезапно стать очень спокойным и внешне контролирующим себя, но при этом говорить на пониженных тонах.

Интенсивность также может выражаться путем включения большего количества доступных семантических примитивов в данный канал, а также с помощью добавления большего числа самих каналов. Например, внутри канала лицевого выражения возрастание интенсивности улыбки будет включать в себя не только сильное поднятие уголков губ, но также движение бровей, а также продолжительность этого изменения/движения. Когда интенсивность будет нарастать дальше, улыбка может смениться смехом, который дополнительно задействует речевой канал и разные вокализации, а также различные функции речи, которые выражаются каналом с помощью просодии (например, тембр, темп и объем речи).

Относительно возникновения и угасания эффектов эмоций выбор механизма будет аналогичен тому, который был сделан для возникновения и угасания интенсивности эмоции в процессе ее генерации. Так или иначе воплощение возникновения и угасания эффектов эмоций и настроений является более сложным из-за вовлеченности в выражение эмоций разнообразных каналов внутри разных модальностей. Динамика разнообразных процессов должна быть продумана одновременно. Во многих случаях эти процессы включают в себя физиологические системы, которые должны работать самостоятельно в случае воздействия триггера и которые обладают связанными с ними рефракторными периодами. Например, возросшее возбуждение, связанное с повысившейся интенсивностью многих эмоций (радость, страх, злость), сопровождается изменениями в концентрации специфичных гормонов и нейротрансмиттеров, которые должны быть расщеплены или нейтрализованы прежде, чем система вернется в свое базовое состояние. Поэтому разные эффекты выражаются с помощью разных временных шкал.

Многие из описанных механизмов могут игнорироваться в существующих агентах / игровых персонажах / роботах. Как и в случае с интенсивностью эмоции, простейшим решением является игнорирование совместного возникновения и угасания показателей. Другим простейшим решением будет передача изменений значения интенсивности напрямую в соответствующие магнитуды внутри областей эффектов, доступных данному агенту. Для отдельного лицевого выражения это будет означать перевод уровня интенсивности выражения в степень изменения положения специфичных лицевых мышц от их нейтрального состояния. Для речи это будет означать перевод разных интенсивностей в магнитуды разных сигнальных функций речи, например объема или темпа. Эти переводы должны принимать во внимание индивидуальные различия в выражении эмоций разными агентами как функцию их личности.

Так или иначе эти простые решения/требования для возрастания аффективной сложности и реализма не будут адекватно работать для случаев разнообразных независимо существующих процессов, сопровождающих эффекты эмоций по разным каналам, где разработчикам аффективной модели потребуется принимать более сложные решения. Кроме того, некоторые аффективные состояния могут быть определены только с помощью отдельных паттернов их аффективной динамики (например, смущение). Создание правдоподобных выражений сложных эмоций у воплощенных агентов является активной сферой исследований.

Как и в случае с вычислением интенсивности для разных личностей агентов могут быть выделены различные паттерны эффектов эмоций и настроений для большего аффективного реализма и повышения вовлеченности пользователя / другого агента. Это может быть сделано с помощью спецификации порогов для отдельных выражений разных эмоций, благодаря чему делается акцент на выражении конкретной эмоции, и/или выражения конкретной эмоции внутри некоторых каналов. Личностные особенности также могут отражаться в значениях появления и угасания разных эффектов. Например, агент с личностью А может обладать быстро возрастающими и интенсивными положительными реакциями с очень медленным угасанием положительных эмоций, и в это же время персонаж с личностью Б может создавать негативные социальные эмоции высокой интенсивности и высокими значениями угасания.

В данной статье были рассмотрены и проанализированы некоторые методологические проблемы моделирования эффектов эмоций: проблема поиска теоретических оснований моделирования эффектов эмоций, различия между внутренними и внешними эффектами эмоций, роль формы телесного (материального) воплощения агента для моделирования эффектов эмоций, взаимоотношения триггеров эмоций и эффектов эмоций, процесс моделирования динамики эффектов эмоций.

В качестве вывода можно привести ряд принципов, которые необходимо учитывать при моделировании эмоций:

1. Необходимо опираться на теорию, подходящую под задачи, и требования, выдвигаемые к моделируемому искусственному агенту, при отсутствии подходящей теории использовать комбинацию нескольких теоретических подходов или разрабатывать уникальный собственный подход.

2. Возможности репрезентации эффектов эмоций определяются той телесной формой, в которую помещен агент.

3. Разработчик должен принимать во внимание состояние развития генерации аффективных выражений для того экспрессивного канала, который планирует использовать.

4. При разработке модели аффективной динамики необходимо продумать темпы появления и угасания эмоций, а также механизм интеграции сложных эмоций.

5. Для достижения наибольшей аффективной реалистичности и человекоподобности агента необходимо включать в модель сложные социальные эмоции.

Эти пять принципов, касающихся генерации эмоций и выражения эффектов эмоций, а также разработки вычислительных моделей эмоций в целом, необходимо положить в основу модели эмоциональной системы для искусственного интеллекта (ИИ), чтобы эмоциональная подсистема у искусственных агентов не только отражала заложенные теоретические конструкторы эмоций, но и стала областью, поддерживающей и участвующей сложной коммуникации, что, в свою очередь, позволит приблизиться к разработке искусственного агента человекоподобного типа или ИИ третьей волны.

БИБЛИОГРАФИЯ

- Andre et al. 2000 – *Andre E., Klesen M., Gebhard P., Allen S., Rist T.* Exploiting Models of Personality and Emotions to Control the Behavior of Animated Interactive Agents, Proceedings of IWAI, 2000. https://www.researchgate.net/publication/2330891_Exploiting_Models_of_Personality_and_Emotions_to_Control_the_Behavior_of_Animated_Interface_Agents.
- Aylett 2004 – *Aylett R. S.* Agents and affect: why embodied agents need affective systems, 3rd Hellenic Conference on AI, Springer Verlag LNAI. P. 3025–3027. 2004.
- Bates et al. 1992 – *Bates J., Loyall A. B., Reilly W. S.* Integrating Reactivity, Goals, and Emotion in a Broad Agent, Proceedings of the 14th Meeting of the Cognitive Science Society. P. 45–49. 1992.
- Becker et al. 2005 – *Becker C., Nakasone A., Prendinger H., Ishizuka M., Wachsmuth I.* Physiologically interactive gaming with the 3D agent Max. Proceedings of the International Workshop on Conversational Informatics at JSAI-05, Kitakyushu. P. 347–351. 2005.
- De Rosis et al. 2003 – *De Rosis F., Pelachaud C., Poggi I., Carofiglio V., De Carolis B.* From Greta’s mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International journal of human-computer studies*. 59 (1–2). P. 81–118. 2003.
- Ekman, Friesen 1978 – *Ekman P., Friesen W. V.* Facial action coding system, Palo Alto, CA: Consulting Psychologists Press, 1978.
- El-Nasr et al. 2000 – *El-Nasr M. S., Yen J., Ioerger T. R.* FLAME – Fuzzy Logic Adaptive Model of Emotions. *Autonomous Agents and Multi-Agent Systems*. 3 (3). P. 219–257. 2000.
- Elliot 1992 – *Elliot C.* The Affective Reasoner: A Process Model of Emotions in a MultiagentSystem. Northwestern University, Evanston, 1992.
- Frijda 1986 – *Frijda N. H.* The Emotions. Cambridge: Cambridge University Press, 1986.
- Gratch, Marsella 2004 – *Gratch J., Marsella S.* A domain independent frame-work for modeling emotion. *Journal of Cognitive Systems Research*, 5(4). P. 269–306. 2004.
- Hudlicka 2007 – *Hudlicka E.* Reasons for Emotions. In W. Gray (Ed.), *Advances in Cognitive Models and Cognitive Architectures*. NY: Oxford. P. 263–278. 2007.
- Loyall 1997 – *Loyall A. B.* Believable Agents: Building Interactive Personalities. CMU, Pittsburgh, 1997.
- Ortony et al. 1988 – *Ortony A., Clore G. L., Collins A.* The Cognitive Structure of Emotions. NY: Cambridge, 1988.

- Prendinger et al. 2002 – *Prendinger H., Descamps S., Ishizuka M.* Scripting Affective Communication with Life-like Characters in Web-based Interaction Systems. *Applied Artificial Intelligence*, 16 (7–8). P. 519–553. 2002.
- Prendinger et al. 2004 – *Prendinger H., Saeyor S., Ishizuka M.* MPML and SCREAM: Scripting the Bodies and Minds of Life-Like Characters. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters: Tools, Affective Functions, and Applications*. Berlin: Springer, 2004.
- Picard 1997 – *Picard R.* *Affective Computing*. Cambridge, MA: The MIT Press, 1997.
- Reilly 1996 – *Reilly W. S. R.* *Believable Social and Emotional Agents*. CMU, Pittsburgh, 1996.
- Reilly 2006 – *Reilly W. S. N.* Modeling What Happens Between Emotional Antecedents and Emotional Consequents, *Proceedings of the ACE 2006*, Vienna, Austria, 2006.
- Scherer 1992 – *Scherer K. R.* Emotions are biologically and socially constituted: A response to Greenwood. *New Ideas in Psychology*, 10 (1). P. 19–22. 1992.
- Sonnemans, Frijda 1994 – *Sonnemans J., Frijda N. H.* The Structure of Subjective Emotional Intensity. *Cognition and Emotion*. 8 (4). P. 329–350. 1994.
- Staller, Petta 1998 – *Staller A., Petta P.* Towards a tractable appraisal-based architecture for situated cognizers. *Proceedings of the 5th International Conference of the Society for Adaptive Behaviour (SAB'98): Grounding Emotions in Adaptive Systems Workshop.*, Zurich, Switzerland, 1998.

Материал поступил в редакцию 31.01.2019