ООО Нейрок Техсофт
г Троицк, Московской обл.
Ул. Нагорная, д.9, офис 6.

Отдел Интеллектуальных Технологий, НейрОК Техсофт

Адаптивные нейросетевые методы в многошаговых играх с неполной информацией

Сергей А. Терехов

Лекция для VII Всероссийской научно-технической конференции “Нейроинформатика -2005”

Введение: От задач оптимизации и оптимального управления к теории игр 3

Специфика теоретико-игровых постановок информационных задач 5

Стратегические игры в матричной форме 7

Равновесие в играх с полной информацией..... 7

Равновесия в смешанных стратегиях. 9

Многошаговые игры.....11

Модель многошаговой игры с неполной информацией 14

Задача о деятельности холдинга консалтинговых компаний 18

Модель бизнес-процессов..... 18

Формализация математической модели многошаговой игры 20

Индивидуальные стратегии игроков 22

Компьютерная система COGITO..... 24

Выборочные результаты моделирования..... 24

Обсуждение 27

Послесловие.....28

Благодарности 29

Литература 29

Приложение. Нейросеть CNLS (Connectionist Normalized Local Splines) 32

Задачи 35

Задача 1. Ошибка в формуле..... 35

Задача 2. Плотность распределения континуума оптимальных решений в игре со стохастической матрицей. 35

Адаптивные нейросетевые методы в многошаговых играх с неполной информацией

Сергей А. Терехов
ООО Нейрок Техсофт

В лекции рассматриваются понятия и постановки задач математической теории игр. Особое внимание уделяется информационным и вычислительным аспектам игрового моделирования. Для важного случая игр нескольких игроков с неполной информацией предложена вероятностная формализация алгоритма обучения игрока оптимальной смешанной стратегии принятия решений. Важную часть лекции составляет подробное описание практического примера теоретико-игрового моделирования в оптимизации совместной бизнес-деятельности группы компаний, предоставляющих сходные консалтинговые услуги. Модель основывается на нейросетевой аппроксимации функции цены состояния игроков. При моделировании используется нейронная сеть специального вида (локальные нормализованные сплайны), алгоритмы и результаты обучения которой описаны в Приложении.

This Lecture introduces basic topics of contemporary game theory. Simulation and informational aspects of games are emphasized. Key concept of games with incomplete information is studied from point of view of proposed probabilistic formalism of Markov decision process. Player payoff probabilities are approximated by neural network model with specific architecture (Connectionist Normalized Local Splines). The model of strategic cooperation and competition in the internal market of holding company is served as a practical application.

> *Разобравшись в деле, убавь то, что должно быть убавлено*
>
> И Цзин, гексаграмма Сунь

Введение: От задач оптимизации и оптимального управления к теории игр

Ключевым аспектом организационной деятельности является принятие решений. В условиях, когда характер деятельности допускает полное подробное описание и всецело зависит только от ваших решений, эти решения, скорее всего, будут относиться к области *оптимизации*. Оптимизация состоит в проведении непосредственных изменений параметров деловой системы с целью максимизации выбранной функции ценности. Примером оптимизационного решения может служить пространственное размещение продукции в складском помещении, уменьшающее затраты на ее отгрузку потребителю [1].

В более сложных случаях, когда непосредственное произвольное изменение желаемых параметров затруднено, лицо, принимающее решения, имеет дело с задачей *управления*. Управляющие решения (разовые или последовательности решений) характеризуются внесениями изменений в некоторые переменные или параметры системы, которые лишь опосредованно отражаются в движении системы по благоприятной траектории или к желаемому состоянию. Широкий класс задач управления основывается на измерении отклика системы на управляющие решения. Возможный пример – система управления качеством бизнес-процессов “6 sigma” [2]. Одним из центральных принципов этого подхода является проведение измерений качества и использование принципа обратной связи.

Новым масштаб сложности возникает в ситуациях, когда отклик бизнес-системы зависит одновременно от решений нескольких лиц, занятых достижением индивидуальных целей. Даже в случаях, когда каждый из участников, принимающих решения, обладает полной информацией о совместно управляемой системе, ее отклик и степень достижения целей оказываются весьма богатыми по своему разнообразию. Индивидуальные наилучшие решения могут не существовать (в математическом смысле), либо быть неустойчивыми – некоторые из участников предпочтут изменить свои решения на следующем этапе управления.

Рациональные способы принятия решений несколькими лицами являются предметом *теории игр* – математической дисциплины, замыкающей цепочку, начатую оптимизацией и теорией управления.

В соответствии с этой логикой выстроена и серия вводных лекций, прочитанных автором на школе по нейроинформатике в МИФИ в 2001-2004 г. В первой из лекций рассматривалась проблема описания неопределенностей в оптимизируемой системе с позиции аппроксимации плотности вероятности ее состояний. Во второй лекции, посвященной Байесовым сетям, обсуждались практические вопросы принятия решений в условиях неопределенности и моделирование вероятностей достигаемых вследствие решений состояний системы. Третья лекция обобщает механизмы принятия решений на случай многошаговых задач управления, с поиском нестационарных управляющих траекторий методами нейро-динамического программирования (обучения с подкреплением).

Новая, четвертая, лекция рассматривает математические имитационные модели принятия рациональных решений в играх, в которых игроки не имеют полной информации о системе (и решениях других игроков). При этом основное внимание будет уделено повторяющимся играм, когда управляющие решения применяются игроками многократно, а целью служит нахождение *стратегий* принятия решений, обеспечивающих для каждого игрока наилучший совокупный исход множества игр.

Базовым исследовательским инструментом, используемым автором во всех лекциях[1], являются искусственные нейронные сети, аппроксимирующие искомые в соответствующих задачах функции (плотность вероятности, частные условные вероятности в Байесовых сетях, функции цены решения в нейродинамическом программировании). Поэтому параллельно основному содержанию, в лекциях обсуждаются также вопросы эффективного синтеза и обучения нейросетевых моделей.

В этой лекции нейронная сеть используется для моделирования распределения вероятности суммарного дисконтированного выигрыша игрока. Примененная архитектура – локально-связанные нейросетевые сплайны (Connectionist Local-

[1] Равно, как и в основной производственной деятельности

Spline Neural Network – CLSNN) была предложена в середине 80-х в Центре исследований нелинейности национальной лаборатории в Лос-Аламосе. Важной особенностью этой нейросети является удачное сочетание локальной и глобальной аппроксимации благодаря специальной нормировке радиальных базисных функций. Подробно эта нейросеть описана в Приложении.

Теория игр, как научное направление, относительно молода (в сравнении с теорией оптимизации по Лагранжу, вариационным исчислением и оптимальным управлением) – фундаментальные положения ее были сформулированы в середине прошлого века Дж. Фон-Нейманом [3]. Примерно в это же время были установлено понятие равновесия стратегий[4] по Нэшу. В дальнейшем были предложены различные подходы к достижению равновесия [5]. Цикл исследований по применениям теории игр в экономических задачах был отмечен Нобелевской премией по экономике (Дж. Нэш, Дж. Харшаньи, Р. Зельтен, 1994). Нужно сказать, что теория игр не занимала в нашей стране такого ведущего положения, как в США (в теории управления, по всей видимости, наблюдалась обратная ситуация), что во многом было вызвано особым отношением к экономическим вопросам, которые составляют основное прикладное содержание теории игр. В последнее время число публикаций и книг по теории игр в России значительно возросло [9-10, и др], многие монографии по оптимизации и управлению включают теоретико-игровые разделы.

Приложения теории игр не ограничиваются экономическими проблемами. К ним относятся такие области, как коллективное принятие решений советом экспертов, эксплуатация совместной собственности, судебная деятельность, международные переговоры, военно-стратегическое планирование, политические и маркетинговые технологии.

Данная лекция, вводная по своей сути, призвана расширить круг интересующихся теоретико-игровыми приложениями в контексте практического использования методов обучения машин и нейроинформатики для моделирования практических игровых ситуаций в деловой и организационной деятельности.

В первом разделе будут в простой форме обсуждены основные понятия теории игр, приведены иллюстративные примеры игр. Второй раздел будет посвящен многошаговым играм, в которых решения принимаются в динамике и с учетом предыдущей истории игр. Многошаговый характер игры позволяет обобщить методы обучения с подкреплением, развитые ранее для случая одного игрока (“игры с природой”). В случае, когда игроки обладают полной информацией, будет предложен вычислительный подход к поиску равновесия на основе нейродинамического программирования.

В третьем, наиболее интересном с прикладной точки зрения, разделе будет описан наиболее близкий к реальности случай Байесовых игр (игр с неполной информацией). Изложение будет построено вокруг имитационной модели деятельности холдинга компаний, оказывающих консалтинговые услуги. Будет показано, как можно обеспечить эффективное управление группой компаний на основе распределения выполняемых бизнес-проектов путем проведения внутренних аукционов.

Специфика теоретико-игровых постановок информационных задач

Прежде чем перейти к изложению основных понятий, с которыми оперирует теория игр, рассмотрим пример “простой” игры, в решении которой содержатся многие ключевые особенности теоретико-игровых задач.

Игра 1 (Дележ, Д). Двое имеют возможность разделить сумму в 100 долларов. Каждый игрок, в тайне от другого игрока, сообщает судье (или помещает в запечатанный в конверт) количество денег, которое он желает получить. Конверты вскрываются одновременно, и если сумма двух значений не превышает 100, то каждый получает столько, сколько он запросил. Однако, если сумма оказывается больше 100, никто не получает ничего. Как следует[2] поступить игрокам?

Допустим, что игроки сделали заявки x и y. Каждому из них сообщен результат его игры (запрошенная сумма или ноль), но не ставка соперника. Следует ли игрокам изменить свои решения, если им представится возможность сыграть еще раз? Как изменится подход[3] к решению, если *обе* ставки становятся известными обоим игрокам?

При размышлении, ответы на эти вопросы становятся все менее очевидными. В целом, понятно, что если сумма ставок оказалась меньше 100, то свою ставку при следующем ходе можно увеличить каждому игроку (и уменьшить, если они не получили ничего). После нескольких ходов игры сумма ставок, вероятно, приблизится к 100. Однако дальнейшее движение не выглядит *рациональным* ни для одного игрока – и уменьшение, и увеличение своей ставки одним из игроков приведут к потерям, каково бы не было достигнутое ранее распределение ставок.

Такие конфигурации решений, при которых каждый игрок не может единолично улучшить свою премию, называются *состояниями равновесия* (по Нэшу). Приведенная Игра 1 имеет бесконечное число состояний равновесия вида $[x, 100 - x]$.

Понятие равновесия является ключевым в теории игр – собственно, результатом теории является обнаружение и классификация состояний равновесия, равновесных *стратегий* (о стратегиях – см. ниже) игроков, а также получаемых ими выигрышей (т.е. вычисление *цены* игры).

Другим существенным моментом является само понятие рационального поведения. Формализация этого понятия проведена Дж. фон-Нейманом [3] в теории рационального выбора, основанной на аксиоматическом подхлде. Она состоит в том, что *рациональный* игрок применяет наилучшее для себя решений A из набора возможных решений, руководствуясь индивидуальной функцией полезности (utility function, U(A)). Значения функции полезности являются абстрактными и имеют только относительный смысл при сравнении двух возможных решений. Так, для двух возможных решений A и B, ситуации {U(A)=2, U(B)=1} и {U(A)=100, U(B)=1} рассматриваются как идентичные, т.е., предпочтение, безусловно, отдается решению A (при этом во втором случае это предпочтение не является ни в каком смысле “более сильным”).

---

[2] Предварительные переговоры между игроками не допускаются. Но даже наличие (словесных, без механизма обеспечения условий договоренности) переговоров между игроками мало что меняет – игроки могут проигнорировать итоги обсуждения.

[3] Игры такого типа весьма близки к проблемам использования совместной собственности. Два фермерских хозяйства располагают общим пастбищем. Каждая ферма решает, какую численность голов скота целесообразно содержать. При превышении некоторого предела по общему поголовью, пастбище теряет возможность восстанавливаться, и оба хозяйства терпят убытки. В пределе полного эгоизма ни одно из хозяйств не желает отступать от достигнутого исторически (одного из возможных) равновесия.

Стратегические игры в матричной форме

Если игра проводится конечное и *известное заранее* число раз, то каждый игрок может предложить набор *ходов* (управляющих решений) для каждой из возможных промежуточных ситуаций в игре. Перечень решений для каждой из ситуаций игры называют *стратегией* игрока. Такая совокупность ходов, понимаемая как один макро-ход, по индукции, может быть выбрана до начала игры, и таким образом игра фактически является одношаговой. Игроки публикуют свои стратегии и вычисляются итоговые выигрыши и цена всей игры.

Игры этого типа называются стратегическими. Выигрыш каждого игрока зависит от стратегий *всех* игроков. При известном (и конечном) наборе стратегий он может быть явно вычислен. Таким образом, все выигрыши данного игрока для всех комбинаций стратегий формируют матрицу (с числом измерений, равным числу игроков). Индексами в каждом измерении служат номера стратегий соответствующих игроков, а матричные элементы равны выигрышам игрока. В общем случае каждый игрок имеет дело со своей матрицей выигрышей.

Для простейшего случая двух игроков матрицы выигрышей представляют собой обычные прямоугольные таблицы чисел. В стратегической игре один игрок выбирает номер строки в матрице, второй игрок – номер столбца. После этого игра считается состоявшейся, выигрыш дается соответствующим матричным элементом.

Таким образом, исход игры с матрицей выигрышей M есть, суть, произведение:

(1) $$\vec{p}_R \cdot M \cdot \vec{p}_C^{\,'}$$

где, $\vec{p}_R$ - вектор решения игрока, выбирающего строку матрицы, все компоненты вектора кроме одной (единичной) равны нулю, $\vec{p}_C^{\,'}$ - аналогичный вектор решения игрока, выбирающего столбец[4].

Равновесие в играх с полной информацией

Действующий рационально игрок *R* должен стремиться к максимизации своего выигрыша (1) при любых вариантах выбора решений другими игроками. В этом разделе рассматривается ситуация, когда оба игрока обладают полной информацией об игре (известны как матрицы выигрышей, так и выигрыши всех игроков, если игра повторяется). Для случая двух игроков оптимальная стратегия состоит в максимизации гарантированного выигрыша:

(2) $$\vec{p}_q^{\,*} = \arg\max\left(\min_{P_C}\left(\vec{p}_R \cdot M_q \cdot \vec{p}_C\right)\right),\ q = \{R,C\}$$

Если каждый игрок ищет единственную наилучшую стратегию k_0

(3) $$p_{qk} = \delta_{kk_0(q)}$$

то полученные решения называются *чистыми* стратегиями.

---

[4] Традиционно, игрока, выбирающего строку, называют R (row), а игрока, выбирающего столбец – C (column). В дальнейшем символы транспонирования для векторов в формулах будут опускаться.

Решение (2) в чистых стратегиях далеко не всегда существует. Наиболее подробно изучен случай *антагонистических* игр двух игроков (когда сумма выигрышей игроков строго равна нулю – что один выиграл, то другой проиграл). В этом случае

(4) $$M_R = -M_C = M$$

Если матрица игры имеет седловую точку, то точным результатом, установленным фон-Нейманом, является равенство цен игр игроков, следующих соответственно, максиминной и минимаксной стратегий:

(5) $$\max_{p_R}\left(\min_{p_C}\left(\vec{p}_R \cdot M \cdot \vec{p}_C\right)\right) = \min_{p_C}\left(\max_{p_R}\left(\vec{p}_R \cdot M \cdot \vec{p}_C\right)\right)$$

В этой лекции будет рассматриваться, в основном, общий случай, когда для трех и более игроков прямого антагонизма нет.

Особенности равновесия в чистых стратегиях иллюстрирует описываемая ниже хрестоматийная стратегическая игра в матричной форме (“Дилемма заключенного”).

> Игра 2 (Дилемма заключенного, ДЗ). Двое задержанных лиц подозреваются в совершении серьезного преступления. Не имея возможности общаться между собой, они поставлены перед выбором, свидетельстовать ли против друг друга или нет? При этом, если один обвинит другого, а тот обвинит первого, то оба получат наказание. Если обвинение будет только со стороны одного из игроков, то он получает свободу, а другой – максимальное наказание. Если же оба откажутся от обвинений (т.е. выберут сотрудничество), то, в силу других обстоятельств дела, они оба получат относительно небольшое наказание. Как поступить рационально?

Несмотря на детективно-развлекательное описание игры, проблема, которая в ней поставлена, является весьма серьезной – а именно, возможно ли возникновение сотрудничества между участниками, преследующими индивидуальные цели? Игра ДЗ имеет более чем полувековую историю, и впервые, по-видимому, обсуждалась в начале 50-х годов в контексте глобальных ядерных стратегий, см. [8]. Она также имеет приложения к проблемам в экономике, политике и биологии.

Матрицы игры ДЗ имеют вид:

		Второй игрок	
		Сотрудничество	Обвинение
Первый игрок	Сотрудничество	R, R	S, T
	Обвинение	T, S	P, P

Таб 1. Матрицы игры “Дилемма заключенного”. Приведены пары выигрышей первого и второго игрока

Важными являются относительные значения выигрышей игроков. Наибольший выигрыш достигается, если поддаться искушению и обвинить другого игрока. При этом обвиненный игрок окажется в наихудшей из возможных ситуаций. Однако, если оба игрока выбирают стратегию сотрудничества, то выигрыш каждого будет больше, чем в случае взаимных обвинений. Поэтому T > R > P > S. Кроме того, оба игрока не должны иметь возможность выйти из дилеммы, по очереди предавая друг друга, т.е. R > (S+T)/2.

Де-факто, обычно используются значения T=5, R=3, P=1, S=0.

Проведем теперь рациональные рассуждения за обоих игроков, используя только чистые стратегии. Предположим, что первый игрок считает, что другой игрок будет придерживаться стратегии сотрудничества. Если первый также будет сотрудничать, то оба получат выигрыши R=3. Если же первый в этих условиях выберет обвинение, то его индивидуальный выигрыш возрастет, и окажется равным T=5. Следовательно, согласно теории рационального выбора, оптимальной стратегий первого игрока, если второй выбрал сотрудничество, будет обвинение.

Пусть теперь первый игрок считает, что второй его обвинит. Если первый в этих условиях выберет сотрудничество (отказ от обвинения), то его выигрыш окажется равным S=0. Однако, если первый также выберет обвинение, то его выигрыш снова возрастет, и окажется равным P=1.

В итоге, при любых предположениях о характере поведения второго игрока, оптимальным для первого игрока будет стратегия отказа от сотрудничества. В этих случаях говорят, что стратегия обвинения *доминирует* над стратегией сотрудничества.

Аналогичной логике следуют рассуждения второго, тоже рационального, игрока. Оптимальные стратегии игроков могут содержаться только среди доминирующих стратегий, и следовательно оба выберут взаимное обвинение, получив при этом выигрыши P=1. Проблемная суть задачи состоит в том, что оба игрока упускают возможность кооперации, которая дала бы им выигрыши R=3. Этот факт и составляет дилемму.

Вопрос о возникновении сотрудничества является крайне важным, и эта игра многократно подвергалась экспериментальной проверке – будут ли реальные люди-игроки всегда следовать равновесию Нэша, являющемуся следствием математической рациональности? Результаты экспериментов [7] варьируются, в зависимости от фактических условий эксперимента (насколько точно удается воспроизвести ситуацию равенства функций полезности для игроков условиям задачи). Во многих случаях люди чаще (от 50% до 94%) выбирают стратегию обвинения, если не допускается предварительных обсуждений перед ходом в игре. Однако, если оппоненты непосредственно наблюдают друг друга, обвинение возникает реже (от 29% до 70% случаев, в зависимости от фактических условий игры). Мы вернемся к экспериментам ниже, при обсуждении повторяющихся игр.

Равновесия в смешанных стратегиях.

Факт обязательного следования игроком чистой стратегии является, в некотором смысле, сковывающим действия игрока, что может быть использовано другими игроками для увеличения их выигрышей. Может ли игрок “скрыть” часть своих намерений и придерживаться разных стратегий в разных актах игры? Ответ на этот вопрос утвердительный. Формальное рассмотрение этого вопроса состоит в введении вероятностей следования игроком той или иной стратегии.

Стратегия, включающую вероятностные комбинации возможных чистых стратегий, называют *смешанной*. Чистые стратегии являются частным случаем смешанных стратегий, когда вероятность одной из стратегий равна единице.

Выигрыш игрока при смещанной стратегии p по-прежнему дается выражением (1), в котором под p теперь следует понимать вектор распределения вероятностей чистых стратегий, составляющих смешанную.

Число возможных смешанных стратегий для конечного набора чистых стратегий бесконечно велико. Однако, это факт, как это часто бывает, лишь упрощает задачу. А именно, для случая смешанных стратегий, имеется фундаментальный результат [4], относящийся к вопросу о существовании равновесия. Напомним, что в случае чистых стратегий, установить в общем случае существование состояний равновесия не удавалось. Для игр с конечным набором состояний доказана следующая теорема:

Теорема (Дж. Нэш, 1950). Каждая финитная игра имеет точку равновесия[5].

Под равновесием в теореме Нэша понимается набор смешанных стратегий, таких, что для каждого игрока значение функции полезности (зависящее от его выигрыша при данном наборе стратегий) не может быть увеличено *индивидуальным* уклонением от равновесной смешанной стратегии.

Фундаментальность этой теоремы состоит в том, что она справедлива как для кооперативных, так и для антагонистических игр с любым числом участников. До теоремы Нэша исследования были основаны на классификациях игр с рассмотрением каждого класса в отдельности (именно так построена основополагающая книга [3]).

С практических воззрений теория Нэша не дает ответа на все вопросы. В частности, в теореме ничего не говорится о существовании пути к равновесию в повторяющихся играх и об устойчивости точек равновесия (существовании областей притяжения). Кроме того, в теореме утверждается существование по-крайней мере одного равновесия, но, фактически, равновесных точек может быть более одной. Почему игроки будут стремиться в своих стратегиях к одной общей для всех точке? А если это не так, то достижимы ли вообще равновесные состояния в играх? Равновесие по Нэшу имеет смысл только тогда, когда каждый игрок информирован о том, каким стратегиям будут следовать остальные. Если же теоретическое равновесие не единственно, то игроки не будут иметь этой информации.

Вопрос выбора равновесия позднее был положительно разрешен в работах Дж. Харшаньи и Р. Зельтена [5] (разделивших с Дж. Нэшем нобелевскую премию). Ими была построена общая теория выбора точек равновесия, которая рекомендует единственную[6] стратегию для каждого игрока. Книга [5] недавно переведена на русский язык и доступна заинтересованному читателю.

Завершая на этом общий экскурс в предмет теории игр хотелось бы остановиться на информационном аспекте, поскольку именно он интересен в контексте тематики школы по нейроинформатике. Информированность участников об условиях и параметрах игры является основой для рационального поведения. В современной деловой активности, характеризующейся высоким уровнем конкуренции, свободное распространение информации, существенной для принятия решения, крайне затруднено. При этом особенно охраняется информация о функции полезности игрока, что, конечно, затрудняет выбор даже при известной матрице игры.

---

[5] Именно так теорема сформулирована в 27-страничной диссертации Дж Нэша, воспроизведенной в фотокопии в книге [4]: “THEO. 1: Every finite game has an equilibrium point.”

[6] Ложка дегтя заключается в том, что теория Харшаньи-Зельтена о выборе единственного равновесия сама не является единственно возможной...

Неполнота информации может быть вызвана не только действиями по ее защите, но и большим объемом самой информации. Так, например, участники биржевых игр имеют дело с невероятным по объему потоком новостей от источников с различной степенью достоверности. Разнообразие реакции участников на доступную им информацию определяет общую ценовую ситуацию на рынке. В итоге, наблюдаемые решения рыночных игр весьма разнообразны и динамичны!

Схожая ситуация наблюдается и на рынке консалтинговых услуг, предоставляемых компаниям экспертами или специализированными фирмами. Фактически, цена контракта для покупателя определяется его функцией полезности, неизвестной продавцу. По этому на практике не существует универсальной “справедливой” стоимости услуг – функции полезности различны у разных компаний, и кроме того, они могут изменяться во времени.

При неполной информации игроки вынуждены принимать решения на основе своих ожиданий, выраженных в форме информационных *моделей* игры, прогнозирующих как поведение других игроков, так и исход игры. На практике сбор информации для уточнения моделей происходит одновременно с самим процессом игры. К этому вопросу мы и переходим в следующем разделе.

Многошаговые игры

Рассматриваемые ранее стратегические игры со смешанными стратегиями игроков понимались, как повторяющиеся игры с независимыми выигрышами. В каждом конкретном акте игры игрок, конечно, должен был выбрать какой-то конкретный ход. При этом объем информации, доступной каждому игроку, о функциях полезности других игроков и матрице игры предполагался постоянным.

Для моделирования процесса получения и использования в игре информации более реалистичным являются *многошаговые* модели игр.

В многошаговых играх игроки заинтересованы в максимизации суммарных выигрышей, полученных на каждом шаге игры. Обычно в качестве целевой рассматривается дисконтированная сумма выигрышей

(6) $$V = \sum_t \gamma^t r(t)$$

где $\gamma < 1$ – фактор дисконтирования, $r(t)$ – выигрыш игрока на шаге t.

Абстракция бесконечного времени игры является существенным отличием от задачи повторяющихся игр и известным конечным числом повторений. Действительно, *последняя* игра в цепочке игр не может повлиять на сложившиеся исходы предыдущих актов, и следовательно, должна рационально рассматриваться игроками, как обычная стратегическая игра. Это одинаково понимается всеми рациональными игроками, и, следовательно, задача сводится к цепочке из (n-1) игр. Это рассуждение по индукции приводит к рассмотрению каждой игры из повторяющейся серии, как игры в независимой стратегической постановке.

В многошаговой игре с дисконтированным суммарным выигрышем стратегия нетривиальным образом зависит от исходов многих игр[7].

Игра “Дилемма заключенного”, описанная в предыдущем разделе, экспериментально исследовалась и в многошаговой постановке[8], при этом в качестве игроков выступали компьютерные программы разных авторов. Турниры между попарно играющими программами были организованы Робертом Аксельродом (Robert Axelrod) в 1980, и позднее повторены спустя несколько лет. На основе стратегий, предложенных в программах, была сформулирована смешанная стратегия, вероятности в которой определялись генетическим оптимизационным алгоритмом. В итоге, в популяции стратегий выживали, в основном, стратегии типа “око-за-око”, в которых игрок принимает решение сотрудничества в первой игре, а далее повторяет ходы оппонента. Однако не нулевые вероятности имели также и другие стратегии, например “обидчивая” стратегия, в которой игрок сотрудничает до первого отказа от сотрудничества со стороны оппонента, а далее всегда выбирает решение обвинения. Подробная информация об этой игре и турнирах Аксельрода имеется в Интернет [16].

Поиск оптимальной стратегии в многошаговой игре может быть представлен как обобщение формализма Марковского процесса решений (Markov Decision Process) на случай нескольких игроков.

Для одного игрока (агента) задача оптимизации поведения была подробно рассмотрена в предыдущей лекции автора [12], поэтому приведем лишь основные соотношения, которые и будут обобщены. Марковский процесс принятия решений представляет собой совокупность из множества состояний среды (игры!) $s \in S$, множества возможных решений агента $a \in A$, функции подкрепления (выигрыша!) агента $r(s,a)$, а также плотности вероятности $T(s'|s,a)$ переходов между состояниями среды.

Выбор решений агента определяется его стратегией $p(a)$, заданной как плотность распределения на пространстве решений. Оптимальная стратегия определяется уравнением Беллмана (см. список литературы в [12]):

(7) $$V(s,p^*) = \max_a \left\{ r(s,a) + \gamma \sum_{s'} T(s'|s,a) \cdot V(s',p^*) \right\}$$

Для оптимизации выбора стратегии удобнее перейти к понятию цены Q пары (состояние, решение), которая равна суммарному дисконтированному подкреплению при принятии в состоянии s решения a, и в дальнейшем следования оптимальной стратегии:

(8) $$Q(s,a) = r(s,a) + \gamma \max_{a'} Q(s',a')$$

[7] В этом смысле, многошаговая игра может рассматриваться, как специальный вариант известной задачи об останове [11]. В игре участвует еще один игрок, ходом которого является вероятностное решение, продолжать дальше многошаговую игру или нет. Если игрокам известна вероятность продолжения γ, то с их точки зрения полученная игра эквивалентна многошаговой игре с фактором дисконтирования γ.

Для практических вычислений широко применяются итерационное Q-обучение и алгоритм *SARSA*, в сочетании с ε-оптимальным поведением агента на оптимизируемой траектории.

В теоретико-игровой постановке цены игр V зависят от решений нескольких игроков:

(9) $$V_q\left(s, p_q^*; \vec{p}_{-q}^*\right) = \sum_t \gamma^t \left\langle r_q\left(s(t), p_q^*; \vec{p}_{-q}^*\right) \middle| s_0 = s \right\rangle$$

В формулах индексом (-q) обозначены наборы решений всех игроков кроме данного, угловые скобки указывают на то, что r является решением соответствующей стратегической игры на каждом шаге. Каждый игрок при поиске оптимальной стратегии должен был бы просто максимизировать свою функцию ценности Q в новом состоянии s':

(10) $$Q_q\left(s, a_q; \vec{a}_{-q}\right) \overset{?}{=} r_q\left(s, a_q; \vec{a}_{-q}\right) + \gamma \max_{a'} Q_q\left(s', a_q'; \vec{a}_{-q}\right)$$

однако непосредственная индивидуальная ее максимизация невозможна, так как функция Q зависит от решений других игроков. Для выхода из противоречия [13] заметим, что $Q_q\left(s', a_q'; \vec{a}_{-q}\right)$ определяет матрицу некоторой игры в состоянии s', и оптимальная стратегия опирается на смешанную стратегию $p\left(a_q^{'}\right)$ в этой игре.

В частном случае строго антагонистической игры агент выбирает максимизацию гарантированного выигрыша:

(11) $$Q_q\left(s, a_q; \vec{a}_{-q}\right) = r_q\left(s, a_q; \vec{a}_{-q}\right) + \gamma \max_{p\left(a_q^{'}\right)} \min_{p\left(a_{-q}^{'}\right)} Q_q\left(s', a_q'; \vec{a}_{-q}\right)$$

В общем случае второе слагаемое в сумме (11) содержит решение матричной игры $\left\{Q_q\left(s', a_q'; \vec{a}_{-q}\right); Q_{-q}\left(s', a_q'; \vec{a}_{-q}\right)\right\}$, которое учитывает игровые матрицы всех игроков:

(12) $$Q_q\left(s, a_q; \vec{a}_{-q}\right) = r_q\left(s, a_q; \vec{a}_{-q}\right) + \gamma \max_{p\left(a_q^{'}\right)} \min_{p\left(a_{-q}^{'}\right)} \left\langle p_q^{'} \cdot Q_q\left(s', a_q'; \vec{a}_{-q}\right) \cdot p_{-q}^{'} \right\rangle$$

Здесь по-прежнему угловыми скобками обозначена цена игры.

Решение (12) отражает характер получения информации при взаимодействии агента с игровой средой: для поиска оптимальной стратегии каждый агент кроме формирования собственной функции цены состояния Q_q должен обучаться моделям матриц выигрышей Q_{-q} всех остальных игроков. При этом их текущие выигрыши должны быть ему известны на каждом шаге (хотя фактические смешанные стратегии других игроков и их фактические матрицы игр остаются скрытыми). В каждый момент времени агент оптимизирует свое поведение в текущем равновесии Нэша для текущей *модели* игры. Это еще раз подчеркивает отличие стохастических многошаговых игр от повторений одной стратегической игры: в многошаговой игре матрицы выигрышей эволюционируют во времени.

Для получения оптимальных стратегий, также как и в случае обычного обучения с подкреплением, агент может следовать ε-оптимальной стратегии на каждом шаге, с

вероятностью ε уклоняясь от текущего оптимального (“жадного”) распределения вероятностей своих решений с целью исследования всего пространства решений.

Сходимость получаемых стратегий в модели типа (12) исследовалась в работах авторов [12] и установлена для ряда специальных предположений о характере игр. Однако вопросы совместного выбора общего равновесия остаются открытыми. Нужно заметить, что полученные модели игр не обязательно сходятся к фактическим матрицам выигрышей, а модели стратегий – к фактическим стратегиям игроков-соперников данного агента.

Вычислительные затраты на реализацию модели (12) значительны. Обучаемые параметры включаю в себя матричные элементы всех матриц игр всех игроков, при этом теоретически сходимость обучения гарантирована только для случая, когда каждое из состояний игры посещается (бесконечно) часто [16]. Для вычисления итерационных поправок к ценам Q требуется на каждом шаге решать задачу серию задач квадратичного программирования для определения векторов вероятностей в смешанных стратегиях каждого игрока. Прямой путь, как и в случае обучения с подкреплением, состоит в использовании нейросетевых аппроксимаций для всех Q-факторов, при этом в качестве оценки ошибки аппроксимации используются разности старых и новых значений Q-факторов на временном шаге (TD-алгоритм, [14]).

Для практических целей в нашей лекции будет рассмотрена упрощенная постановка задачи обучения агента-игрока, в которой неопределенность в ценах игры, вызванная решениями других игроков моделируется статистически. Одновременно с упрощением вычислений предлагаемая модель является более общей, т.к. она включает случай, когда агенту известен только его текущий выигрыш в игре, а значения выигрышей других игроков не известны. Полная теория игровых моделей такого типа далека от завершения, поэтому в лекции обсуждается практический пример из области организации работы холдинга компаний, оказывающих схожие услуги на общем рынке. Пример иллюстрирует общие проблемы, возникающие при игровом моделировании, а также характер результатов, которые можно получить из игровых моделей.

Описание основного вычислительного звена модели – нейросетевой архитектуры, аппроксимирующей стохастическое значение цены игры, вынесено в Приложение.

Модель многошаговой игры с неполной информацией

Рассмотрим ситуацию, когда игрок вынужден принимать последовательность решений в многошаговой игре в условиях, когда матрицы игры и выигрыши других игроков ему не известны. Фактически, в этих условиях неизвестно даже общее число игроков, участвующих в игре. Постановка такой задачи весьма близка к реальной ситуации при игре на бирже или при деловой деятельности в условиях конкуренции на рынке.

До начала игры игроку[8] известно множество ходов, которые он может делать[9] в игре. Единственная информация, которой дополнительно снабжается игрок в

[8] В этой леции мы,прежде всего, интересуемся вопросами принятия решений игроками – компьютерными программами, поэтому в текте не делается различий между терминами *игрок* и *агент*.

[9] В действительности, это требование может быть ослаблено, если игрок будет получать отрицательные выигрыши при невозможных ходах. В этом случае фактические границы области допустимых ходов устанавливаются агентом в процессе обучения в игре.

течение игры – это значения его выигрышей. Игроку также известен *его* текущий счет. В чуть более ослабленной постановке игроку может сообщаться информация о значениях некоторых переменных, характеризующих текущее состояние игры при играх с несколькими состояниями. Эти переменные добавляются к переменным собственного состояния агента (например, его текущий счет, пространственное расположение и т.п. – опредедляется конкретными особенностями игры).

Целью агента по-прежнему является максимизация собственного дисконтированного выигрыша в последовательности игр. Q-фактор агента (максимальное значение суммарного выигрыша при начале игры из состояния s, в котором агент принимает решение a, и далее следует избранной стратегии) зависит от “скрытых” переменных – решений других игроков. Поэтому непосредственное использование итерационных алгоритмов (12) затруднено.

Эффект скрытых переменных, который невозможно при отсутствии информации учесть точно, может быть описан статистически. В предлагаемой вероятностной модели значение Q фактора объявляется случайной величиной с параметрической плотностью распределения, моменты которого зависят только от переменных состояния агента и его решений. Без ограничения общности будем считать параметрическое распределение нормальным (с двумя функциональными параметрами – математическим ожиданием и дисперсией). Обобщение на случай других распределений, как будет видно из дальнейшего, состоит в использовании при обучении других видов функции правдоподобия. Гипотезу о соответствии распределения выбранной форме в каждой практической задаче следует проверить статистическими критериями. Описание соответствующих процедур имеется в стандартных курсах статистики, в данном тексте эти вопросы не затрагиваются.

Собственно, специфика нормальности распределения никак не используется[10], кроме явного вида самой функции и ее производных.

Вероятностная модель функция цены пары {состояние, решение} имеет вид:

$$Q_q\left(s, a_q; \vec{a}_{-q}\right) \propto \tilde{Q}\left(s, a_q\right) = N\left(m\left(s, a_q\right), \sigma\left(s, a_q\right)\right) \tag{13}$$

Моменты Гауссовой плотности распределения N рассматриваются, как независимые функции (от известных переменных), подлежащие определению в процессе обучения. Их удобно представить в виде нейросетевых аппроксимаций. На практике для аппроксимации может использоваться одна и та же нейронная сеть, с двумя выходами.

Теперь необходимо обобщить понятие ожидаемого выигрыша в игре с вероятностной матрицей $\tilde{Q}$. Пусть имеется некоторая j-я реализация значений матрицы в состоянии s для (дискретного) набора их k возможных решений агента

$$\tilde{Q}_j\left(s, \vec{a}\right) = \left[\tilde{Q}\left(s, a_1\right), \tilde{Q}\left(s, a_2\right), \ldots, \tilde{Q}\left(s, a_K\right)\right]_j \tag{14}$$

Ожидаемым исходом j-й реализации игры для агента со смешанной стратегией $p(a)$ является величина

[10] Предпочтительнее использовать распределения, определяемые небольшим числом достаточных статистик, и допускающие эффективную генерацию выборок для метода Монте-Карло.

(15) $$\sum_k p(a_k)\cdot Q_j(s,a_k)$$

Наилучшим ходом в этой реализации игры будет

(16) $$a_j^* = \arg\max_k Q_j(s,a_k)$$

Наилучшей “жадной” стратегией в целом является (выборочный) закон распределения величин $\{a_j^*\}$, которые может оцениваться методом Монте-Карло, как статистика[11] большого числа выборок $\tilde{Q}_j$. Уравнение обучения с подкреплением в форме временных разностей имеет вид:

(17) $$\Delta\tilde{Q}(s,a) = \left[r(s,a) + \gamma\left\langle \max_{a'} \tilde{Q}_j(s',a')\right\rangle - \left\langle \tilde{Q}(s,a)\right\rangle\right]$$

(угловыми скобками обозначено усреднение по реализациям игры). Это уравнение является специальным видом рекуррентного уравнения Беллмана для поиска оптимума аддитивных функций. Колебания величины (17) определяют дисперсию плотности вероятности величины $\tilde{Q}$.

Завершающим шагом описания модели адаптации агента в многошаговой игре является вывод соотношений для построения нейросетевых аппроксимаций среднего и дисперсии распределения $\tilde{Q}$.

Аппроксимация плотности проводится путем максимизации функции правдоподобия (maximum likelihood). Выбор нейросетей как пробных функций продиктован их хорошими аппроксимационными свойствами и возможностью быстрого вычисления градиента по свободным весовым параметрам при решении задачи максимизации.

В этой работе использовалась эффективная нейронная сеть локальных нормализованных сплайнов (Connectionist Normalized Local Splines – CNLS), предложенная в начале 90-х годов специалистами Центра исследований нелинейности[12] в Лос-Аламосской национальной лаборатории [18,19]. Важной особенностью этой нейросети является не только возможность быстрого обучения, но и специфические свойства базисных функций, дающих близкую к кусочно-линейной аппроксимацию с автоматической гладкой “сшивкой” областей. Публикации по применениям этой сети немногочисленны, что, по-видимому, сдерживает ее широкое распространение[13]. Описание функционирования и обучения нейросети CNLS представляет собой самостоятельную тему, и поэтому вынесено в Приложение.

Искомые функциональные зависимости $m(s,a)$ и $\log\sigma(s,a)$ представляются двумя выходами нейронной сети, входами которой являются совокупность

[11] Напомним, что в задаче рассматривается конечное множество возможных решений a_k.
[12] Аббревиатура названия Центра – CNLS (Center for Non-Linear Studies)
[13] Этот факт уже несколько лет приводит автора в недоумение. Нейронная сеть этой архитектуры значительно превосходит и по скорости и по удобству использования популярные нейросети с обратным распространением ошибки (backpropagation). В лекции предоставился удобный случай подробно описать эту нейроархитектуру.

переменных состояния (их может быть несколько) и переменной, описывающей решение агента. Будем считать, что все переменные представлены вещественными числами из некоторой ограниченной области определения. Вместо максимизации функции правдоподобия удобнее минимизировать ее логарифм, имеющий для Гауссового распределения вид:

$$-\log L = \tfrac{1}{2}\log(2\pi) + \log(\sigma(s,a;w)) + \frac{(m(s,a;v) - F(s,a))^2}{2\sigma(s,a;w)^2} \tag{18}$$

где $F(s,a)$ – наблюдаемое значение аппроксимируемой величины, известное в традиционной задаче обучения нейросети с учителем, и *неизвестное* в рассматриваемом здесь случае обучения с подкреплением. На текущем шаге обучения разность значений (m-F) заменяется ее оценкой (17), являющейся разностной производной аппроксимируемой функции по времени.

Обе пробные функции зависят от наборов весовых коэффициентов нейросетевой модели (w и v). Обучение основывается на значении градиента[14] минимизируемого функционала по этим параметрам. Непосредственные вычисления дают:

$$\frac{\partial(-\log L)}{\partial v_i} = \frac{\Delta Q(s,a)}{\sigma(s,a;w)^2} \cdot \frac{\partial m(s,a;v)}{\partial v_i} \tag{19}$$

$$\frac{\partial(-\log L)}{\partial w_j} = \left[1 - \left(\frac{\Delta Q(s,a)}{\sigma(s,a;w)}\right)^2\right] \cdot \frac{\partial \log \sigma(s,a;w)}{\partial w_i} \tag{20}$$

Выражения (19-20) справедливы для дифференцируемых нейросетей любой архитектуры (без рекуррентных связей). Специфика конкретного типа нейросети проявляется в конкретном виде производных выходов по параметрам w и v.

Заметим, что большинство реализаций алгоритмов обучения игнорируют зависимость дисперсии плотности вероятности, даваемую соотношением (20), довольствуясь лишь аппроксимацией матожидания (19). При постоянной дисперсии ее значение не сказывается на направлении градиента и знаменатель в (19) полагают равным единице. В этом пределе формула (18) переходит в метод наименьших квадратов. Максимизация функции правдоподобия с аппроксимацией всех параметров выбранного распределения является более универсальным (и численно устойчивым) подходом. Соотношения вида (19-20) могут быть легко выведены и для других (отличных от Гауссового) законов распределения вероятности. Это становится важным, например, если фактическое распределение уклонений от матожидания является несимметричным, или имеет широкие крылья (что часто встречается в финансовых приложениях и при прогнозировании временных рядов). Большим преимуществом метода является одновременное

[14] При обучении целесообразно использовать более мощные методы оптимизации, чем прямой метод стохастического градиента. Литература по применениям методов оптимизации к обучению нейронных сетей весьма обширна, и читатель без труда найдет необходимые книги и статьи. Методы оптимизации реализованы в большинстве свободно распространяемых и коммерческих компьютерных программ, моделирующих нейронные сети.

построение и самой аппроксимации и оценки ее ошибки в каждой точке, определяемое функциональной зависимостью дисперсии от переменных задачи.

Итак, в этом разделе сформулирован вычислительный подход к нахождению оптимальных смешанных стратегий при принятии решений в многошаговых играх. Основным его содержанием является представление цены стратегии в текущем состоянии игры в параметрической вероятностной форме. Параметры распределения определяются уравнением Беллмана, решения которого находятся в виде нейросетевых аппроксимаций.

Завершающая часть лекции посвящена применению предложенного подхода к решению практической задачи теоретико-игрового моделирования и оптимизации бизнес-процессов на внутреннем рынке холдинга консалтинговых компаний.

Задача о деятельности холдинга консалтинговых компаний

Фундаментальной областью приложений теории игр является проблема коллективного поведения и взаимодействия игроков - участников совместной бизнес-деятельности. Проблема весьма многогранна, и в этой лекции мы ограничимся подробным рассмотрением конкретного примера из области организации деловой активности. В качестве иллюстрации выбрана деятельность холдинга компаний, специализирующихся на предоставлении консалтинговых услуг с нескольких смежных областях, относящихся к информационным технологиям. В этой главе предложна *модель* внутренней конкуренции и кооперации, основанная на аукционных механизмах выработки управляющих решений.

Основу модели составляют нейросетевые методы выбора стратегии рационального поведения для индивидуальной компании – участника холдинга. При этом игровые условия учитывают неполноту и *вероятностный* характер информации, доступной каждому игроку.

Модель бизнес-процессов

Несколько компаний оказывают консалтинговые услуги в области информационных технологий. Экспертные профили компаний близки по тематикам выполняемых проектов, и с целью снижения маркетинговых и других издержек компании объединяются в холдинг, управляемый головной компанией, не занятой в проектах непосредственно. Головная компания имеет фиксированную долю прибыли от каждого успешно проведенного проекта.

Имеется фиксированный набор тематических разделов, однако каждый проект может потенциально содержать задачи из разных разделов. Профиль проекта задается распределением представленных в нем тематик. Стоимость каждого проекта для внешних заказчиков фиксирована (и равна 1). Проект полностью выполняется одной из компаний холдинга.

Каждая компания имеет свой профиль распределения занятых в ней экспертов по тематикам. Этот профиль представлен в виде вероятностей успешного выполнения проекта по каждой из тематик в отдельности. Таким образом, экспертный уровень компании (вероятность успешного выполнения комплексного проекта) равна скалярному произведению профиля проекта и профиля компании. Предполагается, что компании имеют достаточно высокий экспертный уровень, однако их экспертизы по отдельным тематикам разнятся.

Каждая компания использует свой собственный набор методик и программного обеспечения, и фактические экспертные уровни остальных компаний ей

неизвестны. Неизвестными являются также и функции полезности других компаний, определяемые только их уровнем текущих доходов в расчете на один проект. Состояния счетов не публикуются.

При успешном выполнении проекта его финансовое обеспечение, за вычетом доли управляющей компании, целиком поступает на счета компаний. Если проект завершается неудачно, то холдинг несет убытки в размере полной стоимости проекта плюс дополнительные издержки, процент которых также фиксирован. Убытки не затрагивают управляющую компанию.

Компании холдинга, в целом, остаются независимыми, и преследуют, прежде всего, интересы своего собственного бизнеса. Для организации совместной деятельности в холдинге предложена схема деятельности, основанная на внутреннем аукционе. Рациональность этой схемы, а также оптимальная доля прибыли головной компании и являются предметом исследования.

Схема бизнес-процессов состоит в следующей последовательности этапов, составляющих один цикл (временной шаг).

1. На текущем шаге холдинг получает заказ на выполнение нового проекта, случайно выбираемого из фиксированного списка проектов. Профиль тематик проекта которого сообщается всем компаниям-участникам.

2. Каждая компания индивидуально вычисляет свой собственный экспертный уровень относительно данного проекта и владеет информацией о состоянии своего счета.

3. Компании делают одинаковый страховой взнос, суммарная по компаниям величина которого равна стоимости проекта и фиксированных издержек, которые возникнут, если выполнение проекта будет сорвано.

4. Каждая компания назначает свою, внутреннюю для холдинга, цену выполнения проекта (не превышающую стоимость проекта от заказчика). Заявленные цены передаются в управляющую компанию, и держатся в секрете от других участников.

5. Заявки компаний участвуют во внутреннем аукционе, при этом побеждает заявка с минимальной ценой.

6. Если цена победителя не превышает отпускную цену проекта за вычетом фиксированной доли управляющей компании, то компания – победитель аукциона получает возможность выполнить этот проект *по заявленной ею цене*. В противном случае холдинг просто отказывается от выполнения этого проекта, не получая на этом шаге ни доходов, ни убытков. Страховые взносы полностью возвращаются компаниям.

7. С вероятностью, равной своему экспертному уровню относительно данного проекта, компания-победитель аукциона успешно выполняет проект. В этом случае она получает доход, равный заявленной ею аукционной цене, управляющая компания получает фиксированную долю, а остаток средств равномерно распределяется по остальным компаниям холдинга. Страховые взносы полностью возвращаются компаниям. Таким образом, все заинтересованы в успешном выполнении проекта.

8. Если же проект окажется неуспешным, то компании теряют свои страховые взносы, управляющая компания ничего не получает.

После этапа 8 происходит новый цикл деятельности.

Таким образом, компании поровну делят риски совместной деятельности, а также частично и доходы от нее. Управляющая компания, обеспечивающая приток новых проектов, потенциально имеет единственный рычаг долгосрочного управления – ставку своей прибыли. В течение моделируемого периода времени эта ставка остается постоянной.

Каждая компания также имеет одну степень свободы – внутреннюю цену каждого проекта, участвующую в аукционе. Прямой путь повышения собственных доходов – увеличение этой цены, однако, высокие заявки имеют малый шанс победить в аукционе. При низких же ценах компании самой не выгодно выполнять проект – пусть это сделает кто-то другой, а она получит больше при дележе остатка денег. С другой стороны, передача проекта в компанию, имеющую в его тематиках невысокий экспертный уровень, также невыгодна – возрастает риск неудачи проекта. Потери равномерно распределяются между участниками, но это не влияет на выбор компаний, т.к. они не преследуют цель относительной успешности на рынке, важен лишь собственный уровень доходов в расчете на один заказанный холдингу проект. Высокие цены на проект со стороны всех компаний приведут к тому, что проект вообще не будет выполняться (но время упущено, и это учитывается в подсчете числа заказанных проектов).

Целью каждой компании является нахождение стратегии выбора цен в аукционах, максимизирующей суммарный (дисконтированный) доход.

Формализация математической модели многошаговой игры

Эта схема, с математической точки зрения, представляет собой многошаговую игру с несколькими игроками, обладающими неполной информацией. Выигрыши других игроков, их функции полезности и матрицы игры данному игроку неизвестны. Фактически, ему известна только тематика серии проектов, состояние собственного счета и значение его выигрыша (с учетом страхового взноса) на каждом шаге.

Принцип моделирования такой игры был предложен в предыдущем разделе лекции. Для его детализации введем обозначения:

$j(t)$	Номер состояния игры на шаге t. Состояние игры задается типом (распределением тематик) нового проекта. Число типов конечно.
n	Число игроков
e_{qj}	Экспертный уровень компании q в тематиках проекта j. Равен вероятности успешного выполнения проекта этой компанией. Экспертный уровень вычисляется скалярным произведением профиля компании и профиля проекта. Каждой компании известен только ее экспертный уровень.
$u_q(t) = U_q(t)/t$	Относительный доход компании в расчете на один проект (U – текущее значение счета компании)
$h(t) = H(t)/t$	Относительный доход управляющей компании холдинга

$s_q(t) = \lfloor u_q(t), e_q(t) = e_{qj(t)} \rfloor$	Совокупный вектор состояния агента-комнании на шаге *t*
$a_q(t)$	Решение агента (фактическая заявленная ставка на аукционе) на шаге *t*
$p_q(t)$	Смешанная стратегия (распределение аукционной ставки) компании на шаге *q*. Определяется в виде кусочно-постоянной функции на заданном конечном разбиении диапазона возможных ставок [0..1]
$\tilde{Q}_q^p(s,a)$	*Q*-фактор агента при стратегии *p* (случайная функция с локально нормальным распределением значений при каждом выборе аргументов). Математическое ожидание *Q* равено суммарному дисконтированному выигрышу агента при начале игры из состояния *s*, в котором агент принимает решение *a* и далее следует выбранной стратегии *p*
$m_q^p(s,a), \sigma_q^p(s,a)$	Среднее и дисперсия распределения значений *Q*-фактора.
$\tilde{Q}_q^*(s,a)$	*Q*-фактор при оптимальной стратегии p^*
γ	Фактор дисконтирования
ξ	Фиксированная доля дохода управляющей компании в каждом успешном проекте
η	Фиксированный процент дополнительных издержек (сверх суммы проекта) при его неудаче
$b(t)$	Значение заявки, выигравшей аукцион (равно минимальной заявке, сделанной участниками).
$r_q(t)$	Фактический выигрыш агента на данном шаге игры (может быть отрицательным при неудачном проекте)

Каждый агент после получения информации о своем выигрыше на шаге выполняет итерацию обучения своего *Q*-фактора цен состояний.

Прежде чем описать алгоритм математической модели игры, отметим, что в условиях задачи одна или несколько фирм могут разориться (получить отрицательное значение счета) на ранних шагах игры. Фактически, счета фирм испытывают несимметричные случайные блуждания[15] с вероятностью шага вверх, равной экспертному уровню *e* фирмы в проекте, и соотвествующей вероятности

[15] В действительности, это “управляемые” блуждания, поэтому точное решение задачи о вероятности нулевого счета затруднено.

шага вниз (1-*e*). Эта задача сводится к известной вероятностой задаче о раззорении, рассмотренной подробно в курсах теории вероятностей (например, в фундаментальном труде Феллера). Экспертные уровни предполагаются достаточно высокими, и в контексте данной лекции задача о раззорении не рассматривается. Игровые траектории, приводящие к раззорению хотябы одной фирмы, просто исключаются из общей статистики игр.

Рассматриваемая многошаговая игра проводится по следующему алгоритму:

цикл по раундам игр
t := 0
инициализация: $\vec{u} := 0; h := 0; \tilde{Q} := \tilde{Q}_0;$
цикл по шагам игры
  $t := t + 1;$
  разыграть тематику нового проекта $j(t)$;
  для каждого игрока q
    определить состояние $s_q := \lfloor u_q, e_q \rfloor;$
    найти ε-"жадную" стратегию $p_q(t)$;
    разыграть ставку $a_q(t)$ *из распределения* $p_q(t)$;
  провести аукцион $b(t) := \min_q \{a_q\};$
  провести розыгрыш успешности проекта для победителя;
  если (проект успешен)
    обновить счет управляющей компании $H := H + \xi$;
    для каждого игрока q
      найти выигрыш $r_q(t)$; $U_q := U_q + r_q$
  иначе
    для каждого игрока q
      вычесть страховой взнос $U_q := U_q - (1+\eta)/n$;
  для каждого игрока q
    провести цикл обучения $\tilde{Q}_q$;
  накопить статистику шага игры;
накопить статистику раунда игры;
стоп;

Индивидуальные стратегии игроков

Прежде чем приступать к моделированию, обсудим особенности решений и равновесий, которые могут возникать в данной модели, в случае, когда игроки *обладают информацией об экспертном уровне друг друга*[16]. Для числовых оценок примем конкретный набор параметров игры, этот же набор параметров будет в дальнейшем использован в компьютерном моделировании.

$$n = 3; \quad 0.7 < e < 1; \quad \gamma = 0.99; \quad \xi = 0.1; \quad \eta = 0.1$$

[16] В противном случае достижение равновесия рациональным логическим путем не возможно.

Рассмотрим зависимости математического ожидания выигрыша игрока от сложившейся аукционной цены, для двух случаев: когда победителем аукциона является этот игрок, и когда проект выполнялся каким-то другим игроком.

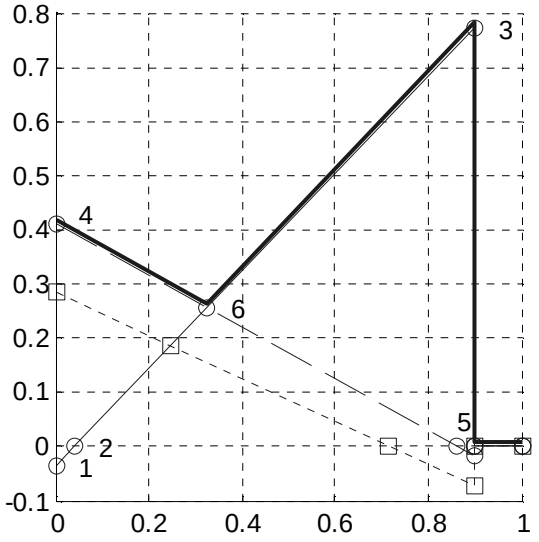


Рис. 1. Зависимость ожидаемого выигрыша игрока от результирующей цены на аукционе.

Сплошная тонкая линия – выигрыш, если победила ставка этого игрока;
Прерывистая линия – выигрыш, если победил игрок, экспертный опыт которого выше, чем у данного игрока;
Тонкий пунктир – выигрыш, если победил игрок с меньшим экспертным опытом;
Сплошная жирная линия – наиболее предпочтительный сценарий для данного игрока.

Отмечены точки выигрышей при нулевой цене (1,4), точки безубыточности (2,5), максимально возможный выигрыш (3) и точка “безразличия” (6).

Важной точкой на кривой выигрышей является точка безразличия (6), в кторой доход не зависит от того, кто фактически победил в аукционе – данный игрок или какой-то другой игрок. На кривой выигрыша имеется (n-1) таких точек – для каждого из оппонентов. Чем сильнее оппонент, тем правее расположена эта точка (напомним, что у сильных оппонентов выше вероятность успешного завершения проекта).

Соотношения для точек координат точек легко получить прямо из условия задачи. Математические ожидания складываются и вероятности успеха проекта с соответствующим дележом выигрышей, и вероятности его провала, который штрафуется.

Для справки, точка “безразличия” (№6) некоторого игрока q по отношению к игроку p имеет абсциссу:

$$\frac{\left(e_p - e_q\right)\cdot\left(1+\eta\right)\cdot\frac{n-1}{n} + e_p\cdot\left(1-\xi\right)}{e_q\cdot\left(n-1\right)+e_p} \qquad (21)$$

Особенность игры состоит в том, что игрок-победитель может влиять на цену аукциона только понижая свою ставку – что, конечно, выгодно всем другим игрокам. *Если игроки обладает полной информацией* об экспертном опыте оппонентов, то рациональным исходом аукциона в чистых стратегиях будет цена безразличия между самым сильным и самым слабым игроком *на диаграмме победителя* (диаграммы разных игроков отличаются наклонами линий, пропорциональными их экспертному опыту). Все, кроме победителя, заинтересованы в понижении этой цены, победитель же предпочет бы ее повысить (но остаться при этом победителем! – это и сдерживает повышение).

Заметим, что эта точка находится ниже “всеобщей средней” цены. Если такая ситуация имеет место для всех типов выполняемых проектов, то чистое равновесное состояние в каждом отдельном шаге игры становится невыгодным самому сильному игроку – он выполняет *все* проекты, но имеет доход ниже среднего. Поэтому он будет вынужден искать смешанную стратегию, в которой время от времени побеждает второй по силе игрок. Низкие защитные заявки слабых игроков в этом случае будут приводить к их победам на аукционе, но достигаемый при этом выигрыш ниже, чем при уступке проекта более сильным игрокам. Таким образом, в многошаговой игре равновесие Нэна в чистых стратегиях для одной стратегической игры (играемой на каждом шаге) неустойчиво.

Таким образом, картина многошаговой игры достаточно сложна даже в случае полной информации. Нас же будет интересовать практически важный случай, когда игроки не обладают полной информацией (в частности, не имеет значений экспертного опыта оппонентов). Рассмотрение такой игры может быть проведено, по-видимому, только в компьютерной модели.

Компьютерная система COGITO

Описанный алгоритм проведения игры реализован в компьютерной системе COGITO. Система позволяет описывать переходы между состояниями многошаговой игры, проводить аукционы, вычислять выигрыши игроков. Модель игрока основывается на предложенном нейросетевом алгоритме аппроксимации плотности распределения его цены состояния, она позволяет проводить обучение нейросети с подкреплением и оценивать текущую оптимальную смешанную стратегию.

Система COGITO позволяет изменять механизмы аукционов (например, в коммерческих аукционах в качестве цены часто выбирается цена *второй* заявки, следующей за ценой победителя), а также использовать ситемы распределения, отличные от аукционов (например, комитетные решения). В COGITO может также “дозироваться” информация, сообщаемая игрокам, например, могут публиковаться выигрыши и уровни экспертизы всех других игроков, или их части.

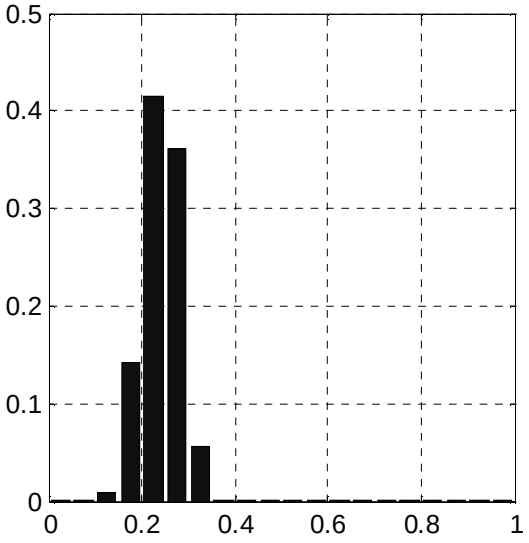
Статистическая информация, собираемая в системе, включает все аспекты каждого шага игры (ставки игроков, цену аукциона, выигрыши и текущие счета игроков).

Важно также, что на аукционные предпочтения игроков влияет ставка ξ прибыли управляющей компании. При высоких ставках прибыли в выполнени проектов включаются более слабые компании, это приводит к понижению вероятности успешности проектов, и, как следствие, к снижению общих доходов управляющей компании. Модель игры позволяет определить *оптимальные значения норм прибыли*, а также выяснить, как они зависят от процента издержек при неудаче в проекте. Таким образом, компьютерная игровая ситема позволяет провести оптимизацию деятельности управляющей компании.

Выборочные результаты моделирования

Ситема COGITO позволяет получать решения игр и смешанные стратегии игроков в различных вариантах многошаговых игр. С практической стороны интересны возможные типы получаемых результатов. Некоторые из них проиллюстрированы ниже на рисунках. В приведенных иллюстрациях с целью экономии времени вычислений игра проводилась для случая трех игроков (не считая управляющей компании, параметры которой предполагались постоянными) и двух типов проектов по двум возможным тематикам с профилями [0,1] и [1,0]. Профили игроков были устроены таким образом, что два из них были безусловными экспертными лидерами (каждый по своему направлению), а третий игрок имел

высокие, но уступающие лидерам, экспертные уровни в каждом из проектов. Выбраны $e_1 = [0.8, 0.8]$, $e_2 = [0.95, 0.75]$, $e_3 = [0.71, 0.99]$.



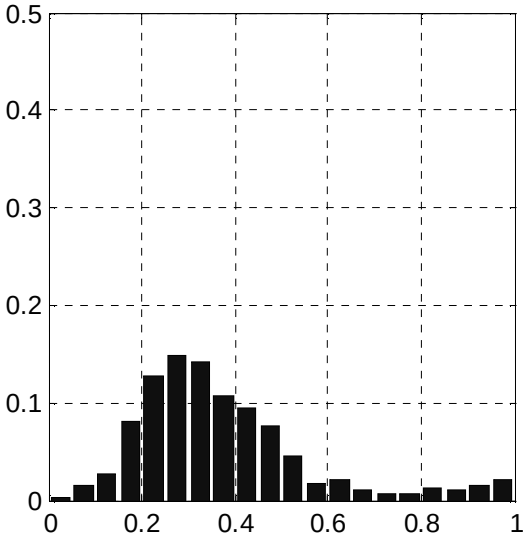


Рис. 2. Смешанные стратегии сильного третьего (верхний рисунок) и “слабого” первого игроков в отношении одного из проектов на поздних стадиях многошаговой игры. Слабый игрок с некоторой вероятностью делает заявки в области точки безразличия сильного игрока, цены слабого игрока в целом выше – при риске стать победителем он предпочитает более высокие ставки. Совсем низкие ставки (с вероятностью около 0.18) у высокого игрока позволяют ему иногда назначать и относительно высокую цену. Вероятность более высокой цены выше вероятности низкой.

Первый интересный вопрос состоит в оценке относительного успеха первого и второго игроков. Причина – суммарная экспертиза первого игрока равна 1.6, что меньше значения 1.7 у второго и третьего участника. Однако первый не имеет очевидных “слабостей”. Что в данных условиях лучше?

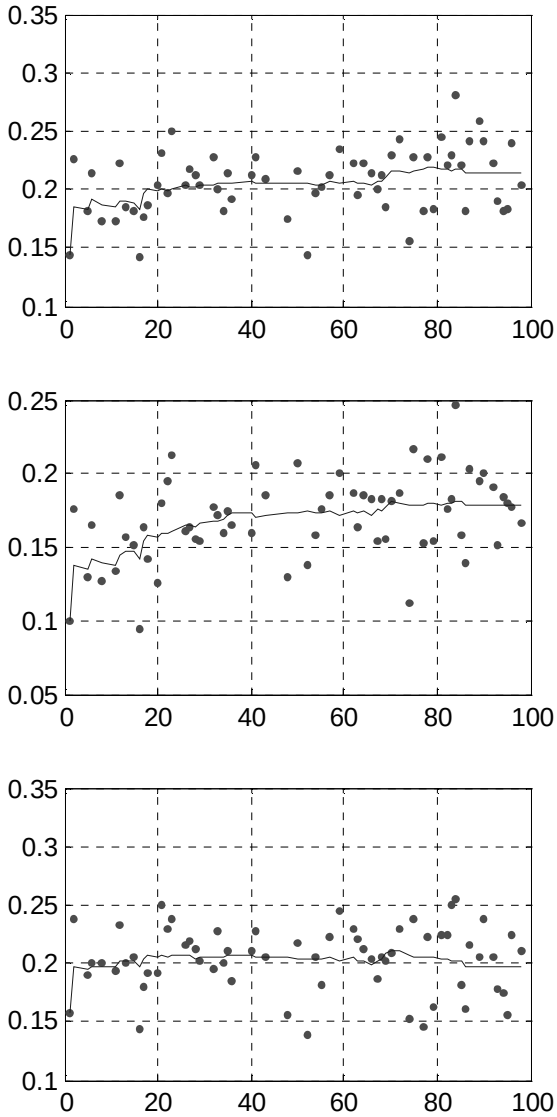
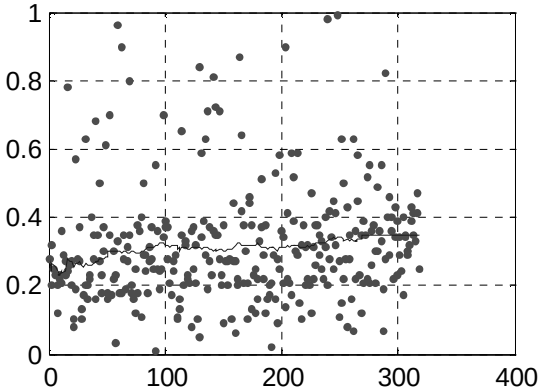


Рис. 3. Динамика индексов доходов (в расчете на один проект) трех игроков в игровых сериях, по 100 шагов каждая.

Результаты являются до определенной степени неожиданными – первый игрок, не являясь безусловным лидером ни в одном из проектов, находит более успешные стратегии, чем каждый из лидеров. Третий игрок успешнее второго (при равенстве суммарных экспертных уровней). Поскольку вероятности проектов обоих типов одинаковы, то это означает, что переход уровня от 0.95 к 0.99 в одном из проектов компенсирует потери 0.75 : 0.71 в другом.

Интересно также посмотреть, как распределения цен игроков сходятся во времени.



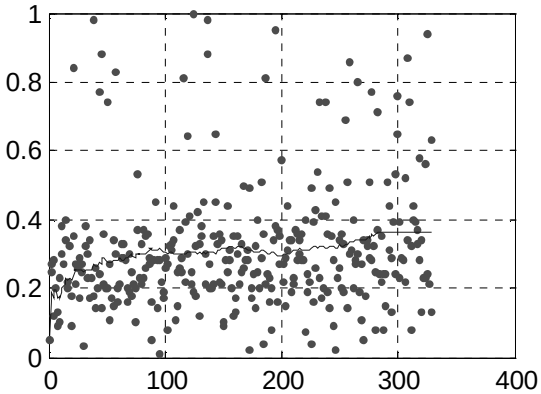


Рис. 4. Динамика аукционных ставок первого игрока. Показана каждая 10-я ставка и скользящее среднее по 50 ставкам.

Видно, что после достаточно долгого равновесия в окрестности цены 0.31 первый игрок начал плавно повышать цены до значений 0.35, одновременно увеличив дисперсию своей смешанной стратегии, что привело к победе над 3-м игроком.

И в заключение этого параграфа приведем вероятности выигрыша каждым из игроков в каждом из типов проектов в течение всего набора игр.

	Проект 1 [1,0]	Проект 2 [0,1]
Игрок 1	0.24	0.25
Игрок 2	0.44	0.46
Игрок 3	0.32	0.29

Таб 2.Вероятности победы в аукционе для различных игроков

Играющий чаще – проигрывает!

Обсуждение

Первое наблюдение состоит в том, в условиях крайне ограниченной информации фирмы-участницы находят способ рационального позиционирования на внутреннем рынке. Равновесные стратегии имеют интерпретируемые распределения. Смешанные стратегии сильных и слабых игроков различаются.

Важный методологический момент – алгоритмы обучения с подкреплением, обобщенные на случай стохастических многошаговых игр, оказываются достаточно эффективными, сходимость достигается за несколько тысяч шагов одиночных игр, что является типичным значением для методик онлайн-обучения.

С прикладной точки зрения, имитационная модель игры COGITO позволяет изучать важные вопросы, касающиеся оптимизации бизнеса:

- Как влияет на доходы управляющей компании изменение ставки ее прибыли?
- Как при этом перераспределяются доходы отдельных компаний?
- Выгодно ли расширять холдинг путем включения компаний с относительно невысоким экспертным уровнем? Каковы пороговые значения для этого уровня?
- Выгодно ли одной компании повышать экспертный уровень другой компании (путем передачи технологий, обучения специалистов и пр.)?
- Каким должен быть экспертный профиль новой компании, чтобы обеспечить максимальную полезность холдингу без дестабилизации его работы?
- Насколько выгодно иметь экспертный уровень 0.999… в одной из тематик?

Список приложений далеко не исчерпан этим перечнем.

Послесловие

Сегодня теория игр – это, в основном, экономическая теория.

Когда говорят о ее (всего лишь) полувековой истории, то под этим понимается возраст *математической* теории, но не самой игровой проблематики. Еще Платон, обращаясь к Сократ, обсуждал проблему о рациональном поведении дозорного на посту перед предстоящем боем. Если дозорный уверен в победе своего войска, то полезность его *личного* участия в бою исчезающе мала, а так как он может пострадать, то рациональнее будет покинуть пост до боя. Если же его наблюдения свидетельствуют о том, что противник сильнее, то тогда тем более разумно сбежать, т.к. это ничего не изменит, но он не пострадает. Этой же рациональной логике должны следовать и все другие воины. Но тогда поражение неизбежно!

На протяжении всей истории возникали проблемы подобного рода, но лишь начиная с работ Дж. фон-Неймана в 30-40-х годах эти вопросы получили обрекли продуктивную математическую форму.

Фактически, теория игр пионерских времен фон-Неймана, в ее математическом содержании, представляла собой совокупность отдельных моделей, описывающих изученные классы игр. Наиболее фундаментальным ее аспектом является теория индивидуальной полезности.

Полезность достаточно просто и корректно определяется для пары альтернатив, в смысле предпочтения одной альтернативы *в сравнении* с другой, но для полной теории ситуация несколько сложнее, поскольку имея только сравнительный порядок довольно не просто выбрать универсальные метрические шкалы, позволяющие сравнивать полезности разных игроков, и тем самым, перейти к математическому понятию объективной полезности.

В теории игр любая информация об игре может иметь значение. Простой пример – должны ли игроки делать ходы в раутах повторяющейся игры одновременно или по очереди? Анализ равновесий Нэша в стратегических играх типа “дилеммы заключенного” показывает, что порядок ходов не влияет на равновесие. Однако вообразите себе (многошаговую игру) шахматы, в которых игроки делают ходы *одновременно* (сообщая их в конвертах в игре; при невозможных парах ходов они игнорируются; троекратный повтор невозможных комбинаций – ничья). Стратегии в ней *кардинально* будут отличаться от обычных шахмат.

Тотальная индуктивная рациональность игроков может приводить к логическим парадоксам, в которых следствия из рационального поведения приводят к заключению об иррациональности игрока (см. примеры в [20]). Для выхода из таких ситуаций требуется отказ от механически точного исполнения рациональных решений (“дрогнувшая рука” Р. Зельтена – в ходах игроков должны допускаться ошибки).

В данной лекции обсуждались, в основном, механизмы обучения при выработке оптимальных стратегий. Другой взгляд на это проблему - эволюционная теория игр. Это направление также интенсивно развивается (см. [8] и цитированную там литературу).

В завершении нужно отметить, что практические применения теории игр широко опираются на вычислительное моделирование. Компьютерные модели, такие как разработанная автором система COGITO, способны учитывать особенности бизнес-процессов, прогнозировать и, что особенно ценно, оптимизировать деловую активность в практических условиях множества одновременно действующих участников.

Фактически, система COGITO является автоматизированной торговой системой, поскольку вырабатываемые решения являются рекомендациями по ценам сделок в рыночных условиях. В контексте этой лекции рынок формируется всеми участниками, каждый их которых своими действиями влияет на складывающуюся цену. Однако используемый аукционный алгоритм механизм может быть обобщен на случай, когда цены определяются небольшим числом крупных агентов, а остальные агенты, по-существу, выступают в роли трейдеров.

Благодарности

Автор выражает глубокую благодарность Ю.В.Тюменцеву, Н.Г.Макаренко, С.А. Шумскому за доброжелательную критику.

Литература

1. М. Интрилигатор. **Математические методы оптимизации и экономическая теория**. М. Айрис-пресс, 2002. 576 с.

2. Майкл Хэрри, Ричард Шредер. **6 SIGMA**. М. Эксмо, 2003. 464 с.

3. Дж. фон Нейман, О. Моргенштерн. **Теория игр и экономическое поведение**. М. Наука, 1970. 707 с.

4. **The Essential John Nash**. ed by Harold W. Kuhn and Sylvia Nasar. Princeton Univ. Press, 2002. (244 pages)

5. Джон Харшаньи, Рейнхарт Зельтен. **Общая теория выбора равновесия в играх.** СПб: Экономическая школа, 2001. 498 с.

6. Р. Айзекс. **Дифференциальные игры**. М. Мир, 1967. 479 с.

7. Martin J. Osborne. **An Introduction to Game Theory**. (in press, OUP). Selected chapters from URL: http://www.chass.utoronto.ca/~osborne/igt/index.html

8. Philip Hingston, Graham Kendall. **Learning versus Evolution in Iterated Prisoner's Dilemma**. URL: http://www.cs.nott.ac.uk/~gxk/papers/cec2004ph.pdf

9. М.В. Губко, Д.А. Новиков. **Теория игр в управлении организационными системами**. М. Синтег, 2002. 139 с.

10. **Экономико-математическое моделирование**. Под ред. И.Н. Дрогобыцкого. М. Экзамен, 2004. 798 с.

11. Лео Брейман. **Задача о правилах остановки**. В кн. Прикладная комбинаторная математика. Сборник статей под ред. Э.Беккенбаха, М. Мир, 1968. 362 с.

12. С. А. Терехов. **Нейродинамическое программирование автономных агентов**. Лекция для Школы-семинара "Современные проблемы нейроинформатики". М. МИФИ, 2004. Часть 2., с.111-139.

13. Junling Hu and Michael P. Wellman**. Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm**.
URL: http://ai.eecs.umich.edu/people/wellman/index.html

14. Gerald Tesauro. **Extending Q-Learning to General Adaptive Multi-Agent Systems**. Advances in Neural Information Processing Systems 16 (NIPS'2003). MIT Press, Cambridge, MA, 2004.
URL: http://books.nips.cc/papers/files/nips16/NIPS2003_CN16.pdf

15. Richard Sutton and Andrew Barto. **Reinforcement Learning: An Introduction**. MIT Press, 1998.

16. URL: http://plato.stanford.edu/entries/prisoner-dilemma/

17. Yoav Shoham, Rob Powers, Trond Grenager. **Multi-Agent Reinforcement Learning: a critical survey.** Stanford Univ., 2003. URL: http://robotics.stanford.edu/~shoham/www%20papers/MALearning_ACriticalSurvey_2003_0516.pdf

18. W. C. Mead, S. K. Brown, R. D. Jones, P. S. Bowling, and C. W. Barnes. **Optimization and Control of a Small-Angle Negative Ion Source Using an On-Line Adaptive Controller Based on the Connectionist Normalized Local Spline Neural Network**. Nuclear Instruments and Methods Vol. A352, 309 (1994).

19. R. D. Jones, Y. C. Lee, S. Qian, C. W. Barnes, K. R. Bisset, G. M. Bruce, G. W. Flake, K. Lee, L. A. Lee, W. C. Mead, M. K. O'Rourke, I. Poli, and L. E. Thode

Nonlinear Adaptive Networks: A Little Theory, A Few Applications., in Proceedings of the First Los Alamos Conference on Cognitive Modeling in System Control, June 10-14, 1990, Santa Fe, NM.

20. Don Ross. **Game Theory**. In Stanford Encyclopedia of Philosophy. URL: http://plato.stanford.edu/entries/game-theory/

Приложение. Нейросеть CNLS (Connectionist Normalized Local Splines)

Статистическая модель распределения значений многомерной случайной функции основывается на параметрическом представлении ее плотности. Моменты распределения являются вещественными многомерными функциями, и для их эффективных аппроксимаций по конечной выборке применяются нейросетевые методы.

Модель локальных нейросетевых сплайнов основана на следующих построениях. Выберем систему нормируемых функций радиального базиса

(П1) $$\varphi_k(\vec{r}) = \frac{1}{\|\varphi\|}\varphi\left(\|\vec{r} - \vec{c}_k\|^2\right)$$

таких, что их сумма нигде в области аппроксимации не обращается в ноль, и каждая из функций локализована в окрестности своего максимума $\vec{r} = \vec{c}_k$. Простейший пример – функции Гаусса.

Для гладкой (почти везде) функции *f* запишем тождества:

(П2) $$f(\vec{r}) \equiv \frac{f(\vec{r})\sum_k \varphi_k(\vec{r})}{\sum_j \varphi_j(\vec{r})} \equiv \frac{\sum_k f(\vec{r})\cdot\varphi_k(\vec{r})}{\sum_j \varphi_j(\vec{r})}$$

Далее перейдем к приближениям. Поскольку каждый член суммы в числителе (П2) представляет собой произведение гладкой функции *f* и локализованной базисной функции, то *f* может быть заменена ее отрезком ряда Тейлора. Если сохранить только члены вплоть до линейны, то получим представление

(П3) $$f(\vec{r}) \approx \frac{\sum_k \left(a_k + \vec{b}_k(\vec{r} - \vec{c}_k)\right)\cdot\varphi\left(\|\vec{r} - \vec{c}_k\|^2\right)}{\sum_k \varphi_k(\vec{r})}$$

В этом соотношении формально можно считать значения набора коэффициентов $\left\{a_k, \vec{b}_k, \vec{c}_k\right\}$ произвольными. В итоге мы имеем дело с аппроксимационной моделью

(П4) $$f(\vec{r}) \approx \sum_k \left(a_k + \vec{b}_k(\vec{r} - \vec{c}_k)\right)\cdot\Phi_k(\vec{r};\vec{c}_1..\vec{c}_K)$$

которая по аналогии с моделями радиальных базисных функций названа авторами [18-19] нейросетевой. “Рецептивные поля” базисных нейронов вследствие процедуры нормализации существенно отличаются от исходных базисных функций φ, по которым проводилось разложение:

(П5) $$\Phi_k = \frac{\varphi\left(\|\vec{r} - \vec{c}_k\|^2\right)}{\sum_j \varphi\left(\|\vec{r} - \vec{c}_j\|^2\right)}$$

В областях, где центры базисных функций распределены относительно разреженно, вблизи максимума φ_k лишь одна эта функция существенно отлична от нуля. В этом случае она доминирует в сумме в знаменателе, и, следовательно, значение нейронной функции $\mathbf{\Phi}_k$ близко к константе. В итоге, в области влияния этой функции общая аппроксимация, даваемая нейросетью, близка к линейной регрессии. Кусочно-линейные фрагменты автоматически гладко “сшиваются” при переходе к области соседней нейронной функции.

В областях высокой концентрации центров нейронная сеть CNLS фактически переходит в радиально базисную сеть RBF, поскольку из-за высокой степени перекрытия сумма в знаменателе близка к константе.

При использовании полученной аппроксимации нужно учитывать, что нейронные функции $\mathbf{\Phi}_k$, располагающиеся на периферии области, занятой данными из моделируемой выборки, имеют пределы на бесконечности, равные 1.

Выбор значений параметров нейросетевой модели проводится в традиционной постановке обучения с учителем. Для обучающей выборки пар $\{\vec{r}_s, y_s\}$ формулируется функция правдоподобия, которая максимзируется с использованием градиента по параметрам. Для Гауссового распределения ошибок аппроксимации формулы для градиента получены в основном тексте лекции (18)-(20). Для аппроксимации математического ожидания и дисперсии используется пара соотношений (П4), при этом коэффициенты $\vec{c}_k$, отвечающие положениям центров нейронных функций для простоты могут выбираться одинаковыми. В этом случае каждая из аппроксимируемых функций определяется своей парой (a, b). Приведем здесь явный вид производных нейросети:

(П6)

$$\begin{cases} \dfrac{\partial net(\vec{r})}{\partial a_k} = \dfrac{\varphi\left(\left\|\vec{r} - \vec{c}_k\right\|^2\right)}{\sum\limits_j \varphi\left(\left\|\vec{r} - \vec{c}_j\right\|^2\right)} \\ \dfrac{\partial net(\vec{r})}{\partial b_{kn}} = \dfrac{(r_n - c_{kn})\varphi\left(\left\|\vec{r} - \vec{c}_k\right\|^2\right)}{\sum\limits_j \varphi\left(\left\|\vec{r} - \vec{c}_j\right\|^2\right)} \\ \dfrac{\partial net(\vec{r})}{\partial c_{kn}} = -\dfrac{b_{kn} \cdot \varphi\left(\left\|\vec{r} - \vec{c}_k\right\|^2\right)}{\sum\limits_j \varphi\left(\left\|\vec{r} - \vec{c}_j\right\|^2\right)} + \dfrac{2(r_{kn} - c_{kn}) \cdot \varphi'\left(\left\|\vec{r} - \vec{c}_k\right\|^2\right)}{\sum\limits_j \varphi\left(\left\|\vec{r} - \vec{c}_j\right\|^2\right)}\left[net(\vec{r}) - \left(a_k + \vec{b}_k \cdot (\vec{r} - \vec{c}_k)\right)\right] \end{cases}$$

В приложениях для ускорения вычислений выбор положений $\vec{c}_k$ может проводиться внешним алгоритмом без использования производных (П6). В частности, могут использоваться случайные выборки из обучающих данных, центры кластеров карты Кохонена или *k-means*, и другие алгоритмы. Положение центров может также диктоваться требованиями к точности аппроксимации.

Выбираемые при автоматическом обучении центры нейронов группируются в областях больших градиентов аппроксимируемой функции, что, конечно, является несомненным достоинством этой нейронной сети.

В завершении приведем пример аппроксимации функции моделью CNLS зашумленных данных для случая одной переменной. Обучение сети проводилось алгоритмом *RProp*.

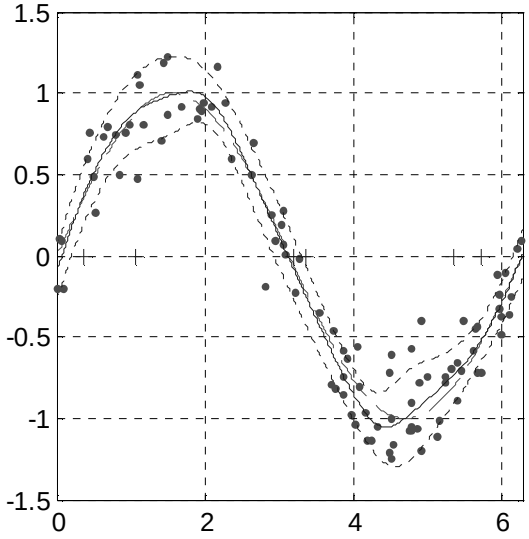


Рис. П1. Аппроксимация зашумленной синусоиды CNLS – сетью из 7 нейронов. Пунктирные линии соответствуют прогнозируемым нейронной сетью значениям дисперсии. Крестиками отмечены положения центров нейронов, выбранные алгоритмом обучения. Видно, что нейроны адаптивно располагаются в областях максимальных градиентов функции.

Задачи

Задача 1. Ошибка в формуле

По статистике в трети опубликованных в математических журналах статей со сложными выкладками, в последствии были найдены ошибки. Пусть читателю известно, что в формулах для обучения нейронной сети (18)-(20), (П6) этой лекции с вероятностью 0.3 содержится ошибка. Читатель, который знакомится с текстом с целью

А) общего любопытства

Б) выбора и дальнейшего совершенствования нейросетевых моделей

В) реализации нейросетевой модели в программе, которую предполагается продавать на рынке

сталкивается с дилеммой – {проверять, не проверять} выкладки. Сформулируйте проблему выбора в форме стратегической игры двух игроков (читатель, автор) в матричной форме для случаев А), Б) и В). Автор придерживается смешанной стратегии на своем множестве решений {ошибка, нет ошибки}, в которой вероятность ошибки p=0.3. Какова оптимальная стратегия читателя в трех вариантах? Является ли эта стратегия чистой или смешанной? Как изменится решение, если стратегия автора неизвестна?

Задача 2. Плотность распределения континуума оптимальных решений в игре со стохастической матрицей.

Алгоритм поиска “жадной” стратегии на шаге обучения с подкреплением (16)-(17) включает, как одну из возможностей, поиск плотности распределения значений аргументов, при которых стохастическая функция имеет максимум. Прямым путем для численной реализации такого поиска для случая конечного разбиения области определения (т.е. конечного дискретного набора возможных значений аргументов функции) является использование метода Монте-Карло.

Переход к более общему случаю континуума требует известной аккуратности при обращении с пределами. В этой связи предлагается следующая задача, являющаяся прелюдией к проблематике *бесконечных* игр.

Рассмотрим отрезок $[a,b]$, на котором задана пара непрерывных функций $m(x)$ и $\sigma(x)$, определяющих в каждой точке отрезка случайную величину с нормальным распределением $N(m(x),\sigma(x))$. Таким образом, на отрезке задана стохастическая функция (или случайный процесс, см. например, *Е.С. Вентцель, Л.А. Овчаров. Теория случайных процессов и ее инженерные приложения. М., Наука, 1991*).

Разобьем исходный отрезок на множество отрезков длины ε, на каждом из них будем считать функции матожидания и дисперсии постоянными. Реализация случайной функции включает результат розыгрыша н.р. случайной величины для каждого из отрезков ε-разбиения. Найдем максимум из полученного конечного набора чисел. Пусть значение аргумента, при котором реализуется максимум *в j-ой реализации* случайной функции равно x_j^*. Нас интересует плотность распределения величин x_j^* на отрезке $[a,b]$. Путем генерации новых реализаций исследуемой стохастической функции мы получаем выборочную оценку (гистограмму) $\rho_{J\varepsilon}(x)$

этой плотности, которая зависит от двух параметров – подробности разбиения ε и числа реализаций J.

Первый вопрос – корректен ли предельный переход $\rho_\varepsilon(x) \overset{?}{=} \lim\limits_{J\to\infty} \rho(x)$. Как будут устроены предельные функции (если они существуют)? Напомним, что нас интересует распределение значений *аргументов* максимумов, значения самих максимумов даже для нормального закона распределения в точке расходятся (убедитесь в этом).

Второй вопрос – при конечном числе реализаций J существует ли предел выборочной плотности при $\varepsilon \to 0$?

В какой метрике следует изучать эти пределы? Что можно сказать о задаче одновременного перехода к обоим пределам $J \to \infty, \varepsilon \to 0$?