МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«Ярославский государственный университет им. П.Г. Демидова»

Курсовая работа

Моделирование коррупции на основе игры в развернутой форме

02.04.01. Математика и компьютерные науки

Исполнитель: Сычев Р.С.

гр. МКН-21 МО

Руководитель: к.ф.-м.н. Глазков Д.В.

Ярославль 2022

Содержание

Введение

В настоящее время под явлением коррупции понимается неправомерное использование должностных привилегий в личных целях. Например, к коррупции относят присвоение ренты и получение взяток. При этом, некоторые из этих процессов могут быть описаны с помощью моделей теории игр. Одной из таких моделей является модель совместного преступления[1]. Такой подход подразумевает эндогенную природу игровых ситуаций и ассиметричные выплаты. Похожий вопрос, связанный с механизмом повторной проверки, в достаточно большом объеме изучен для однородного набора из нескольких проверяемых лиц[2]. В то же время, проверяемые лица могут иметь свою организационную структуру, наличие которой может вносить значительные коррективы в распределение информации на различных этапах игры. Например, за счет иерархии можно реализовать механизм дополнительных проверок[3]. Имеет смысл построение и анализ более обобщенных с точки зрения организационной структуры и распределения информации моделей.

Подобные модели могут быть представлены как игры в развернутой форме. При такой постановке задачи можно близким к естественному способом отразить в игровой форме структуру последовательного принятия решений набором участников в конфликтной ситуации.

Существуют методы позволяющие достаточно эффективно сформировать приближенное кореллированное равновесие для заданной игры в развернутой форме с неполной информацией. Довольно популярен итеративный алгоритм минимизации контрафактических сожалений (Conterfactual Regret Minimization)[4] и его модификация предусматривающая использование метода Монте-Карло (MCCFR)[5]. Данные алгоритмы появились не так давно, но на их основе уже получен ряд недостижимых до этого по сложности результатов.

Целью данной работы является изучение возможности применения данного алгоритма в задачах моделирования корррупции и его отработка на практике.

В соответствии с темой работы были поставлены следующие задачи:

— выбор конкретной формы модели с использованием уже существующихи улучшил предыдущие подобные модели;
— выбор алгоритма для поиска решения и анализа модели;
— проведение расчетов и анализ результатов.

Данная работа может быть интересна людям желающим ознакомится с некоторыми современными техниками решения игр с неполной информацией.

1 Первая глава. Описание алгоритмов

1.1 Игры в развернутой форме и равновесие

Игра в развернутой форме представляют компактную общую модель взаимодействий между агентами и явно отражает последовательный характер этих взаимодействий. Последовательность принятия решений игроками в такой постановке представлена деревом решения. При этом, листья дерева отождествлены с терминальными состояниями, в которых игра завершается и игроки получают выплаты. Любой нетерминальный узел дерева представляет точку принятия решения. Неполнота информации выражается в том, что различные узлы игрового дерева считаются неразличимыми для игрока. Совокупность всех попарно неразличимых состояний игры называется информационными состояниями. Приведем формальное определение.

Определение 1: Конечная игра в развернутой форме с неполной информацией содержит следующие компоненты:

— Конечное множество игроков N;

— Конечное множество историй действий игроков H, такое, что $\emptyset \in H$ и любой префикс элемента из H также принадлежит H. $Z \subseteq H$ представляет множество терминальных историй (множество историй игры на являющихся префиксом). $A(h) = \{a \colon (h, a) \in H\}$ — доступные после нетерминальной истории $h \in H$ действия;

— Функция $P \colon H \setminus Z \to N \cup \{c\}$, которая сопоставляет каждой нетерминальной истории $h \in H \setminus Z$ игрока, которому предстоит принять решение, либо игрока c представляющего случайное событие;

— Функция f_c, которая сопоставляет всем $h \in H$, для которых $P(h) = c$, вероятностное распределение $f_c(\cdot|h)$ на $A(h)$. $f_c(a|h)$ представляет вероятность выбора a после истории h;

— Для каждого игрока $i \in N$ $\mathcal{I}_i$ обозначает разбиение $\{h \in H \colon P(h) = i\}$, для которого $A(h) = A(h')$ всякий раз когда h и h' принадлежат одному члену разбиения. Для $I_i \in \mathcal{I}_i$ определим $A(I_i) = A(h)$ и $P(I_i) = i$ для всех $h \in I_i$. $\mathcal{I}_i$ называют информационным набором игрока i, а $I_i \in \mathcal{I}_i$ информационным состоянием игрока i;

— Для каждого игрока $i \in N$ определена функция выигрыша $u_i \colon Z \to R$. Если для игры в развернутой форме выполняется $\forall z \in Z \sum_{i \in N} U_i(z) = 0$, то такую игру называют игрой с нулевой суммой. Определим $\Delta_{u,i} = max_{z \in Z}\ u_i(z) - min_{z \in Z}\ u_i(z)$ для диапазона выплат игрока.

Отметим, что информационные наборы могут использоваться не только для реализации правил конкретной игры, но и могут быть использованы для того, чтобы заставить игрока забыть о предыдущих действиях. Игры в которых игроки не забывают о действиях называют играми с полной памятью. В дальнейшем мы будем рассматривать конечные игры в развернутой форме с полной памятью.

Стратегия игрока i — это функция σ_i, которая ставит в соответствие каждому информационному состоянию $I_i \in \mathcal{I}_i$ вероятностное распределение на $A(I_i)$. Обозначим за Σ_i множество всех стратегий игрока i. Стратегический профиль σ содержит стратегии для каждого игрока $i \in N$. При этом за σ_{-i} обозначим σ без σ_i.

Обозначим за $\pi^\sigma(h)$ вероятность того, что игроки достигнут h руководствуясь σ. Мы можем представить π^σ как $\pi^\sigma = \prod_{i \in N \cup \{c\}} \pi_i^\sigma(h)$, выделяя вклад каждого игрока. В таком случае, $\pi_i^\sigma(h)$ обозначает вероятность принятия совокупности решений игрока i, ведущих от $\emptyset$ к h. Иными словами

$$\pi_i^\sigma(h) = \begin{cases} \prod_{h \sqsubset h' \land P(h') = i \land h \sqsubset (h', a)} \sigma(h')(a) & \{h' | h \sqsubset h' \land P(h') = i\} \neq \emptyset \\ 1 & \text{иначе.} \end{cases}$$

Запись $h \sqsubset h'$ означает, что h' является префиксом h. Обозначим за $\pi_{-i}^\sigma(h)$ вероятность достижения истории h всеми игроками (включая c) за исключением i. Для $I \subseteq H$ определим $\pi^\sigma(I) = \sum_{h \in I} \pi^\sigma(h)$. Аналогично, введем $\pi_i^\sigma(I)$ и $\pi_{-i}^\sigma(I)$.

Ожидаемое значение выплаты для игрока i обозначим как $u_i(\sigma) = \sum_{h \in Z} u_i(h) \pi^\sigma(h)$.

Традиционным способом решения игр в развернутой форме для двух игроков является поиск равновесного профиля стратегий σ, который удовлетворяет следующему условию

$$u_1(\sigma) \geqslant \max_{\sigma_1' \in \Sigma_1} u_1(\sigma_1', \sigma_2) \qquad u_2(\sigma) \geqslant \max_{\sigma_2' \in \Sigma_2} u_2(\sigma_1, \sigma_2'). \tag{1.1}$$

Такой стратегический профиль называют равновесием по Нэшу. В случае, если стратегический профили σ удовлетворяет условию

$$u_1(\sigma) + \epsilon \geqslant \max_{\sigma_1' \in \Sigma_1} u_1(\sigma_1', \sigma_2) \qquad u_2(\sigma) + \epsilon \geqslant \max_{\sigma_2' \in \Sigma_2} u_2(\sigma_1, \sigma_2') \tag{1.2}$$

его называют ϵ – равновесием.

Для рассматриваемых далее алгоритмов наиболее интересен вариант игры с нулевой семмой для двух игроков. Именно для него имеется строгое математическое обоснование сходимости к равновесию нэша.

1.2 Контрафактические сожаления и их минимизация

Минимизация сожалений является популярным концептом, для построения итеративных алгоритмов приближенного решения игр в развернутой форме [6]. Приведем связанные с ней определения. Рассмотрим дискретный отрезок времени T включающий T раундов от 1 до T. Обозначим за σ_i^t стратегию игрока i в раунде t.

Определение 1 *Средним общим сожалением игрока i на момент времени T называют величину*

$$R_i^T = \frac{1}{T}\max_{\sigma_i^* \in \Sigma_i} \sum_{t=1}^{T} (u_i(\sigma_i^*, \sigma_{-i}^t) - u_i(\sigma^t)) \tag{1.3}$$

В дополнении к этому, определим $\bar{\sigma}_i^T$ как среднюю стратегию относительно всех раундов от 1 до T. Таким образом для каждого $I \in \mathcal{I}_i$ и $a \in A(I)$ определим

$$\bar{\sigma}_i^T(I) = \frac{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)\sigma^t(I)(a)}{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)}. \tag{1.4}$$

Теорема 1 *Если на момент времени T средние общие сожаления игроков меньше ϵ, то σ является 2ϵ равновесием.*

Говорят, что алгоритм выбора σ реализует минимизацию сожалений, если средние общие сожаления игроков стремятся к нулю при t стремящимся к бесконечности. И как результат, алгоритм минимизации сожалений может быть использован для нахождения приближенного равновесия по Нэшу, в случае игр двух игроков с нулевой суммой. Однако, стратегии сформированные для игр с большим числом игроков могут также успешно применяться на практике [7].

Понятие контрафактического сожаления служит для декомпозиции среднего общего сожаления в набор дополнительных сожалений, которые могут быть минимизированы независимо, для каждого информационного состояния.

Обозначим через $u_i(\sigma, h)$ цену игры с точки зрения истории h, при условии, что h была достигнута, и игроки спользуют в дальнейшем σ.

Определение 2 *Контрафактической ценой $u_i(\sigma, I)$ назовем ожидаемую цену, при условии, что информационное состояние I было достигнуто, когда все игроки кроме i играли в соответствии с σ. Формально*

$$u_i(\sigma, I) = \frac{\sum_{h \in I, h' \in Z} \pi_{-i}^{\sigma}(h)\pi^{\sigma}(h, h')u_i(h')}{\pi_{-i}^{\sigma}(I)}, \tag{1.5}$$

где $\pi^{\sigma}(h, h')$ — вероятность перехода из h в h'.

Обозначим за $\sigma^t|_{I \to a}$ стратегический профиль идентичный σ за исключением того, что i всегда выбирает a в I.

Немедленным контрафактическим сожалением назовем

$$R_{i,imm}^T = \frac{1}{T}\max_{a \in A(I)} \sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)(u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I)). \tag{1.6}$$

Интуитивно это выражение можно понимать, как аналог среднего общего сожаления в терминах контрафактической цены. Однако вместо рассмотрения всевозможных максимизирующих стратегий рассматриваются локальные модификации стратегии. Положим $R_{i,imm}^{T,+}(I) = \max(R_{i,imm}^T(I), 0)$ Связь немедленных контрафактических сожалений и общих средних сожалений раскрывает следующая теорема.

Теорема 2 $R_i^T \leqslant \sum_{I \in \mathcal{I}_i} R_{i,imm}^{T,+}(I)$.

Таким образом, минимизация немедленных контрафактических сожалений минимизирует общие сожаления. В свою очередь минимизация немедленного контрафактического сожаления может происходить за счет минимизации выражений под функцией максимума. Таким образом, мы приходим к понятию контрафактического сожаления

$$R_i^T(I, a) = \frac{1}{T}\sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)(u_i(\sigma^t|_{I\to a}, I) - u_i(\sigma^t, I)). \tag{1.7}$$

Контрафактическое сожаление рассматривает действие в информационном состоянии. В свою очередь, для минимизации контрафактических сожалений можно применить алгоритм приближения Блэквела, который, применимо к рассматриваемым сожалениям, приведет к следующей последовательности стратегий

$$\sigma_i^{T+1}(I)(a) = \begin{cases} \frac{R_i^{T,+}(I,\, a)}{\sum_{a\in A(I)} R_i^{T,+}(I,\, a)} & \text{если } \sum_{a\in A(I)} R_i^{T,+}(I,\, a) > 0, \\ \frac{1}{|A(I)|} & \text{иначе.} \end{cases} \tag{1.8}$$

Другими словами, действие выбирается в пропорции соотношения позитивных контрафактических сожалений не выбора этого действия. Обоснование сходимости полученного решения и оценку ее скорости предоставляет следующая теорема.

Теорема 3 *Если игроки придерживаются стратегий, заданных выражением (1.8), то* $R_{i,imm}^T(I) \leqslant \Delta_{u,i}\sqrt{|A_i|}/\sqrt{T}$ *и следовательно* $R_i^T \leqslant \Delta_{u,i}|\mathcal{I}_i|\sqrt{|A_i|}/\sqrt{T}$, *где* $|A_i| = \max_{h\colon P(h)=i}|A(h)|$.

2 Вторая глава. Описание моделей

2.1 Модель раскрытия совместного преступления

Модель раскрытия совместного преступления, предложенная Спенглером [1], предполагает игру в развернутой форме для трех игроков с эндогенным характером формирования игровых историй. Игра построена следующим образом. В игре участвуют три игрока: клиент(C), чиновник(O) и инспектор(I). Игру начинает клиент. Клиент может подкупить чиновника с вероятностью γ или нет с вероятностью $1-\gamma$. Чиновник, в случае подкупа, может ответить взаимностью с вероятностью β или нет с вероятностью $1-\beta$. Взаимность определяется как акт возврата благосклонности за взятку, то есть возвращение некоторых привелегий (например, государственный контракт) клиенту. Коррупция, как взаимное взяточничество, может произойти только при совместных усилиях клиента и чиновника. Игра включает в себя четыре штрафа, один за подкуп p_L и один для получения взаимности q_L (штрафы клиента), а также один для принятия взятки и взаимности q (штраф чиновника). Это позволяет использовать асимметричное распределение штрафов. Штрафы применяются с вероятностью инспектирования, которая представлена действием инспектора. Таким образом инспектор может провести проверку с вероятностью α, либо не проводить с вероятностью $1-\alpha$. При этом награда инспетора зависит факта инспектирования и наличия преступления.

Для подкупа чиновника клиент тратит b на взятку и получает выгоду от взаимности чиновника в размере v. Чиновник в случае подкупа получает взятку в размере b , а неотвечая взяимностью получает r. Параметр r играет роль нейтральной выплаты, для подкрепления непринятия взятки и может быть расценен как моральное удавлетворение от несовершения преступления. В случаях выявленного преступления клиент и чиновник должны выплатить соответствующие штрафы. Распределение выплат инспектору показано в таблице 2.1.

История игры	Провести проверку: α	Не проводить: $1-\alpha$
Взаимная взятка: $\gamma\beta$	$x+\Delta x$	x
Невзяимная взятка: $\gamma(1-\beta)$	$y+\Delta y$	y
Не было взятки: $(\gamma-1)$	z	$z+\Delta z$

Таблица 2.1 — Схема распределения выплат инспектору

Предполагается, что проверка приводит к лучшим для инспектора результатам в случае (взаимного) взяточничества, чем в случае отсутствия взяточничества и наоборот: $0<\Delta x,\Delta y,\Delta z$.

Для клиента мы предполагаем, что взяточничество является прибыльным, если оно встречает взаимность, но не с проверкой, где b - взятка, а v - это выгода от взаимного обращения с клиентом. Это подразумевает, что $0 < b < v$ и $0 < p_L$, p_H и $v - b - p_L - p_H < 0$. Для инспектора предполагется, что проверка является прибыльной, если по крайней мере один правонарушитель совершает правонарушения, но обходится дорого, если нет. Это подразумевает, что $x < 0 < x + \Delta x$ и $y < 0 < y + \Delta y$, но $z < 0 < z + \Delta z$. Эта настройка отражает интуицию, что успешный осмотр стоит того, поскольку он приводит к продвижению или аналогичным выгодам, в то время как безуспешная проверка просто стоит усилий. Сложность модели требует, чтобы мы сделали некоторые дополнительные предположения о выплатах инспектора. Мы предполагаем, что проверка взаимного взяточничества является более прибыльной, чем проверка простого взяточничества, и, аналогично, что не проверки взаимного взяточничество несет большую потерю, чем не проверка простого взяточничества из-за более высокой альтернативной стоимости неинспекции. Это подразумевает, что $0 < y + \Delta y < x + \Delta x$ и $x < y < 0$. Для чиновника мы предполагаем, что получение взятки является прибыльным, пока нет проверки, подразумевая, что $0 < r < b < q$.

Последовательность выборов игроков и распределение выплат данной игры могут быть представлены в развернутой форме. Развернутая форма игры представлена на рисунке 2.1.

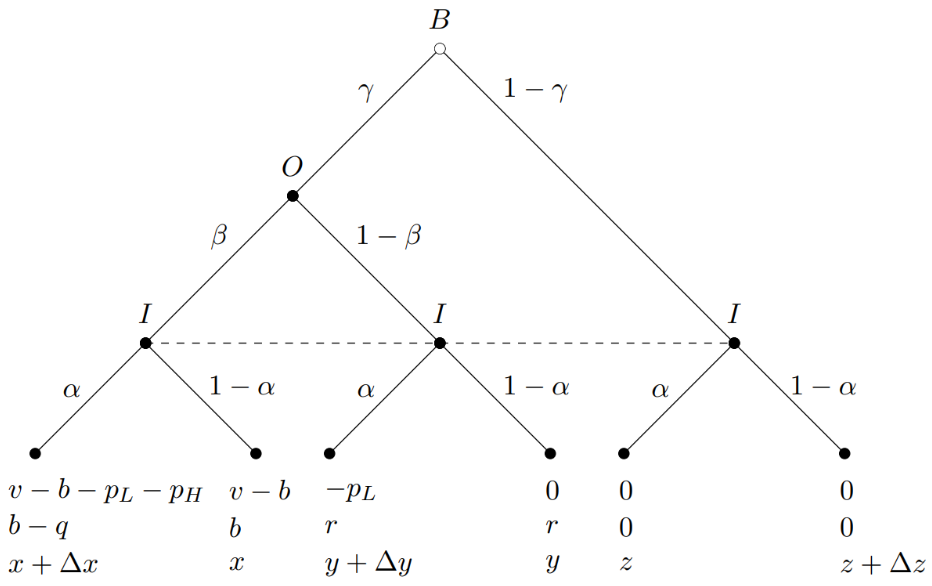

Рисунок 2.1 — Развернутая форма игры

Чтобы алгебраически выразить равновесие, для каждого игрока приравняем выплаты для каждой стратегии. Используя выплаты на рисунке 2.1, мы получаем уравнения (2.1), (2.2) и (2.3), связанные с выплатами клиента, чиновника и инспектора соответственно.

$$\beta(v - b - \alpha(p_L + p_H)) + (1 - \beta)(-\alpha p_L) = 0 \tag{2.1}$$

$$\alpha(b - q) + (1 - \alpha)b = \alpha r + (1 - \alpha)r \tag{2.2}$$

$$\gamma\beta(x+\Delta x)+\gamma(1-\beta)(y+\Delta y)+(1-\gamma)z=\gamma\beta x+\gamma(1-\beta)y+(1-\gamma)(z+\Delta z) \quad (2.3)$$

Мы получаем следующие вероятности равновесия для v, β и α из предыдущих уравнений:

$$\beta=\frac{\alpha p_L}{v-b-\alpha p_H} \quad (2.4)$$

$$\alpha=\frac{b-r}{q} \quad (2.5)$$

$$\gamma=\frac{\Delta z}{\beta(\Delta x-\Delta y)+\Delta y+\Delta z} \quad (2.6)$$

Дальнейшие преобразования (2.4-2.6) позволяют получить выражения β и γ заисящие от параметров:

$$\alpha=\frac{(b-r)p_L}{(v-b)q-(b-r)p_H} \quad (2.7)$$

$$\gamma=\frac{((v-b)q-(b-r)p_H)\Delta z}{(b-r)p_L(\Delta x-\Delta y)+((v-b)q-(b-r)p_H)(\Delta y+\Delta z)} \quad (2.8)$$

Полученные таким образом значения α (2.5), β(2.7) и γ(2.8) образуют равновесный профиль стратегий для данной модели при условии соблюдения указанных ограничений.

2.2 Модель коррупции в иерархической структуре

3 Третья глава. Программная реализация алгоритма

3.1 Общая схема вычислений

В рассмотренных далее примерах рассматривалась вероятностная реализация алгоритма MCCFR. При использовании данного метода значения случайных событий генерируются перед началом каждой обучающей итерации. Данный подход позволяет сократить обьем памяти и ускорить вычисления в некоторых случаях[5]. При реализации примеров были выделены следующие компоненты:

— описание правил игры (зависит от настроек);
— модуль с реализацией алгоритма относительно определенных правил.

Настройки игры, например, по возможности могут включать число игроков, состав костяшек домино и т.п.

Правила игры включают структуру игрового дерева, механизм распределения событий и функцию выплат. Игровое дерево строится с применением узлов – обьектов с информацией о историии игры, о игроке и о возможных действиях.

Сам расчет итераций CFR происходит в выделенном модуле, на основе, определенных для каждого конкретного случая, правил игры. Работа алгоритма начинается с создания игрового дерева. Далее происходит расчет заданного числа итераций. После любой итерации можно получить средние стратегии игроков, которые представляют из себя приближенное коррелирующее равновесие.

3.2 Первый пример. Разоблачение совместного преступления

На рисунке **??** представлен график приближенной эксплуатируемости стратегий. На приведенном графике, и в дальнейшем, горизонтальная ось содержит экспоненциальные отметки о числе итераций по основанию 2. Основная часть кода программы представлена в приложении А.

3.3 Второй пример. Иерархическая модель коррупции

В данной работе в качестве основного объекта исследования была выбрана игра «Иерархическая модель коррупции».

Заключение

В данной работе был продемонстрирован один из лучших на данный момент алгоритмов для приближенного решения игр с неполной информацией. Но несмотря на наличие вполне обобщенных методов, приходится уделять большое внимание разбору частных случаев. Основной проблемой при решении подобных задач является экспоненциальный рост сложности вычислений в зависимости от увеличения числа возможных действий игроков. В связи с этим приходится идти на различные ухищрения с целью получить практически ценный аналог оригинальной задачи. К подобным приемам относят использование метода монте-карло, построение игровых абстракций и многие другие оптимизации.

В качестве объекта исследования была выбрана вполне популярная настольная игра домино. Однако, данной игре уделено довольно мало внимания в контексте рассматриваемого алгоритма. Автор данной работы постарался частично исправить данный недостаток, хотя полученные результаты пока что скромны. Был рассмотрен сильно упрощенный, по сравнению с спортивным, вариант игры с минимумом абстракций. Однако, даже подобный упрощенный вариант раскрывает широкое разнообразие смешанных стратегий, а теоретическая база позволяет говорить о строгой математической обоснованности полученных решений. Для непосредственного расчета стратегий была реализована компьютерная программа.

Дальнейшим развитием данной темы может служить построение более общих абстракций для данной игры. Решения в подобной сфере могут быть полезны как с точки зрения концепции, так и с точки зрения частных методов и оптимизаций.

Список использованных источников

1. *Spengler, D.* Detection and Deterrence in the Economics of Corruption: a Game Theoretic Analysis and some Experimental Evidence. / D. Spengler. — University of York, 2014.

2. *Kumacheva, S. Sh.* The Strategy of Tax Control in Conditions of Possible Mistakes and Corruption of Inspectors. Contributions to Game The ory and Management (Petrosyan, L. A., Zenkevich, N. A. eds), Vol. 6 / S. Sh. Kumacheva. — pp. 264–273. St. Petersburg University, St. Petersburg, 2013.

3. *Orlov I. M., Kumacheva S. Sh.* Hierarchical Model of Corruption: Game theoretic Approach [in Russian]. Mathematical Control Theory and Its Applications Proceedings / Kumacheva S. Sh. Orlov I. M. — pp. 269–271. CSRI ELEKTROPRIBOR, St. Petersburg, 2020.

4. *Martin Zinkevich Michael Johanson, Michael Bowling Carmelo Piccione.* Regret minimization in games with incomplete information. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems 20 / Michael Bowling Carmelo Piccione Martin Zinkevich, Michael Johanson. — MIT Press, Cambridge, 2008, pages 1729–1736.

5. *Marc Lanctot Kevin Waugh, Martin Zinkevich Michael Bowling.* Monte carlo sampling for regret minimization in extensive games. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, Advances in Neural Information Processing Systems 22 / Martin Zinkevich Michael Bowling Marc Lanctot, Kevin Waugh. — MIT Press, Cambridge, 2009, pages 1078–1086.

6. *Hart, Sergiu.* A simple adaptive procedure leading to correlated equilibrium / Sergiu Hart, Andreu Mas-Colell. — Econometrica, 68(5), September 2000, pages 1127–1150.

7. *Brown, Noam.* Supplementary Materials for Superhuman AI for multiplayer poker / Noam Brown, Tuomas Sandholm. — Science First Release DOI: 10.1126/science.aay2400, 11 July 2019.

Приложение А Алгоритм решения

Приложение Б Реализация правил игры разоблачения совместного преступления