МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«Ярославский государственный университет им. П.Г. Демидова»

Ккрсовая работа

Поиск приблеженного равновесия в играх с неполной информацией

02.03.01. Математика и компьютерные науки

Исполнитель: Сычев Р.С.
гр. МКН-41 БО
Руководитель: к.ф.-м.н. Глазков Д.В.

Ярославль 2021

Содержание

Введение

В последнее время математическая теория игр с неполной информацией находит все большее применение в таких отраслях как теория операций, экономика, кибербезопасность и физическая безопасность. Не в последнюю очередь это происходит благодаря постоянному совершенствованию алгоритмов и увеличению производительности вычислительной техники. Частным случаем таких игр являются 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.

Существенным шагом в развитии данного направления является использование алгоритма подсчета сожалений (regret matching). Алгоритм предполагает итеративное вычисление последовательности стратегий в среднем сходящейся к оптимальному стратегическому профилю. Открытие этого метода привело к появлению ряда алгоритмов для поиска приближенного решения в играх с неполной информацией.

Данная работа посвящена рассмотрению одного из популярных в настоящее время итеративных алгоритмов - контерфактической минимизации сожалений (Conterfactual Regret Minimization) и его модификации приедусматривающей использование метода монте карло (MCCFR). Данные алгоритмы появились не так давно, но на их основе уже получен ряд недостижимых до этого по сложности результатов.

Целью данной работы является описание и отработка на практике приведенных выше алгоритмов.

В соответствии с темой работы поставлены следующие задачи:

— подготовить теоритическое описание алгоритмов;
— выделить некоторые игры в развернутой форме для последующего решения;
— реализовать алгоритм и произвести расчет стратегического профиля для приведенных игр.

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 Нэша

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 состояния, 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 наборами. 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 разбиением игрока i, а $I_i \in \mathcal{I}_i$ информационным набором 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 нулевой суммой и полной памятью.

Стратегия игрока i — это функция σ_i, которая ставит в соответствие каждому информационному набору $I_i \in \mathcal{I}_i$ 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

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

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 [ссылка]. 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. 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 [4].

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

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. Контрафактической ценой $u_i(\sigma, I)$ назовем ожидаемую цену, при условии, что информационный набор I был достигнут, 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 наборе. 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* $|A_i| = \max_{h:\, P(h)=i} |A(h)|$.

1.3 Метод Монте-Карло

2 Вторая глава. Программная реализация алгоритма

2.1 Общая схема вычислений

2.2 Первый пример. Покер Куна

2.3 Второй пример. Домино

Заключение

Реализация современных криптографических алгоритмов, безусловно, требует высокой квалификации разработчика. Это связанно с тем, что подобные алгоритмы должны работать как можно более эффективно для обеспечения быстродействия обслуживаемых ими систем. Бывает весьма трудоемко оптимальное распределение переменных по регистрам процессора и уровням кэша. Также, особого рассмотрения требует использование встроенных типов. Подобные проблемы часто бывают узкоспециализированны в зависимости от архитектуры ЭВМ и выливаются в широкий спектр прикладных вопросов.

В данной работе были рассмотрены наиболее общие и надежные высокоуровневые языковые конструкции, призванные сформировать у читателя представление о процессах практического вычисления функции хэширования согласно ГОСТ Р 34.11-2012 и реализации процессов генерации и проверки ЭЦП согласно ГОСТ Р 34.10-2012.

В дополнение к изложенному, в рамках данной работы был разработан ряд приложений для операционной системы Windows с графическим интерфейсом, которые позволяют непосредственно использовать описанные механизмы.

Приложения реализуют:

— вычисление хэш-функции от файла;
— редактирование параметров и ключей схемы цифровой подписи:
— генерацию цифровой подписи;
— проверку цифровой подписи.

Список использованных источников

1. *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, pages 1729–1736 / Michael Bowling Carmelo Piccione Martin Zinkevich, Michael Johanson. — MIT Press, Cambridge, 2008.

2. *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, pages 1078–1086 / Martin Zinkevich Michael Bowling Marc Lanctot, Kevin Waugh. — MIT Press, Cambridge, 2009.

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

4. *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.