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 1
1 Первая глава. Описание алгоритмов 2
1.1 Игры в развернутой форме и равновесие Нэша 2
1.2 Теорема Блэквела и минимизация сожалений 5
1.3 Контерфактические сожаления и их минимизация 5
1.4 Метод Монте-Карло 5
2 Вторая глава. Программная реализация алгоритма 6
2.1 Общая схема вычислений 6
2.2 Первый пример. Покер Куна 6
2.3 Второй пример. Домино 6

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 бу-

дем 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 хъ вероятность того, что игроки достигнут хъ руководствуясь хъ. Мы можем представить хъ как хъ, выделяя вклад каждого игрока. В таком случае, хъ обозначает вероятность принятия совокупности решений, ведущих от префикса хъ к префиксу хъ для участков в которых хъ. Аналогично обозначим за хъ вероятность достижения истории хъ всеми игроками за исключением хъ. Формально можно определить предыдущие величины следующим образом

Хъ

Для хъ определим хъ. Аналогично введем хъ и хъ.

Ожидаемое среднее значение выплат для игрока хъ обозначим как хъ.

Традиционным способом решения игр в развернутой форме является поиск равновесного профиля стратегий хъ, который удовлетворяет следующему условию.

Хъ

Такой стратегический профиль называют равновесием по Нэшу.

В случае, если некий стратегический профили хъ удовлетворяет условию

Хъ

Его называют хъ – равновесием

Контрафактические сожаления и их минимизация

Минимизация сожалений является популярным концептом, для построения итеративных алгоритмов приближенного решения игр в развернутой форме. Приведем связанные с ней определения. Рассмотрим дискретный отрезок времени T включающий 1, T раундов. Обозначим за хъ стратегию игрока хъ в раунде хъ.

Определение. Средним общим сожалением игрока хъ на момент времени T называют величину

Хъ

В дополнении к этому определим хъ как среднюю стратегию относительно всех раундов от 1 до T. Таким образом для каждого хъ и хъ введем

Хъ

Теорема. Если на момент времени T средние общие сожаления игроков меньше хъ, то хъ является 2хъ равновесием.

Говорят, что алгоритм выбора хъ 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.

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

Обозначим через хъ цену игры с точки зрения истории хъ, при условии, что хъ была достигнута, и игроки спользуют в дальнейшем хъ.

Определение. Контрафактической ценой хъ назовем ожидаемую цену при условии, что информационный набор хъ был достигнут, когда все игроки кроме хъ играли в соответствии с хъ. Формально

Хъ,

где хъ вероятность перехода из хъ в хъ.

Обозначим за хъ стратегический профиль идентичный хъ за исключением того, что хъ всегда выбирает хъ попадая в хъ.

Немедленным контрафактическим сожалением назовем

Хъ

Интуитивно это выражение можно понимать, как аналог среднего общего сожаления в терминах контрафактической цены. Однако вместо рассмотрения всевозможных максимизирующих стратегий рассматриваются локальные модификации стратегии. Хъ. Связь немедленных контрафактических сожалений и общих средних сожалений раскрывает следующая теорема.

Теорема. Хъ

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

Хъ.

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

Хъ

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

Теорема. Если игроки придерживаются стратегий, заданных выражением (), то хъ

1.2 Теорема Блэквела и минимизация сожалений

1.3 Контерфактические сожаления и их минимизация

1.4 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