Инструкция

Данное приложение моделирует процесс случайного поиска. То есть частица блуждает по поверхности, состоящей из N квадратных ячеек, закрашивая пройденные ячейки. Некоторые ячейки закрашены зеленым цветом (их количество – N_g), они соответствуют целям поиска. Процесс продолжается до тех пор, пока не будут найдены все цели. Также в реальном времени строится график энтропии, которая совершает небольшой скачок вниз при посещении (закрашивании) новой ячейки, и большой скачок вниз при нахождении цели. У пользователя есть возможность управлять следующими параметрами программы:

1. Количество целей поиска N_g.
2. Количество блуждающих точек для ускорения набора статистики.
3. Скорость симуляции.
4. Размер поля N.
5. Распределение, из которого генерируются смещения точки. При смени распределения или его дисперсии, плотность нового отображается на специальном графике, что помогает пользователю понять, какой характер будет иметь блуждание.
6. Также в случае поиска одной цели есть возможность перейти в режим поиска по одной оси, при котором если точка находит одну из координат цели, то дальнейшие блуждания начинают осуществляться только по второй оси. Предполагается, что данный режим будет использоваться для ускорения завершения работы программы.

Теоретические сведения

Заметим, что случайный поиск тесно связан с энтропией. Действительно, чем больше точек мы посетили, тем больше вероятность того, что найдены искомые ячейки. Таким образом, важно изучить, как влияют на скорость минимизации энтропии используемые для получения приращений распределения. У нас имеется сетка из N ячеек, N_g из которых являются целями поиска, пусть к моменту времени t алгоритм посетил k ячеек, d из которых соответствуют искомым. Тогда энтропия вычисляется по формуле:

$$S(t) = -(N_g - N_d) \cdot \log{(N - N_k)}$$

В классических алгоритмах для случайного поиска используются распределения с тяжелыми хвостами, мы же сравним два распределения – симметричное продолжение на вещественную прямую экспоненциального распределения и смесь гауссиан (распределение с несколькими пиками). Плотности этих распределений записываются следующим образом:

$$f_1(x;\lambda) = \frac{1}{2}\lambda e^{-\lambda x}[x \geq 0] + \frac{1}{2}\lambda e^{\lambda x}[x < 0]$$

$$f_2(x) = p_1 f(x;\mu,\sigma) + p_2 f(x;0,\sigma_0) + p_1 f(x;-\mu,\sigma)$$

где

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad 2p_1 + p_2 = 1$$

Заметим, что совершая большие прыжки, частица уходит из области, где уже много клеток закрашено, в свободную область и начинает закрашивать ее. В результате энтропия уменьшается быстрее. Но при этом использование только больших скачков не способствует непосредственному нахождению целевой ячейки. Этим объясняется использование описанных выше распределений.

Информационный смысл моделирования

Мы уже убедились, что эта задача тесно связана с энтропией, поэтому можем рассмотреть ее с информационной точки зрения. Действительно, если выражать энтропию в тех же единицах, что и информацию, то энтропию можно считать недостатком информации о состоянии системы. Также известно, что попадание частицы в маленький объем дает больше информации, чем ее попадание в большой объем. Таким образом, если поставить в соответствие маленьким объемам целевые ячейки, а за большой объем считать все остальные ячейки, то получим, что программа показывает, сколько информации мы получаем о системе с течением времени.

Связь с биологией

Процесс случайного поиска моделирует некоторые биологические явления. Так, некоторые виды животных (например, голубые и китовые акулы) имеют два "режима" поиска места с пищей: локальный и глобальный. В локальном режиме их поведение описывается небольшими перемещениями в заданной области, что соответствует небольшим случайными прыжкам. Но когда ничего не находят, животные перемещаются на большее расстояние (что соответствует редким длинным случайным прыжкам), где опять возвращаются к локальному поиску. Тогда наши искомые точки дискретной сетки соответствуют областям с большим количеством пищи, а блуждающие точки моделируют поведение животных.

Математическая постановка задачи случайного поиска

Задача случайного поиска является методом оптимизации, в котором минимизация функционала $f(x, y)$ осуществляется путем случайных перемещений точки и последующим выбором минимального значения из найденных. Обычно выход такого алгоритма используется в качестве начального приближения для других методов оптимизации (например, для градиентного спуска в случае его применимости). Опишем задачу подробнее для двумерного случая.

$$f(x, y) \to \min_{x,y}$$

Вводится равномерная сетка $x_1, ..., x_m$, $y_1, ..., y_n$, разбивающая всю область на квадраты (тогда наше поле является визуализацией такого разбиения). Тогда задача алгоритма – найти такие x_i и y_j, что:

$$f(x_i, y_j) \le f(x_k, y_l) \quad \forall x_k, y_l$$

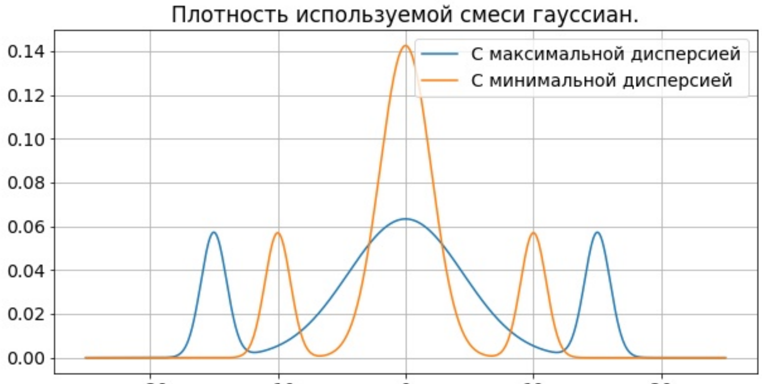

Рис. 1: Плотность используемой смеси гауссиан.

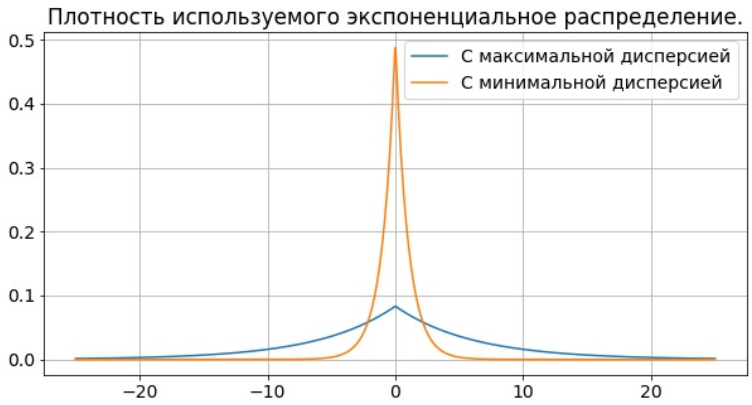

Рис. 2: Плотность используемого симметричного продолжения экспоненциального распределения.

Поиск будем осуществлять следующим образом: генерируется случайная точка, затем на каждой итерации вычисляются независимые приращения по x и y из некоторого распределения. Затем определяется, в какую область дискретной сетки попадает сдвинутая точка и вычисляется значение функционала в ней, оно сравнивается с предыдущим минимумом, и при необходимости происходит обновление минимизирующей точки. Алгоритм повторяется некоторое заданное наперед количество итераций.

Интересные наблюдения, связанные с программой

- При больших скачках энтропия падает сильнее, чем при локальном поиске.
- Смесь гауссиан быстрее минимизирует энтропию, чем экспоненциальное распределение, так как наличие нескольких пиков в распределении способствует более частым дальним скачкам, а на них происходит более сильное падение энтропии.
- Увеличение дисперсии ускоряет убывание энтропии, так как появляются более дальние скачки, а при них энтропия уменьшается сильнее.
- В среднем со временем скорость падения энтропии становится меньше, что обосновывается тем, что чем больше точек мы посетили, тем меньше информации мы начинаем получать от блуждания.

  Также при использовании воображения можно находить какие-то образы, которые рисует точка в процессе блуждания. Например, на приложенном рисунке мы видим черепашку. Нахождение таких образов – довольно интересное развлечение.

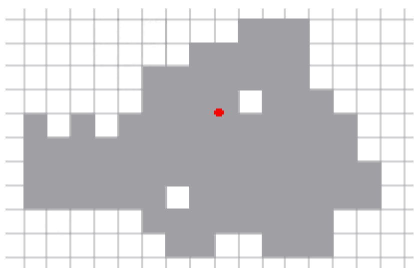

Рис. 3: Пример образа, который можно увидеть.

- Еще к развлечениям можно отнести попытки угадать, найдет ли точка в ближайшее время какую-либо цель.