Алгоритм A*

Донской Юрий

27 октября 2022 г.

Оглавление

Глава 1

Введение

1.1 Описание алгоритма

Алгоритм A* (*англ. A star*) — алгоритм поиска, который находит во взвешенном графе маршрут наименьшей стоимости от начальной вершины до выбранной конечной.

1.2 История

В 1964 году Нильс Нильсон изобрёл эвристический подход к увеличению скорости алгоритма Дейкстры. Этот алгоритм был назван A1. В 1967 году Бертрам Рафаэль сделал значительные улучшения по этому алгоритму, но ему не удалось достичь оптимальности. Он назвал этот алгоритм A2. Тогда в 1968 году Питер Э. Харт представил аргументы, которые доказывали, что A2 был оптимальным при использовании последовательной эвристики лишь с незначительными изменениями. В его доказательство алгоритма также включён раздел, который показывал, что новый алгоритм A2 был возможно лучшим алгоритмом, учитывая условия.

1.3 Применение алгоритма

Алгоритм A* применяют в различных областях от машинного обучения до разработки игр, например, при выстраивании маршрута по сложной местности с препятствиями.

Глава 2

Описание алгоритма

2.1 Описание

В процессе работы алгоритма для вершин рассчитывается функция $f(v) = g(v) + h(v)$, где

- $g(v)$ — наименьшая стоимость пути в v из стартовой вершины,
- $h(v)$ — эвристическое приближение стоимости пути от v до конечной цели.

Фактически, функция $f(v)$ — длина пути до цели, которая складывается из пройденного расстояния $g(v)$ и оставшегося расстояния $h(v)$. Исходя из этого, чем меньше значение $f(v)$, тем раньше мы откроем вершину v, так как через неё мы предположительно достигнем расстояние до цели быстрее всего. Открытые алгоритмом вершины можно хранить в очереди с приоритетом по значению $f(v)$.

2.2 Свойства

Чтобы A* был оптимален, выбранная функция$h(v)$ должна быть допустимой эвристической функцией.

Определение: *Говорят, что эвристическая оценка $h(v)$ допустима, если для любой вершины v значение $h(v)$ меньше или равно весу кратчайшего пути от v до цели.*

Допустимая оценка является оптимистичной, потому что она предполагает, что стоимость решения меньше, чем оно есть на самом деле. Второе, более сильное условие — функция $h(v)$ должна быть монотонной.

Определение: *Эвристическая функция $h(v)$ называется монотонной (или преемственной), если для любой вершины v_1 и ее потомка v_2 разность $h(v_1)$ и $h(v_2)$ не превышает фактического веса ребра $c(v_1, v_2)$ от v_1 до v_2, а эвристическая оценка целевого состояния равна нулю.*

Теорема: *Любая монотонная эвристика допустима, однако обратное неверно.*

Доказательство: Пусть $k(v)$ — длина кратчайшего пути из вершины v до цели. Докажем индукцией по числу шагов до цели, что $h(v) \leqslant k(v)$.

Если до цели расстояние 0, то v — цель и $h(v) = 0 \leqslant k(v)$.

Пусть v находится на расстоянии i от цели. Тогда существует потомок v', который находится на кратчайшем пути от v до цели и v' лежит на расстоянии $i-1$ шагов до цели. Следовательно, $h(v) \leqslant c(v, v') + h(v')$. По предположению, $h(v') \leqslant k(v')$. Следовательно, $h(v) \leqslant c(v, v') + k(v') = k(v)$.

Таким образом, монотонная эвристика $h(v)$ допустима.

Утверждение: *Если $h(v)$ монотонна, то последовательность значений $f(v)$ на любом пути неубывает.*

Доказательство следует из определения монотонности. Пусть v' — потомок v, тогда $g(v') = g(v) + c(v, v')$. Следовательно, $f(v') = g(v') + h(v') = g(v) + c(v, v') + h(v') \geqslant g(v) + h(v) = f(v)$.

Утверждение: *Алгоритм A^* является оптимальным, если функция $h(v)$ монотонна.* Последовательность вершин "развёрнутых"во время работы алгоритма находится в неубывающем порядке значений f. Поэтому очередная выбираемая вершина должна представлять собой оптимальное решение, поскольку все дальнейшие узлы будут, по меньшей мере, столь же дорогостоящими.

2.3 Примеры эвристик

Поведение алгоритма сильно зависит от того, какая эвристика используется. В свою очередь, выбор эвристики зависит от постановки задачи. Часто A* используется для моделирования перемещения по поверхности, покрытой координатной сеткой.

- Если мы можем перемещаться в четырех направлениях, то в качестве эвристики стоит выбрать манхэттенское расстояние

$$h(v) = |v.x - v'.x| + |v.y - v'.y|$$

- Расстояние Чебышева применяется, когда к четырем направлениям добавляются диагонали:

$$h(v) = max(|v.x - v'.x|, |v.y - v'.y|)$$

- Если передвижение не ограничено сеткой, то можно использовать евклидово расстояние по прямой:

$$h(v) = \sqrt{(v.x - v'.x)^2 + (v.y - v'.y)^2}$$

Также стоит обратить внимание на то как соотносятся $f(v)$ и $h(v)$. Если они измеряются в разных величинах (например, $g(v)$ — это расстояние в километрах, а $h(v)$ — оценка времени пути в часах) A* может выдать некорректный результат.