COS30019 Semester 2024

ASSIGNMENT 1 TREE-BASED SEARCH

Proposal by:

Tran Quoc Dung - 103803891

Contents

1.	. Instructions	2
2.	. Introduction	3
3.	. Search Algorithms	4
	3.1. Depth-First Search (DFS)	4
	3.2. Breath-First Search (BFS)	5
	3.3. Greedy Best-First Search (GBFS)	6
	3.4. A Star (A*)	7
	3.5. Custom Search Strategy 1 (CUS1)	8
	3.6. Custom Search Strategy 2 (CUS2)	10
	3.7. Advantages & Disadvantages of Algorithms	11
4.	. Testing	11
5.	. Features	14
6	Conclusion	14

1. Instructions

Step 1:

Firstly, my project code is developed & implemented by Visual Studio. Therefore, to open and run our project, Visual Studio is required to be installed.

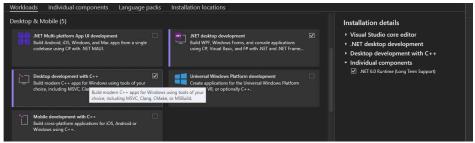
If you have not download Visual Studio, click here to install.

Step 2:

If you just downloaded Visual Studio, remember to choose the "Desktop Development with C++" workload, which provide virtual libraries for my project code.

If you have already downloaded it, you can check the workloads by choosing *Tools => Get Tools* and Features

Remember, this workload "Desktop Development with C++" must be installed, so the code can run smoothly.

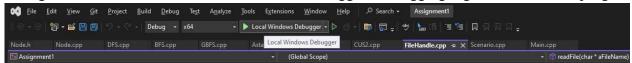


Step 3: Open Visual Studio. Select "Open a project or solution", which will open a project through .sln file.



Then access the "Assignment1.sln" file through this following path: Assignment 1/Assignment1.sln

Step 4: Once the project code is open, you can comprehensively observe the code files, along with running them. Press F5, or this Local Windows Debugger/ Debugging logo here to run the program.



After that, you can see the output of the project, showing respective illustrations of each algorithm.

2. Introduction

The Robot Navigation Problem is a program where a robot must explore an NxM grid to reach a goal located in a random node within the grid. The robot starts on an empty node and moves one node at a time, with only four directions that are right, left, up and down. The grid contains walls that the robot cannot pass through, requiring it to navigate around obstacles. For this assignment, the grid size, agent location, goals, and walls are predefined. The focus is on showcasing theoretical and practical AI skills by illustrating the resolution of intricate pathfinding obstacles in a grid setting using C++ and tree-based search algorithms. By prioritizing algorithmic nuances and strategic AI decision-making, the objective of the document is to offer perspectives on addressing the Robot Navigation issue through detailing the implementation and efficiency of algorithms.

Basic concepts of Robot Navigation:

- Graphs: A collection of vertices connected by edges is referred to as a graph. In the context of the NxM Puzzle, each state can be represented as a vertex, and the edges represent possible transitions between states.
- Trees: The tree is a special kind of graph that starts at a root node and spreads out to leaf nodes. It is a hierarchical structure without any cycles. Trees are essential for representing the search space of the NxM Puzzle, where one node can result in several subsequent states, similar to a branching process.

Glossary:

- Node/Vertex: The fundamental unit in graphs and trees, representing states in the context of NxM Puzzles.
- Edge: A connection between graph nodes, indicating possible moves or transitions within the puzzle.
- Leaf node: A childless node signifying a possible solution conclusion during the search.
- Heuristic: An approach to calculating the cost of achieving a goal from a specific state, used to prioritize paths in intelligent search algorithms.

3. Search Algorithms

3.1. Depth-First Search (DFS)

- Definitions: DFS is an algorithm used for navigating or searching through tree or graph data structures. The algorithm investigates each branch as far as it can after beginning at the root node and then turns around. DFS is unique by means of using a stack with recursion, to manage the nodes under study. Though it may not always identify the shortest path in weighted graphs or the entire exploration in infinite graphs, it performs well when solving puzzles, mazes, and cycles.
 - Image 3.1: Depth-First Search (DFS) source code:

```
#include "Scenario h
void Scenario::DFS(Node* aSource) const
    //Marking the source node as visited, preventing revisitation
aSource->fvisited = true;
    //If current node is the goal, print the path & stop the search if (aSource->fColor == 'G')
         stringstream path;
         printPath(aSource, path);
          // Counting number of steps (By converting "path" to istringstream)
         istringstream is(path.str());
auto numOfSteps = distance(istream_iterator<string>(is), istream_iterator<string>());
         cout << "Number of steps: " << numOfSteps - 4 << endl;
         cout << path.str() << endl;
cout << "\n";
    if (isValid(aSource->fX - 1, aSource->fY))
          fNodes[aSource->fX - 1][aSource->fY].fPrevious = aSource;
         DFS(&fNodes[aSource->fX - 1][aSource->fY]);
fNodes[aSource->fX - 1][aSource->fY].fPrevious = NULL; //Reseting fPrevious
    if (isValid(aSource->fX, aSource->fY - 1))
         fNodes[aSource->fX][aSource->fY - 1].fPrevious = aSource;
DFS(&fNodes[aSource->fX][aSource->fY - 1]);
fNodes[aSource->fX][aSource->fY - 1].fPrevious = NULL; //Reseting fPrevious
    if (isValid(aSource->fX + 1, aSource->fY))
         fNodes[aSource->fX + 1][aSource->fY].fPrevious = aSource;
         DFS(&fNodes[aSource->fX + 1][aSource->fY]);
fNodes[aSource->fX + 1][aSource->fY].fPrevious = NULL; //Reseting fPrevious
    if (isValid(aSource->fX, aSource->fY + 1))
         fNodes[aSource->fX][aSource->fY + 1].fPrevious = aSource;
DFS(&fNodes[aSource->fX][aSource->fY + 1]);
          fNodes[aSource->fX][aSource->fY + 1].fPrevious = NULL; //Reseting fPrevious
```

- In my project, DFS is implemented as a function *Scenario::DFS(Node* aSource) const*, which takes coordinates of *aSource* as inputs, then implements Depth-First Search from that node.
- The code assumes that the grid/maze is represented as a 2D array (fNodes) of Node objects. Each node is identified by its coordinates (fX, fY), a route tracking pointer (fPrevious), its type and color (fColor), and a flag (fVisited) that shows if it has been visited.
- The line "aSource->fVisited = true" appears as a repeated state checking statement, marking the state of the beginning node as visited.
- The *if* statement, if (aSource->fColor == "G"), means the path, number of steps will be displayed when the agent reaches the goal node.
- Next, the function iteratively explores all nodes by turning up, down, left, or right, that are legitimate moves from the current node. It uses the *isValid* function to determine whether a move is permitted and inside boundaries.
- The *isValid* function is necessary for the method to avoid examining nodes that shouldn't be visited due to the limitations of the challenge or out-of-bounds areas.

3.2. Breath-First Search (BFS)

- Definitions: BFS is also a tree-based search algorithm, quite similar to DFS but with different implementations. Before going on to the nodes at the next depth level, it investigates every node at the current depth starting with the tree root.

- Image 3.2: Breath-First Search (BFS) source code:

- A queue named *Search* oversees managing the nodes throughout the exploring stage. The source node is marked as visited to prevent it from being re-visited and enqueued from the start of the search.
- The algorithm loops back on itself if there still are nodes in the queue. This implies that the search will go on for as long as there are still accessible nodes that haven't been investigated. The neighbor is noted as visited and added to the search queue. This is a compulsory step in level-by-level exploration process of BFS, since it guarantees that each node at the current depth is examined before moving on to the next level.
- The node at the front of the queue is removed once it has been established that it is the node to examine at this time (temp). If the current node is the goal node (*if* statement), the procedure builds the path to the current node using a *stringstream* and outputs it before exiting the loop.
- The function iteratively explores all nodes by turning up, down, left, or right, that are legitimate moves from the current node. It uses the *isValid* function to determine whether a move is permitted and inside boundaries.

3.3. Greedy Best-First Search (GBFS)

- Definitions: GBFS is an informed search method, which ignores the edge weights in a weighted graph by just considering the heuristic value, with the evaluation function precisely equal to the heuristic function. It expands the node that is closest to the target, as calculated by the heuristic function, to seek for a goal node.

- Image 3.3: Greedy Best-First Search (GBFS) source code:

- According to the algorithm, the agent tends to explore the nodes with the closest distance consecutively, according to the heuristic that defines the cost to the goal node.
- A queue called *priority_queue* is utilized for storing arranged nodes, based on their heuristic ranking.
- The line aSource > aCost = cost(aSource, aDestination) helps calculating the distance between the source node to the destination node, which will be used for ranking in queue.
- Nodes in the queue are checked by the algorithm continuously, especially the one with lowest cost to the goal node will be chosen and removed from queue. This process is implemented until the queue is empty.
- This algorithm establishes whether a node eliminated from the queue is the goal node. If so, the path from the source to the destination is created and printed using the *printPath* function.
- If the present node is not the goal node, the agent will continue exploring four neighbors by turning up, down, left, and right. Each case's heuristic cost to the destination is determined, the neighbor is added to the priority queue, the neighbor is marked as visited, and the present node is positioned as the neighbor's predecessor (to trace the path).

3.4. A Star (A*)

- Definitions: A* is an informed search algorithm, which helps finding the shortest path in a graph, from the start node to the goal node. A* algorithm utilizes the features of both GBFS and Uniform Cost Search. By calculating the cost from a node to the goal based on heuristics, A* maximizes efficiency and ensures completeness in its search for the objective.

- Image 3.4: AStar Search (A*) source code:

```
Secile Section 1 (Activity and Section 1) (Act
```

- According to the above code, *fCost* stands for the heuristic estimated cost between the current node & the destination, while *fDistance* is the true cost between them.
- The nodes in the priority queue is determined by a specially constructed comparator structure, which adds the heuristic estimated cost (fCost) to the true cost (fDistance) to reach a node.
- The cost function is applied to the source node, which has a distance of 0 and is marked as visited, to evaluate the heuristic cost to the destination. The source node is then added to the priority queue.
- To ensure the node with the total lowest cost, which is fCost + fDistance, is explored first, a priority queue is also be used to manage the nodes in order, similar to GBFS.
- The *isValid* function is also be used for checking the validity of the agent's move. In addition, A* assess the heuristic expense for the goal node, then it backtracks the neighbor's path and adds the neighbor to the queue. At long last the neighbor's distance is processed as the current node in addition to one, for each valid neighbor.
- The objective is the point at which the goal node is taken out from the queue, indicates the shortest way has been found, ends the search. From that point forward, the path is printed out subsequent to being backtracked, utilizing the *fPrevious* pointers from the goal to the source node.

3.5. Custom Search Strategy 1 (CUS1)

- Definitions: CUS1 algorithm is one of my customize tree-based search algorithms, which utilizes a priority-based search method to investigate the nodes in a graph. The calculation oversees node investigation in view of cost, which is progressively resolved by using node colours, utilizing a priority queue.

- Image 3.5: Custom Search Algorithm 1 (CUS1):

- In the priority queue, lower costs will equally mean higher priority, which is used to evaluate the nodes. Moreover, essential component is added to pathfinding respectively based on the color of the nodes, by working out the cost between two nodes and leaning toward moves between nodes of a similar color with a lower cost (1 for matching colors, 2 otherwise).
- The source node is set as visited normally, then added to the priority queue. For each node, the path from the source to the objective is made, printed, and the search continues in case the node is the objective (fColor == 'G'). This makes it conceivable to distinguish at least one objective.
- The calculation then, at that point, takes a gander at the ongoing node's symmetrical neighbors (up, down, left, right), preventing from slanting moves (symmetry is guaranteed by the way that abs(i) + abs(j)! = 1).
- For each valid neighbor, CUS1 marks the visitation of it. For path tracking, this marks the current node as the neighbor's ancestor (*fPrevious*). Then, the cost to reach this neighbor is evaluated by means of the cost function.
- The manner in which CUS1 decides the cost of moving between nodes while representing the color comparability, which might address different landscapes or zones in a circumstance where moving inside comparable zones is more beneficial or more affordable, makes it unique & special. Verifies that nodes with lower total cost are explored first, coordinating the search toward conceivably more powerful paths to the goal.

3.6. Custom Search Strategy 2 (CUS2)

- Definitions: CUS2 is my second customize tree-based search algorithm. CUS2 algorithm is an exceptionally planned search methodology that incorporates way cost and distance measures to track down the most proficient way from a source node to an objective node.

- Image 3.6: Custom Search Algorithm 2 (CUS2):

```
Finclude "Scenario.h"

road Scenario:CUS2(Node* aSource, Node* aDestination) comet

{
    //Custes comparison function for surting modes based on the sum of their cost and distance
    ands comp = [[coss theore hode* h]
    {
        return (a->fCost + a->fDistance) > (b->fCost + b->fDistance);
    }

    //Open List to stare modes that are yet to be explained

std::Vector*Node** openList;

    //Start with the source mode
    openList.push_back(ASource);

    // Continus until there are no modes left to explain

white (iopenList empty())

{
    //Index of the mode with the minimum cost * distance
    int Ninimum = 0;
    // (continus until there are no modes left to explain

white (iopenList(i)->fCost * openList(i)->fDistance) < (openList[Minimum]->FCost * openList(Minimum]->FDistance))

{
    // Index of the mode minimum cost * distance
    Nodes temp = openList(Minimum];

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum);

    // Nodes with the minimum cost * distance
    Nodes temp = openList(Minimum
```

- Focuses on nodes as per the absolute of their crossing cost (*fCost*) and heuristic distance (*fDistance*) to the goal, by involving a lambda function for nodes correlation in the open rundown. This heuristic strategy really coordinates the inquiry towards the goal.
- To identify nodes that require further investigation, I utilize an open list (which is implemented as a *std::vector*). This list is dynamically updated with nodes that are investigated and discovered.
- Iteratively, the node with the lowest combined cost and distance is selected from the open list to be investigated. Using the specified comparator criteria, the open list is linearly searched to find the most promising node.
- Valid, undiscovered neighbors are processed after calculating their *fCost* for the destination, modifying the *fDistance* from the starting point, setting respective *fPrevious* pointer for path recording, and storing them to the open list for additional investigation.
- The technique builds the path from the final point back to the source using the *fPrevious* pointers after discovering the destination node. It then emits this route, showing the journey taken, using a *stringstream*.

3.7. Advantages & Disadvantages of Algorithms

Algorithms	Advantages	Disadvantages
DFS	- Simple to be used with recursion	- Not always the shortest path
	- Requires less memory than BFS	- Without repeated state check, the
	- Most suitable for solving puzzles,	agent might be stuck in infinity loops.
	which requires only finding a way to the	
	goal state.	
BFS	- Seeks for shortest path quicker than	- Requires more memory than DFS
	DFS	- If the solution requires deep explore
	- Can prevent unnecessary paths	in the tree, the performance might be
		slower
GBFS	GBFS can be more efficient than DFS or	Not always the quickest path, since
	BFS, by means of utilizing the heuristics	the heuristic quality can totally affect
	towards the goal node.	the performance of the algorithm.
A*	- Guarantees the shortest path displayed	- Requires efficient heuristic
	- Combining both the completeness of	- Insufficient heuristics might
	BFS & heuristic cost of GBFS	negatively affect the performance
		- Might require much memory
CUS1 &	It custom fitted to specific issue	- Requires more fine-tuning or testing
CUS2	limitations or goals, perhaps giving more	to optimize the output.
	powerful solutions in certain conditions.	- Custom solutions probably won't
		transfer well to different sorts of
		issues or circumstances.

4. Testing

To finish the program's Testing Procedures, it requires a testing methodology. As a result, I believe **Metamorphic Testing** is the best option that I can use to approach the testing phase.

By specifying relationships between inputs and outputs, **Metamorphic Testing** is a software testing approach that helps create and generate related test cases and solve the test oracle problem. A test oracle is defined as a mechanism or technique that compares the program's outputs to its test cases to determine the program's correctness. **Metamorphic Relations** are equally essential notions needed to implement **Metamorphic Testing**.

For verifying the program implementation under these conditions, which include different linked inputs and their outputs, **Metamorphic Relations (MR)** are essential qualities. New test cases can be produced by creating related inputs based on the identification of metamorphic relations.

The following actions will enable us to begin building Metamorphic Testing:

- Carrying out several arbitrary test scenarios.
- Determine the characteristics of the issue, which will lead to the identification of MR.
- Using the discovered MR to generate new test cases.
- Confirming the metamorphic relations using the calculated results.

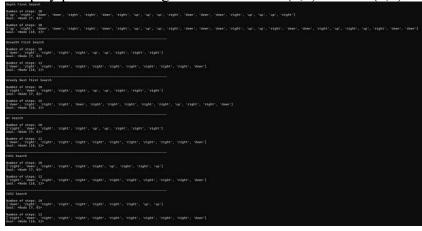
Test case 1: Displaying path from customized start node, goal nodes

MR: Different coordinate of start node might lead to different and respective paths. Input 1.2 is generated from input 1.1 by adjusting the coordinates of the start node, where the agent begins.

Input 1.1: Original inputs from the assignment



Output 1.1: Display paths of each algorithm from Node(0,1) to Node(7,0) & Node(10,3)



Input 1.2: New coordinates of the agent/the start node



Output 1.2: New paths are display respectively, from Node(0,3) to Node(7,0) & Node(10,3)

```
| Part |
```

Conclusion: After adjusting the coordinates of the agent, the path changes accurately. So, the test case passes the requirements.

Test case 2: Checking whether goal node is reachable or not

MR: Different coordinate of goal nodes might lead to less paths displayed, since several nodes with inappropriate coordinates cannot be reached.

Input 2.1: Original inputs from the assignment



Output 2.1: Display paths of each algorithm from Node(0,1) to Node(7,0) & Node(10,3)



Input 2.2: New coordinates of the goal nodes



Output 2.2: New paths are display respectively, from Node(0,3) to Node(7,0) & Node(10,3)



Conclusion: After adjusting the coordinates of one goal node, that unreachable goal node was notified to the user. The remaining paths to the remaining goal node are still displayed ordinarily.

5. Features

- Depth-First Search (DFS): Extremely skilled at finding new paths or locations inside a grid or graph. use a backtracking technique to find new paths and trace back its movements.
- Breath-First Search (BFS): Explore all the nodes from the source at the same level first, ensuring that the nearest nodes are extensively searched first, using a queue to monitor the next nodes to probe in a First In, First Out (FIFO) fashion.
- Greedy Best-First Search (GBFS): Determines the cost or distance to the destination and applies this heuristic to prioritize nodes, hence focusing the search. tries to choose the most direct route to find a solution as soon as possible, often requiring less investigation.
- A Star (A*): Assures the quickest route to the destination by combining heuristic estimates with real path expenses. successfully finds a middle ground between pursuing intriguing directions and getting closer to the goal.
- Custom Search Strategy 1 (CUS1): By altering the cost function, this algorithm tailors the search to specific issue constraints. focuses on cost-effective routes while supervising the node discovery process using a priority queue.
- Custom Search Strategy 2 (CUS2): This algorithm decides which node will be inspected through a unique interaction, considering total information, by utilizing an open list to accommodate node insertion and elimination throughout the searching processes.

6. Conclusion

Which search calculation is best depends on the issue's aims and size: A*, or alternatively BFS, is used for small areas that need meticulous planning; GBFS, or customized approaches, are used for large, effectiveness-focused areas; and DFS is used for simpler, memory-cognitive tasks. The selection characteristics emphasize how important it is to understand the problem, test out several computations, and choose which one best satisfies the objectives by comparing memory requirements, optimality, and proficiency.

7. References

- A fast algorithm for finding better routes by AI search techniques: https://ieeexplore.ieee.org/abstract/document/396824/
- The PN*- Search Algorithm: https://www.sciencedirect.com/science/article/pii/S0004370201000844
- The variants of the harmony search algorithm: https://link.springer.com/article/10.1007/s10462-010-9201-y