1 Suchbäume

Im Folgenden betrachten wir ein typischen Szenario einer studentischen Abschlussarbeit: Nach langem Forschen liegen die Daten des Experiments vor. Leider ist kaum noch Zeit bis zum Abgabetermin. Die Daten müssen also möglichst schnell aufbereitet werden. Wir suchen aus den möglichen Verarbeitungsschritten das schnellste Vorgehen.

Die Verarbeitung besteht aus 3 Hauptschritten: Der Aufbereitung der Daten, den statistischen Tests und der abschließenden Visualisierung. Wir gehen davon aus, dass 3 mögliche Toolboxen für Ihre Tests verfügbar sind (z.B. Matlab, Excel und R). Nicht jede dieser Toolboxen kann mit jeder Aufbereitungstechnik und jeder Visualisierung verwendet werden. Wir gehen weiter davon aus, dass Sie 3 geeignete Aufbereitungstechniken (A1, A2, A3) und 2 Visualisierungstechniken kennen (V1, V2).

Wir kommen also auf folgende Möglichkeiten:

- Excel (T1): Nimmmt Daten der Aufbereitungstechniken A1 und A2 an. Gibt Daten aus, die mit Visualisierungstechnik V1 kompatibel sind.
- Matlab (T2): Nimmt Daten der Aufbereitungstechniken A1, A2 und A3 an, kann Daten an die Visualisierungen V1 und V2 ausgeben.
- R (T3): Nimmt nur Daten der Aufbereitungstechnik A3 an, gibt nur an Visualisierungstechnik V2 aus. Mit einer Konvertierung ("K") können aber auch Daten an V1 weiter gegeben werden.
- Man kann die Aufbereitungstechniken A2 und A3 hintereinander ausführen: Erst A2, dann A3. Die Ergebnisse werden dann wie Ergebnisse von A3 weiter behandelt.

Es sind nun folgende Aufgaben zu erfüllen:

- In einem ersten Schritt zeichnen Sie nun bitte die Struktur des Problems auf. Daraus sollte ersichtbar werden, auf welchen Wegen Sie von den rohen Daten zum fertigen Ergebnis kommen. Kennzeichnen Sie den Startknoten mit "D" und das Ziel mit "E".
- Nachdem Sie so das Problem visualisiert haben, müssen wir einen Suchbaum aufstellen. Nehmen Sie hierfür am Besten ein DIN A4 Bogen quer und verwenden Sie nur die Kürzel (A1, T1, etc.), um genug Platz zu haben.
- Nun sollten wir das Problem also als Suchbaum vorliegen haben. Eine Tiefensuche und eine Breitensuche sind hier gute Fingerübungen. Bei der Expansion eines Knotens werden die Kinderknoten in der Reihenfolge Ihrer Markierung eingetragen (Hierarchie von links nach rechts fallend: A1, A2, A3, K, T1, T2, T3, V1, V2). Führen Sie diese Suchen aus und notieren Sie den jeweils gefundenen Ablauf.

Als ich die Aufgabe entwarf, hatte ich ursprünglich einen möglichen Übergang von T3 auf A2 erwogen.

- Warum habe ich diese Idee wohl wieder verworfen?

Tiefensuche und Breitensuche sind offensichtlich für unsere Szenario nicht sonderlich gut geeignet. Wir gehen nun davon aus, dass Sie schon Erfahrungen im Umgang mit allen verwendeten Tools und Techniken haben. Dazu zählen auch Erfahrungswerte, wie lange Sie typischerweise nach Abschluss eines Arbeitsschrittes bestenfalls für den restlichen Ablauf bis zum Vollenden der Abschlussarbeiten brauchen. Daneben kennen Sie natürlich auch die Zeiten, die Sie für die Durchführung eines Arbeitsschrittes brauchen würden. Diese Werte sind in der folgenden Tabelle gegeben:

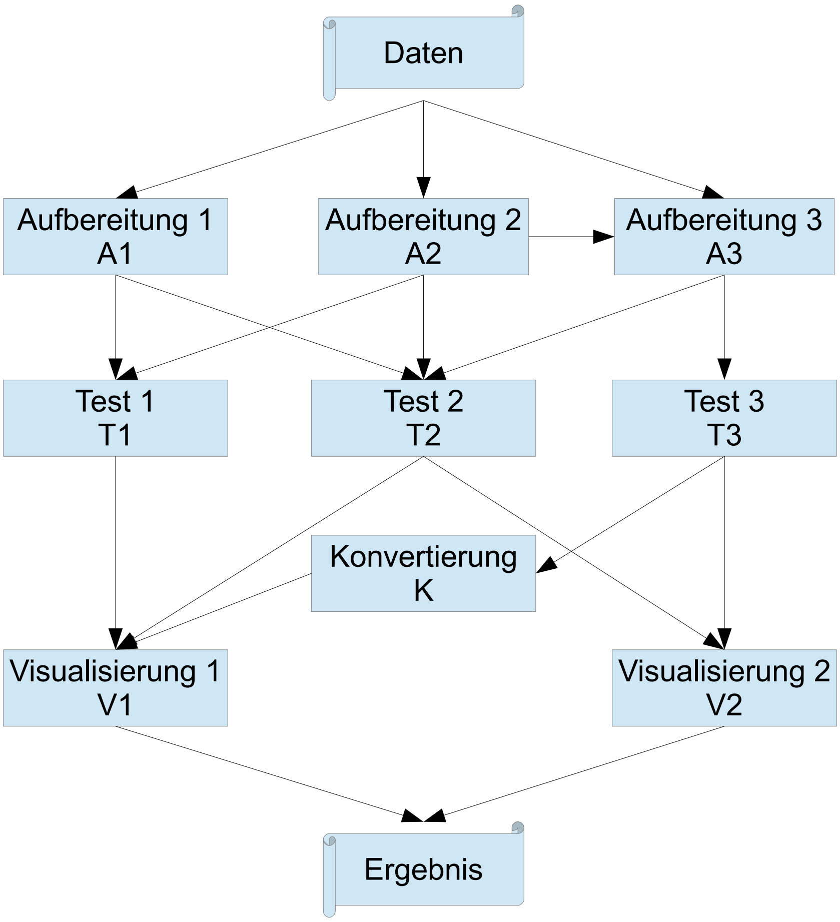
Knoten	Durchführung (min)	Verbleibend (min)
D	0	140
A1	40	170
A2	30	130
A3 (ohne Vorverarbeitung A2)	50	120
A3 (mit Vorverarbeitung A2)	10	120
T1	80	70
T2	70	50
T3	20	110
K	5	100
V1	80	27
V2	90	5
E (Zusammenschreiben etc.)	30	0

Verwenden Sie also diese Werte für eine $A^\star$ Suche nach dem schnellsten Lösungsweg. Gehen Sie davon aus, dass Ihre Erfahrungswerte den tatsächlich verbleibenden Zeitaufwand bis zum fertigen Ergebnis unterschätzen.

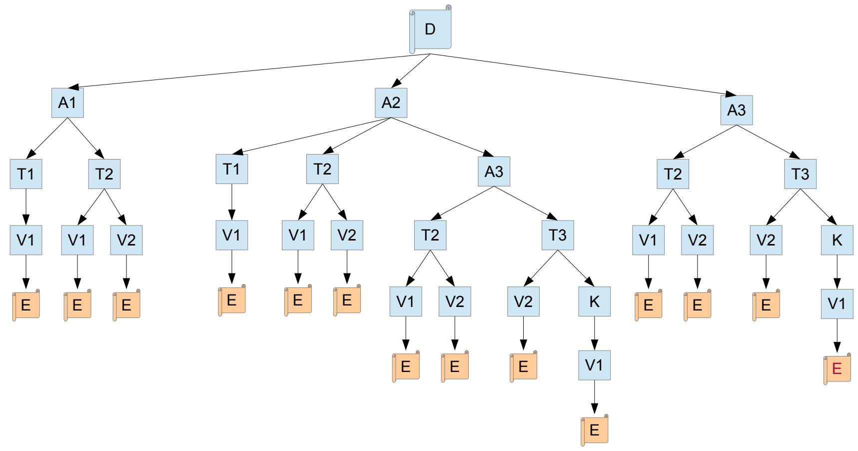
- Führen Sie die $A^\star$ Suche durch.

2 Musterlösung: Suchbäume

Struktur des Problems:



Suchbaum:



Breitensuche:

$$\text{(I)}: \begin{bmatrix} D \Leftarrow \end{bmatrix} \qquad \text{(II)}: \begin{bmatrix} A1(D) \Leftarrow \\ A2(D) \\ A3(D) \end{bmatrix} \qquad \text{(III)}: \begin{bmatrix} A2(D) \Leftarrow \\ A3(D) \\ T1(A1, D) \\ T2(A1, D) \end{bmatrix}$$

$$\text{(IV)}: \begin{bmatrix} A3(D) \Leftarrow \\ T1(A1, D) \\ T2(A1, D) \\ T1(A2, D) \\ T2(A2, D) \\ A3(A2, D) \end{bmatrix} \qquad \text{(V)}: \begin{bmatrix} T1(A1, D) \Leftarrow \\ T2(A1, D) \\ T1(A2, D) \\ T2(A2, D) \\ A3(A2, D) \\ T2(A3, D) \\ T3(A3, D) \end{bmatrix} \qquad \text{(VI)}: \begin{bmatrix} T2(A1, D) \Leftarrow \\ T1(A2, D) \\ T2(A2, D) \\ A3(A2, D) \\ T2(A3, D) \\ T3(A3, D) \\ V1(T1, A1, D) \end{bmatrix}$$

$$\text{(VII)}: \begin{bmatrix} T1(A2, D) \Leftarrow \\ T2(A2, D) \\ A3(A2, D) \\ T2(A3, D) \\ T3(A3, D) \\ V1(T1, A1, D) \\ V1(T2, A1, D) \\ V2(T2, A1, D) \end{bmatrix} \qquad \text{(VIII)}: \begin{bmatrix} T2(A2, D) \Leftarrow \\ A3(A2, D) \\ T2(A3, D) \\ T3(A3, D) \\ V1(T1, A1, D) \\ V1(T2, A1, D) \\ V2(T2, A1, D) \\ V1(T1, A2, D) \end{bmatrix} \qquad \text{(IX)}: \begin{bmatrix} A3(A2, D) \Leftarrow \\ T2(A3, D) \\ T3(A3, D) \\ V1(T1, A1, D) \\ V1(T2, A1, D) \\ V2(T2, A1, D) \\ V1(T1, A2, D) \\ V1(T2, A2, D) \\ V2(T2, A2, D) \end{bmatrix}$$

$$\text{(X)}: \begin{bmatrix} T2(A3, D) \Leftarrow \\ T3(A3, D) \\ V1(T1, A1, D) \\ V1(T2, A1, D) \\ V2(T2, A1, D) \\ V1(T1, A2, D) \\ V1(T2, A2, D) \\ V2(T2, A2, D) \\ T2(A3, A2, D) \\ T3(A3, A2, D) \end{bmatrix} \qquad \text{(XIII)}: \begin{bmatrix} T3(A3, D) \Leftarrow \\ V1(T1, A1, D) \\ V1(T2, A1, D) \\ V2(T2, A1, D) \\ V1(T1, A2, D) \\ V1(T2, A2, D) \\ V2(T2, A2, D) \\ T2(A3, A2, D) \\ T3(A3, A2, D) \\ V1(T2, A3, D) \\ V2(T2, A3, D) \end{bmatrix} \qquad \text{(XIV)}: \begin{bmatrix} V1(T1, A1, D) \Leftarrow \\ V1(T2, A1, D) \\ V2(T2, A1, D) \\ V1(T1, A2, D) \\ V1(T2, A2, D) \\ V2(T2, A2, D) \\ T2(A3, A2, D) \\ T3(A3, A2, D) \\ V1(T2, A3, D) \\ V2(T2, A3, D) \\ V2(T3, A3, D) \\ K(T3, A3, D) \end{bmatrix}$$

$$\text{(XV)}: \begin{bmatrix} V1(T2, A1, D) \Leftarrow \\ V2(T2, A1, D) \\ V1(T1, A2, D) \\ V1(T2, A2, D) \\ V2(T2, A2, D) \\ T2(A3, A2, D) \\ T3(A3, A2, D) \\ V1(T2, A3, D) \\ V2(T2, A3, D) \\ V2(T3, A3, D) \\ K(T3, A3, D) \\ E(V1, T2, A1, D) \end{bmatrix}$$

Sind wir jetzt fertig? Letztlich ist es eine Frage der Implementierung. Wir nehmen an, dass bei der Eintragung der Kinder in die Liste kurz geprüft wird, ob es sich um den Zielknoten handelt. Alternativ ließe sich auch festlegen, dass erst bei Expansion des Knotens geprüft wird. In dem Fall müßten wir zunächst noch die gesamte Liste abarbeiten, bevor wir merken dass wir schon längst das Ziel vor der Nase haben. *In der Klausur kann von der frühen Prüfung ausgegangen werden: Bei Eintragen in die Liste würde der Zielknoten schon als gefunden betrachtet werden.*

Tiefensuche:

$$\text{(I)}: \begin{bmatrix} D \Leftarrow \end{bmatrix} \qquad \text{(II)}: \begin{bmatrix} A1(D) \\ A2(D) \\ A3(D) \Leftarrow \end{bmatrix}$$

$$\text{(III)}: \begin{bmatrix} A1(D) \\ A2(D) \\ T2(A3, D) \\ T3(A3, D) \Leftarrow \end{bmatrix} \qquad \text{(IV)}: \begin{bmatrix} A1(D) \\ A2(D) \\ T2(A3, D) \\ V2(T3, A3, D) \\ K(T3, A3, D) \Leftarrow \end{bmatrix}$$

$$\text{(V)}: \begin{bmatrix} A1(D) \\ A2(D) \\ T2(A3, D) \\ V2(T3, A3, D) \\ V1(K, T3, A3, D) \Leftarrow \end{bmatrix} \qquad \text{(VI)}: \begin{bmatrix} A1(D) \\ A2(D) \\ T2(A3, D) \\ V2(T3, A3, D) \\ V1(K, T3, A3, D) \Leftarrow \end{bmatrix}$$

$$\text{(VII)}: \begin{bmatrix} A1(D) \\ A2(D) \\ T2(A3, D) \\ V2(T3, A3, D) \\ E(V1, K, T3, A3, D) \Leftarrow \end{bmatrix}$$

Und somit hätte auch die Tiefensuche eine Lösung gefunden. Sowohl die Tiefensuche als auch die Breitensuche ignorieren dabei allerdings völlig unseren Zeitdruck. Für den gegebenen Anwendungsfall sind sie also nicht geeignet.

Im Folgenden stellte ich die Frage, warum ich die Querverbindung von T3 auf A2 wieder verwarf. Wenn man versucht, hierfür den entsprechenden Suchbaum zu zeichnen, würde man in eine unendliche Schleife gelangen. Um solche Fälle zu vermeiden, müßten wir noch zusätzliche Regeln einführen, was diese Aufgabe nur unnötig verkomplizieren würde.

Schließlich wollen wir noch die $A^\star$ - Suche durchführen. Wir beginnen wie gewohnt am Start-Knoten D. Anders als in der Breiten- oder Tiefensuche berechnen wir hier die Kosten jedes Knotens aus den gegebenen Werten:

$$\begin{aligned} f[n] &= g[n] + h[n] \\ &= \left(\sum_{p \in P} g[p] + g[i_n] \right) + h[n] \end{aligned}$$

Hier wird also für $g[n]$ die Durchführungszeit $g[i_n]$ des aktuellen Schrittes zu der kumulierten Dauer der vorherigen Schritte $\sum_{p \in P} g[p]$ addiert, wobei P die Gesamtheit vorheriger Schritte bis zum Knoten n bezeichnet. Diese Werte können aus der linken Spalte der oben gegebenen Tabelle abgelesen werden. Die Schätzung $h[n]$ der verbleibenden Restarbeitszeit ist jeweils in der rechten Spalte gegeben.

Im ersten Schritt bzw. Knoten "D", den vorliegenden Daten, ist noch keine Bearbeitung notwendig. Die Durchführungszeit $g[D_n]$ ist also Null, vorherige Schritte gibt es auch keine. Somit ist $g[D] = 0$. Für $h[n]$ lesen wir aus der Tabelle ab: $h[D] = 140$. Wir expandieren den Knoten mit den geringsten Kosten (und in diesem Fall auch der einzige Knoten der Liste) und berechnen die Kosten $f[A1(D)]$, $f[A2(D)]$und $f[A3(D)]$ der Folgeknoten. Der expandierte Knoten wird aus der Liste gestrichen.

$$f[A1(D)] = (0 + 40) + 170 = 210 \tag{1}$$
$$f[A2(D)] = (0 + 30) + 130 = 160 \tag{2}$$
$$f[A3(D)] = (0 + 50) + 120 = 170 \tag{3}$$

Wir suchen nach Doppeleinträgen, finden hier aber keine. Es werden somit keine Einträge der Liste gestrichen. Danach expandieren wir den Knoten mit den geringsten Kosten, in diesem Fall $A2(D)$:

$$f[A1(D)] = \text{keine Neuberechnung notwendig} = 210 \tag{4}$$
$$f[A3(D)] = \text{keine Neuberechnung notwendig} = 170 \tag{5}$$
$$f[A3(A2, D)] = (0 + 30 + 10) + 120 = 160 \tag{6}$$
$$f[T1(A2, D)] = (0 + 30 + 80) + 70 = 180 \tag{7}$$
$$f[T2(A2, D)] = (0 + 30 + 70) + 50 = 150 \tag{8}$$

Hier sehen wir nun einen Doppeleintrag: $A3(A2, D)$ und $A3(D)$ führen über verschiedene Wege zum selben Knoten $A3$. Wir streichen also den teureren Weg, bevor wir den günstigsten Knoten expandieren. Damit haben wir die wesentlichen Elemente des $A^\star$ Algorithmus versammelt:

- Expandiere den günstigsten Knoten des vorherigen Schrittes und schreibe sie zusammen mit den alten, nicht expandierten Einträgen in eine neue Liste. Die alten Einträge können einfach kopiert werden, eine Neuberechnung ist nicht notwendig.

- Berechne für jeden Folgeknoten die Kostenfunktion $f[n]$ aus dem kumulierten bisherigen Weg $g[n]$ und der Heuristik $h[n]$.
- Streiche bei Doppeleinträgen den teureren Weg.

Im Folgenden ist nun die vollständige $A^\star$-Suche für unser Problem gegeben:

$$\text{(I):}\ \begin{bmatrix} D & 140 \Leftarrow \end{bmatrix} \qquad \text{(II):}\ \begin{bmatrix} A1(D) & 210 \\ A2(D) & 160 \Leftarrow \\ A3(D) & 170 \end{bmatrix}$$

$$\text{(III):}\ \begin{bmatrix} A1(D) & 210 \\ \cancel{A3(D)} & 170 \\ A3(A2,D) & 160 \\ T1(A2,D) & 180 \\ T2(A2,D) & 150 \Leftarrow \end{bmatrix} \qquad \text{(IV):}\ \begin{bmatrix} A1(D) & 210 \\ A3(A2,D) & 160 \Leftarrow \\ T1(A2,D) & 180 \\ V1(T2,A2,D) & 207 \\ V2(T2,A2,D) & 195 \end{bmatrix}$$

$$\text{(V):}\ \begin{bmatrix} A1(D) & 210 \\ T1(A2,D) & 180 \\ V1(T2,A2,D) & 207 \\ V2(T2,A2,D) & 195 \\ T2(A3,A2,D) & 160 \Leftarrow \\ T3(A3,A2,D) & 170 \end{bmatrix} \qquad \text{(VI):}\ \begin{bmatrix} A1(D) & 210 \\ T1(A2,D) & 180 \\ V1(T2,A2,D) & 207 \\ V2(T2,A2,D) & 195 \\ T3(A3,A2,D) & 170 \Leftarrow \\ \cancel{V1(T2,A3,A2,D)} & 217 \\ \cancel{V2(T2,A3,A2,D)} & 205 \end{bmatrix}$$

$$\text{(VII):}\ \begin{bmatrix} A1(D) & 210 \\ T1(A2,D) & 180 \\ V1(T2,A2,D) & 207 \\ \cancel{V2(T2,A2,D)} & 195 \\ V2(T3,A3,A2,D) & 155 \Leftarrow \\ K(T3,A3,A2,D) & 165 \end{bmatrix} \qquad \text{(VIII):}\ \begin{bmatrix} A1(D) & 210 \\ T1(A2,D) & 180 \\ V1(T2,A2,D) & 207 \\ \cancel{V2(T2,A2,D)} & 195 \\ K(T3,A3,A2,D) & 165 \Leftarrow \\ E(V2,T3,A3,A2,D) & 180 \end{bmatrix}$$

$$\text{(IX):}\ \begin{bmatrix} A1(D) & 210 \\ T1(A2,D) & 180 \\ \cancel{V1(T2,A2,D)} & 207 \\ E(V2,T3,A3,A2,D) & 180 \\ V1(K,T3,A3,A2,D) & 172 \Leftarrow \end{bmatrix} \qquad \text{(X):}\ \begin{bmatrix} A1(D) & 210 \\ T1(A2,D) & 180 \\ \cancel{E(V2,T3,A3,A2,D)} & 180 \\ E(V1,K,T3,A3,A2,D) & 175 \Leftarrow \end{bmatrix}$$

Es ergibt sich also folgender Weg als schnellste Lösung: Aufbereitung 2 gefolgt von Aufbereitung 3, danach statistische Tests in der Statistik-Software “R” (T3), anschließende Konvertierung (K) zur Visualisierung mit Technik 1 (V1) und zum Abschluss das Zusammenschreiben der Ergebnisse (E). Oder kurz: $D \to A2 \to A3 \to T3 \to K \to V1 \to E$. **Achten Sie bitte stets darauf, nach all den Berechnungen auch Ihr Ergebnis kenntlich zu machen und den Weg aufzuschreiben!!**

Wir haben in dieser Aufgabenstellung glücklicherweise nicht die Situation, zwischen zwei Knoten mit gleichen Kosten wählen zu müssen. In dem Fall wäre es eine Frage der Implementierung des Algorithmus, welchen der verfügbaren Knoten man wählen würde. Zu Zwecken der Übung und der Klausur werden solche Situationen aber bewusst vermieden.

Ein Tipp: Man kann sich wertvolle Rechenzeit sparen, wenn man sich die kumulierten Kosten bis zu jedem Knoten merkt. Wenn man die Kinder dieses

Knotens expandiert, muss man damit nicht die kumulierten Kosten neu aufsummieren. Stattdessen schaut man einfach in die Liste und fügt nur die Kosten des neuen Knotens hinzu!

3 Do It Yourself

Eigentlich ganz einfach: Nehmen Sie ein Blatt, malen Sie 5-8 Kreise drauf und markieren Sie die Kreise mit Buchstaben. Fügen Sie nun ein paar gerichtete Kanten zwischen den markierten Kreisen hinzu. Das ganze entspricht nun einer "Landkarte". Wählen Sie einen Startknoten und einen Endknoten und zeichnen Sie den entsprechenden Suchbaum. Sie werden recht schnell merken, welche Strukturen zu endlosen Schleifen, welche zu wenigen und welche zu unüberschaubar vielen Pfaden führen. Experimentieren Sie, bis Sie einen überschaubaren Suchbaum erstellt haben.

Wenn Sie also einen nicht zu komplexen Suchbaum erstellt haben, können Sie daran einfach die Tiefen- und Breitensuche üben.

Sobald Sie sich mit der Tiefen- und Breitensuche wohl fühlen, können Sie versuchen, den Kanten auf der Landkarte "Gewichte", also Kosten $g[n]$, zuzuordnen. Hiermit können Sie eine einfache "Greedy-Search" ausprobieren, wobei immer nur der günstigste Knoten verfolgt wird und keine Liste geführt wird.

Für die $A^\star$-Suche fehlt allerdings noch ein Ausdruck für die erwarteten Restkosten $h[n]$. Versuchen Sie, den Knoten entsprechende Restkosten zuzuordnen, so dass die tatsächlichen Kosten bis zum Zielknoten tatsächlich unterschätzt werden (nicht immer ganz einfach!). Zum Abschluss können Sie nun den $A^\star$-Algorithmus auf Ihr selbst-gebautes Suchproblem anwenden.