

1.7 Hauptkomponenten MMK: - Bedienkonzept: Aufbau der Dialoge und Darstellung der Menüs -> Softwaretechnik, Ergonomie, intelligente Systeme
- Schnittstellentechnologie: Technologie, mit der Mensch & Maschine kommunizieren (z.B. Tastatur, Spracherkennung) -> Mustererkennung, Signal-, Sprach- & Bildverarbeitung

Typische Eigenschaften intelligenter Systeme:

- Fortgeschrittenes Ein-/Ausgabetechniken (z.B. Sprache, Handschrift, Gestik)
- Intelligentes Systemverhalten und maschinelle Lernfähigkeiten
- Ausgeprägte Dialogfähigkeit und benutzerfreundliches MMI

Wichtigste Komponenten eines interaktiven Systems:

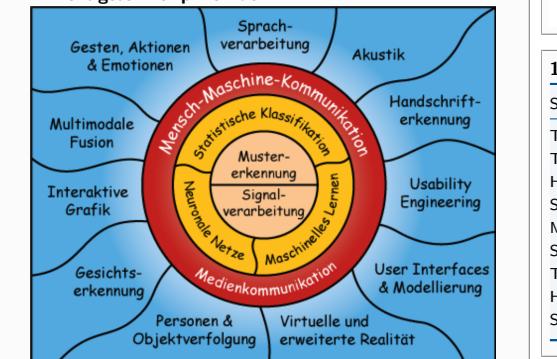
- Dialogkomponente (für Navigation im Menü)
- Benutzerschnittstelle (für Ein-/Ausgabe)

1. Allgemeine Einführung

1.1. Grundbegriffe der MMK

Interaktion	Kommunikation zwischen Mensch und Maschine.
Interaktives System	System, das auf Eingaben reagiert und gegebenenfalls auch Ausgaben generiert.
HCI	Human-Computer Interaction.
MMI	Mensch-Maschine-Interface.
Usability	Gebrauchstauglichkeit bzw. Eignung eines Produkts.
Usability Engineering	Gestaltung und Testen eines Produktes mit dem Ziel optimaler Bedienbarkeit durch die Mensch-Maschine-Schnittstelle.
Software-Ergonomie Medium	Wissenschaft über die Gestaltung von Programmen mit benutzerfreundlicher Mensch-Maschine-Schnittstelle.
Multimedia	Datenverarbeitung und -darstellung unter Nutzung verschiedener Medien, z.B. Text, Grafik und Audio und Video.
Modalität	Ein-/Ausgabekanal der menschlichen Kommunikation und Sinneswahrnehmung, z.B. Sprache, Zeigen, Gestik, Tastatur.

1.2. Wichtigste Disziplinen der MMK



1.3. Trends in der MMK

- Steigerung der Leistungsfähigkeit
- Reduzierung der Kosten
- Erweiterung der Funktionalität
- Verbesserung der Bedienbarkeit

Umrechnung zwischen L_p und N:

Nur bei frequenzreinen Signalen möglich ($f = \text{const}$)
In der Realität i.d.R. nur Annäherungen/Richtwerte angegeben

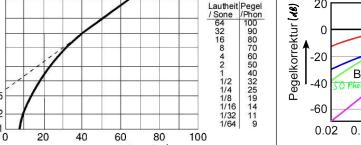
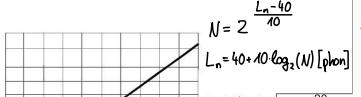
Gründe gegen dB(A):

dB(A) berücksichtigt physiolog. Bedingungen nur ungenügend
-> Mensch empfindet best. Frequenzen störender und Mensch reagiert mit abnehmender Schallintensität empfindlicher -> Lautheit N besser, da dort auch das Spektrum des Schalls berücksichtigt wird

$$L_p = 40 + 10 \log_2(N) [\text{phon}]$$

Lautheit N (Sone):

Subjektive Lautheit von Schall, welche zur proportionalen Abbildung des menschlichen Lautsinnempfindens dient. 1 sone: empfundene Lautstärke bei $f=1\text{kHz}$ und $L=40\text{dB}$. Die Lautheit verdoppelt sich, wenn ein Schall doppelt so laut empfunden wird. Dabei erhöht sich der Schalldruckpegel um 10dB. Breitbandrauschen wird bis zu dreimal lauter wahrgenommen als Schmalbandrauschen mit gleichem Pegel.

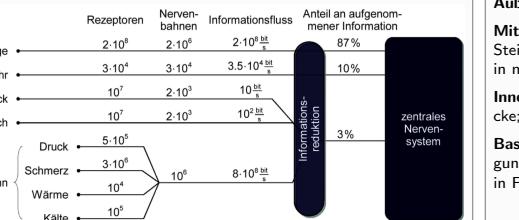


1.4. Übersicht über Sinnesmodalitäten

Sinnesbezeichnung	Modalität	Bemerkung
Sehen	visuell	
Hören	auditiv	
Riechen	olfaktorisch	„5 Sinne“
Schmecken	gustatorisch	
Tasten	taktil ("berührt werden")	
Druck	haptisch ("berühren")	mechanische Modal.
Kraft		
Berührung	taktil	oberflächen-sensitiv
Vibration		
Temperatur	thermorezeptorisch	
Bewegung und Orientierung	kinästhetisch	
Gleichgewicht	vestibular	

Verwendung mehrerer Modalitäten -> Multimodalitäten

1.5. Die Sinne des Menschen und ihre Datenraten



1.6. Datenraten gängiger Systeme der MMK

System	Verhalten	Rate (KByte/sec)
Tastatur (ungeübt)	Eingabe	0.01
Tastatur (geübt)	Eingabe	0.025
Handschrift	Eingabe	0.0025
Spracheingabe	Eingabe	0.01-0.02
Maus	Eingabe	0.02
Sprachausgabe	Ausgabe	0.6
Text lesen	Ausgabe	0.03-0.3
Hören (CD)	Ausgabe	40
Sehen (Video)	Ausgabe	20000

2. Sprachkommunikation

Ermittlung der geäußerten Wortfolge aus einem vorliegenden Sprachsignal und Verarbeitung dieser Information. Die Sprachkommunikation hat größtes Potential aller Eingabemethoden, da sie auch beim Menschen die häufigste und natürlichste Kommunikationsform ist.

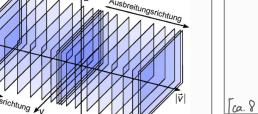
2.1. Physikalische Wellen

Transversalwelle: (z.B. Licht)



Licht: elektromagn. Welle, Schall: Druckwelle

Longitudinalwelle (z.B. Schall):



Bewertungsfilterkurven:

Durch die frequenzbewertenden Eigenschaften des Ohrs hört man niedrigere Frequenzen nicht so „laut“ -> werden nach unten korrigiert

dB(A): Bewerteter Schalldruckpegel (L_p) nach Filterkurve A, welche von der 20-Phon-Kurve abgeleitet ist

dB(B): 20 Phon. die frequenzbewertende Eigenschaft des Ohrs berücksichtigt wird und Filterkurve B zu relativ leisen Geräuschen passt

dB(C): 20 Phon. die frequenzbewertende Eigenschaft des Ohrs berücksichtigt wird und Filterkurve C zu relativ leisen Geräuschen passt

dB(D): 20 Phon. die frequenzbewertende Eigenschaft des Ohrs berücksichtigt wird und Filterkurve D zu relativ leisen Geräuschen passt

Filterkurve A ist die 20-Phon-Kurve abgeleitet

Frequenz <-> Tonhöhe

Amplitude <-> Lautheit

Psychoakustische Größen: Berücksichtigen die Eigenschaften der Wahrnehmung und bieten ein subjektives/bewertetes Maß für das menschliche Ohr. Zur Ermittlung der Empfindungsfunktion werden Versuchsreihen durchgeführt zur Messung der quantitativen Empfindungen. Normierung via Mittelung über viele Versuche und Testpersonen.

2.2. Schallquellen und ihre typischen Pegel

Schall	ggf. Entfernungssangabe	Pegel
Düsengejäger, 30 m		$L = 140 \text{ dB} \equiv p = 2 \cdot 10^{-2} \text{ Pa}$
lautes Händeklatschen, 1 m		$L = 130 \text{ dB} \equiv p = 63 \text{ Pa}$
Trillerpfeife, 1 m		$L = 120 \text{ dB} \equiv p = 20 \text{ Pa}$
Discman		$L = 110 \text{ dB} \equiv p = 6.3 \text{ Pa}$
Presslufthammer, 10 m		$L = 100 \text{ dB} \equiv p = 2 \text{ Pa}$
lauter Fabrikhalle		$L = 90 \text{ dB} \equiv p = 6.3 \cdot 10^{-1} \text{ Pa}$
starker Straßenverkehr		$L = 80 \text{ dB} \equiv p = 2 \cdot 10^{-1} \text{ Pa}$
Staubsauger, 1 m		$L = 70 \text{ dB} \equiv p = 6.3 \cdot 10^{-2} \text{ Pa}$
normale Sprache, 1 m		$L = 60 \text{ dB} \equiv p = 2 \cdot 10^{-2} \text{ Pa}$
Kühlschrank, 1 m		$L = 50 \text{ dB} \equiv p = 6.3 \cdot 10^{-3} \text{ Pa}$
normale Wohngeräusche, 1 m		$L = 40 \text{ dB} \equiv p = 2 \cdot 10^{-3} \text{ Pa}$
Flüstersprache		$L = 30 \text{ dB} \equiv p = 6.3 \cdot 10^{-4} \text{ Pa}$
mechanischer Wecker, 1 m (Ticken)		$L = 20 \text{ dB} \equiv p = 2 \cdot 10^{-4} \text{ Pa}$
Blätterrauschen in der Ferne		$L = 10 \text{ dB} \equiv p = 6.3 \cdot 10^{-5} \text{ Pa}$

2.3. Psychoakustik

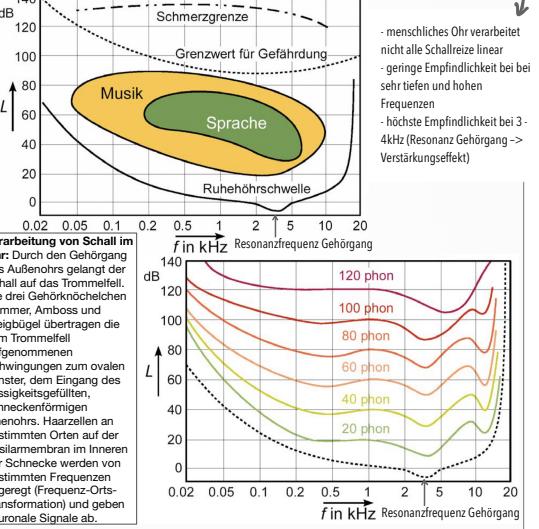
- Empfindlich von etwa 20 Hz - 20 kHz (≈ 10 Oktaven)
- Starke Dämpfung für sehr niedrige und sehr hohe Frequenzen
- Resonanzfrequenz des Gehörgangs bei ca. 3...3.4 kHz;
- Lauteinheit in [sone] 1 sone ≡ Lautheit eines 1kHz Sinus mit 40 dB
- Verhältnistonalhöhe [mel] 1000 mel ≡ 1000Hz

Psychoakustik	Physik		
Bezeichnung	Einheit	Bezeichnung	Einheit
Tonheit Z	Bark	Frequenz f	Hz
Verhältnistonalh. V	Mel	Schalldruck p	$\frac{N}{m^2} = Pa$
		Schallschnelle v	$\frac{m}{s}$
		Schallintensität I	$\frac{W}{m^2}$
Lautstrk.pegel L_n	Phon	Schalldruckp. L	dB , $20 \log(p/p_0)$
Lautheit N	sone	Schalleist. P_ak	$W = \frac{Nm}{s}$

$$\text{Bezugsschalldruck } p_0 = 2 \cdot 10^{-5} \frac{N}{m^2} = 20 \mu Pa$$

$$\text{Bezugsintensität } I_0 = 1 \cdot 10^{-12} \frac{W}{m^2}$$

Hörfläche Jener Frequenz- und Pegelbereich von Schall, der vom menschlichen Gehör wahrgenommen werden kann:



Frequenzbewertung Verfahren zur frequenzabhängigen Anpassung von Schalldruckpegeln an das menschliche Hörempfinden (nichtlinear zur Lautstärke). Hierfür werden verschiedene Filterkurven verwendet: A(20-40 phon), B(50-70 phon), C(80-90 phon), D(sehr hohe Schalldrücke) mit gleichem Lautstärkeindruck. Lautheit N in Sone ist angepasstes Schema.

Frequenzgruppen Begrenztes Frequenzauflösungsvermögen des Gehörs -> best. Frequenzbereiche werden zu sog. Frequenzgruppen zusammengefasst. Jede Frequenzgruppe (insg. 24) nimmt die gleiche Länge auf Basilarmembran ein (1.3mm). Frequenzgruppenbreite: Für $f < 500\text{Hz}$: $\Delta f = 100\text{Hz}$; $f > 500\text{Hz}$: Kleine Terz, d.h. Frequenzverhältnis von 1.19 bzw. 23% der Mittelfrequenz. Bark-Frequenzskala: Im wesentlichen Nummerierung der Frequenzgruppen. Mel-Frequenzskala: Lässt sich aus Tonhöhenempfindung ableiten. Zusammenhang: 1.31 Bark = 131 Mel = 131 Hz.

Verdeckungen Hörschwelle bei Vorhandensein eines Störschalls (Maskierer) (Töne unterhalb der Mithörschwelle sind nicht wahrnehmbar für menschl. Gehör).

- Spektrale: verbreitert sich mit steigendem Pegel überproportional.
- Zeitliche: Vorverdeckung; Simultanverdeckung; Nachverdeckung (etwa hundert ms).

Kompression Audiodecoder: Ausnutzung psychoakustischer Effekte, wie zeitliche und spektrale Verdeckung (-> Berechnung der Mithörschwelle). MP3 ab 160 kBit/s.

Schalldruck p [Pa]: Druckschwankungen in schallübertragendem Medium (Kraft/Fläche) [$p = N/m^2 = Pa$]

Schalldruckpegel L [dB]: Druckschwankungen im gesamten Schalldruckpegel p , $L = 20 \log(p/p_0) \text{ dB}$

Lautstärkepegel L_n [phon]: Ohrgemäße /Empfundene Lautstärke abh. von f und L . 1 kHz Sinus mit $L = 40\text{dB}$ hat $L_n = 40\text{phon}$.

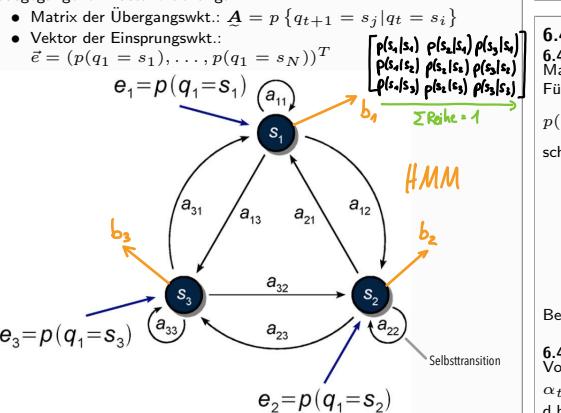
phon. x phon heißt für Töne mit anderen Frequenzen: Ein Sinus mit 1kHz muss x dB haben, um gleich „laut“ zu sein.

6. Hidden-Markov-Modelle und Algorithmen

Wahrscheinlichkeit Statistischer Klassifikator. Liefert Wahrscheinlichkeit p , dass eine Beobachtung einer bestimmten Klasse zuordnet werden kann. Klassifizieren ganze Sequenzen (dynamische Folgen). „Finde diejenige Klasse, die die Beobachtung $o = (o_1, o_2, \dots, o_t)$ am besten nachbilden kann.“.

6.1. Markov-Modelle (MM)

Abbildung stochastischer Prozesse, deren aktueller Zustand nur vom vorangegangenen Zustand abhängt.



6.2. Hidden-Markov-Modelle (HMM)

Stochastische Version eines endlichen Zustandsautomaten; Zustandsübergänge und Symbolemissionen nicht deterministisch.

- Matrix A und Vektor \vec{e} siehe MM
- Beobachtungsfolge: $\vec{o} = (o_1, \dots, o_T)^T$
- Alphabet: $\vec{v} = (v_1, \dots, v_M)^T$
- Beobachtungswahrscheinlichkeiten: $b_{mi} = p(v_m | s_i)$
- Matrix der Beobachtungswahrscheinlichkeiten:

$$B = \begin{bmatrix} 1 & 2 & 3 \\ p(v_1 | s_1) & \dots & p(v_1 | s_N) \\ \vdots & \ddots & \vdots \\ p(v_M | s_1) & \dots & p(v_M | s_N) \end{bmatrix} \quad \sum \text{Spalte} = 1$$

Zusammengefasste Parameter des HMMs: $\lambda = (\vec{e}, A, B)$

Beobachtungs- bzw. Produktionswkt.: $p(\vec{o} | \lambda)$

Dabei durchlaufene (verborgene/hidden) Zustandsfolge:

$$\vec{q} = (q_1, \dots, q_T)$$

HMM - Eigenschaften

Ergodisches HMM: Es kann aus jedem Zustand in jeder andere Zustand erreicht werden; A ist voll besetzt $\forall i, j: a_{ij} > 0$

Links-Rechts-HMM: keine Rücksprünge; kausal; A hat rechte obere Dreiecksform; Graphisch nach rechts aufsteigend $a_{ij} = 0$ für $i > j$

6.2.1. Klassifizierung mit HMM

Pro Klasse ein HMM; das HMM welches die größte Produktionswahrscheinlichkeit $p(o | \lambda_k)$ liefert, repräsentiert die gesuchte Klasse k_x ;

6.2.2. Training von HMM

Kompensation von Störungen; Bed.: geeignete Parameter λ_k ; Training mit iterativen Verfahren; \Rightarrow Baum-Welch-Algorithmus

Tonheit Z ($[Z]$ = Bark) ist die psychoakustische Empfindungsgröße für die Tonhöhe. Sie unterscheidet sich von der physikalischen Frequenz und leitet sich von den Frequenzgruppen auf der Basilarmembran ab.

Verhältnishöhe V ($[V] = \text{Mel}$) dient wie die Tonheit zur Bestimmung der Tonhöhe. Sie ist ebenfalls eine psychoakustische Messgröße.

Frequenz f ($[f] = \text{Hz}$) ist die quantitativ messbare Größe für die Tonhöhe.

Schallbeschleunigung v gibt an, mit welcher Geschwindigkeit die Teilchen eines schallleitenden Mediums schwingen. Sie ist demnach nicht mit der Schallgeschwindigkeit c zu verwechseln ($c_{\text{Luft}} = 340 \text{ m/s}$, während für die Bezugsschallneule $v_0 = 5.0 \cdot 10^{-8} \text{ m/s}$ gilt).

Schallintensität I ($I = \text{W/m}^2 = N(\text{s}\cdot\text{m})$) errechnet sich aus dem Effektivwert des Produktes des Schalldrucks $p(t)$ und der Schallschallneule v_0 : $I = 1/2 \cdot p(t) \cdot v_0^2$.

Schalleistung P ($P = \text{J/s}$) beschreibt die pro Zeiteinheit von einer Schallquelle abgegebene Energie. $P_{\text{ak}} = p \cdot A$ (A : durchsetzte Fläche der Schallquelle)

6.3. HMM in der Spracherkennung

Cepstrum; Merkmalsextraktion; 12D Merkmalsvektor;

6.3.1. Modelle

Einzelwortverkerner vs. fließende Sprache; Phoneme, kleinste bedeutungsunterscheidenden Lauteinheiten; HMM pro Phonem; Pausen;

6.3.2. Training

Zusammensetzung der Phonen HMM zu einem HMM;

6.3.3. Erkennung

Wörterbücher, Grammatiken, Wahrscheinlichkeiten bestimmter Phonemkombinationen, Sprachmodelle für Wortkombinationen;

6.4. HMM-Algorithmen

6.4.1. Trellis / Brute-Force

Mathematische Formel zur Berechnung der Beobachtungswkt. Für verschiedene Wege g gilt:

$$p(\vec{o}, \vec{q} | \lambda_k) = e_{q_1} b_{q_1}(o_1) \prod_{t=2}^T a_{q_{t-1} q_t} b_{q_t}(o_t) \quad \text{Beobachtungswkt.}$$

Benötigte OPS $\sim 2T \cdot N^T$ (sehr rechenintensiv)

6.4.2. Vorwärts-Algorithmus

Vorwärts-Wahrscheinlichkeit:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = s_i | \lambda_k)$$

d.h. die Wahrscheinlichkeit, dass die Teilbeobachtung o_i emittiert werden und das sich das HMM zu t im Zustand s_i befindet;

Vorwärts-Algorithmus (Rekursiv)

1. Initialisierung:

$$\alpha_1(i) = e_i b_i(o_1), \quad 1 \leq i \leq N \quad N: \text{Anzahl Zustände}$$

$$\psi_1(i) = 0$$

2. Induktion:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) = \left[\sum_{i=1}^N \alpha_t(i) \cdot p(s_j | s_i) \right] p(o_{t+1} | s_j)$$

$$1 \leq t \leq T-1; \quad 1 \leq j \leq N; \quad \alpha_{t+1} = (A^T \cdot \alpha_t) \times b(o_{t+1})$$

3. Terminierung:

$$P(o | \lambda_k) = \sum_{i=1}^N \alpha_T(i)$$

Benötigte OPS $\sim T \cdot N^2$ $N: \text{Anzahl Zustände}; T: \text{Länge der Beobachtung}$

Hilfsgrößen für Baum-Welch:

$$\gamma_t(i) \equiv \text{„alle Aufenthalte im Zustand } s_i“ \quad \sum_{t=1}^{T-1} \xi_t(i, j) \equiv \text{„alle Übergänge vom Zustand } s_i \text{ zum Zustand } s_j“$$

Parameterschätzung mit Baum-Welch:

$\hat{\epsilon}_t \equiv \text{„Wahrscheinlichkeit zum Zeitpunkt } t \text{ im Zustand } s_i \text{ zu sein“} = \gamma_t(i)$

$$\hat{a}_{ij} \equiv \frac{\text{„alle Übergänge vom Zustand } s_i \text{ zum Zustand } s_j“}{\text{„alle Aufenthalte im Zustand } s_i“} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

6.4.3. Baum-Welch-Algorithmus

Rückwärtswahrscheinlichkeit:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = s_i, \lambda_k); \quad \text{d.h. Wahrscheinlichkeit, die restlichen Teilbeob. zu emittieren};$$

Rückwärts-Algorithmus (Rekursiv)

1. Initialisierung

$$\beta_T(i) = 1 \quad 1 \leq i \leq N$$

2. Induktion

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

$$t = T-1, T-2, \dots, 1 \quad 1 \leq i \leq N$$

Baum-Welch-Algorithmus (Training/Bestimmung der Parameter (a,b))

Wahrscheinlichkeit, dass sich das HMM zu t im Zustand s_i befindet und o emittiert wird; Summe drüber \Rightarrow „alle Aufenthalte im Zustand s_i “

$$\frac{p(q_t = s_i | \vec{o}, \lambda_k)}{p(\vec{o} | \lambda_k)} = \gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)} = p(q_t = s_i | \vec{o}, \lambda_k)$$

Wahrscheinlichkeit, dass sich das HMM zu t in s_i und zu $t+1$ in s_j befindet; Summe drüber \Rightarrow „aller Übergänge von s_i zu s_j “

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)} = \frac{P(q_t = s_i, q_{t+1} = s_j | \vec{o}, \lambda_k)}{P(\vec{o} | \lambda_k)}$$

7. Suchverfahren

Formulierung und Darstellung eines Problems im Zustandsraum; Graphen-Darstellung; Suchbaum;

Zyklische Wiederholungen unterbinden (gerichtete Kanten im Baum).

7.1. Allgemeiner Algorithmus für Suche

Suchalgorithmus

(Tree search, d.h. kein Explored-Set)

1. Initialisiere Queue

2. Schreibe Startknoten in Queue

3. Wiederhole:

Art des Algorithmus bestimmt die Art der Queue, und damit die Update-Funktion:

Suchalgorithmus

Art der Queue

Breitensuche FIFO-Queue

Tiefensuche LIFO-Queue (Stack)

A-Suche Prioriti-Queue

A*-Suche Prioriti-Queue mit heuristischen Kosten als Priorität

Dijkstra Prioriti-Queue mit bisherige Weg als Heuristik

7.2. Tiefensuche und Breitensuche

1. einelementige Liste mit Wurzelknoten

2. bis Liste leer / Ziel erreicht:

Voraussetzung: Elemente der Warteliste werden systematisch erzeugt; Suchtiefe wird geeignet groß festgesetzt / ausgewertete Suchbaum muss gespeichert werden;

7.3. Heuristische Suche / A-Algorithmus

Verarbeitung zusätzlicher Informationen; Bewertungsmöglichkeit für Erfolgsaussichten eines bestimmten Pfades; Entscheidungen ordnen; Vielversprechende Alternative zuerst, „dem atm billigsten folgen“; Heuristik besteht in Definition einer geeigneten Bewertung (Kostenfunktion) $f(n)$:

$$f(n) = g(n) + h(n)$$

$g(n)$: bisherige Kosten des Knotens

$h(n)$: geschätzte Kosten bis zum Ziel

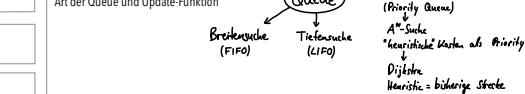
Falls $h(n) = 0$ gewählt wird identisch zur Dijkstra-Algorithmus (wenn dazu $g(n) = \text{Tiefe des Knotens}$ identisch zur Breitensuche)

7.4. A*-Algorithmus

tatsächliche Kosten zum Erreichen des Zielknotens von Knoten n aus Schätzfunktion $h(n)$ monoton, d.h. \downarrow werden nicht überschätzt; terminiert wenn Zielknoten gefunden und keine geringere Kostenabschätzung existiert; A* somit optimaler Pfad; wird die optimale Kostenfkt. $h^*(n)$ verwendet, so wird kürzester Pfad auf Anhieb gefunden (sprich: informierte Suche); Liste mit allen Elementen erstellen + sortieren; dem insg. billigsten folgen; nix verworfen. $0 \leq h(n) \leq h^*(n)$ $h^*(n)$: tatsächliche Kosten bis zum Zielenknoten

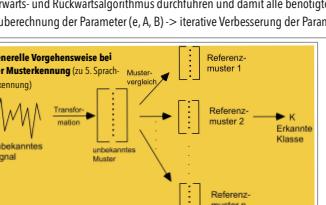
Unterschiede zwischen den Algorithmen:

Art der Queue und Update-Funktion



Tiefensuche: Die Liste der noch zu traversierenden Knoten wird im LIFO-Prinzip abgearbeitet, d.h. der zuletzt eingefügte Knoten wird zuerst expandiert.

Breitensuche: Die Liste der noch zu traversierenden Knoten wird im FIFO-Prinzip abgearbeitet, d.h. der zuerst eingefügte Knoten wird zuerst expandiert.



Moderne Methoden der Sprachverarbeitung:

RNN-Netze, MLP-Netze, Rekurrente NNs (z.B. LSTM), graphische Modelle

Typische Fehler (allgemein) - auf Folgendes sollte geachtet werden:

- Bei Suche Ergebnis (Pfad) immer angeben.

- Einsturzspur bei Automaten/HMMs nicht vergessen.

8. Logik und Theorembeweisen

Wissen algorithmisch darstellen; Fakten ableiten; Behauptungen bestätigen / widerlegen;

8.1. Aussagenlogik

atomare Aussagen; wahr oder falsch; UND , ODER, NICHT; Implikation \Rightarrow :

8.2. Prädikatenlogik

Analyse und Bewertung von Beziehungen und logischen Verknüpfungen
1. Ordnung \Rightarrow nur Veränderung von Objekten, nicht Prädikaten
Prädikate und Funktionen, Konstanten, Variablen, Funktionen, Negation, Disjunktion, Konjunktion, Existenz-Quantor, All-Quantor, Implikation, Äquivalenz.

Beispiel: „In jeder Stadt gibt es einen Bürgermeister“
 $(\forall x) \{ \text{Stadt}(x) \Rightarrow (\exists y) [\text{Mensch}(y) \cdot \text{Bgm}(x, y)] \}$

Regeln und Zusammenhänge aufstellen; \Rightarrow Regelwerk (Axiome); Frage (Theorem); Beweis durch Wahrheitstabelle oder Umformen der Regeln und Schlussfolgerungen (Resolution, Unifikation - effektiver);

a	w	w	f	t	Ausdruck	Name	Umgangssprachlich
b	w	f	w	w	$\neg a$	Negation	Nicht A
w	f	f	w	w	$a \cdot b$	Konjunktion	A und B
w	w	w	f	w	$a + b$	Disjunktion	A oder B
f	w	w	f	w	$a \Leftrightarrow b$	Exkl. Disjunktion	Entweder A oder B
w	w	f	w	w	$a \Rightarrow b$	Implikation	Wenn A, dann B
w	w	f	w	w	$a \Leftrightarrow b$	Bijunktion	A genau dann, wenn B
w	w	w	w	w	Tautologie	Stets wahr	
f	f	f	f	f	Kontradiktion	Stets falsch	

Umformregeln

1. Doppelte Negation $\neg\neg A \equiv A$
2. Idempotenz $A + A \equiv A$ und $A \cdot A \equiv A$
3. Kommutativität $A + B \equiv B + A$ und $A \cdot B \equiv B \cdot A$
4. Assoziativität $A + (B + C) \equiv (A + B) + C$ und $A \cdot (B \cdot C) \equiv (A \cdot B) \cdot C$
5. Distributivität $A + (B \cdot C) \equiv (A + B) \cdot (A + C)$ und $A \cdot (B + C) \equiv (A \cdot B) + (A \cdot C)$
6. De Morgan $\neg(A \cdot B) \equiv \neg A + \neg B$ und $\neg(A + B) \equiv \neg A \cdot \neg B$
7. Kontraposition $A \Rightarrow B \equiv \neg B \Rightarrow \neg A$
8. $A \Rightarrow B \equiv \neg A + B$
9. $A \Rightarrow B \equiv (A \Rightarrow B) \cdot (B \Rightarrow A) \equiv (A \cdot B) + (\neg A \cdot \neg B)$
10. $\neg(\forall x)A(x) \equiv (\exists x)(\neg A(x))$
11. $\neg(\exists x)A(x) \equiv (\forall x)(\neg A(x))$
12. $(\forall x)(A(x) \cdot B(x)) \equiv (\forall x)A(x) \cdot (\forall y)B(y)$
13. $(\exists x)(A(x) + B(x)) \equiv (\exists x)A(x) + (\exists y)B(y)$

8.3. Standardformen

Konjunktive Normalform (KNF):

$$(A_1 + A_2 + \dots) \cdot (B_1 + B_2 + \dots) \cdot \dots$$

Disjunktive Normalform (DNF):

$$(A_1 \cdot A_2 \cdot \dots) + (B_1 \cdot B_2 \cdot \dots) + \dots$$

Regeln zur Umformung in Normalform:

1. Eliminierung aller Äquivalenzen (# 9)
2. Eliminierung aller Implikationen (# 8)
3. Einziehung der Negation nach innen (#6, #10, #11)
4. Einführung neuer Variablen für jeden Quantifizierer
5. Eliminierung aller Existenz Quantoren mithilfe von Skolem-Funktionen
6. Ausklammern der All-Quantoren und Entfernen dieser
7. Anwendung des Distributivgesetzes zur Transformation in Konjunktive Normalform (#5)
8. Eliminierung der UND-Verknüpfungen durch Auflistung der Klauseln
9. Einführung getrennter Variablen für jede Klausal

8.4. Theorembeweis mit Resolutionsverfahren

Allgemeines Resolutionsgesetz:

$$(X + A) \cdot (\neg X + B) \equiv (X + A) \cdot (\neg X + B) \cdot \underbrace{(A + B)}_{\text{Resolvente}}$$

Spezielles Resolutionsgesetz:

$$(X + A) \cdot (\neg X + A) \equiv A$$

Absorptionsgesetz:

$$(A + B) \cdot A \equiv A$$

Weitere Sonderfälle:

$$\begin{aligned} 1. \quad A &\Rightarrow B \equiv \neg A + B & R \equiv B \\ 2. \quad A + B &\quad \neg A + B & R \equiv B + B = B \\ 3. \quad A &\quad \neg A & R \equiv NIL \\ 4. \quad A \Rightarrow B \equiv \neg A + B && \\ B \Rightarrow C \equiv \neg B + C && R \equiv \neg A + C \equiv A \Rightarrow C \end{aligned}$$

Anwendung beim Theorembeweis:

Geg.: Satz von n existierenden und bewiesenen Axiomen $S = \{S_1 \dots S_n\}$; Es gilt T zu beweisen

Vorgehen: Erweiterung von S zu $S^* = \{S_1 \dots S_n, \neg T\}$ Und Resolutionen bis leere Klausal erzeugt wird.

Erklärung: Statt Beweis wird Unerfüllbarkeit seines Gegenteils gezeigt.

Tautologie beweisen

Beweis durch Widerspruch: $\neg w \equiv f$:

1. Negiere Aussage
2. Bringe negierte Aussage in KNF
3. Zeige Kontradiktion durch Resolutionsverfahren

9. Wissensrepräsentation

effizient speichern; strukturiert darstellen; Menge von Fakten, Regeln, Prozeduren, Modellen, Daten, Heuristiken; interpretierbar mit Hilfe von Repräsentationsmechanismen;

9.1. Prädikatenlogik

Aufteilung in Fakten und Regeln; Standardisiert durch KNF; Resolution als Inferenzmechanismus; Formulierung aufwändig und unnatürlich; zwingend Umformung in KNF;

9.2. Produktionsregeln

keine Umformung in KNF; Wenn-Dann bleibt erhalten; Vorwärts-Rückwärtsverkettung als Inferenzmechanismus; Darstellung im UND/ODER-Graphen; Fakten als Blatt, Regeln als Verzweigung;

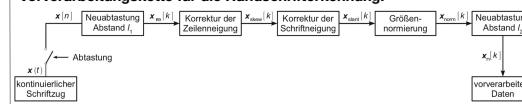
9.3. Semantische Netze

Graphische Modelle zur Darstellung von Wissen über Beziehungen zw. Objekten; entsprechen etwa Fakten der Prädikatenlogik; Knoten = Objekte; Kanten = Prädikate; Verwendung bei natürlichsprachlichen Systemen; keine 2 Knoten gleicher Beschriftung; Richtung der Kanten von Bedeutung;

9.4. Rahmen / Frame

Darstellung der Zerlegung von Objekten oder Situationen in ihre Bestandteile; Ähnlichkeit zu semantischen Netzen, wesentlich mächtiger und flexibler; FrameName - zentraler Knoten, Slots - Kanten, Filler - Knoten; 1. Suchverfahren zur Ermittlung von Beziehungen; 2. „Rahmen-Abgleich“; Fakten als Fragezeichen markiert; mit aktuellen Daten auffüllen;

Vorverarbeitungskette für die Handschrifterkennung:



10. Handschrifterkennung

10.1. Vorverarbeitung

Eingabemethoden

1. freie Eingabe (hohe Vorverarbeitung)
2. liniengeführte Eingabe
3. feldgeführte Eingabe

Eingangssignal: $\vec{x}(t) = (x(t), y(t), p(t))^T$

x(t)	x-Koordinate
y(t)	y-Koordinate
p(t)	Druck (des Stifts)



Schriftgröße

1. Schätzen der Referenzlinien
2. Berechnung der Kernhöhe
3. Normierung des Schriftzuges

W: Höhe der Bins, P: Projektionsprofil

Oberlängenlinie: $y_{ober} = y_{max}$, Unterlängenlinie: $y_{unter} = y_{min}$

Kernlinie: $y_{kern} = \operatorname{argmin}_{\Delta x} (\frac{d}{dx} P_y(j)) - 0.5)W + y_{min}$

Basislinie: $y_{grund} = \operatorname{argmax}_{\Delta x} (\frac{d}{dx} P_y(j)) - 0.5)W + y_{min}$

Kernhöhe: $h_{kern} = |y_{kern} - y_{grund}|$

Normierung:

$$x_{norm}[k] = \frac{1}{h_{kern}} \left[y[k] - (y_{grund} + \frac{h_{kern}}{2}) \right]$$

10.2. Merkmalsextraktion

Extraktion aus dem normalisierten Schriftzug $Sgn(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$

$$\theta[k] = \frac{\pi}{2} + \begin{cases} \arctan(\frac{\Delta y}{\Delta x}) - \frac{\pi}{2} \operatorname{sgn}(\Delta x) & \text{für } \Delta x \neq 0 \\ \frac{\pi}{2} (1 - \operatorname{sgn}(\Delta x)) & \text{für } \Delta x = 0 \end{cases}$$

$$\Delta x = x_{norm}[k+1] - x_{norm}[k], \Delta y = y_{norm}[k+1] - y_{norm}[k]$$

Richtungsänderung:

$$\Delta\theta[k] = \theta[k+1] - \theta[k]$$

$$\text{5-dim. Merkmalsvektor: } \vec{m}[k] = \begin{bmatrix} \sin(\theta[k]) \\ \cos(\theta[k]) \\ \sin(\Delta\theta[k]) \\ \cos(\Delta\theta[k]) \\ p[k] \end{bmatrix}$$

10.3. Erkennung

Training und Erkennung läuft über Hidden-Markov-Modelle (HMM) mit Graphemen (z.B. Buchstabe, Sonderzeichen od. Ziffern) als kleinste Einheit

Training: Baum-Welch-Algorithmus

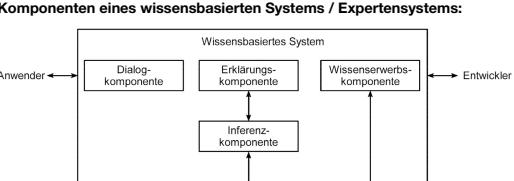
Erkennung: Viterbi-Algorithmus

11. Dialogsystem (Anhang)

- fortgeschrittene intuitive Ein-/Ausgabetechniken
- Hohes Maß an Interaktivität durch Benutzerfreundlichkeit und ausgeprägte Dialogfähigkeit
- Intelligentes Systemverhalten, selbstständig logische Schlüsse ziehen;

Teilgebiete der KI: Maschinelles Lernen, Bildverstehende Systeme, Expertensysteme, Robotik, Logik und automatisches Beweisen, natürlichsprachliche Systeme;

Komponenten eines wissensbasierten Systems / Expertensystems:



Verfahren zur Neuabtastung:

- Lineare Interpolation [mit euklidischem Abstand, $d(x,y) = \|y - x\|_2$] \rightarrow Sinnvoll bei hohen Abstraktionen
- Polynomielle Interpolation (z.B. quadratische). Berechnung des Abstands l der Punkte mit Wegintegral $\int_{x_1}^{x_2} \sqrt{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2} dt$

Lineare Interpolation (eukl. Abstand):

