Rekursions- und Lerntheorie

WiSe 2010/11; Univ. Trier

Henning Fernau
Universität Trier
fernau@uni-trier.de

Rekursions- und Lerntheorie Gesamtübersicht

1. Einführung: Grundsätzliche Betrachtungen
2. Berechenbarkeitstheorie: Die Churchsche These (4 VL)
3. Ausblicke auf weitere Ergebnisse der Rekursionstheorie
4. Lerntheorie: Modelle und Aussagen

Motivation

Lernen ist eine der offensichtlichen Ausprägungen von Intelligenz.
(Mathematische) Modelle für “Lernen” können hilfreich bei der Erklärung von Lernverhalten bei Menschen und auch anderen Lebewesen sein.
$\leadsto$ Zusammenhänge mit Neurologie, Psychologie, Pädagogik usw.

Können wir “intelligente Maschinen” bauen?
Lerntheorie will helfen, Möglichkeiten und Grenzen dieses Vorhabens im Zusammenhang mit “Lernen” auszuloten.
Denkbare Anwendungen:
— Entwicklung autonomer Systeme
— Unterstützung beim Erkennen von kausalen Zusammenhängen, insbesondere im Bereich der Naturwissenschaften.

Aufgaben aus dem Bereich der Naturwissenschaften

Ein Forscher mag an dem Zusammenhang zweier physikalischer Größen in einem Experiment interessiert sein.
Seine Messreihe liefert ihm Zahlenpaare, z.B.:
$(2,4)$, $(0,0)$, $(3,9)$, $(1,1)$, ...
Die bisherigen Ergebnisse könnten den Forscher zu der Hypothese leiten, er habe es mit einem "quadratischen Zusammenhang" der Form (n, n^2) der beiden Messgrößen zu tun.
Es ist wünschenswert, solcherlei Hypothesen automatisch gestützt zu erzeugen.

Was benötigt man, um ein Lernparadigma anzugeben?

(1) Welche Wirklichkeiten sind theoretisch möglich?
(2) Welche Art von Hypothesen dürfen geäußert werden?
(3) Von welcher Natur sind die zur Verfügung stehenden Daten über die Wirklichkeit?
(4) Welche Möglichkeiten besitzt der Forscher?
(5) Wann gilt die Arbeit des Forschers als erfolgreich?

Lernen von Funktionen

(1) (Total) rekursive Funktionen
(2) Gödelnummern (für partiell rek. Fkt.)
(3) Wertepaare (Auflistung)
(4) Am einfachsten wären prinzipiell beliebige “Macht” für den Forscher oder aber die Modellierung des Forschers durch eine Turing-Maschine.
(5) Wir werden einen geeigneten “Grenzwertbegriff” benutzen.

Mögliche Anwendungen

- Modellierung eines Naturwissenschaftlers (s.o.)
- Spracherwerb: Zusammenhang zwischen Wörtern und deren Bedeutung
- automatische Programmsynthese (Programming by Examples)
  Inwieweit ist es möglich, ein Programm aus dem beobachteten Ein- / Ausgabeverhalten zu rekonstruieren?
  Diese Ideen werden heute auch industriell eingesetzt, um Schaltkreise (meist aufgefasst als endliche Automaten) zu überprüfen.

Sprachenlernen von sog. Informanten

In der Vorlesung über Lernalgorithmen wird auf Sprachenlernen fokussiert.
Dort wird auch ein Lernmodell betrachtet, bei welchem dem Lerner sowohl positive als auch negative Beispiele zur Verfügung gestellt werden.
Dies bedeutet formal einen Beispielstrom von mit 0 bzw. 1 markierten Wörtern.
In der Sichtweise des Funktionenlernens entspricht dies der charakteristischen Funktion der Menge als Lernziel.

Dies ist eine weitere Möglichkeit, Funktionenlernen anzuwenden im Bereich des Spracherwerbs.

Zur Darstellung von *Beispielströmen*:

Ein *(Funktionen-)Text* sei eine Abbildung $T : \mathbb{N} \mapsto \mathbb{N} \times \mathbb{N}$ mit der Eigenschaft:
$(T(i) = (a, b) \wedge T(j) = (a, c)) \implies b = c$.
—Die Bedingung modelliert den funktionalen Zusammenhang.
—Wir verzichten also auf ein Pausenzeichen.

Der *Inhalt* von T: $\mathrm{Inh}(T) = \{T(i) | i \in \mathbb{N}\}$.
T heißt *Text für* $f : \mathbb{N} \to \mathbb{N}$, falls $\mathrm{Inh}(T) = \mathrm{Graph}(f)$.
$\mathrm{Graph}(f) = \{(n, f(n)) \mid n \in \mathbb{N}\}$ ist der *Graph* der Funktion $f : \mathbb{N} \to \mathbb{N}$.

Hinweis: Für jede nicht-konstante Funktion gibt es unendlich viele Texte, denn
(1) die Anordnung ist beliebig, und (2) Wiederholungen sind zulässig.

Setze $T[n] = (T(0), \ldots, T(n-1))$.
$SEG = \{T[n] | T \text{ Funktionen-Text}, n \in \mathbb{N}\}$ ist die Menge der Graphen aller endlichen Funktionen.

Die *Länge* von $\sigma \in SEG$ werde mit $|\sigma|$ bezeichnet.
Bem.: T beginnt mit σ genau dann, wenn $T[|\sigma|] = \sigma$.

Bem.: $T[n]$ ist die Datenmenge, die den Lerner bis zum n-ten Zeitpunkt zur Verfügung steht.

Gehen wir von einer fixierten Nummerierung aller Turingmaschinen aus, so können wir einen *Lerner*, auch *Inferenzmaschine (IM)* genannt, als (zunächst beliebige) partielle Funktion $F : SEQ \text{—}\circ\!\!\rightarrow \mathbb{N}$ verstehen.
Notation: $F(T[n]) \downarrow$: F ist definert auf $T[n]$, das heißt, der Lerner äußert eine Hypothese nach Beobachtung von $(T(0), \ldots, T(n-1))$.
Zugehörige Funtion bzw. deren Graph: $h(F(T[n]))$.

F *konvergiert* auf Text T, falls es einen *Grenzwert* $i \in \mathbb{N}$ gibt, so dass $\exists\, n_0 \in \mathbb{N}\ \forall\, n \geq n_0 : F(T[n]) = i$.
In Zeichen: $F(T) \downarrow = i$.

F *identifiziert* Text T genau dann, wenn
$\exists i \in \mathbb{N} : F(T) \downarrow = i \wedge h(i) = Inh(T)$.

F *identifiziert* $L \in \mathcal{E}$ genau dann, wenn
$\forall$ Texte $T : Inh(T) = h \Rightarrow F$ identifiziert T.

F *identifiziert* eine Funktionenklasse $\mathcal{F} \subseteq \mathcal{R}$ genau dann, wenn
$\forall h \in \mathcal{F} : F$ identifiziert h.
$\mathcal{R}$ ist die Klasse der rekursiven Funktionen.

Wir werden sehen, dass es keine nicht identifizierbaren Funktionenklassen $\mathcal{F} \subseteq \mathcal{R}$ gibt.

Alles ist möglich . . .

Satz 1: $\mathcal{R}$ ist identifizierbar.

Beweis: Für $\sigma \in \mathrm{SEG}$ sei $F(\sigma)$ die kleinste Zahl $i \in \mathbb{N}$, sodass:
(1) h_i ist total;
(2) $\mathrm{Inh}(\sigma) \subseteq \mathrm{Graph}(h_i)$.
F rät also das "erste" Element von $\mathcal{R}$, das konsistent mit σ ist.
Es sei T ein Text für die Zielfunktion $f \in \mathcal{R}$.
Ist die Hypothese $F(T[n])$ falsch, d.h., $h := h(F(T[n])) \neq f$, so gibt es $x, y, z \in \mathbb{N}$ mit $f(x) = y$ und $h(x) = z$ mit $y \neq z$.
Da F konsistent mit den gesehenen Daten, wurde (x, y) noch nicht präsentiert.
Spätestens mit der Präsentation von (x, y) muss die Hypothese wieder geändert werden.
Da irgendwann einmal alle Zahlen bis zur kleinsten Gödelnummer m von f durchprobiert wurden, konvergiert das Verfahren mit $\lim_{n\to\infty} F(T[n]) = m$ für jeden Text T für f.

Konsequenzen, Bemerkungen und Beispiele

Folgerung: Jede Teilklasse von $\mathcal{R}$ ist identifizierbar.

Folgerung: Jede Teilklasse der rekursiven Sprachen kann durch Informanten gelernt werden.

Bemerkung: Die Definition des Lerners im vorigen Beweis ist weder effizient noch effektiv. Tatsächlich ist es von entscheidender Bedeutung, dass (vorläufig) unsere Forscher beliebig mächtig sind.

Beispiel: Weiß man mehr über die konkrete Funktionenklasse, so kann der Forscher natürlich viel effizienter arbeiten. Gibt ihm seine Theorie beispielsweise vor, dass es sich um eine quadratische Funktion handeln muss, genügt die Angabe dreier Stützpunkte.

Effektive Forscher

Wir wollen jetzt eine (recht natürliche?) Einschränkung für “Forscher” betrachten:
Im Folgenden seien Forscher *effektiv*, d.h., sie seien durch Turing-Maschinen modellierbar.

Wir werden sehen:
Diese scheinbar harmlose Einschränkung bedeutet, dass erheblich weniger Sprachen bzw. Funktionen identifizierbar sind.

Es bezeichne $\mathbf{Lang}$ die Klasse aller identifizierbaren Sprachfamilien und $\mathbf{TxtEx}$ die Klasse aller *rekursiv-identifizierbaren*, also durch TM identifizierbaren Sprachfamilien.

Satz 2: Es gibt eine identifizierbare, aber nicht rekursiv-identifizierbare Sprachfamilie.
Kurz: $\mathbf{TxtEx} \subsetneq \mathbf{Lang}$.

Wir werden für diesen wichtigen Satz zwei Beweise deutlich unterschiedlicher Natur zeigen.

Der erste Beweis ist kürzer und arbeitet mit einem Kardinalitätsargument, liefert aber im Gegensatz zum zweiten Beweis kein konkretes Beispiel für die behauptete Echtheit der Inklusion.

Ein “beispielloser” Beweis für Satz 2

Für $i \in \mathbb{N}$ und $X \subseteq \mathbb{N}$ definiere: $L_{i,X} = \{\langle i, x\rangle \mid x \in X\}$.
Zu $Q \subseteq \mathbb{N}$ def. Sprachfamilie $\mathcal{L}_Q = \{L_{i,\mathbb{N}} \mid i \in Q\} \cup \{L_{i,D} \mid i \notin Q, D \subset \mathbb{N}, |D| < \infty\}$.
(1) $\mathcal{L}_Q \in \mathbf{Lang}$ für jedes Q:
Das erste Beispiel $z = \langle i, x\rangle$ gestattet bereits den Test “$i \in Q$?”;
im positiven Fall, d.h., $i \in Q$, wird $L_{i,\mathbb{N}}$ als Hypothese geäußert und
im negativen Fall ein adaptierter FIN-Lerner benutzt, um das richtige D herauszufinden.
(2) Kein effektiver Forscher kann $\mathcal{L}_Q$ und $\mathcal{L}_P$ identifizieren für $Q, P \subseteq \mathbb{N}$, $Q \neq P$.
Betrachte (o.E.) einen effektiven Forscher M für $\mathcal{L}_Q$ und ein $i \in Q \setminus P$.
Also identifiziert M insbesondere $L_{i,\mathbb{N}}$ für das soeben fixierte i.
Nach dem Satz von Blum und Blum gibt es eine Schlussfolge σ für $L_{i,\mathbb{N}}$ und M.
Es gibt daher eine endliche Menge D, sodass $\mathrm{Inh}(\sigma) \subseteq L_{i,D}$.
Da σ Schlussfolge, konvergiert M auf einem σ erweiternden Text für $L_{i,D}$ gegen einen Index für $L_{i,\mathbb{N}}$.
Daher kann M nicht die Sprache $L_{i,D} \in \mathcal{L}_P$ identifizieren und mithin nicht $\mathcal{L}_P$.
(3) Um $\mathcal{L}_Q$ zu identifizieren, bräuchte man also einen Spezialisten-Forscher M_Q.
Wir benötigen also überabzählbar viele Spezialisten-Forscher M_Q, um alle $\mathcal{L}_Q$ zu identifizieren.
Das widerspricht der Abzählbarkeit der Menge aller Turingmaschinen, falls alle Spezialisten TM sein sollen.
Hinweis: Der Beweis funktioniert auch z.B. für höhere Schichten der arithmetischen Hierarchie.

Ein konkretes Beispiel für die Echtheit von $\mathbf{TxtEx} \subsetneq \mathbf{Lang}$.

Es sei $\mathcal{K} = \{ n \in \mathbb{N} \mid \text{die } n\text{-te TM hält, angesetzt auf } n \}$.
Betrachte $\mathcal{K}^* = \{ \mathcal{K} \cup \{x\} \mid x \in \mathbb{N} \}$.
$\mathcal{K}^*$ ist identifizierbar durch

$$F(\sigma) = \begin{cases} \text{Index von } \mathcal{K}, \text{ falls } \operatorname{Inh}(\sigma) \subseteq \mathcal{K}, \\ \text{Index von } \mathcal{K} \cup \{x\}, \text{ falls } \exists x \notin \mathcal{K},\ x \in \operatorname{Inh}(\sigma). \end{cases}$$

Wir zeigen, dass $\mathcal{K}^*$ nicht rekursiv identifizierbar ist.
Angenommen, ein berechenbarer Lerner F identifiziert $\mathcal{K}^*$.
Dann gibt es eine Schlussfolge σ_K für F auf $K \in \mathcal{K}^*$.
Fixiere eine berechenbare Aufzählung $y_0, y_1, y_2, \ldots$ von $\mathcal{K}$.
Für jedes $x \in \mathbb{N}$ bezeichne T^x den Text $\sigma_K x y_0 y_1 y_2 \ldots$.

Es gilt:
(a) $\forall x \in \mathcal{K}:\ T^x$ ist Text für K.
(b) $\forall x \in \overline{\mathcal{K}}:\ T^x$ ist Text für $\mathcal{K} \cup \{x\} \neq \mathcal{K}$.
Also: Nur Texte "für" $\mathcal{K}^*$!
Aus (a) folgt (denn σ_K ist Schlussfolge für $\mathcal{K}$):
$\forall x \in \mathcal{K} \forall n > |\sigma_K| : F(T^x[n]) = F(\sigma_K)$.
Aus (b) folgt (denn F identifiziert $\mathcal{K}^*$):
$\forall x \in \overline{\mathcal{K}} \exists n > |\sigma_K| : F(T^x[n]) \downarrow \wedge F(T^x[n]) \neq F(\sigma_K)$.
Wäre F eine (partiell) rekursive Funktion, so auch

$$\chi : \mathbb{N} \text{—}\circ\!\!\to \mathbb{N},\ \chi(x) = \begin{cases} \min\{S(n) \mid n > |\sigma_K| \wedge F(T^x[n]) \neq F(\sigma_K)\}, & \text{falls def.!} \\ \uparrow & \text{sonst.} \end{cases}$$

Hierbei: $S(n)$ sei Schrittzahl einer F realisierenden TM bei Eingabe von $T^x[n]$.
Der Definitionsbereich von χ wäre $\overline{\mathcal{K}}$, d.h., $\overline{\mathcal{K}}$ wäre rekursiv aufzählbar. Widerspruch! □

Satz 3: Es sei $F^0, F^1, \ldots$ eine Aufzählung aller Turing-berechenbaren (möglicherweise partiellen) Lerner.
Dann gibt es eine total-rekursive Funktion $f : \mathbb{N} \to \mathbb{N}$, so dass
(a) $F^{f(i)}$ total ist und
(b) $\forall L \in \mathcal{E}$: falls F^i identifiziert L, dann gilt: $F^{f(i)}$ identifiziert L.

Beweis: $F^{f(i)}$ soll F^i simulieren, aber definiert sein auf σ, falls $F^i(\sigma) \uparrow$. Es sei $\sigma_{f(i)}$ der längste Präfix von $\sigma \in SEQ$ (falls dieser existiert), auf dem F^i nach $|\sigma|$ Schritten fertig ist.

$$F^{f(i)}(\sigma) = \begin{cases} F^i(\sigma_{f(i)}), & \text{falls } \sigma_{f(i)} \text{ definiert ist.} \\ 0, & \text{sonst.} \end{cases}$$

Sei T ein Text für L, und F^i identifiziere L, d.h., es gibt ein n_0, so dass für alle $n \geq n_0$ $F^i(T[n]) = F^i(T[n_0])$ ein Index für L ist.
Sei s die Laufzeit zur Berechnung von $F^i(T[n_0])$.
$\forall n \geq \max\{n_0, s\}$: $F^{f(i)}(T[n]) = F^i(T[n_0])$, das heißt:
$F^{f(i)}$ konvergiert auf T gegen einen Index von L. □

Zwei nicht-triviale Beispiele für effektive Funktionenlerner

$SD = \{f \in \mathcal{R} \mid h_{f(0)} = f\}$: Die Klasse der *selbstbeschreibenden Funktionen*.

Lernverfahren:

Warte, bis $(0, z)$ als Eingabe erscheint. Dann liefere z zurück.

Beobachte, $SD \neq \emptyset$, denn der zweite Kleenesche Fixpunktsatz liefert zu $\psi(x, y) = x$ einen Index e mit $h_e(y) = \psi(e, y) = e$, für den gilt: $h_e \in \mathcal{R}$ und $h_{h_e(0)} = h_e$.

$AEZ = \{f \in \mathcal{R} \mid |\{f(n) \mid f(n) \neq 0\}| < \infty\}$: Die Klasse der fast überall verschwindenden Funktionen.

Lernverfahren:

Speichere die Paare (n, m) mit $m \neq 0$ und äußere als Hypothese die Funktion, die überall Null ist bis auf die gespeicherten Paare.

Die Klasse $\mathbf{Ex}$: durch TM identifizierbare Mengen rekursiver Funktionen

Formal: $\mathbf{Ex} = \{\mathcal{F} \subseteq \mathcal{R} \mid (\exists \text{berechenbarer Lerner } M)[M \text{ identifiziert jede } f \in \mathcal{F}]\}$

Wir sagen, Text T ist *in kanonischer Ordnung* für f gdw. $T(n) = (n, f(n))$ für alle n.
Entsprechend heißt $\sigma \in SEG$ in kanonischer Ordnung, falls $\sigma = (0, x_0), \ldots, (n, x_n)$ für $n = |\sigma| - 1$.
INIT sei die Menge aller $\sigma \in SEG$ in kanonischer Ordnung.

Für $\sigma \in SEG$ sei $F_A(\sigma)$ das längste Wort $\tau \in$ INIT mit $Inh(\tau) \subseteq Inh(\sigma)$.
Im Wesentlichen durch Sortieren von σ sieht man, dass F_A berechenbar ist.
Damit lässt sich auch ein Text in kanonische Ordnung bringen:

Lemma: : Ist T ein beliebiger Text für $f \in \mathcal{R}$, dann ist $\bigcup_{m \in \mathbb{N}} F_A(T[m])$ der kanonische Text für f.

Eine Charakterisierung von $\mathbf{Ex}$

Satz 4: Sei $\mathcal{F} \subseteq \mathcal{R}$ vorgelegt.
$\mathcal{F} \in \mathbf{Ex}$ gdw. Es gibt einen berechenbaren Forscher, der den kanonischen Text jeder Funktion aus $\mathcal{F}$ identifiziert.

Beweis: Die Richtung $\Longrightarrow$ ist offenkundig.

Nehmen wir an, M würde den kanonischen Text jeder Funktion aus $\mathcal{F}$ identifizieren.

Definiere den Forscher M^* wie folgt:

Für $\sigma \in \mathrm{SEG}$, setze $M^*(\sigma) = M(\mathrm{F}_A(\sigma))$.

Als Komposition berechenbarer Fkt. ist M^* berechenbar.

Nach dem Lemma konvergiert M^* gegen i auf Text T gdw. M konvergiert gegen i auf dem entsprechenden kanonischen Text.

Da M (erfolgreich) kanonische Texte identifiziert, wird M^* bel. Texte ebenso identifizieren. $\square$

Auch Funktionenlerner dürfen total sein

Wir halten ein Analogon zu Satz 3 fest:

Satz 5: Es sei $F^0, F^1, \ldots$ eine Aufzählung aller Turing-berechenbaren (möglicherweise partiellen) Lerner.
Dann gibt es eine total-rekursive Funktion $f : \mathbb{N} \to \mathbb{N}$, so dass
(a) $F^{f(i)}$ total ist und
(b) $\forall g \in \mathcal{R}$: falls F^i identifiziert g, dann gilt: $F^{f(i)}$ identifiziert g.

Folgerung: Es gibt eine rek. aufzählbare Folge total-berechenbarer Forscher $M_0, M_1, \ldots$, sodass für jede Funktionenmenge $S \in \mathbf{Ex}$ einen Forscher M_i gibt, der alle Funktionen aus S identifiziert.

Satz 6 (Nicht-Vereinigungsabschluss-Theorem von Blum & Blum):
$\mathrm{SD} \in \mathbf{Ex}$, $\mathrm{AEZ} \in \mathbf{Ex}$, aber $\mathrm{SD} \cup \mathrm{AEZ} \notin \mathbf{Ex}$.

Folgerung (Gold): $\mathcal{R} \notin \mathbf{Ex}$.

Beweis von Satz 6: Nach den Vorüberlegungen bleibt zu zeigen: $\mathrm{SD} \cup \mathrm{AEZ} \notin \mathbf{Ex}$.
Betrachte einen berechenbaren Forscher M, der (wenigstens) AEZ identifiziert.
Nach Satz 4 ist M o.E. total.
Wir werden zeigen: Zu jedem kanonischen σ gibt es ein τ mit $\sigma \sqsubseteq \tau$, sodass $M(\sigma) \neq M(\tau)$.
Dieses lässt sich algorithmisch finden.
Diese Aussage benutzen wir zu einer Stufenkonstruktion:
Setze $\sigma_e^0 = (0, e)$.
Ist σ_e^s def., so finde σ_e^{s+1} mit $\sigma_e^s \sqsubseteq \sigma_e^{s+1}$, sodass $M(\sigma_e^s) \neq M(\sigma_e^{s+1})$.
Die wie folgt def. Funktion ψ ist berechenbar (durch Suche nach geeignetem s):
$\psi(e, x) = y$, falls es ein s gibt mit $(x, y) \in \mathrm{Inh}(\sigma_e^s)$.
Der zweite Kleenesche Fixpunktsatz liefert ein e mit $\phi_e(x) = \psi(e, x)$.
Nach Konstruktion gilt: $\phi_e \in \mathrm{SD}$.
ϕ_e wird nicht identifiziert, weil nach Konstruktion unendlich viele Hypothesenwechsel durch M bei der Präsentation des kanonischen Textes von ϕ_e erfolgen. □

Wir müssen noch folgende Behauptung zeigen:

Zu jedem kanonischen σ gibt es ein τ mit $\sigma \sqsubseteq \tau$, sodass $M(\sigma) \neq M(\tau)$. Dieses lässt sich algorithmisch finden.

Haben wir die (reine) Existenz bewiesen, so können wir uns dies durch erschöpfende Suche algorithmisch zunutze machen.

Nehmen wir an, alle kanonischen Erweiterungen τ von σ erfüllten $M(\sigma) = M(\tau)$.
Dann kann M höchstens eine Funktion identifizieren, deren kanonischer Text mit σ startet.
Andererseits gibt es unendlich viele Funktionen in AEZ, deren kanonischer Text mit σ startet.
Also kann M nicht alle Funktionen aus AEZ identifizieren, im Gegensatz zu unserer anfänglichen Annahme. □

Einordnung und Ausblick

Wir haben begonnen, Lerner / Forscher als TM zu untersuchen.

Die “Lernfähigkeit” wird dadurch erheblich eingeschränkt.

Andererseits kann sich kein TM-mächtiger “Gegner” dies wirklich zunutze machen:

Mitteilung: Es gibt keine (totale) berechenbare Funktion $f : \mathbb{N} \to \mathbb{N}$, die zu jedem $i \in \mathbb{N}$ eine Funktion $h_{f(i)} \in \mathcal{R}$ liefert, die von M_i nicht identifiziert werden kann. (Die M_i beziehen sich auf die Aufzählung der Forscher gemäß der Folgerung nach Satz 5.)

In der nächsten VL werden wir beginnen, verschiedene Lernstrategien zu untersuchen.