Seminar Kolmogorovkomplexität

Universität Potsdam
Wintersemester 2009/10

Kolmogorovkomplexität

- **Kolmogorovkomplexität** (auch "**Algorithmische Komplexität**") ist der zentrale Begriff der **Algorithmischen Informationstheorie** (AIT).
- Kombiniert Informations- und Berechenbarkeitstheorie, um den **Informationsgehalt eines einzelnen Objekts** –seine Komplexität– absolut und objektiv beschreiben zu können
- Gleichzeitig ergibt sich daraus ein objektiver Begriff für die **Zufälligkeit eines Objekts**
- AIT ermöglicht so die vielseitige Verwendung von "Information" als mathematisches Beweismittel.

Kolmogorovkomplexität

- Die **algorithmische Komplexität eines Objekts** ist die **Länge des kürzesten Programms**, das dieses Objekt erzeugt (also die **kürzeste effektiv berechenbare Beschreibung**)
- Je mehr **Regelmäßigkeiten** ein Objekt (meist codiert als binäre Zahlenfolge) aufweist, desto kürzer lässt es sich beschreiben, desto stärker **komprimierbar** ist es
- Je **weniger Struktur** eine Zahlenfolge aufweist, desto weniger komprimierbar ist sie und damit **zufälliger**

Kolmogorovkomplexität

- Bis auf eine additive Konstante **unabhängig vom Maschinenmodell**; meist Betrachtung des Programms als Eingabe für eine optimale universelle Turingmaschine
- Einfache algorithmische Komplexität ist **nicht berechenbar**, **nur approximierbar**
- Wichtige Varianten:
  - **Algorithmische Präfix-Komplexität**: Die kürzeste Beschreibung, die nicht Präfix einer Beschreibung eines anderen Objekts ist
  - **Ressourcen-beschränkte Komplexität**: Die kürzeste in vorgegebenen Zeit-, bzw. Platzschranken berechenbare Beschreibung

Unterschied zur klassischen Informationstheorie

Klassische (Shannon'sche) Informationstheorie beschäftigt sich mit **Kommunikation und Zufallsvariablen**, bietet jedoch keine Antwort auf folgende Fragen:

- Was ist der **Informationsgehalt eines einzelnen Objekts**?
- Wann ist ein **einzelnes Objekt zufällig**?

Was ist der Informationsgehalt eines Objekts?

Klassischer Informationstheoretischer Ansatz:

- Objekt wird betrachtet als **Element einer zuvor festgelegten Menge** mit einer bestimmten **Wahrscheinlichkeitsverteilung**
- Der **Informationsgehalt** des Objekts hängt nur von seiner **Wahrscheinlichkeit** ab, **nicht vom Objekt selbst**

Kolmogorovkomplexität beschreibt den **Informationsgehalt eines Objekts in Abhängigkeit vom Objekt selbst**: Wie lang ist die **kürzeste Beschreibung** dieses Objekts?

Wann ist ein Objekt/ eine Folge von Ereignissen zufällig?

- Betrachte die **Ergebnisse eines (fairen) Münzwurf-Experiments**: Jede Folge von n Münzwürfen hat wahrscheinlichkeitstheoretisch die gleiche Wahrscheinlichkeit von $\frac{1}{2^n}$
- **Intuitiv** erscheint eine Folge von n Nullen als **weniger wahrscheinlich** als eine Zeichenfolge ohne erkennbares Muster
- Aber was ist mit $\pi = 3,14....$ und Champernownes Zahl $0,12345678910111213 14\ldots$?
- Diese Folgen können durch sehr **kurze Programme** erzeugt werden - damit ist ihre algorithmische Komplexität sehr niedrig

Wann ist ein Objekt/ eine Folge von Ereignissen zufällig?

- Klassische Wahrscheinlichkeitstheorie erlaubt **keine Aussage über die Zufälligkeit eines individuellen Objekts**, sondern nur über Erwartungen bezüglich der Ergebnisse zufälliger Prozesse
- AIT bietet eine "echte" Definition von Zufälligkeit: Nur **Objekte, die nicht wesentlich kürzer als durch buchstäbliches Hinschreiben beschrieben werden können, sind zufällig**

Anwendung

- **Viele Anwendungsbereiche**: Mathematik, Physik, Informatik, Philosophie, Biologie, ...
- Wesentliche Eigenschaft für **Beweise anhand algorithmischer Komplexität**: Komprimierbarkeit, bzw. Inkomprimierbarkeit von Objekten
- z.B. **Inkomprimierbarkeitsmethode**:
  - Ähnlich **allgemeine Beweismethode** wie das Taubenschlag-Prinzip oder die probabilistische Methode
  - Die **meisten** Zahlenfolgen sind **nicht effektiv komprimierbar** und daher geeignet als **"typische Objekte"** bestimmter Klassen
  - **Methodik**: Zeige, dass ein solches inkomprimierbares Objekt eine bestimmte Eigenschaft besitzen muss, da es sonst komprimierbar wäre

Konkretes Anwendungsbeispiel

Theorem

Für unendlich viele natürliche Zahlen n gilt, dass es mindestens $\frac{\log n}{\log \log n} - o(1)$ *Primzahlen gibt, die kleiner als n sind.*

- Die Länge der Binärdarstellung einer Zahl n ist $\log n$; die Zahl der Primzahlen, die kleiner als n sind, sei m und $p_1, \ldots, p_m$ eine Liste dieser Primzahlen
- Dann gilt $n = p_1^{e_1} p_2^{e_2} \ldots p_m^{e_m}$ und n lässt sich durch den Vektor der Exponenten $e_1, \ldots, e_m$ eindeutig darstellen, wobei jeder der Exponenten maximal $\log n$ groß sein kann
- Jeder Exponent ist somit binär durch $\log \log n$ Bits darstellbar und $e_1, \ldots, e_m$ folglich mit Länge $m * \log \log n$

Konkretes Anwendungsbeispiel (Forstsetzung)

- Wenn die Blocklänge $\log\log n$ bekannt ist, lässt sich diese Beschreibung parsen, so dass sich eine effektive Beschreibung von n ergibt; dazu wird $\log\log n$ der Beschreibung in Präfixcodierung hinzugefügt, wofür $(1 + o(1)) \log\log\log n$ Bits nötig sind
- Für jede ganze Zahl $l > 0$ gibt es mindestens eine Zahl n, deren Binärdarstellung $l \approx \log n$ ist und die sich nicht kürzer darstellen lässt
- Für jedes solche (und somit unendlich viele) n gilt also $(1 + o(1)) \log\log\log n + m * \log\log n \geq \log n$, und da $\lim_{n\to\infty} \frac{(1+o(1)) \log\log\log n}{\log\log n} = 0$, ergibt sich $m \geq \frac{\log n}{\log\log n} - o(1)$

□

Geschichte

- **Algorithmische Komplexität**:
  - Solomonoff 1964: Universelle a priori Wahrscheinlichkeit
  - Kolmogorov 1965, Chaitin 1966: Definiere Informationsgehalt eines Objektes als Länge des kürzesten Programms, das dieses Objekt beschreibt
  - Levin 1970: Ausarbeitung der mathematischen Details
- **Algorithmische Präfix-Komplexität**:
  - Levin 1974, Gacs 1974, Chaitin 1975
- **Resourcen-beschränkte Komplexität**:
  - Kolmogorov 1965, Solomonoff 1964: Hinweis auf die Problematik der Nicht-Berechnenbarkeit
  - z.B. Daley 1973, 1977, Feder 1992, Ko 1986, Pintado 1997, Schmidhuber 2002: Verschiedene berechenbare und/ oder Zeit- oder Platzbeschränkte Komplexitäten/ Wahrscheinlichkeiten

Organisatorisches

- Literatur:
  Ming Li, Paul Vitanyi: An Introduction to Kolmogorov Complexity and its Applications. 3. Auflage, Springer 2008
- Seminar (alle Studienordnungen); Vortragende müssen sich **vor** ihrem Vortrag zur Prüfung anmelden!
- Sowohl Vortrag an der Tafel als auch mit Folien möglich, Zeitrahmen variabel; wichtig sind:
  - Wirkliches Interesse am Thema
  - Gründliche und verständliche Darstellung der mathematischen Inhalte
  - Kompetente Leitung der Diskussion bei auftretenden Fragen
- Terminverlegung? (Vorschlag: Dienstag 14 Uhr)

Leistungsbewertung

- Grundlage ist hauptsächlich die Qualität der Vorträge:
  - Wie verständlich und sorgfältig war die Darstellung der entsprechenden Inhalte?
  - Wurden die üblichen Regeln wissenschaftlichen Arbeitens wie Fairness und vollständige Quellenangaben eingehalten?
- Aktive Teilnahme:
  - Anwesenheit und Aufmerksamkeit
  - Sichtbares Interesse am Thema
  - Bereitschaft, Fragen zu stellen, bzw. zu beantworten
- Ausarbeitung entfällt zugunsten einer umfangreichen Vorbereitung auf die Themen

Themen

1. Informations- und Codierungstheoretische Grundlagen (1.11)
2. Algorithmische Komplexität I (2.1)
3. Algorithmische Komplexität I (2.2–2.3)
4. Algorithmische Präfix-Komplexität (3.1, 3.3–3.4)
5. Anwendungen Algorithmischer Komplexität I (4.4, 6.1)
6. Anwendungen Algorithmischer Komplexität II (6.2–6.3)
7. Anwendungen Algorithmischer Komplexität III (6.4–6.6)
8. Anwendungen Algorithmischer Komplexität IV (6.7–6.9)
9. Anwendungen Algorithmischer Komplexität V (6.10–6.12)
10. Anwendungen Algorithmischer Präfix-Komplexität I (1.6,1.10,5.1)
11. Anwendungen Algorithmischer Präfix-Komplexität II (5.2)
12. Anwendungen Algorithmischer Präfix-Komplexität III (5.4)

Informations- und Codierungstheoretische Grundlagen

- Der Shannon'sche Entropie-Begriff
- Präfix-Codes
- Kraftsche Ungleichung
- Optimale Codes / Noiseless Coding Theorem
- Universelle Codes

Algorithmische Komplexität I–II

I.

- Definition der algorithmischen Komplexität
- Der Invarianz-Satz

II.

- Inkomprimierbarkeit
- Zufälligkeitsdefizit
- Algorithmische Komplexität als Ganzzahlfunktion

Algorithmische Präfix-Komplexität

- Definition der algorithmischen Präfix-Komplexität
- Der Invarianz-Satz
- Inkomprimierbarkeit
- Algorithmische Präfix-Komplexität als Ganzzahlfunktion

Anwendungen der Algorithmischen Komplexität I–III

I.

- Universelle durchschnittliche Komplexität von Algorithmen
- Die Inkomprimierbarkeitsmethode: Einstiegsbeispiele

II.

- Eigenschaften mit hoher Wahrscheinlichkeit
- Kombinatorik

III.

- Kolmogorov-zufällige Graphen
- Kompaktes Routing
- Analyse des durchschnittlichen Verhaltens von Sortieralgorithmen

Anwendungen Algorithmischer Komplexität IV–V

IV.

- Längste gemeinsame Teilfolge
- Formale Sprachen
- Online Erkennung kontextfreier Sprachen

V.

- Laufzeitkomplexität von Turingmaschinen
- Paralleles Rechnen
- Switching Lemma

Anwendungen der Algorithmischen Präfix-Komplexität I–III

I.

- Wahrscheinlichkeitstheorie
- Induktives Schließen: Historischer Hintergrund

II.

- Universelle Wahrscheinlichkeit: Solomonoffs Theorie der Voraussage

III.

- Identifizierung von Hypothesen mit anhand der minimalen Beschreibungs-Länge ("MDL" = "Minimum Desription Length")