ETH Zürich
Dept. Informatik

Prof. Markus Gross
WS 05/06

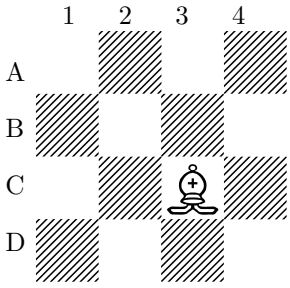
Informationstheorie

Übung 5

Ausgabe: 5. Dezember 2005
Abgabe: 12. Dezember 2005

5.1 Schach als Informationsquelle (aus dem Vordiplom Herbst 2000)

Betrachten Sie dieses reduzierte 4×4 Schachbrett mit einem Läufer.



Ein gültiger Zug eines Läufers besteht gemäss Schachregel aus einer Verschiebung entlang einer seiner beiden Diagonalen. Zum Beispiel darf sich der weisse Läufer im Bild auf die Felder A1, B2, D4, B4 oder D2 bewegen.

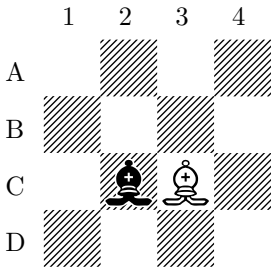
Betrachten Sie nun die Informationsquelle gegeben dadurch, dass der Läufer jeweils einen zufälligen gültigen Zug ausführt (beachten Sie, an Ort bleiben ist *kein* gültiger Zug).

a) Wie viele mögliche Zustände hat diese Informationsquelle?

b) Bestimmen Sie die stationäre Verteilung dieser Informationsquelle. Tipp: Symmetrien ausnutzen.

c) Wie gross ist die Entropierate dieser Informationsquelle?

d) Betrachten Sie nun das 4×4 Schachbrett mit zwei Läufern, einem auf einem weissen und einem auf einem schwarzen Feld, und die Informationsquelle gegeben dadurch, dass jeweils beide Läufer zusammen einen zufälligen gültigen Zug ausführen.



Beantworten Sie (a), (b) und (c) für diese Informationsquelle.

5.2 Informationsdichte der Schweizer Presse

In dieser Aufgabe soll die Entropie pro Buchstabe in einem Zeitungsartikel bestimmt werden. Eine reine Häufigkeitsanalyse von einzelnen Buchstaben, Paaren oder Tripeln usw. würde wegen der starken Abhängigkeiten auf einen zu hohen Wert der Entropie führen.

Eine Vorgehensweise ist die folgende: Ein intelligenter menschlicher Codierer transformiert den Text in eine Folge von Zahlen. Der Codierer versucht, den jeweils nächsten Buchstaben im Text zu erraten, wobei ihm nur gesagt wird, ob der Versuch richtig oder falsch war. Für einen Klartext der Länge t erhält man so eine Sequenz $M_1, \ldots, M_t$ von t Zahlen, wobei M_i die Anzahl Versuche für den i-ten Buchstaben ist. Wir nehmen an, der Codierer folge einem deterministischen Algorithmus und darf nur Fragen von der Form „Ist der Buchstabe ein x?“ stellen.

a) Wie kann aus der Sequenz $M_1, \ldots, M_t$ der Text wieder hergestellt werden?

b) Wie verhält sich die Entropie der Zahlensequenz zur Entropie des Textes?
Fakultativ: Führen Sie das beschriebene Experiment durch. [1]

c) Überlegen Sie, wie Sie aus der Sequenz $M_1, \ldots, M_t$ die Entropie pro Buchstaben abschätzen können. Nehmen Sie an, dass alle M_i unabhängig und identisch verteilt sind.

d) Die Annahme, dass alle M_i unabhängig sind, stimmt nur näherungsweise. Wieso? Wie lässt sich die Schätzung von c) noch verbessern?

5.3 Fussball als Informationsquelle

Die beiden Fussballmannschaften FCB und GC treffen aufeinander. Vereinfachend nehmen wir an, dass der Ball 6 Zustände annehmen kann: Er kann in der Verteidigung sowie im Angriff der beiden Mannschaften sein sowie in einem der Tore landen. Ist die Verteidigung des FCB in Ballbesitz, so gelingt es dieser mit Wahrscheinlichkeit 0.6, den Ball in den Angriff zu bringen. Mit Wahrscheinlichkeit 0.4 allerdings geht der Ball im Zweikampf von der Verteidigung des FCB an die Stürmer des GC verloren. Die entsprechenden Werte für die Defensive des GC betragen 0.8 und 0.2. Sind die Stürmer des FCB erst einmal im Ballbesitz, erzielen sie mit Wahrscheinlichkeit 0.1 ein Tor, andernfalls (0.9) bleibt das Leder in der GC-Verteidigung hängen. Umgekehrt betragen diese Werte 0.2 und 0.8. Ist schliesslich der Ball im Tor einer Mannschaft gelandet, hat diese Mannschaft (und zwar die Defensive) das Recht auf Wiederanstoss. Nehmen Sie für die folgenden Aufgaben an, dass das Spiel sehr lange dauert.

a) Was ist die stationäre Verteilung der Informationsquelle, die aus der Folge der Zustände gebildet wird?

b) Welche Mannschaft wird voraussichtlich mit welchem Torverhältnis das Spiel gewinnen?

c) Bestimmen Sie die Entropierate der Zustandsfolge.

[1] Wählen Sie einen kurzen Zeitungsartikel (rund 150 Buchstaben) und ermitteln Sie zu zweit die Werte $M_1, \ldots, M_t$. Hinweise: Im Deutschen bestehen folgende (recht grobe) Relationen in der Häufigkeit der einzelnen Buchstaben: $e > n > i, s, r, a, t > d, h, u, l, c, g, m, o, b, w, f, k, z > \ldots$. Ausserdem beginnen Wörter häufig mit d oder s.

5.4 One-Time Pad

Der One-Time Pad ist ein perfekt sicheres Verschlüsselungssystem, welches auch tatsächlich verwendet wurde, z.B. beim "Roten Telefon" zwischen den Präsidenten der USA und der UdSSR.

Für einen Klartext $M \in \{0,1\}^N$ wird ein zufälliger Schlüssel $K \in \{0,1\}^N$ generiert. Dabei ist es wichtig, dass K gleichverteilt aus $\{0,1\}^N$ gewählt wird, und dass K nur einmal verwendet wird (also nur für die Nachricht M). Das Chiffrat ist $C := M \oplus K$ ($\oplus$ ist hier die bitweise Addition modulo 2). Beweisen Sie, dass der One-Time Pad eine perfekt sichere Verschlüsselung ist, d.h. dass $I(M;C) = 0$ gilt. Zeigen Sie die Behauptung anhand des Entropiediagramms für die Zufallsvariablen M, C und K. Tipp: Gewisse Flächen werden Sie mit 0, gewisse Flächen mit N beschriften können. Zeigen Sie, dass $H(C) \leq H(K)$.