Übungsblatt 2 – Algorithmisches Lernen

Abgabe 08.05.2001 vor der Vorlesung

Aufgabe 2.1

a) Die Bewertung von Beispielen wird bei diesem variierten Weighted Majority Algorithmus von einer Runde von n Experten übernommen. Dabei wird jedem Experten ein Gewicht w_i zugeordnet, das seine Glaubwürdigkeit beschreibt. Zu Anfang des Verfahrens sind alle $w_i = 1$.

Der Algorithmus klassifiziert zunächst ein Beispiel als hypothesenerfüllend h(x) = 1, wenn sich die gewichtete Mehrheit der Experten für *Wahrheit* entscheidet, sonst als h(x) = 0. Anschließend werden die Gewichte der Experten, die eine falsche Bewertung abgegeben haben, um die Hälfte reduziert. Die Gewichte der Experten mit richtiger Einschätzung bleiben gleich. Diese Abwertung erfolgt, solange Gegenbeispiele geliefert – also Fehler gemacht – werden.

Wie viele Fehleinschätzungen macht der Algorithmus aufgrund der falschen Mehrheitsentscheidung? Und wie viele Fehler macht der beste Experte mindestens? Für die Beantwortung dieser Fragen lohnt es, das Gewicht des besten Agenten w_b sowie das Gesamtgewicht W zu betrachten:

$$W = \sum_{i=1}^{n} w_i \tag{1.1}$$

Wenn der Algorithmus einen Fehler macht, dann hat die Mehrheit der Experten eine falsche Einschätzung gegeben. Damit reduziert sich das Gesamtgewicht um mindestens um den Faktor ¼. Gehen wir von insgesamt k Fehlern des Algorithmus aus, erhalten wir für das Gesamtgewicht bei n Experten

$$W \leq n \cdot \left(\frac{3}{4}\right)^k \tag{1.2}$$

Die Anzahl der Fehler der besten Experten sei m_b. Dann ergibt sich für sein Gewicht bei m_b Fehlern

$$w = \left(\frac{1}{2}\right)^{m_b} \tag{1.3}$$

Das Gesamtgewicht W ist immer größer gleich dem Gewicht eines Experten w_i. Dementsprechend ist auch das Gewicht des besten Experten w_b kleiner gleich W. Setzt man die Aussagen (1.2) und (1.3) gegenüber, ergibt sich nach Umformen

$$
\begin{aligned}
w_b &\le W \\
\left(\frac{1}{2}\right)^{m_b} &\le n \cdot \left(\frac{3}{4}\right)^{k} \quad \left|\log\right. \\
m_b \cdot \log\left(\frac{1}{2}\right) &\le \log n + k \cdot \log\left(\frac{3}{4}\right) \quad \left|: (-1)\right. \\
m_b &\ge k \cdot \log\left(\frac{4}{3}\right) - \log n
\end{aligned}
\tag{1.4}
$$

Für Abschätzung erhält man dann mit c = log (4/3)

$$m_b \ge \Omega(k) - O(\log(n)) \tag{1.5}$$

was zu zeigen war. Im Folgenden noch ein Screenshot von einer simulierten Expertenrunde mit hier 5 Experten und 100 Gegenbeispielen. Der beste Experte machte in diesem Lauf 44 Fehler, wobei die Schranke bei

$$m_b \ge 18 \tag{1.6}$$

Fehlern für den besten Experten lag.

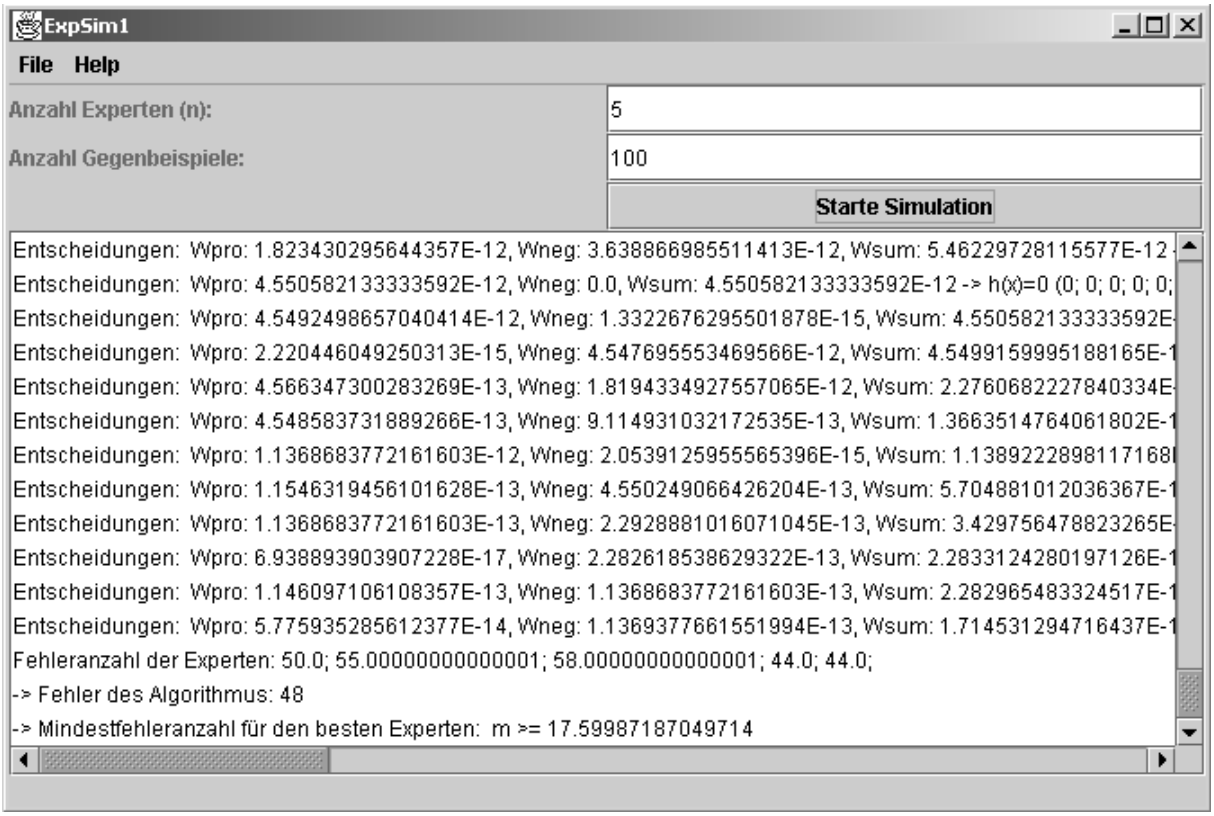

Abbildung 1 - Screenshot von simulierter Expertenrunde

Aufgabe 2.2

Um den in Aufgabe 1 angewendeten Algorithmus stabiler gegen falsche Mehrheitsentscheidungen zu machen, kann er randomisiert werden. Der Algorithmus

wird ein Beispiel jetzt als hypothesenerfülllend klassifizieren, wenn ein zufällig gewählter Experte dies vorschlägt. Die Wahrscheinlichkeit des Experten E_i ist laut Aufgabenstellung mit Gesamtgewicht W und n Experten

$$\frac{w_i}{W} = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{1.7}$$

In der bisherigen Version wurden die Expertengewichte w_i bei Falschbewertung halbiert. Alternativ soll hier ein Faktor $0 < \beta < 1$ verwendet werden.

Die Frage ist, wie sich die Mindestanzahl von Fehlern des besten Experten bei diesen Veränderungen verhält. Da die Expertenwahl zufällig erfolgt, wird der Erwartungswert der Fehleranzahl des Algorithmus betrachtet

$$\mathrm{E}M = \sum_{i=1}^{k} F_i \tag{1.8}$$

wobei F_i den Teil der Fehler am Gesamtgewicht W_i beim i-ten Beispiel und k die Anzahl der bewerteten Beispiele darstellt.

Bei jedem Beispiel sinkt das Gesamtgewicht W um den Faktor $\left(1-\left(1-\beta\right)\cdot F_i\right)$. Nach k Beispielen erhält man dann für W das Produkt

$$W = n \cdot \prod_{i=1}^{k} \left(1-\left(1-\beta\right)\cdot F_i\right) \tag{1.9}$$

Die Mindestanzahl von Fehlern des besten Experten sei wie in Aufgabe 2.1 m_b und damit das Gewicht des besten Experten

$$w_b = \beta^{m_b} \tag{1.10}$$

Wobei jetzt der Faktor β anstatt des konstanten ½ benutzt wird. Nach Umformung erhalten wir im Folgenden die in 2.1 bereits bestätigte asymptotische Schranke

$$\begin{aligned}
w_b &\le W \\
\beta^{m_b} &\le n \cdot \prod_{i=1}^{k} \left(1-\left(1-\beta\right)\cdot F_i\right) \quad \left|\ln\right. \\
m_b \cdot \ln\beta &\le \ln n + \sum_{i=1}^{k} \ln\left(1-\left(1-\beta\right)\cdot F_i\right) \quad \left|\cdot(-1)\right. \\
m_b \cdot \ln\left(1/\beta\right) &\ge -\ln n - \sum_{i=1}^{k} \ln\left(1-\left(1-\beta\right)\cdot F_i\right) \quad \left|\ln(1-x) \le -x\right. \\
m_b \cdot \ln\left(1/\beta\right) &\ge \left(1-\beta\right)\sum_{i=1}^{k} F_i - \ln n \quad \left|: \ln\left(1/\beta\right)\right. \\
m_b &\ge \frac{\left(1-\beta\right)}{\ln\left(1/\beta\right)} M - \frac{1}{\ln\left(1/\beta\right)} \ln n
\end{aligned} \tag{1.11}$$

Für die Abschätzung der minimalen Fehleranzahl ergibt sich also auch hier

$$\begin{aligned} m_b &\geq c_1 M - c_2 \log n \\ &\geq \Omega(k) - O(\log(n)) \end{aligned} \tag{1.12}$$

Aufgabe 2.3

a) Eine totale Ordnungen auf n Elementen stellt bei dieser Problemstellung ein Konzept dar. Entsprechend sei die Menge aller totalen Ordnungen, die sich auf n Elementen bilden lassen, die Konzeptklasse C_n:

$$C_n = \{\text{alle totale Ordnungen auf n Elementen}\} \tag{1.13}$$

Wobei es n! verschiedene totale Ordnungen auf n Elementen gibt und C_n somit die Kardinalität n! hat

$$|C_n| = n! \tag{1.14}$$

Ein Konzept i der Konzeptklasse C_n besteht somit aus disjunkten, geordneten Paaren der n Elemente (weil totale Ordnung!)

$$c_i = \left\{ (a,b)^t \mid t = \binom{n}{2} \right\} \tag{1.15}$$

Die VC-Dimension ist definiert als die Beispielmenge S mit der größten Kardinalität, die von der Konzeptklasse C zertrümmert wird. *Zertrümmern* ist dabei so definiert, dass wir eine Menge S suchen müssen, deren Potenzmenge P(S) durch den Schnitt der Konzepte aus C und der Beispielmenge entsteht.

$$P_x = \{c \cap S \mid \forall c \in C_n\} \tag{1.16}$$

Durch den Schnitt zweier Mengen erhalten wir eine Schnittmenge, deren Mächtigkeit höchstens so groß ist wie die der größeren der beiden geschnittenen Mengen. Die Beispielmenge S enthält geordnete Paare. Da nur n! Konzepte zur Verfügung stehen, ist die Mächtigkeit der Schnittmenge nach oben hin begrenzt mit

$$|P_x| \leq |C_n| = n! \tag{1.17}$$

Somit gilt als obere Schranke der VC-Dimension

$$\begin{aligned} 2^{|S|} &\leq n! \\ |S| &\leq \frac{\log(n!)}{\log 2} = \log(n!) \\ VC(C_n) &\leq \log(n!) \end{aligned} \tag{1.18}$$

Wie sieht es aber mit der unteren Schranke für $VC(C_n)$ aus? Die größte Kardinalität einer Beispielmenge S, deren Potenzmenge kleiner gleich n! viele Paare enthält, wäre mit

$$\left|S^*\right| = \left\lfloor \log_2 \left|C_n\right| \right\rfloor = \left\lfloor \log_2 \left|n!\right| \right\rfloor \tag{1.19}$$

gegeben. Können wir also garantieren, dass diese |S*| Beispielpaare geschnitten mit den n! Konzepten mindestens $\log_2|C_n|$ Schnittmengen produzieren? Dem scheint so zu sein :)

b) Hier kommt das zu Beginn der Vorlesung angewandte Programmschema zum Einsatz, dass ein Konzept c erlernt, indem von von einer nicht erfüllbaren Hypothese ausgehend mit jedem Gegenbeispiel eine neue Hypothese aufgestellt wird, die mehr und mehr dem zu erlernenden Konzept c entspricht.

$$h_0 = \{(1,2),(2,1),(1,3),(3,1),(2,3),\ldots,(n,n-1)\} \tag{1.20}$$

Ein Gegenbeispiel ist ein geordnetes Paar (a,b), das nicht durch Konzept erfüllt wird. Mit jedem präsentierten Gegenbeispiel nimmt die Anzahl der erfüllenden Konzepte um mindestens die Hälfte ab, so dass nach

$$Gegenbeispiel_H(C_n) \leq \log(n!) \tag{1.21}$$

der Algorithmus terminiert und ein Konzept erlernt wurde.