

Algoritmen voor multichain Markovbeslissingsketens

Begeleider: Lodewijk Kallenberg

Bij Markov beslissingsketens hebben we niet te maken met één Markov keten, maar met een aantal Markov ketens. We zullen veronderstellen dat de ketens eindig veel toestanden hebben. Formeel is het model als volgt:

- S is een eindige *toestandsruimte*, zeg $S = \{1, 2, \dots, N\}$;
- in toestand $i \in S$ wordt een actie uit een eindige *actieverzameling* $A(i)$ gekozen;
- indien in toestand i actie $a \in A(i)$ wordt gekozen, dan gebeurt het volgende: er is een *directe opbrengst* $r_i(a)$ en met *overgangskans* $p_{ij}(a)$ is het systeem op het volgende beslissingstijdstip in toestand j .

Een (*deterministische*) *strategie* f is een beslisregel: het geeft aan welke actie in toestand i wordt gekozen; de actie die in toestand i wordt gekozen noteren we met $f(i)$. Iedere strategie genereert een Markov keten met overgangsmatrix $P(f)$, waarbij $[P(f)]_{ij} = p_{ij}(f(i))$ voor alle $i, j \in S$.

Als minstens één zo'n Markov keten meerdere recurrente klassen heeft, dan spreken we van een *multichain* Markovbeslissingsketen. Voor dergelijke ketens bestond tot voor kort geen *waarde-iteratie methode* om de gemiddelde opbrengst te optimaliseren. In 2007 hebben de Japanners Iki, Horiguchi en Kurano een waarde-iteratie methode voorgesteld voor multichain Markovbeslissingsketens.

Het project houdt het volgende in:

1. Het en en ander bestuderen over Markovbeslissingsproblemen.
2. Het artikel van Iki, Horiguchi en Kurano (en enkele andere artikelen waarop dit gebaseerd is).
3. Een implementatie maken van deze methode.