Solving large structured Markov Decision Problems for perishable inventory management and traffic control

Haijema, R.

Publication date
2008

Nederlandse samenvatting

In dit proefschrift staan grote gestructureerde Markov beslisproblemen (MBP) centraal. Met ‘groot’ wordt bedoeld dat het aantal toestanden in het MBP model te groot is om in een acceptabele rekentijd een optimale oplossing te bepalen. Met ‘gestructureerd’ wordt bedoeld dat de toestandsruimte van een MBP een probleem-specifieke structuur kan hebben die benut kan worden in het oplosproces.

Het grote aantal toestanden is vaak het gevolg van de dimensionaliteit van de toestandsruimte. Voor iedere toestand is er een optimaal besluit, of optimale actie, te bepalen. Met behulp van (stochastisch) dynamisch programmeren kunnen de nodige berekeningen efficient plaatsvinden. Voor veel vraagstukken uit de praktijk, is het aantal toestanden echter dermate groot dat men zich tevreden heeft te stellen met een benadering van de optimale oplossing. Daartoe dient men de aanwezige probleemstructuur te benutten.

Voor een tweetal optimalisatievraagstukken uit de praktijk wordt geïllustreerd hoe de kennis en oplostechnieken van MBPs ingezet kunnen worden om tot benaderingen van een optimale oplossing te komen. Het eerste vraagstuk betreft het productie-voorraadbeheer van bloedplaatjespools (BPPs). Het tweede vraagstuk betreft de drukte-afhankelijke regeling van verkeerslichten. De twee vraagstukken vergen een verschillende aanpak om tot een benadering te komen daar ze verschillen in probleemstructuur.

Deel I – Voorraadbeheer van bloedplaatjespools (H2 tot H4)

Bloedbanken en ziekenhuizen houden bloedproducten op voorraad om aan de slecht voorspelbare vraag naar transfusies te kunnen voldoen. Hoge voorraadniveaus garanderen een hoge beschikbaarheid, echter leiden vaak in de praktijk tot hoge vervalpercentages. De duurste en tevens kortst-houdbare bloedproducten zijn trombocytenconcentraten. De meeste trombocytenconcentraten worden verkregen door de plaatjes van vijf volbloeddonaties samen te voegen tot een bloedplaatjespool (BPP).

Zowel in Europa (Veihola e.a. [155]) als in de Verenigde Staten (Whitaker and Sullivan [163]) wordt gemiddeld 15 à 20% van de geproduceerde BPPs weggegooid, voornamelijk als gevolg van het verlopen van de uiterste houdbaarheidsdatum van de BPPs. Door voorraadniveaus verstandiger vast te stellen, kan het verval gereduceerd worden zonder dat dit ten koste gaat van de beschikbaarheid van BPPs.

In dit proefschrift wordt onderzocht hoe een optimale productie- of bestelregel er uitziet, en in welke mate een makkelijke regel gehanteerd kan worden in de praktijk. Daartoe wordt een MBP model opgesteld, waarin het onderscheid naar bloedgroepen in eerste instantie achterwege gelaten wordt. Het uitgiftebeleid ten aanzien van BPPs wordt middels eenvoudige regels (FIFO, LIFO, of mengvormen) gemodelleerd. In 2004 kenden BPPs in Nederland een houdbaarheid van maximaal vijf dagen. Het aantal BPPs op voorraad dient daarom opgesplitst te worden in vijf leeftijdscategorieën. De voorraad wordt in een MBP model aldus middels een vector aangeduid: $\mathbf{x} = (x_1, \ldots, x_5)$, met x_r is het aantal BPPs met een resterende houdbaarheid van r dagen. Als er iedere dag hoogstens 9 BPPs geproduceerd worden, dan zijn er reeds 10^5 mogelijke toestanden ten aanzien van de voorraad te beschouwen. Daarnaast maakt de dag van de week (Maandag t/m Zondag) onderdeel uit van de toestandsruimte. Huidige ontwikkelingen staan het toe om BPPs 7 dagen op voorraad te houden, het aantal voorraadtoestanden is dan 10^7.

De productievolumes liggen in de praktijk van de Nederlandse bloedbanken aanzienlijk hoger dan 9, waardoor het aantal toestanden al snel te groot wordt om nog een optimale oplossing te bepalen. De toestandsruimte van het MBP kan nog enigzins gereduceerd worden doordat er in weekeinden niet geproduceerd wordt en door toestanden met een extreem voorraadniveau uit te sluiten in de berekeningen. Zelfs dan, is veelal het aantal toestanden te groot om een optimale oplossing te berekenen.

De structuur van de toestandsruimte staat het toe om toestanden te aggregeren. Door de vraag en de toegestane productievolumes te modelleren in batches van, zeg, 4 pools, kan het aantal toestanden gereduceerd worden met een factor 4^m, waarbij m de maximale houdbaarheid is, gemeten in dagen. Als de maximale houdbaarheid 5 dagen is ($m = 5$), wordt de toestandsruimte gereduceerd met ruwweg een factor 1000. Schaling van het MBP vergt echter het zorgvuldig aanpassen van de toestandsovergangen en met name van de overgangskansen.

Eventueel na schaling, kan er een optimale strategie berekend worden. De optimale strategie kan vrij complex zijn daar deze afhankelijk is van de leeftijd van de op voorraad liggende BPPs. Op basis van de optimale bestelstrategie, wordt indien mogelijk een eenvoudige regel afgeleid. Daartoe wordt gekeken welke toestanden vaak voorkomen onder de optimale strategie en welke bestelvolumes daarbij horen. Met dit doel wordt een frequentietabel opgesteld door simulatie van de optimale strategie.

Hieruit blijkt dat voor data van één van de Nederlandse bloedbanken, een aanvulregel met vaste aanvulniveaus voor iedere werkdag goed aansluit bij de optimale strategie. De prestatie van de aanvulregel kan geevalueerd worden door numerieke evaluatie van de onderliggende Markov ketens of middels simulatie. Het blijkt dat voor de gegeven dataset de regel inderdaad vrijwel optimaal presteert. Voor praktisch gebruik worden de aanvulniveaus van deze regel teruggeschaald, zodat voor het ongeschaalde probleem optimale productievolumes benaderd kunnen worden.

De aanvulregel is vervolgens toegepast in een simulatiemodel waarin patiënten en donoren van verschillende bloedgroepen onderscheiden worden. Hieruit blijkt het gerechtvaardigd te zijn om het onderscheid van bloedgroepen in het MBP model achterwege te laten. De bloedgroepen zijn immers in ruime mate uitwisselbaar en in de Nederlandse situatie blijven er voldoende bloedplaatjes over na het verwerken van de volbloeddonaties tot rode bloedcelconcentraten (RBCs).

Vele gevoeligheidsanalyses zijn verricht en alternatieve bestelregels zijn afgeleid met het oog op situaties bij andere bloedbanken en ziekenhuizen, alwaar:

- een ander uitgifte beleid gehanteerd kan worden,
- de vraag opgesplitst kan worden in een categorie voor patiënten die de jongste BPPs op voorraad krijgen en een categorie die de oudste plaatjes krijgen,
- een kortere of langere maximale houdbaarheid van BPPs van toepassing kan zijn,
- een andere afweging gemaakt kan worden tussen verval en beschikbaarheid,
- de vraag meer of minder onzeker kan zijn,
- de problematiek zich op een kleinere schaal kan afspelen, of
- er vaste kosten gemoeid kunnen gaan met het opstarten van een productierun of met het plaatsen en afleveren van een bestelling.

In eerste instantie is het model opgezet voor een periodiek stationair MBP. Later is de methode uitgebreid opdat niet-stationaire perioden, zoals rondom vakantiedagen, in de planningshorizon opgenomen kunnen worden. Uit de gedetailleerde simulatieresultaten blijkt dat het verval van BPPs aanzienlijk teruggebracht kan worden van 15 à 20% in de huidige praktijk tot slechts een paar procent of zelfs minder dan 1%. De beschikbaarheid van BPPs hoeft zelden een probleem te zijn.

Bloedbankmanagers zijn dermate positief over de ontwikkelde methode, dat nu een gebruikersvriendelijke versie van de ontwikkelde programmatuur draait bij een van de Nederlandse bloedbanken.

Deel II – Dynamische regeling van verkeerslichten (H5 tot H8)

Verkeerslichten zijn geïntroduceerd om het wegverkeer veiliger en efficiënter te maken. Sommige verkeersstromen kunnen tegelijkertijd groen genieten, mits ze elkaar niet kruisen. Verkeersdeelnemers in de andere verkeersstromen zullen dan echter moeten wachten totdat zij groen krijgen. De wachttijd wordt beheerst door een softwarematig gestuurde regeling van de verkeerslichten. Informatie over het aantal wachtende auto's op iedere rijbaan kan aan de softwarematige regelaar meegegeven worden. Stel er zijn F rijbanen, het aantal wachtende auto's kan dan worden aangeduid met de vector $\mathbf{q} = (q_1, q_2, \ldots, q_F)$.

De uitdaging om de lange-termijn-gemiddelde wachttijd voor automobilisten te minimaliseren op een enkel kruispunt kan geformuleerd worden als een MBP. Het MBP model met $(x, \mathbf{q})$ als toestandsbeschrijving, waarbij x de toestand ('kleur') van de verkeerslichten aangeeft. In eerste instantie veronderstellen we een stationair aankomstpatroon van auto's. In het MBP worden andere verkeersstromen, zoals voetgangers, fietsers en het openbaar vervoer buiten beschouwing gelaten. Niettemin worden uitbreidingen besproken in Hoofdstuk 6.

Het potentieel aantal wachtrijtoestanden groeit exponentieel met het aantal wachtrijen. Ter illustratie, als er 4 rijbanen samenkomen op een kruispunt, en op iedere rijbaan maximaal 9 auto's zullen wachten is het aantal mogelijke vectoren $\mathbf{q}$ reeds $10^4 = 10.000$. Als er echter 12 rijbanen zijn, dan zijn er reeds 1 biljoen toestanden mogelijk. Voor een optimale strategie heeft men voor iedere mogelijke toestand een optimaal besluit ten aanzien van het verkeerslicht te berekenen.

De optimale oplossing van een MBP met tientallen miljoenen toestanden kan niet in afzienbare tijd bepaald worden. Daarom zal men benaderingen moeten accepteren. Als het probleem van een enkel kruispunt een benadering vergt, dan heeft men voor het regelen van een netwerk van kruispunten zeker benaderingen nodig. In de praktijk treft men veelal heuristsche regels aan. De vaste cyclus, afgekort met FC voor Fixed Cycle, is wellicht de meest bekende: iedere stroom krijgt gedurende een vaste tijd groen en de stromen worden in een vaste cylische volgorde afgewerkt. Daarnaast zijn er drukte-afhankelijke regelingen zoals de cyclische uitputtende regeling, XC voor cyclic exhaustive control, waarbij een licht op groen blijft totdat er geen verkeer meer is nabij de stopstreep van de betreffende rijbaan.

FC is een interessante basisstrategie om tot een benadering van een optimale regeling te komen. FC doorloopt een cyclus van zeg D tijdslots, die we kunnen nummeren van 1 tot D. De lengte van een tijdslot is gelijk gesteld aan 2 seconden. Stel dat verkeersstroom 1 groen heeft gedurende slots 1 t/m 5, dan zijn gedurende slots 6 en 7 de lichten oranje, en gedurende slots 8 t/m D zal stroom 1 rood licht hebben. De kleur van de verkeerslichten kan onder FC dus eenvoudigweg bepaald worden op basis van de positie in de cyclus: het nummer van het tijdslot. De 'roodtijd' van een stroom is dus onafhankelijk van het aantal auto's dat staat te wachten op de andere rijbanen. De gemiddelde wachttijd per auto kan voor FC bepaald worden door de onderliggende Markov keten numeriek te evalueren. Een toestandsbeschrijving van de Markov keten voor stroom f bevat het slotnummer $t \in \{1, \ldots, D\}$ en het aantal auto's dat op de rijbaan staat te wachten q_f.

Middels een één-stapsverbeteralgoritme, kan een betere strategie dan FC bepaald worden. Daartoe zijn de relatieve waarden van toestanden voor een goede vaste cyclus nodig. Daar de toestandsruimte onder FC slechts twee-dimensionaal is, kunnen voor alle verkeersstromen f de relatieve waarden van alle toestanden (t, q_f) snel berekend worden. Hierbij moet wel rekening gehouden worden dat de Markov keten voor FC periodiek is, en dat in de berekening van de relatieve waarden de in principe oneindige horizon wordt afgekapt. De relatieve waarde van een toestand $(t, q_1, \ldots, q_F)$ onder FC, is de som van F relatieve waarden: één voor iedere verkeersstroom.

De verbeterde strategie heet RV (voor 'Relative Value'). RV kent geen vaste groen perioden en hoeft ook niet per se cyclisch te zijn. RV doorbreekt FC door de positie in de cyclus aan te passen, waardoor de groenperiodes van de verkeersstroom dynamisch verlengd, verkort, of beëindigd worden. Naast RV zijn er andere strategiën afgeleid op basis van de relatieve waarden onder FC.

De enige manier om de kwaliteit van RV te bepalen is door een aantal kruispunten te simuleren onder verschillende omstandigheden, bijvoorbeeld ten aanzien van de aankomstintensiteiten. RV is aldus vergeleken met eenvoudige regels zoals FC en een aantal varianten van de uitputtende regeling (XC). Voor infrastructuren met slechts een klein aantal stromen, zeg $F \leq 4$, kan de optimale oplossing voor het MBP berekend worden. De resultaten van een simulatiestudie zijn veelbelovend: in alle (volledig) symmetrische gevallen die beschouwd zijn, is voor alle stromen de gemiddelde wachttijd op lange termijn voor RV 20 procent lager dan voor FC. In Hoofdstuk 6 blijkt dat de gemiddelde wachttijd voor RV zelfs meer dan 70 procent lager kan zijn dan voor FC en XC.

Moderne technologie is in staat om de positie te bepalen van auto's die op weg zijn naar een kruispunt. Schattingen van de verwachte aankomsttijden van deze auto's kunnen in de toestandbeschrijving worden opgenomen. De toestandsruimte per stroom, bestaat dan uit de toestand van het verkeerslicht, het aantal wachtende auto's en een vector met de aankomstinformatie voor de komende, zeg, vijf tijdslots (10 seconden). Het aantal toestanden neemt daarmee drastisch toe. Een optimale oplossing op basis van een MBP model kan dan niet bepaald worden. Daarom is de RV regel uitgebreid met aankomstinformatie. In verscheidene simulaties blijkt dat het toevoegen van 10 seconden aankomstinformatie, leidt tot een significante verbetering van RV (zie Hoofdstuk 7). Zelfs als de aankomstinformatie niet perfect is doordat auto's in het simulatiemodel met verschillende snelheden kunnen rijden, geeft RV met aankomstinfo goede resultaten.

In een netwerk van kruispunten worden lichten vaak statisch geregeld middels een netwerk FC, daar de lichten dan in principe op elkaar afgesteld kunnen worden zodanig dat een groene golf ervaren kan worden in 1 of meerdere richtingen. Vaak gaan groene golven verloren doordat auto's niet allemaal even hard rijden en doordat auto's vertraagd worden door andere auto's die nog staan te wachten. Vanuit het oogpunt van minimalisatie van de gemiddelde wachttijd kan het beter zijn om ook in een netwerk de lichten lokaal en drukte-afhankelijk te regelen. De RV regels met en zonder aankomstinformatie worden in Hoofdstuk 8 gesimuleerd voor verschillende soorten verkeersaders en netwerken.

De gemiddelde wachttijd wordt vergeleken met de best-gevonden netwerk cyclus (FC). Het blijkt dat een RV regel met aankomstinformatie een gemiddelde wachttijd geeft die veelal lager is dan die voor de gesynchroniseerde netwerk FC. Door afslaand verkeer en snelheidsverschillen blijkt de gemiddelde wachttijd in een netwerk van negen kruispunten voor FC meer dan 40% hoger te zijn dan die voor RV met aankomstinformatie. Een ander voordeel van RV is dat het bepalen van een goede basis cyclus veel makkelijker is dan het bepalen van de beste netwerk FC.

Conclusie

De twee praktische optimalisatievraagstukken hebben de neiging dermate omvangrijk te zijn, dat een optimale strategie niet berekend kan worden. Het aantal toestanden in het onderliggende MBP model is (te) groot doordat de zogenaamde toestandsruimte multidimensionaal is. De structuur van de toestandsruimte voor de twee problemen is verschillend, waardoor een verschillende aanpak nodig is.

In Deel I wordt de structuur van een voorraadprobleem bij bloedbanken bestudeerd en blijkt de *aggregatie* van toestanden mogelijk te zijn: toestanden worden geaggregeerd door bloedgroepen samen te nemen, door de vraag naar en de productie van BPPs te modelleren in batches, en tenslotte door leeftijdsgroepen samen te nemen in een vaste aanvulregel.

In Deel II wordt strucuur aan het verkeerslichtenprobleem toegevoegd door een speciale strategie als uitgangspunt te nemen. *Decompositie* van de toestandsruimte blijkt dan mogelijk waardoor een *één-stapsverbeteralgoritme* toegepast kan worden. Decompositie vindt plaats op twee niveaus: voor de regeling van een netwerk van kruispunten wordt het netwerk eerst gedecomponeerd in de afzonderlijke kruispunten, en per kruispunt wordt het rekenwerk opgesplitst in rekenwerk voor iedere verkeersstroom afzonderlijk.

Grote gestructureerde Markov beslisproblemen kunnen aldus succesvol opgelost worden door de structuur van het probleem te benutten. Een uniforme aanpak van hoog-dimensionale Markov beslisproblemen bestaat wellicht niet; elk probleem vraagt om een probleemspecifieke aanpak. Dit proefschrift geeft inzicht dat kan bijdragen tot de oplossing van andere grote MBPs.