



Inhoud

Status	1
Disseminatie	3
Gebruikers- & interessegroep	
Haalbaarheidstudies	
Werktuigendagen 2011	
Aankondigingen	
Projectresultaten	3
Nieuwe ontwikkelingscases	
Iteratieve feedback tuning	
Energie- vs. tijdsefficiënte controle	
Identificatie op basis van binaire sensordata	
Controle op basis van binaire sensordata	

Vragen? Opmerkingen? Contacteer ons!

FMTC
Wim Symens / Gregory Pinte
Campus Arenberg
Celestijnenlaan 300D - bus 4027
B-3001 Heverlee
Tel.: +32 16 32 80 50
Fax: +32 16 32 80 64
E-mail: info@fmtc.be

Indien U deze nieuwsbrief niet meer wenst te ontvangen, gelieve dit te laten weten op bovenstaand adres.

Beste,

Het IWT-SBO project LeCoPro, waarin de mogelijkheden van lerende controle in mechatronische toepassingen onderzocht worden, is halfweg. Tijdens de eerste twee jaren van het project zijn we erin geslaagd verschillende innovatieve lerende strategieën te ontwikkelen voor de controle van machines. Deze algoritmische ontwikkelingen bij de verschillende onderzoeksgroepen hebben reeds geleid tot een aantal concrete implementaties op echte machines. Denk bijvoorbeeld maar aan de autonome lerende tractor die gepresenteerd werd op de Werktuigendagen in Oudenaarde! Met deze nieuwsbrief willen we de geïnteresseerde bedrijven informeren over de behaalde resultaten en de geplande ontwikkelingen in dit project.

Veel leesgenot!

Wim Symens
FMTC – Coördinator LeCoPro

Status

Twee jaar geleden, in september 2009, ging het LeCoPro project van start met als doel om in Vlaanderen een kennisplatform op te richten rond lerende stuurstrategieën voor productiemachines. De door dit platform ontwikkelde technieken moeten de Vlaamse industrie toelaten de lerende machines van de toekomst op de markt te brengen.

Tijdens de eerste twee jaren van het project werkte het LeCoPro consortium, dat bestaat uit FMTC en onderzoeksgroepen van de KULeuven, VUB en UGent, voornamelijk rond de lerende controle van individuele subsystemen. Hierbij werden nieuwe modelgebaseerde en niet-modelgebaseerde lerende technieken ontwikkeld voor de controle van zowel systemen met een complexe dynamica als systemen die gebruikt worden in variërende omgevingscondities. Op het gebied van modelgebaseerde lerende

technieken werden bestaande concepten zoals Iteratief Lerende Controle en Model Predictieve Controle verder uitgebouwd maar werden ook nieuwe algoritmes ontwikkeld zoals Iteratieve Optimalisatie. Op het gebied van niet-modelgebaseerde lerende technieken lag de focus voornamelijk op de uitbreiding van Reinforcement Lerende en andere Automatisch Lerende technieken zodat deze ook bruikbaar worden voor de controle van mechatronische toepassingen.

De praktische validatie van de ontwikkelde algoritmes gebeurt op een aantal relevante, generieke ontwikkelingscases. Deze cases werden gedefinieerd in nauw overleg met de industriële leden van de gebruikers- en interessegroep om de ontwikkelingen zo goed mogelijk af te stemmen op de noden van de industrie. Zo werden vijf ontwikkelingscases geselecteerd waarop gewerkt wordt binnen het project: een koppeling in een power-shiftrtransmissie, een lineaire actuator, een badminton robot, een autonome tractor en een hydrostatische aandrijflijn.

Recent werd ook gestart met de ontwikkeling van lerende sturingen voor interagerende subsystemen. In het huidige regelontwerp wordt de interactie tussen verschillende subsystemen immers vaak niet in rekening gebracht wat leidt tot een suboptimale globale machineperformantie. Daarom worden binnen LeCoPro lerende regelalgoritmes ontworpen die de lokale controleacties voor de verschillende subsystemen op elkaar leren afstemmen zodat een globaal doel kan geoptimaliseerd worden. De ontwikkeling van deze globale lerende algoritmes vormt samen met de verdere praktische validatie van de algoritmes de belangrijkste uitdaging tijdens de volgende fase van het project.



De toetreding van een groot aantal bedrijven tot de gebruikers- en interessegroep tijdens de eerste fase van het project toont het belang van lerende controle voor de Vlaamse industrie. Op dit moment bestaat de gebruikersgroep uit 12 Vlaamse machinebouwers (Bekaert, CNH, Picanol, Van de Wiele, Sirris, Dana, Atlas Copco, LVD, Tenneco, LMS, Reintjes en Arodo). Daarnaast volgt een interessegroep van bedrijven uit andere sectoren (Induct, MESWare, OptimalAction en Ipcos) het project op. Tijdens de volgende twee jaar van het project zullen de projectresultaten in detail met deze bedrijven besproken worden. Op basis hiervan zal getracht worden om opvolgprojecten op te starten, waarin de ontwikkelde algoritmes op machines van deze bedrijven geïmplementeerd zullen worden.

Disseminatie

Gebruikersgroep & Interessegroep

Toetreden tot de gebruikersgroep of interessegroep van het project kan nog steeds! Wilt u meewerken aan het project? Laat het ons weten!

Bedrijven, die wensen toe te treden, kunnen:

- een gedetailleerd overzicht krijgen van bestaande en ontwikkelde lerende stuurtechnieken,
- inzicht opbouwen in voordelen van lerende sturingen,
- als eerste inzage krijgen in behaalde projectresultaten,
- feedback geven over de industriële relevantie van de ontwikkelingen in het project,
- deelnemen aan workshops en seminars over het project,
- actief participeren door het aanbieden van ontwikkelingscases voor de validatie van de ontwikkelde technieken,
- een haalbaarheidstudie laten uitvoeren van één van de ontwikkelde lerende technieken voor de sturing van hun machines (zie verder).

Haalbaarheidstudies

Leden van de gebruikersgroep kunnen een haalbaarheidsstudie laten uitvoeren van de ontwikkelde lerende algoritmes in het project op één van hun machines. Tijdens zulke studie bespreekt een onderzoeker van het LeCoPro consortium hoe een lerende controletechniek kan geïmplementeerd worden op een machine van het betrokken bedrijf. Daarnaast kunnen een aantal verkennende metingen uitgevoerd worden om de toepasbaarheid van de techniek in meer detail na te gaan. Zulke haalbaarheidstudie, die typisch enkele dagen in beslag neemt, is gratis.

Werktuigendagen 2011

Eén van de aandachtstrekkers op de afgelopen Internationale Werktuigendagen in Oudenaarde, een beurs over machines voor land- en tuinbouw, was ongetwijfeld de autonome tractor ontwikkeld in het LeCoPro project. Talrijke bezoekers van de beurs kwamen de tractor, die officieel werd voorgesteld, bewonderen terwijl hij autonoom over het veld rondreed.



Tractor in actie op de Werktuigendagen 2011

De afdeling MeBioS van de KULeuven bouwde tijdens het project in samenwerking met FMTC een tractor van New Holland om tot een autonoom rondrijdende veldrobot. Lineaire aandrijvingen werden aangebracht om de snelheid en de stuurhoek van het voertuig automatisch te kunnen aansturen. Daarnaast werd de tractor uitgerust met een GPS systeem om zijn exacte positie te kunnen opmeten met een nauwkeurigheid van enkele centimeters.

Na de ombouw van de tractor kon aan de belangrijkste uitdaging begonnen worden: het vinden van een gepaste sturing die de rol van de bestuurder overneemt. Een specifieke moeilijkheid voor de ontwikkeling van een autonoom landbouwvoertuig zijn de veranderende bodemkarakteristieken van het veld. De tractor moet zowel op een harde als een natte ondergrond kunnen rondrijden. Traditionele aanstuurtechnieken houden hier geen rekening mee en dienen anders ingesteld te worden voor elk type ondergrond. Hiervoor is de tussenkomst van de operator noodzakelijk. In het LeCoPro project is daarom een sturing ontwikkeld waarbij de tractor de bodemcondities aanleert en een inschatting maakt van het verwachte slipgedrag van de wielen. Op basis van een model van de tractor wordt vervolgens de optimale snelheid en stuurhoek berekend voor de huidige bodem. Door deze lerende sturing behaalt de ontwikkelde tractor een nauwkeurigheid van enkele centimeters.

Zulke nauwkeurigheid bij de sturing van landbouwmachines is van groot belang door de opkomst van de biologische landbouw. Hier is een accurate positionering van het gereedschap onontbeerlijk. Daarenboven kan door landbouwmachines autonoom te laten rondrijden, de steeds toenemende operatorcost aanzienlijk gereduceerd worden. Ook de afregeling van de sturing bij wijzigende bodemcondities wordt overbodig door het zelflerende gedrag van de robot.



Volgend filmpje toont hoe de ontwikkelde tractor autonoom rondrijdt over een veld:

www.lecopro.org/tractor.wmv

Speciale sessies op ICSTCC 2011 en 'Benelux Meeting on Systems and Control 2011'

Op vrijdag 14 oktober 2011 werd het LeCoPro consortium uitgenodigd op de ICSTCC conferentie in Roemenië om hun activiteiten rond de controle van koppelingen toe te lichten. Tijdens deze internationale conferentie, waarop ontwikkelingen op het gebied van systeemtheorie en controle voorgesteld worden, werd een speciale sessie rond de modellering en controle van aandrijflijnen georganiseerd door prof. Lazar van de Gheorge Asachi universiteit uit Iasi in samenwerking met prof. De Keyser van de SYSTeMS onderzoeksgroep. Prof. Lazar is de coördinator van een Roemeens project, SICONA (<http://www.sicona.ac.tuiasi.ro/>), waar in samenwerking met Continental Automotive real-time systeemcontrollers ontwikkeld worden voor aandrijflijnen in automobieltoepassingen. Verdere samenwerkingsmogelijkheden tussen het LeCoPro en SICONA consortium worden momenteel onderzocht.

Ook op de 30^{ste} 'Benelux Meeting on Systems and Control' werden in een special sessie de realisaties van het LeCoPro project toegelicht aan academische specialisten in dit domein. Dit heeft toegelaten de LeCoPro ontwikkelingen op te lijnen met het onderzoek rond lerende controle in Nederland.

Aankondigingen

Date	Place	Event
27 januari 2012	FMTC, Leuven	Workshop voor de gebruikersgroep
10-15 juni 2012	Brisbane, Australia	Speciale sessie "Real-world applications of Reinforcement Learning" op de IEEE-INNS International Joint Conference on Neural Networks

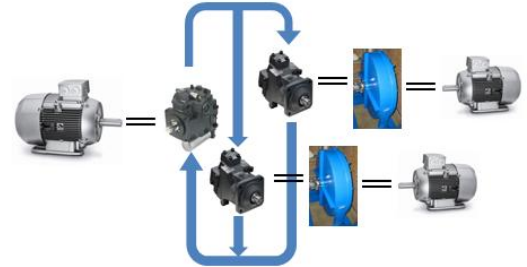
Projectresultaten

Ontwikkelingscases

Naast de koppelingen in een power-shifttransmissie, de badminton robot en de autonome tractor, die in de vorige nieuwsbrief voorgesteld werden, zijn er in het tweede jaar van het project twee nieuwe ontwikkelingscases gedefinieerd:

- **Hydrostatische aandrijving:** Dergelijke aandrijving wordt vaak gebruikt voor de tractie van zwaar belaste mobiele voertuigen. Doordat de vermogensoverdracht hydraulisch gebeurt via vaste buizen of flexibele slangen, is er geen nood aan

aandrijfassen en differentiëlen. Door de slagvolumes van de pomp en/of de hydromotor elektronisch te verstellen kan een continu variabele transmissie (CVT) gerealiseerd worden. Zo is een mechanische versnellingskast overbodig.



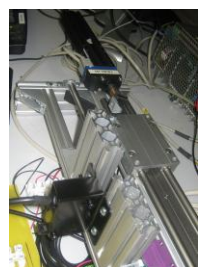
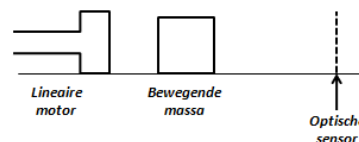
Conceptuele opbouw van de hydrostatische aandrijflijn

In de gebouwde testopstelling (zie bovenstaande figuur) drijft een elektromotor via een pomp met een variabel slagvolume twee hydromotoren met een variabel slagvolume aan. Elke hydromotor is verbonden met een vliegwiel, dat via een elektromotor kan afgeremd worden. In dit systeem is er een sterke koppeling tussen elektrische, hydraulische en mechanische onderdelen. Hiermee dient rekening gehouden te worden bij de sturing van de snelheid van elektromotor en de slagvolumes van de pomp en hydromotoren.



Hydrostatische testopstelling in het FMTC labo

- **Lineaire actuator:** In een tweede, meer conceptuele testopstelling duwt een lineaire actuator een massa vooruit. Het doel is om deze massa op een bepaald tijdstip te laten passeren voor een optische sensor.



Lineaire actuator met optische sensor

Hiertoe zijn identificatie- en controle-algoritmes ontwikkeld die gebruik maken van binaire sensordata. Vermits de exacte positie van de massa niet wordt opgemeten, kan op elk tijdstip immers



enkel geverifieerd worden of de massa zich vóór of achter de optische sensor bevindt. De op deze testopstelling ontwikkelde algoritmes zijn bruikbaar in situaties waarbij een continue sensor teveel kost of moeilijk te installeren is.

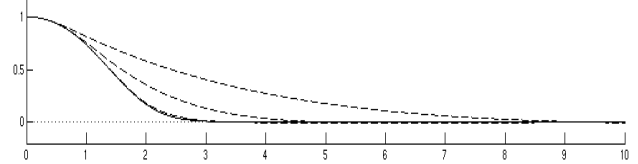
Iteratieve feedback tuning

De overgrote meerderheid van regelaars in de industrie hebben een vrij eenvoudige structuur en worden ingesteld door het vastleggen van enkele parameters. Het bekendste voorbeeld is de PID-regelaar. Teneinde een goede regeling te bekomen, wordt tijdens open lus experimenten op de machine het in- en uitgangssignaal opgemeten om het systeemgedrag te schatten. Aan de hand van een aantal vuistregels worden op basis hiervan de optimale regelparameters berekend worden aan de hand van een aantal vuistregels. Verdere fijnstelling gebeurt dikwijls op basis van ervaring met de machine en trial-and-error. In de praktijk worden deze regelaars, wanneer ze eenmaal ingesteld zijn, niet meer bijgesteld, ofwel omdat de kennis ontbreekt bij de gebruiker, ofwel omdat het gewoon praktisch niet mogelijk is om nog experimenten in open lus uit te voeren. Nochtans is bijstelling noodzakelijk in veel machines om optimaal regelgedrag in uiteenlopende omstandigheden te kunnen garanderen.

Iteratieve feedback tuning (IFT) biedt een antwoord op deze problematiek. Het is een automatische manier om regelingen bij te stellen, waarbij identificatie van het systeem en berekenen van optimale regelparameters in een iteratieve procedure gecombineerd zijn. De regellus van het systeem blijft gedurende de hele procedure gesloten, wat de techniek geschikt maakt om regelingen continu te blijven bijsturen. Eén iteratie bestaat uit een observatie- en een excitatiefase. Tijdens de observatiefase wordt het in- en uitgangssignaal van het geregelde systeem enkel geobserveerd. Op basis hiervan wordt een excitatiesignaal berekend en aangelegd tijdens de excitatiefase. Aan de hand van geobserveerd in- en uitgangsgedrag gedurende beide fases, is het mogelijk om die regelparameters te bepalen die de afwijking tussen het uitgangs- en referentiesignaal minimaliseren.

Een toepassing van IFT is het optimaliseren van de stuurregeling van de autonome tractor. Hierbij past men een PID-regeling toe op de afwijking van de tractor van een gegeven rechte lijn traject. Na het omkeren op het einde van een veld staat de tractor meestal niet juist gealigneerd met het te volgen traject. Om productieverlies te vermijden, is het belangrijk dat deze afwijking zo snel mogelijk weggeregeld wordt. Hiertoe is een juiste instelling van de parameters van de PID-regelaar noodzakelijk. Gezien het stuurgedrag van de tractor en bijgevolg ook de optimale regelparameters in sterke mate bepaald worden door de bodemcondities van het veld, is IFT een geschikte methode om de stuurregeling hieraan aan te passen. Dit wordt geïllustreerd in onderstaande figuur. Terwijl met de initiële

regelparameters de tractor een 10-tal meter nodig heeft om het gewenste rechte lijn traject na een bocht terug te volgen, rijdt de tractor na een convergentie van het IFT algoritme van 10 iteraties reeds na minder dan 3 meter terug recht.



Het volgedrag van de autonome tractor na verschillende iteraties van het IFT algoritme (stippellijnen: na resp. 0, 5 en 10 iteraties; volle lijn: na 15 iteraties)

Energie- vs. tijdsefficiënte controle

Op eenzelfde machine moeten, afhankelijk van de toepassing, bewegingen soms zo snel mogelijk en soms zo energie-efficiënt mogelijk uitgevoerd worden. Dit leidt tot een volledig verschillende aansturing van de machines. In het kader van het LeCoPro project heeft PMA zowel tijds- als energie-optimale controle-algoritmes ontwikkeld en de performantie en het energieverbruik van beide vergeleken op de FMTC badminton robot.

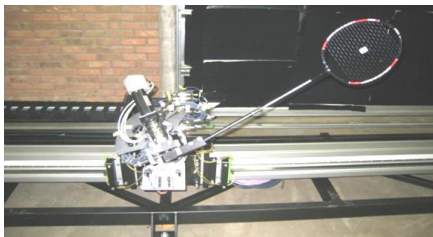
In een tijdsoptimaal algoritme wordt het stuursignaal berekend dat een systeem een bepaalde beweging zo snel mogelijk laat uitvoeren zonder de opgelegde systeembepalingen (bv. maximale snelheid, versnelling en stroom) te overschrijden. Hiertoe wordt een optimalisatieprobleem opgesteld waarbij de duur van de beweging de doelfunctie is, die moet geminimaliseerd worden. De systeemdynamica en de systeembepalingen worden als randvoorwaarden van het optimalisatieprobleem geformuleerd. Twee types tijdsoptimale algoritmes zijn ontwikkeld om dit optimalisatieprobleem op te lossen: iteratieve algoritmes, die het volledige optimale aanstuurtraject telkens berekenen voor de machine een nieuwe beweging aanvat en dit stuursignaal vervolgens in feedforward uitsturen, en feedback algoritmes, die het aanstuurtraject berekenen en bijsturen tijdens de beweging zelf.

In de iteratieve algoritmes wordt de dynamica van het gecontroleerde systeem tussen twee opeenvolgende bewegingen iteratief geschat en worden op basis hiervan de randvoorwaarden van het op te lossen tijds-optimalisatieprobleem aangepast. Na convergentie van de geschatte systeemdynamica naar de reële dynamica wordt de minimale duur voor een bepaalde beweging en het bijhorende stuursignaal gevonden. Bij de iteratieve algoritmes is er tussen twee opeenvolgende bewegingen voldoende tijd beschikbaar om het tijds-optimalisatieprobleem op te lossen. Bij het ontwerp van feedback algoritmes is de snelle oplosbaarheid van dit probleem wel een belangrijke factor bij het ontwerp van de feedback algoritmes. Deze algoritmes lossen immers elke tijdstap een optimalisatieprobleem op om zo te kunnen



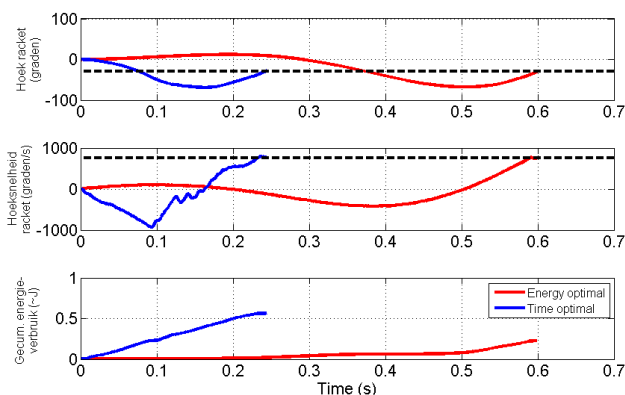
anticiperen op onverwachte storingen of niet-gemodelleerd systeemgedrag. Omwille van de snelle dynamica wordt de bemonsteringsperiode van de controllers voor veel mechatronische systemen vrij kort gekozen (in de grootteorde van een tiental milliseconden) en is de beschikbare rekentijd beperkt. Door een slimme herformulering van het tijds optimalisatieprobleem kon ervoor gezorgd worden dat het oplosbaar werd binnen enkele milliseconden.

Naast deze tijds optimale algoritmes werden ook algoritmes ontwikkeld, waarbij het energieverbruik geminimaliseerd wordt. Hierbij wordt het energieverbruik gedurende de volledige beweging uitgedrukt als een functie van het aangelegde stuursignaal. Vervolgens wordt een optimalisatieprobleem opgelost waarbij het stuursignaal gezocht wordt dat leidt tot een minimaal energieverbruik. Typisch wordt als extra randvoorwaarde (naast de systeemdynamica en de systeembependingen) wel opgelegd dat de beweging binnen een bepaalde tijd dient uitgevoerd te worden. Ook hiervoor werden zowel iteratieve als feedback algoritmes ontwikkeld.



FMTC badminton demonstrator met opslagmechanisme

De tijds- en energie-optimale controle-algoritmes zijn beide geïmplementeerd in de sturing van de badminton robot. De algoritmes worden gebruikt om de robot een opslag te laten uitvoeren. Wanneer een shuttle van op een bepaalde hoogte losgelaten wordt uit een dropmechanisme, moet de robot vanuit een vaste beginpositie bewegen om de shuttle in een bepaalde richting weg te slaan.



Hoekverdraaiing, hoeksnelheid en het energieverbruik tijdens een tijds- en energie-optimale beweging van de badminton robot

In geval van het tijds optimale algoritme blijft de robot zo lang mogelijk stilstaan om dan op een zo kort mogelijke tijd de shuttle weg te slaan. Het energie-optimale algoritme laat de beweging van de robot vroeger starten, maar beperkt de

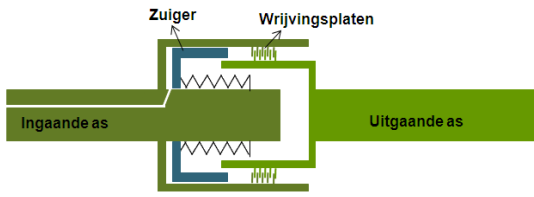
snelheid en versnelling van de verschillende motoren. Op deze manier worden zowel de koperverliezen in de motoren als de wrijvingsverliezen in de geleidingen beperkt. Zoals blijkt uit bovenstaande figuur, kan het tijds optimale algoritme de duur van de beweging met ongeveer 60% beperken maar leidt dit wel tot een extra energieverbruik van 125%.

Identificatie op basis van binaire sensor data

Binnen het LeCoPro project hebben PMA en FMTC identificatie-algoritmes ontwikkeld die enkel gebruik maken van binaire metingen van de systeemuitgang. Binaire meetsignalen treden op in systemen waarvan de uitgang niet direct kan opgemeten worden, maar er wel gedetecteerd kan worden of de uitgang zich boven of onder een bepaalde grenswaarde bevindt. Een typisch voorbeeld is een systeem met een optische sensor, die detecteert of een bewegend object een bepaald punt passeert. Terwijl traditionele algoritmes typisch zeer lange en/of periodische excitatiesignalen vereisen, kan de modelidentificatie in de ontwikkelde algoritmes ook gebeuren op basis van korte, onafhankelijke inputsignalen. Zulke korte signalen kunnen meestal tijdens normale machinewerking opgemeten worden en vereisen bijgevolg geen bijkomende identificatie-experimenten.

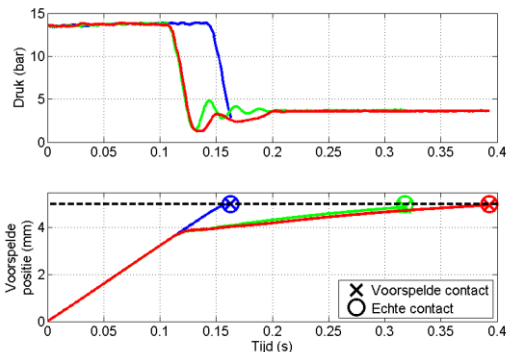
In de ontwikkelde identificatie-algoritmes wordt een modelstructuur vooropgesteld die de dynamica van het geïdentificeerde systeem beschrijft tussen het aangelegde inputsignaal en de niet opgemeten systeemuitgang. Deze modelstructuur kan bijvoorbeeld een transferfunctie van een bepaalde orde zijn, of een impulsresponsie van een bepaalde lengte. De parameters van deze modellen worden vervolgens geschat door het verschil te minimaliseren tussen de effectief opgemeten binaire sensoruitgang en de door het model voorspelde binaire sensoruitgang. Dit optimalisatieprobleem kan door een herformulering van de kostfunctie met lokale optimalisatietechnieken aangepakt worden en zo binnen een beperkte rekentijd opgelost worden.

Een industrieel relevant voorbeeld is de schatting van de zuigerpositie in een natte platenkoppeling. Tijdens de sluiting van zulke koppeling is de zuiger in een eerste fase, de vulfase, nog niet in contact met de wrijvingsplaten en wordt er geen koppel overgedragen tussen de ingaande en uitgaande as van de koppeling. Wanneer de oliedruk in de leiding hoog genoeg is om de kracht van de veer te overwinnen, duwt de zuiger de wrijvingsplaten tegen elkaar en wordt er koppel overgedragen. Vermits de koppeling een roterend element is, is het duur om sensoren aan te brengen die de positie van de zuiger meten. Snelheidssensoren die de rotationele snelheid van de in- en uitgaande as opmeten zijn typisch wel geïnstalleerd en kunnen gebruikt worden om in te schatten wanneer de zuiger in contact komt met de wrijvingsplaten.



Doorsnede van een koppeling

Op basis van bovenstaande identificatietechnieken kan dan een model worden opgesteld dat de positie van de zuiger tijdens de vulfase voorspelt. Het geschatte model is in staat om de positie te voorspellen met een nauwkeurigheid van 1%. Dit wordt geïllustreerd op onderstaande figuur, waarop voor drie willekeurige controlesignalen de geschatte zuigerpositie en het werkelijke contactpunt met de wrijvingsplaten wordt weergegeven.



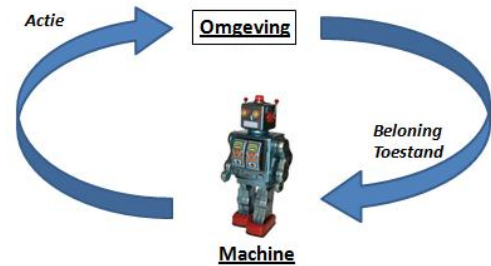
Geschatte zuigerpositie voor drie willekeurige druksignalen naar de koppeling

Controle op basis van binaire sensor data

Binnen het LeCoPro project zijn niet enkel identificatie- maar ook controle-algoritmes ontwikkeld voor systemen met binaire uitgangen. Hiertoe kan enerzijds een beroep gedaan worden op traditionele modelgebaseerde controle-algoritmes, die verder bouwen op de geschatte modellen uit de identificatie. Anderzijds kunnen ook model-vrije lerende technieken gebruikt worden om deze systemen te controleren. Zo heeft COMO in samenwerking met FMTC Reinforcement Lerende (RL) algoritmes ontwikkeld, die bruikbaar zijn voor controle op basis van binaire sensordata.

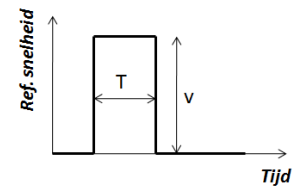
RL staat voor het leren uit beloningen. Het is een model-vrije techniek waarbij niet uitgegaan wordt van a priori systeemkennis maar geleerd wordt door middel van interactie met de omgeving. Afhankelijk van de toestand waarin het systeem zich bevindt, bepaalt het RL algoritme welke actie het systeem moet ondernemen. Hierbij is het belangrijk exploitatie (kiezen voor goede acties) af te wisselen met exploratie (het uitproberen van acties die 'minder' lijken, maar die misschien op termijn wel een beter systeemgedrag kunnen opleveren). Na de uitvoering van een actie komt het systeem in een andere toestand terecht en krijgt het afhankelijk van de kwaliteit van de gekozen actie een hogere of lagere beloning. Op basis hiervan past het RL algoritme vervolgens het gecontroleerde systeemgedrag aan. Cruciaal bij het opstellen van een RL algoritme is de

definitie van een juiste beloningsfunctie, die de kwaliteit van een actie afleidt uit beschikbare sensormetingen op en rond de machine.



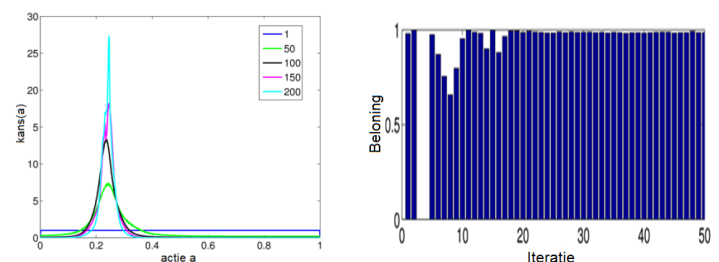
Concept van Reinforcement Leren

In tegenstelling tot traditionele feedback en feedforward controlealgoritmes, die een continu sensorsignaal vereisen, bieden RL algoritmes een uiterst geschikt framework aan om binaire sensorinformatie te verwerken. Binaire sensorinformatie kan immers typisch op een eenvoudige manier opgenomen worden in het beloningssignaal. Dit kan mooi geïllustreerd worden voor de controle van de lineaire motor, die een massa op een bepaald tijdstip voorbij een optische sensor dient te duwen. Het RL algoritme wordt dan gebruikt om de parameters van een geparametriseerd controlesignaal voor de motor te berekenen.



Geparametriseerd controle signaal naar motor

In de definitie van het RL algoritme wordt dan een lagere beloning toegekend wanneer het tijdstip van detectie van de massa door de optische sensor verder afwijkt van het gewenste tijdstip. Initieel legt het RL algoritme random controlesignalen aan en past op basis van de beloning voor deze actie de kans aan om deze actie in een volgende iteratie nogmaals aan te leggen. Zo verandert de initieel uniforme kansverdeling waaruit de parameters van de controlesignalen gekozen worden in een Gaussiaans verdeelde kansverdeling rond de parameters, die de massa op het gewenste tijdstip voor de optische sensor laat passeren (zie onderstaande figuur links). Bijgevolg neemt de beloning toe naarmate de massa meer is weggeduwd.



RL voor de controle van de motor: de kansverdeling voor de parameters van de controlesignalen na een verschillend aantal iteraties (links) en de behaalde beloning als functie van het aantal iteraties (rechts)