

Ro(bo)tterne og spildevandet

Kunstig intelligens, Deep learning, Internet of Things og Big Data – er det ikke noget med talegenkendelse og selvkørende biler? Jo, men det er også nye værktøjer, der lover at kunne løse mange af de problemer branchen kæmper med – manglende overblik og usikre målinger. I anvendelsen er machine learning, der måske bedst oversættes til dansk som ”automatiseret læring”, ofte en uigennemsigtig og lidt magisk sort boks, men resultaterne kan efterprøves fysisk og forståelsesmæssigt.

PETER RASCH

Indledning

Vandbranchen står over for et paradigmeskift med digitalisering – mener nogen. Andre siger, at det hele er opreklameret og gammel vin på nye flasker. Vi mener at begge dele er rigtigt og vil herunder argumentere for, hvordan det kan hænge sammen. Samtidig giver vi med nogle konkrete eksempler fra Odense, København og Aarhus vores bud på, hvordan forsyningerne og deres rådgivere i højere grad kan tænke data science ind i det daglige arbejde.

Data og machine learning

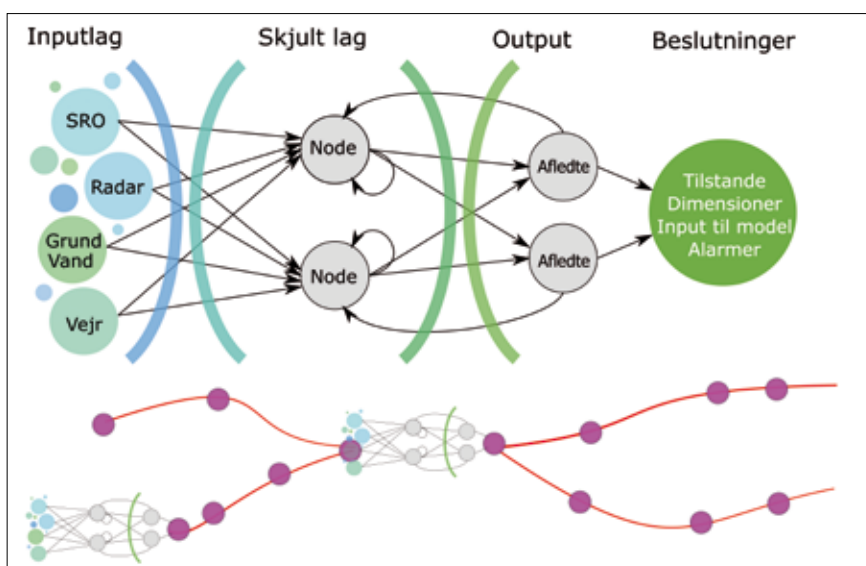
Der findes mange gode guider og introduktioner til kunstig intelligens og ”machine learning” i krydsfeltet mellem data, statistik, modeller og algoritmer. Udfordringen er ofte at få styr på begreberne og finde relevante lærings-eksempler med det rette faglige niveau. Hos InforMetics har vi valgt en pragmatisk tilgang bredt funderet på machine learning ud fra erkendelsen af, at mængden af data – også i vandsektoren – mangedobles hvert år. Som ingeniører, hydraulikere, fysikere og biologer må erkende at vi skal have maskinernes hjælp til ikke kun at filtrere og håndtere data, men også i højere og højere grad til at finde de sammenhænge i data, der fører til ny viden om dynamikken i de tekniske vandssystemer og som danner basis for at kunne tage operationelle beslutninger.

Machine learning – statistik og modellering

Konkret kan man tænke på machine learning som en sort boks, der meget hurtigt tilpasser et sæt matematiske og statistiske algoritmer til at finde en sammenhæng mellem to eller flere inputtidsserier på den ene side og et ønsket output på den anden side. I praksis er det meget lig arbejdet med tilsvarende deterministiske eller statistiske modeller, forskellen er blot at modellen ikke først skal bygges, men i stedet trænes på et eksisterende sæt af input og output. Det smarte er for det første at alle typer af inputs kan bruges – uden hensynta-

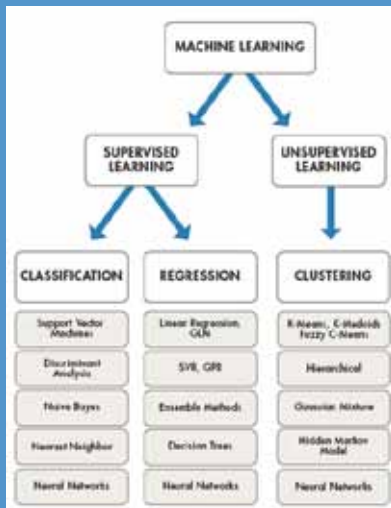
gen til enheder eller fysiske sammenhænge – og for det andet at det går stærkt, både i mandetimer og computertid, selv med relativt mange inputs og frihedsgrader i systemet. Begrænsningen består primært i at de historiske inputs der trænes på, ikke må adskille sig alt for væsentligt fra de nuværende og fremtidige data hvor de fundne relationer skal anvendes. Der er dog den fordel, at læringen kan gentages igen og igen så systemet har mulighed for at tilpasse sig en ny virkelighed.

Nedbørsrelaterede effekter har en stor grad af gentagelse – selvom to hændelser ikke er helt ens, så er der nogle klare sammen-



Et typisk data flow for en machine learning applikation med input fra diverse datakilder, et antal skjulte noder der forbinder inputs – og genererer et eller flere outputs. Nederst et eksempel på hvordan et neuralt netværk kan kombineres med en 1D model som input, eller for at udskifte en sektion af nettet.

Typer af Machine learning



Et eksempel på klassificering af machine learning værktøjerne. I praksis smelter disse værktøjer dog sammen og kan alle anvendes igennem en fælles grænseflade – hvilket dog også forstærker fornemmelsen af at arbejde med en sort boks.

hængemellem nedbørstidsserien i tid og rum og responsen i afløbssystemet. Dette gælder uanset hvor man er i systemet, og også selvom systemet ikke er helt ens mellem følgende hændelser. Netop evnen til at håndtere ikke perfekte data, er en af de helt store styrker ved machine learning.

Integration med eksisterende værktøjer

Branchen bruger allerede i høj grad deterministiske modeller (som fx MIKE Urban/Flood) og statistiske værktøjer, og fordi arbejdsgangen med machine learning værktøjer ikke er væsentlig forskellig fra disse, er det relativt nemt at integrere i det daglige. Herudover er det også oplagt at kombinere machine learning med de eksisterende værktøjer – f.eks. ved at lade et neuralt netværk øve sig på forholdet mellem nogle opstrøms betingelser med en nedstrøms betingelse, og så erstatte den gren i den hydrologiske model. Dermed opnås en alternativ og hurtig beskrivelse i et område der måske ikke er opmålt i tilstrækkelig grad, og maskinerne har i princippet lavet en autokalibrering med mulighed for løbende forbedring.

I de projekter InforMetrics har arbejdet med, har det vist sig, at selvom det er vigtigt at have den fulde årstidsvariation i nedbør og indløbsdata med, så er der allerede efter 3 måneder i et typisk forår tilstrækkeligt mange

nedbørshændelser til at give et fornuftigt bud på vandføringer nedstrøms i afløbssystemet.

Eksempel: Modelkalibrering af indsvinning til hydrologisk model – VCS

Det første eksempel på anvendelse af machine learning i vandsektoren kommer fra Odense. Vandcenter Syd (herefter VCS) kører hver midnat deres MIKE Urban model for hele oplandet med dagens målte nedbør, og sammenligner løbende det beregnede indløb til renseanlæggene med det målte. Ud fra resultaterne var det tydeligt at indsvinning af uvedkommende vand – dvs. indløb til renseanlægget uden for regnhændelserne – blev underbestemt betragteligt af MIKE Urban modellen. VCS har flere regnmålere, og derfor et godt udgangspunkt for at anvende machine learning til at bestemme korrelationen mellem nedbør og det manglende indløb i modellen. Som udgangspunkt anvendtes en standard afløbsmodel med et fordampningsled og en årstidsvariation der beskriver jordens evne til at holde på nedbøren. Dette netværk blev trænet med nedbør som input og vandføring som output. Med den optimerede indsvinnings beskrivelse blev MIKE Urban modellens gennemsnitlige fejl forbedret fra ca. 50% i vintermånederne til ca. 20% som det fremgår af figuren.

Eksempel: Parameterisering af ny model – Aarhus Vand

Ved Viby i Aarhus kan en meget stor del af indløbet til renseanlægget tilskrives indsvinning. Oplandet har en naturlig hældning ned mod renseanlægget, og dermed er der meget få pumpestationer og dermed også meget få målinger at gå ud fra til at adskille indsvinningen fra nedbør fra den der skyldes høj-grundvandsstand. Ideen med at anvende machine learning var at det kunne være muligt at bruge grundvands målinger fra borerer uden for oplandet – i dette tilfælde på den anden side af Aarhus under antagelse af at variationerne i grundvandsspejlet er homogene over et større område. I dette projekt gjorde machine learning det muligt at sammenholde grundvandsspejlinger og nedbørsmålinger uden for området korrelerede med indsvinning til Viby renseanlæg, og bestemme styrken af relationen. Med andre ord blev machine learning indirekte brugt til at bekræfte en hypotese om at variationer i grundvandsspejlet er homogene over et større område, og at en stor del af indsvinningen varierer netop med grundvandsspejlet. Konkret blev det simple neurale netværk til beskrivelse af indsvinning fra projektet i

Odense udvidet med et ekstra grundvandsled. Input til beregningen var nu tidsserier fra fem grundvandsmålere og fire regnmålere for 4 år. Træningen handler i dette tilfælde om at finde de koefficienter der bedst matcher indløbet baseret på input tidsserierne. Resultatet af den færdige model består af to dele: dels et samlet bedste estimat af indløbet der i middel forklarer 90% af signalet som en funktion af nedbørsmålere og grundvandsspejlinger langt væk, og dels forholdet mellem den grundvand og nedbørs bestemte indsvinning der ender på 50,5%. Det forhold stemmer godt overens med forventningerne ud fra erfaringerne i planafdelingen hos Aarhus Vand, men er endnu ikke blevet valideret direkte.

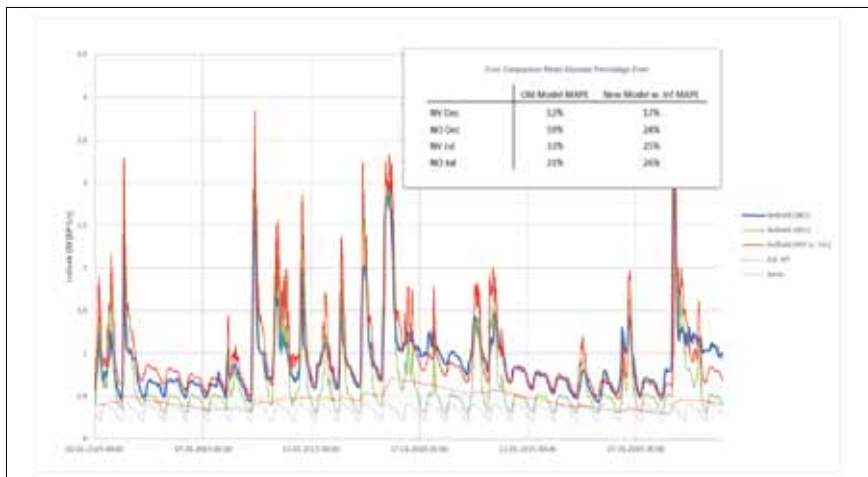
Eksempel: Udsigter for indløb til renseanlæg, med indbygget usikkerhedsestimat – BIOFOS

Forudsigelser af mængden af vand, der løber ind til renseanlæg 1-12 timer forud har stor værdi for planlægningen af vandrensning og service i det daglige og ikke mindst for udnyttelse af kapaciteten i ekstreme situationer. Der anvendes i dag mange forskellige modeller til at forudsige indløb afhængig af oplandets kompleksitet, og mulighederne for at ændre gennemløbet eller kapaciteten på renseanlægget. I København er afløbssystemet meget komplekst og forsinkelsen mellem nedbør og afløb til kloak er kompliceret at beregne og dermed tager det ofte for lang tid at regne systemet igennem med en dynamisk deterministisk model, så der er behov for andre metoder. For at teste potentialet for forudsigelser af indløb ved hjælp af machine learning, blev et neuralt netværk trænet på et års data fra fem regnmålere, tre vandføringsmålere, og vandniveauerne fra to forsinkelsesbassiner. Netværket blev trænet til at forudsige indløbet til tre forskellige renseanlæg, og efterfølgende kørt med realtidsdata hvert 5. minut. Modellen blev trænet til at forudsige indløbet 15 min, 1 time, 2 timer og 4 timer ud i fremtiden – hvilket i praksis svarer til 4 forskellige neurale netværk. Huller i data, både historisk og i realtid, er en generel udfordring i machine learning algoritmer. I dette projekt blev netværket derfor specifikt trænet til at håndtere disse huller og alligevel anvende de resterende inputs. Metoden hedder "dropout" og er en teknik der konceptuelt svarer til at træne en hel række af netværk samtidigt, hvor hver især udelader det input eller output der mangler for det pågældende tids-skridt. Det neurale netværk blev også udvidet med et indbygget usikkerhedsestimat som viser, i hvor høj grad den nuværende

nedbørssituation svarer til noget, netværket har set før. Usikkerhedsestimater er beregnet separat for de forskellige tidshorisoner (15 min, 1 time osv). Det fungerer på den måde, at netværket også trænes til at se hvor genkendeligt det nuværende mønster er og desuden får fejlen fra dens tidligere estimat tilbage som input.

Kvaliteten af indløbsudsigterne er i figuren vist som afvigelserne til det faktiske indløb. Det er ikke overraskende at jo længere udsigtslængde der anvendes jo større afvigelse er der, men det er der ikke så stor afvigelse på 15 minutter og 1 time hvilket indikerer at der er nogle underliggende usikkerheder i systemet som ikke er bestemt af længden på udsigten. Hvis udsigten skal forbedres er næste skridt altså at finde nogle tidsserier der kan være korrelerede med de processer der ikke er regn-afhængige, og det er her machine learning kan inspirere til at kigge i andre retninger end de traditionelle. På dette anlæg ved vi at en del af støjen hidrører fra et tilbageløb der er pumpe styret – måske kunne strømforbruget for anlægget derfor være et input der skulle prøves, eller det kunne være korreleret med et niveau inden på anlægget.

For fejlen i usikkerhedsestimater der vises på den anden graf er det specielt interessant at usikkerhedsestimater er dårligere for de mere ekstreme hændelser med større indløb per time. Det giver mening fordi der i datasæt-



Et eksempel der viser en validerings periode, dels det målte indløb (blå), den beregnede infiltration (stiplet rød), den afledte basis kurve, og MU med (rød) og uden (grøn) den beregnede infiltration.

tet er færre af de kraftige hændelser, og dermed er det med til at bekræfte at metoden fungerer. Den endelige bekræftelse kommer dog først når og hvis operatørerne begynder at tro på usikkerhedsestimater.

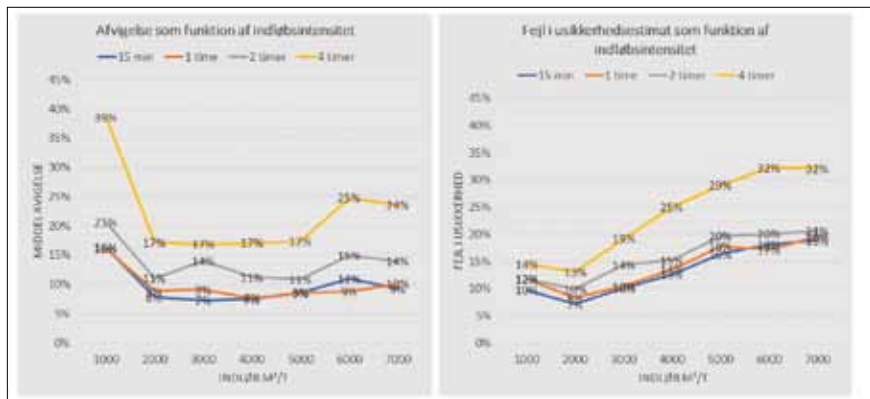
Næste skridt er at anvende radardata som input til netværket og dermed få både bedre rumlig opløsning i inputtet og samtidig få målinger udenfor oplandet og dermed forhåbentlig forbedre kvaliteten af netværket.

Den fremtidige anvendelse

Som illustreret i eksemplerne herover er machine learning lidt af et multi-værktøj. Det er

relativt let at gå til, er hurtigere at bruge end de fleste traditionelle modelleringsværktøjer og indbyder derfor til at arbejde iterativt og kontinuerligt forbedre resultaterne. Som med de fleste multi-vidunder-værktøjer der kan løse det hele på den halve tid skal der betales en pris – og i dette tilfælde er det dels at man skal acceptere at værktøjet ikke er gennemsigtigt – det er en sort boks, dels skal man mobilisere al sin erfaring og sunde skepsis netop fordi resultaterne ikke er garanteret fysisk korrekte og endelig at tid brugt på at kvalitets-sikre sine data ved brug af fysisk forståelse og korrelation er givet mindst ligeså godt ud som ved traditionel modellering. Machine learning er kommet for at blive – også i vandbranchen – nu skal vi fokusere på at finde de rigtige problemer hvor værktøjet passer ind.

PETER RASCH har med en baggrund i geofysik og numerisk modellering af klimaændringer igennem de sidste 15 år været involveret i projekt- og produktudvikling og ledelse. Fokus har været på data, numerisk modellering og realtidsløsninger til analyse og beslutningsstøtte, både på DHI (15 år) og DMI og siden 2016 som direktør for InforMetics. Kontaktes på pr@informetics.com.



Kvalitet i forudsigelse som funktion af hændelses intensitet til venstre, og den tilsvarende fejl i usikkerhedsestimater til højre.