

Viktiga attribut i modell för ledningsunderhåll

600 miljarder svenska kronor motsvarar två tredjedelar av den svenska statsbudgeten, och det är vad Sveriges totala ledningsnät är värderat till [1]. Ledningarna håller inte för evigt och underhåll och förnyelse måste därför ske på ett effektivt sätt. Det här arbetet undersökte därför vilka ledningsattribut som är viktigast i en ANN-modell för att utvärdera vattenledningsnätets underhållsbehov. Den initiala modellen beaktade 19 ledningsattribut, men studien visade att nästintill lika hög noggrannhet kunde erhållas med endast 10 attribut.

För att upprätthålla kvaliteten på Sveriges ledningsnät måste det underhållas och förnyas regelbundet. I och med ledningsnätets stora värde måste detta göras på ett effektivt sätt där nyttan av de resurser som investeras maximeras. Vattenledningsnät är trycksatta vilket försvårar utvärderingen av ledningars kondition. För att kunna underhålla och förnya vattenledningsnät effektivt måste i stället underhåll och förnyelse planeras utifrån tumregler, indirekta observationer och teoretiska samband [1]. Många tumregler baseras på ålder, men ålder har konstaterats vara ett trubbigt mått. Dessutom ger flera metoder endast områdesvisa uppskattningar av underhållsbehovet. För att komplettera dessa metoder kan en ANN-modell för vattenledningsnätet användas; en sådan modell inkorporerar en mängd andra ledningsattribut utöver ålder för att utvärdera risken för läckage, och ger dessutom ett värde för läckagerisken för respektive ledning som analyseras.

ANN är en förkortning av *artificiellt neuronnät*, vilket är en metod för att identifiera samband i indata för att skapa meningsfull utdata. Ett användningsområde för ANN-modeller är bildigenkänning, där modellen kan lära sig att känna igen exempelvis träd. Detta fungerar genom att modellen får se jättemånga bilder på träd, och bilder som inte är träd, om och om igen. Fundamentalt är att modellen i detta skede har tillgång till ett facit – varje gång modellen får se en bild, får den också veta vad bilden föreställer. Modellen lär sig då till slut att känna igen hur ett träd ser ut, och kan användas för att identifiera bilder utan facit som träd eller inte träd.

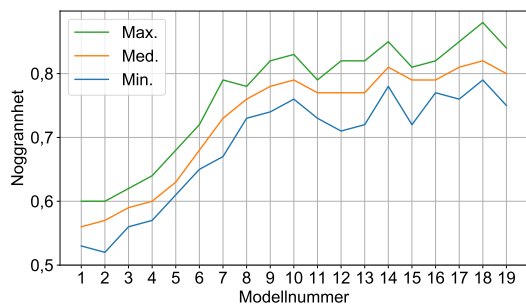
En ANN-modell för ett vattenledningsnät fungerar på exakt samma sätt, men i stället för att se bilder på träd kan man tänka sig att modellen får se bilder på vattenledningar. Dessa ”bilder” är dock inga egentliga bilder, utan en bild är i stället en lista av ledningsattribut. Med ledningsattribut åsyftas faktorer som påverkar en vattenledning, såsom material, omkringliggande jord och ålder. Som facit används läcka eller inte läcka. Den ANN-modell som använts i den här studien är baserad på en modell utvecklad av Rehn och Giertz [2] på Stockholms vatten och avfall.

En inneboende egenskap för ANN-modeller är att ju mer data de har (både kvalitet och kvantitet), desto bättre kan de prestera. När det gäller vattenledningar, och även ledningar i allmänhet, råder det dock ofta brist på data då kunskapen om ledningarna i marken ofta är låg där osäkerhet råder kring attribut: hur gamla är ledningarna? vilket material är de av? vilken jordtyp ligger de i? och så vidare. Dessutom finns det endast ett begränsat antal ledningar, vilket kan jämföras med trädbilder där det nästan går att skapa hur många bilder som helst. Det räcker inte heller med att ha mycket information om ledningar, utan det behövs även ledningar med observerade läckor vilket minskar datamängden ytterligare.

Det här arbetet fokuserade därför på att identifiera vilka attribut som är viktigast för att få en ANN-modell för ett vattennät att prestera väl. Dels underlättar det genom att mindre data behöver samlas in och fokus kan då läggas på att kvalitetssäkra de nödvändiga attributen, dels kan det medföra att fler ledningsobjekt har tillräckligt mycket information för att inkorporeras i modellen. En begränsning med studien är att endast attribut tillgängliga i Umeå kunde utvärderas, men förhoppningen är att resultatet ändå ska vara applicerbart på andra kommuner.

För att utvärdera viktiga attribut användes främst attributurvalsalgoritmen ReliefF, som tilldelar varje attribut en vikt utifrån hur viktiga det är. ReliefF-algoritmen involverar dock inte ANN-modellen vid utvärderingen och därför måste utfallet från utvärderingen testas på modellen. Figur 1 visar hur ANN-modellens prestation förändrades då fler och fler attribut lades till i den ordning ReliefF-algoritmen indikerade. Modellens prestation mäts bland annat i

noggrannhet – ifall noggrannheten är ett klassar modellen all indata korrekt. Med tio attribut uppnåddes en noggrannhet på 0,79, jämfört med 0,80 med 19 attribut – det kan alltså vara tillräckligt att samla in de tio första attributen.



Figur 1 – Hur modellen presterar då attribut läggs till i den ordning ReliefF föreslår.

Vilka de tio viktigaste attributen enligt ReliefF-algoritmen var redovisas i tabell 1. Många av dessa attribut har direkt påverkan på ledningen, exempelvis Jordtyp, Material och Serviskoppling. Vissa är snarare indirekta: risken för läcka ökar inte för att en ledning är situerad i en speciell stadsdel, i stället är det en eller flera andra faktorer som påverkar ledningen och som fångas in av tillhörande stadsdel. I en stadsdel kan exempelvis ledningarna ha ungefär samma ålder, vara av samma material, lagda med samma metodik och vara belägna i samma jordtyp. De indirekta attributen har flera nackdelar, varav en stor är att det inte framgår vad som faktiskt påverkar risken för läckage.

Tabell 1 – De tio högst rangordnade attributen enligt ReliefF-algoritmen i fallande ordning.

Attribut
Jordtyp
Tryckzon
Markanvändning
Befolkningsförändring
Stadsdel
Trafiklast
Innerdiameter
Material
Korsande fjärrvärme
Serviskoppling

Ett annat mål med studien var att utvärdera möjligheten att applicera ANN-modellen på en mindre kommun, och då är diskussionen om direkta och indirekta attribut viktiga. Genom en diskussion med

Petter Walan på Vakin¹ identifierades ledningsattribut som det bör vara möjligt att erhålla för Vindelns kommun, vilket är en liten kommun sex mil utanför Umeå. Utöver att attributen ska vara tillgängliga i Vindeln bör attributen även vara generella. Med generella attribut avses attribut som är desamma mellan orter, det vill säga exempelvis Jordtyp, men inte Stadsdel. Jordtyperna kan skilja sig åt, men friktionsjord är friktionsjord oavsett ort. I de flesta fall motsvarar alltså de generella attributen direkta attribut. Att generella attribut används möjliggör tränandet av modellen på en ort med mycket data, men applicerandet av den på en annan ort. Detta förenklar användandet av modellen ytterligare eftersom modellen kan appliceras på en kommun som annars hade haft för lite data för att träna modellen. Detta undersöktes inte i praktiken på grund av tidsramen för arbetet, utan i stället gjordes en teoretisk utvärdering. Ledningsdata för Vindelns kommun är i vissa avseenden begränsat och därför bör många av de attribut som är tillgängliga för Vindeln även vara tillgängliga i andra kommuner. De attribut som ansågs vara enkla att erhålla och som dessutom är generella redovisas i tabell 2.

Tabell 2 – Attribut som bör gå att få fram för Vindeln.

Attribut tillgängliga i Vindeln
Anläggningsår
Bergtyp
Dimensionsförändring
Innerdiameter
Jordtyp
Markanvändning
Material
Maxhöjd
Minimihöjd
Närliggande järnväg
Serviskoppling
Trafiklast
Ventilkoppling
Ålder

Med dessa attribut presterade modellen sämre än med de tio viktigaste attributen enligt ReliefF, men skillnaden var inte alltför stor: 0,75 med attributen tillgängliga för Vindeln, och 0,79 med de tio främsta enligt ReliefF. Det kan alltså vara möjligt att applicera en ANN-modell på en ort med begränsad datatillgång.

Hur läckfrekvensen varierade med olika attribut undersöktes också. Läckfrekvensen baseras på prediktioner

¹Petter Walan, Utredningsingenjör Vakin

med en tröskel på 0,75 för läcka. Vad detta innebär behandlas senare i den här artikeln. I åtminstone tre metoder från Svenskt vatten används ålder för att bedöma vilka ledningar som behöver förnyas [1], men i figur 2 visas att för Umeå är det inte de stadsdelar med högst medelålder på ledningsnätet som har högst läckfrekvens. Siffrorna ovanför staplarna är ledningsnätets medelålder för respektive stadsdel. Det är viktigt att poängtera att detta baseras på resultatet från modellen, och inte från tidigare identifierade läckor.

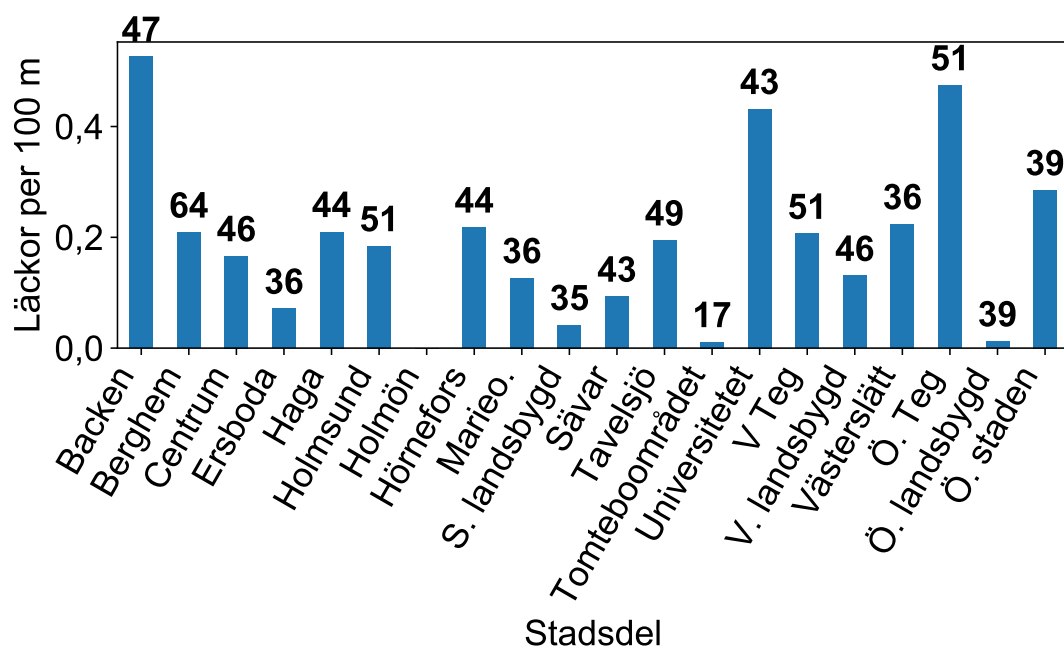
En intressant upptäckt var att läckfrekvensen var högre för ledningar utan anslutande serviser, se figur 3. Detta är märkligt eftersom intuitivt borde varje anslutning medföra en svag punkt. För att säkerställa att det inte var ett fel i modellen som ledde till resultatet undersöktes också läckfrekvensen för tidigare identifierade läckor. Läckfrekvensen är betydligt lägre för historiska läckor, men samma förhållande mellan ej serviskoppling och serviskoppling råder. Varför detta samband uppvisas måste undersökas noggrannare. Anledningen skulle exempelvis kunna vara för att läckor vid kopplingen hänförs till servisen och inte huvudledningen.

En annan spännande upptäckt var att ledningar som korsas av fjärrvärmeledningar har förhöjd läckfrekvens (figur 4). Till detta finns flera förklaringar. Marken runt fjärrvärmeledningar är varmare och på vintern kan tjälen vara mindre utbredd runt dessa. Då kommer vattenledningen på den punkt där den korsas av fjärrvärmeledningen vara mindre påverkad av tjäle än resterande ledning. Eftersom marken ”sväller” av tjäle bör detta inducera spänning i ledningen. En annan faktor kan vara att vattenledningen påverkas vid anläggningen av fjärrvärmeledningen. Om fjärrvärmeledningen lagts efter vattenledningen kan det ha schaktats, borrats och så vidare i anslutningen vilket kan ha påverkat ledningsbädd, sättningar och så vidare.

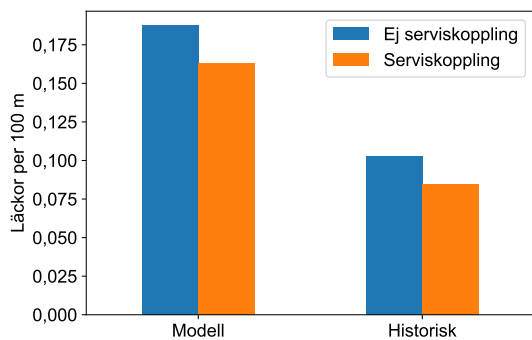
Det prestationsmått som angetts ovan är noggrannhet för valideringsdata. Valideringsdata inom ANN-modellering är det dataset som används vid tränandet av modellen för att utvärdera modellens prestation. Eftersom valideringsdata snarast motsvarar träningsdata för vanlig modellering är det vanskligt att utvärdera modellens prestation utifrån valideringsdata. Därför används något som kallas testdata, men i den här studien förbisågs testdata för att tillgänglig data i Umeå var begränsad och kvantitet är viktigt inom ANN-modellering.

Det som tog absolut mest tid i studien var att samla in data till modellen – att köra ANN-modellen gick relativt fort. Anledningen till att insamlandet av data tog så lång tid var för att även om data fanns tillgängliga, så var de inte samlade på en plats. Dessutom kunde kvaliteten på inrapporterad data var bristfällig varpå arbete krävdes för att förbättra dessa. Alla dessa brister kan leda till en försämrad modell. Det kan vara något så uppenbart som att det står att en ledning haft en läcka fastän den inte haft det, men det kan också vara att rörmaterial inte angetts. Ytterligare ett problem är att det mestadels är okänt vilken kondition en ledning har, eller till och med ifall en ledning läcker eller inte; detta är varför ANN-modellen behövs, men det försvårar också träningen av ANN-modellen eftersom vissa ledningar klassades om Ej läcka kan ha en läcka i praktiken.

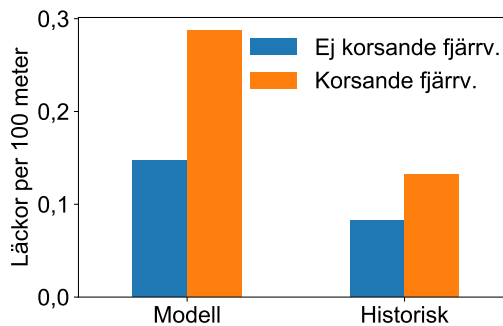
Detta, tillsammans med att testdata exkluderades, medförde att modellens prestation behövde utvärderas på något annat sätt. Detta gjordes på olika sätt, och en variant utgick från relationen mellan predikterade läckor och historiska läckor. Denna utvärdering visas i figur 5. När modellen tränades klassades alla prediktioner över 0,5 som läcka, och alla under 0,5 som ej läcka. Modellen kan tolkas som att ju närmare ett en prediktion är, desto säkrare är modellen på att ledningen har en läcka. Därför är det intressant att inte endast undersöka hur väl den klassar historiska läckor, utan även hur detta varierar med olika tröskelvärden för läcka. De två linjerna har samma täljare: antalet predikterade läckor som också är historiska läckor, men nämnaren varierar. Den gröna linjen har antalet predikterade läckor som nämnare, medan den orange linjen har det totala antalet historiska läckor som nämnare. Att den orange linjen ständigt minskar är inte konstigt: om modellen klassar alla ledningar som läcka kommer andelen korrekt klassade ledningar vara ett eftersom måttet inte tar hänsyn till hur många läckor modellen predikterar. Måttet är ändå intressant eftersom det direkt visar andelen historiska läckor modellen identifierar. Den gröna linjen beaktar i stället hur många läckor modellen predikterar och kan därför vara en bättre mått på modellens prestation. Samtidigt skulle modellen vara ”perfekt” om den klassar en ledning som läcka, och den ledningen också haft en historisk läcka, vilket också är missvisande. Modelles prestation utifrån figur 5 är svårtolkad. Med ett tröskelvärde på 0,95 för läcka är ca 50 procent av de predikterade läckorna historiska läckor. Förutsatt att det finns ledningar med oidentifierade läckor kan



Figur 2 – Läckfrekvens för Umeås olika stadsdelar. Siffrorna ovanför staplarna motsvarar ledningsnätets medelålder för de olika stadsdelarna.

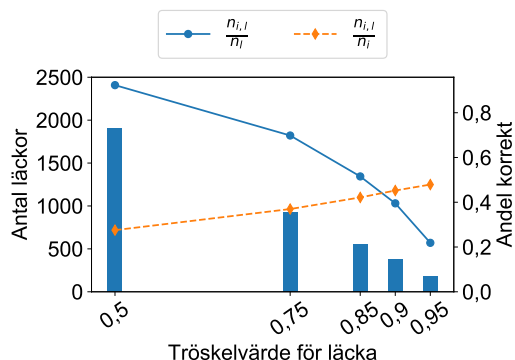


Figur 3 – Läckfrekvens för ledningar som har, och inte har, anslutande serviser. Frekvensen visas både för läckor enligt modellen, och tidigare identifierade läckor.

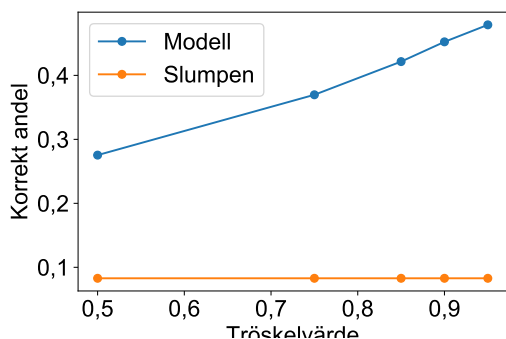


Figur 4 – Läckfrekvens för ledningar som har, och inte har, korsande fjärrvärme. Frekvensen visas både för läckor enligt modellen, och tidigare identifierade läckor.

fler än 50 procent av ledningarna var korrekt klassade, men det går inte att säga hur bra modellen presterar. Figur 6 visar hur väl modellen identifierade läckor jämfört med slumpen. Det här måttet är betydligt tydligare – modellen presterar mycket bättre än slumpen. Noteras bör att prediktionerna för den enkla och generella modellen inte har utvärderats på detta sätt, utan där har endast noggrannheten utvärderats.



Figur 5 – Olika mått på modellens prestation. Staplarna motsvarar det antalet predikterade läckor beroende på tröskelvärdet för läcka.



Figur 6 – Jämförelse mellan andelen historiska läckor utifrån predikterade läckor som modellen respektive slumpen identifierar.

Denna studie har endast skrapat på ytan av hur ANN-modeller kan användas för ledningsnätet. I den här studien undersöktes det teoretiskt ifall en ANN-modell kan användas på en kommun med begränsat dataunderlag, men det skulle behöva utredas i praktiken också. Dessutom skulle resultatet av ANN-modellen behöva utvärderas noggrant – om några år det möjligt att utvärdera hur väl modellen predikterade ledningar med läckor. I detta nu (våren 2020) arbetas det i projektet *Ordning i Rörann* med att ta fram en modell som kan anpassas på hela Sverige. Det ökar mängden

tillgängliga data för att träna på modellen på vilket är viktigt för att uppnå en välpresterande modell och är därför väldigt intressant.

För att sammanfatta det hela kan det återigen nämnas att det var möjligt att uppnå jämförbar noggrannhet mellan en modell med 10 attribut och en modell med 19 attribut. Med attribut som de flesta kommuner bör kunna få fram eller uppskatta uppnåddes en noggrannhet på 0,75. Detta är lägre, men inte mycket lägre än noggrannheten med 10 och 19 attribut (0,79 respektive 0,80).

Referenser

- [1] Annika Malm m. fl. *Handbok i förnyelseplanering av VA-ledningar*. Tekn. rapport 2011-12. Svenskt Vatten, 2011.
- [2] David Rehn och Tommy Giertz. *En AI-modell för vattenledningsnätet*. Tekn. rapport. Stockholm Vatten och Avfall, 11 jan. 2019.