

Pressmeddelande 2015-06-18

Konsten att skapa algoritmer för säkra förutsägelser

Numera samlas det in sådana mängder av data att det ofta är omöjligt att göra manuella analyser. Det kan till och med vara svårt att få datorer att analysera all data och exempelvis kunna förutse skeenden. Tuwe Löfström, doktorand i informatik vid Högskolan i Borås och Stockholms Universitet, har nu disputerat på hur algoritmer kan användas för säkrare analyser.

När datamängderna blir för stora och komplexa för att analyseras av människor behöver datorer läras upp för att bistå i arbetet med analyserna. Algoritmer är ett slags instruktioner för datoriserade beräkningar. Tuwe Löfström har utvecklat befintliga algoritmer och kartlagt hur de bäst kan användas för att med större säkerhet kunna förutse exempelvis risk för biverkningar av läkemedel. När han har undersökt och utvecklat algoritmerna har han utgått från datamängder där det redan finns ett svar, med en historik som visar hur utfallet blev. Ett exempel på en datamängd som han kommer arbeta med i den närmaste framtiden är just data om hur många som har drabbats av biverkningar av ett visst läkemedel. På sådana datamängder har han testat olika varianter och kombinationer av algoritmer, jämfört sina resultat med det faktiska resultatet för att identifiera vilka algoritmer som fungerar bäst.

– Det kallas att träna modeller när man använder en algoritm för att söka igenom och identifiera vilka samband som kännetecknar datamängderna. Syftet är att de modeller som algoritmerna skapar ska kunna förutse ny data från samma domän, säger Tuwe Löfström. Själva forskningsområdet kallas maskininlärning.

– Man kan träna flera olika modeller och kombinera dem för att till exempel kunna få hjälp att identifiera om en patient har drabbats av läkemedelsbiverkningar, säger han. På så sätt kan man skapa en modell som kallas ensemble, som tillsammans blir bättre än de enskilda modellerna som den är uppbyggd av.

**Viktigt att kunna dra nytta av stora datamaterial**

Ett användningsområde för algoritmer som kan göra säkra förutsägelser är att beräkna vilken risk en speciell patient har att drabbas av biverkningar av ett läkemedel. Även om det är känt att sex procent av en viss patientgrupp drabbas, kan det vara viktigt att exakt kunna beräkna risken för en specifik patient.

– Då har det visat sig vara bra att använda en ensemble, bestående av flera enklare modeller. På så vis minskar antalet slumpmässiga fel som modellen gör. Men det gäller att skapa ensembler på bästa sätt.

I sin forskning har Tuwe Löfström kört en stor mängd experiment i datorn, tränat och testkört olika ensemblemodeller gång på gång, för att slutligen kunna jämföra modellerna och dra slutsatser om vad som fungerar bäst.

Användningsområdena för dessa kunskaper är redan nu enorma, och de fortsätter stadigt att växa. Han förväntar sig en kraftig ökning av efterfrågan på bättre förutsägelser.

– Vi behöver kunna dra nytta av de stora datamaterial som genereras, och kunnighet i det kommer att bli en nyckelkompetens i framtiden. Maskininlärning kan användas inom en väldigt stor mängd olika tillämpningsområden, som till exempel för att i förväg förutse biverkningar av läkemedel. Men det kan också behövas när en butik gör en annonskampanj och vill veta hur stora mängder av en specifik vara de bör ta hem för att ha lagom många. Eller för att göra prognoser om bränsleförbrukningen i ett fordon.

**Bäst att träna modellerna**

Att Tuwe Löfström fastnade för maskininlärning beror på att han gillar problemlösning och har ett intresse för matematik. Här får han använda sin kreativitet för att söka lösningar på ständigt nya problem.

Kreativiteten har kommit till användning i avhandlingsarbetet, och nu har han hittat lösningar på hur man kan skapa en så bra ensemble som möjligt. Två metoder har studerats: att träna många modeller och sedan utifrån olika kriterier välja ut vilka som ska ingå i ensemblen, eller att träna så många modeller som möjligt och kombinera allihop.

- Min forskning visar tydligt att metoden där alla modeller används i ensemblen är klart bättre, säger han. Det är alltså bättre att lägga energin på att träna många modeller som alla får ingå i ensemblen än på att försöka räkna ut vilka modeller som är bäst och enbart låta dem vara med i ensemblen.

Text och foto: Lena M Fredriksson

Fakta

Avhandlingen On Effectively Creating Ensembles of Classifiers: Studies on Creation Strategies, Diversity and Predicting with Confidence

Av doktoranden Tuwe Löfström, universitetsadjunkt inom datavetenskap vid Högskolan i Borås

Disputation vid Stockholms Universitet den 11 juni 2015