Inbrottsdetektering med hjälp av maskininlärning

Linus Lindgren & Mikael Lindholm

Hårdvarans prestanda har ökat avsevärt det senaste decenniumet och det har gjort det möjligt att använda sig av så kallade maskinlärningsmetoder för att lösa komplicerade problem. Maskininlärning syftar på att ett datorprogram på något sätt lär sig göra något, antigen genom att analysera data eller upprepa en procedur och granska resultatet. I vårt fall använder vi oss av maskininlärning för att hitta ett sätt att särskilja på två olika typer av händelser, inbrott och ofarlig händselse. Vi har arbetat med en porttelefon som innehåller en accelerometer som då är kapabel att detektera acceleration i tre riktningar. Man kan bryta upp porttelefonens yttre hölje och detta kan vi detektera, när man upptäcker ett intrång kan man förslagsvis ringa polisen.

Problemet är att accelerometern också känner av att man exempelvis stänger dörren och då vill man inte ringa polisen. Problemet är att särskilja dessa händelser. För att göra detta så har vi samlat in en stor mängd data ifrån olika dörrar som både täcker dörrstängningar och inbrottsförsök. Maskinlärningsmetoderna vi har använt oss av använder då olika objektiva mått för att försöka hitta typiska beteenden för de olika händelserna. Exempel på mått vi använder oss av är en signals maximum, minimum och dens frekvensuppdelning. Målet är då att använda det som maskinlärningsmetoden lär sig för då avgöra om en ny signal är något ofarligt eller att någon försöker bryta upp det yttre höljet. Vi har då samlat 2119 exempel-data ifrån 13 olika dörrar varav 1401 är ofarliga och 718 är intrång. Vår metod lyckas då gissa rätt i 99.8% av fallen.

Vad innebär det då att en maskininlärningsmetod lär sig av data? Vi har använt oss av metoden linjär support vector machine som är enkel att förstå. Figur 1 visar hur 2 enkla mått används för att fördela den insamlade datan.

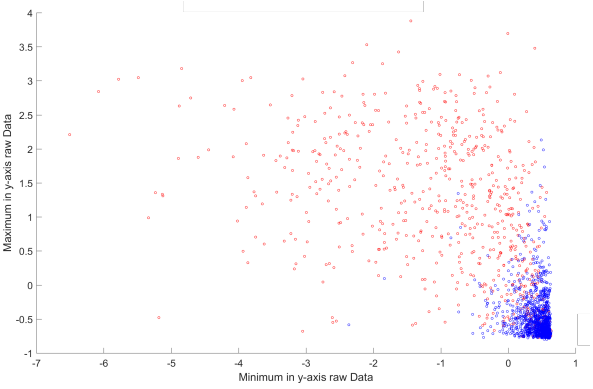

Figure 1: Den insamlade datan ritas ut i ett 2-dimensionellt plan som spänns upp av maximum och minimum i axeln y. De röda punkterna är data ifrån intrång och de blåa punkterna är data ifrån ikke-intrång. Idén är då att försöka särskilja på de röda och blå punkterna.

Ett enkelt sätt att dela upp punkterna är då att dra ett rakt sträck i planet. Om en ny punkt hamnar på ena sidan om sträcket gissar man att det är en typ av händelse och vice versa. För att detta ska gå så bra som möjligt vill man då så klart dra sträcket på ett fiffigt sätt. Detta är då det som linjär support vector machine gör. I praktiken använder man ofta mer än två mått men konceptet är det samma även om det är svårare att föreställa sig.