

Populärvetenskaplig sammanfattning

När mjukvara uppdateras och installeras på Axis kameror är det viktigt att hålla koll på om den nya mjukvaran orsakat förändringar i kamerornas prestanda. Förändringarna kan sedan tolkas av en människa, som kan säga om det är en förbättring eller försämring. Genom att ständigt hålla uppsikt över olika metriker (olika prestandamått) hos kamerorna finns möjligheter att upptäcka förändringar som uppstår på grund av mjukvaru-uppdateringar.

I nuläget görs denna övervakning manuellt men vi har fått i uppdrag att se hur det kan automatiseras.

Vi har använt oss av både maskininlärningsmetoder och av rent statistiska metoder för att hitta dessa förändringar. Vi har testat många metoder, bl.a. binär klassificering, oövervakad maskininläring, samt *CUSUM* och enkla stegdetektionsalgoritmer. De algoritmer som visade sig prestera bäst var *Theta Forecasting* som är en typ av oövervakad maskininläring och *CUSUM*, som är en statistisk metod.

För att hitta dessa förändringar i beteende hos kamerorna behövde vi även fastställa "Vad är en förändring?". I början av detta projekt fick vi en samlad tabell med kända förändringar som samlats in under de senaste 6 åren. Tyvärr visade det sig att tabellen inte var så användningsbar för de metoder vi skulle prova. Det visade sig också att tabellen inte riktigt stämde med hur metrikdiratån såg ut. Vi hittade förändringar som inte var med i tabellen och ibland innehöll tabellen markeringar för förändringar som vi inte kunde se i datan. Att använda detta som facit hade lärt algoritmen att leta efter helt fel grejer.

Predikament vi möttes av var något vi inte var beredda på, men har också fått oss att inse att detta kan vara ett vanligt problem. Utanför kontrollerade labbmiljöer och akademiska experiment är optimala förutsättningar, vad det gäller datakvalité, mer sällsynta. Trots denna motgång var vi bestämda att problemet behöver lösas, så vi hittade en väg runt problemet.

Vi spånade fram två idéer fram som skulle komplettera varandras svagheter. Första tanken var att generera data som har liknande egenskaper som datan vi har tillgång till, men även generera en tabell av förändringar. Alltså utöva en form av *data generering*. Andra tanken var att, för hand, gå igenom några tidsserier med data och markera ut var förändringar skett, vilket är det traditionella sättet att få träningsdata till maskininläring. Den genererade datans svaghet är att den kanske inte säger något om den *verkliga* användbarheten av vår modell, men det löses genom att de sista testet sker på den *verkliga* datan. Den handmarkerade datans svaghet är att det är kostsamt att, för hand, gå igenom all tillgänglig data. Detta gör att mängden *verklig* data är begränsad, vilket löses av mängden genererad datan, som är stor.

Våra metoder fick sen optimeras mot den genererade datan och sen testas mot våra handgjorda markeringar för att se hur de presterade på riktig data.

Resultaten vi fått under projektet talar klarspråk; det går helt klart att automatisera upptäckandet av förändringar i data och vi rekommenderar varmt *Theta Forecasting* och *CUSUM*!