

EXAMENSARBETE Clustering and Classification of Test Failures Using Machine Learning**STUDENT** Andy Truong & Daniel Hellström**HANDLEDARE** Erik Larsson (LTH), Lars Viklund (Axis Communications AB)**EXAMINATOR** Flavius Gruian (LTH)

Kan en dator lära sig att hitta orsaken till varför tester går fel?

POPULÄRVETENSKAPLIG SAMMANFATTNING **Andy Truong & Daniel Hellström**

Att bestämma varför tester går fel görs i nuläget genom att gå igenom miljontals rader av kryptiska loggfiler. Detta arbete undersöker hur maskininlärning kan tillämpas för att automatisera denna process.

Att noggrant undersöka ritningar för att hitta fel i dem är en viktig process inom design av hårdvarukomponenter och kallas för verifiering. De viktiga frågorna i verifiering är att identifiera att något är fel och vad som är orsaken till felet. Målet med verifiering är att identifiera fel i ritningarna så tidigt som möjligt för att slippa onödiga kostnader senare.

I dagsläget görs verifiering genom att komponenters ritningar simuleras och testas för att se hur komponenterna kommer bete sig i verkligheten. Om någonting går fel i ett test så är det viktigt att snabbt komma fram till vad som gick fel och hur det kan åtgärdas. Detta görs genom att manuellt läsa igenom loggfiler som skapas under testexekveringen och leta efter små detaljer som kan ge ledtrådar om vad som gick fel. Att manuellt läsa igenom miljontals rader av information är tidskrävande och kan lätt bli långtråkigt. Det finns därför ett behov av att kunna automatisera denna process.

Lyckligtvis är datorer väldigt bra på att snabbt hantera stora mängder data. Faktum är att en teknik som kallas maskininlärning faktiskt blir bättre ju mer data som finns tillgänglig. Grundkonceptet inom maskininlärning är att låta datorer titta på ett mycket stort antal par av indata och utdata för att lära sig vilka samband som

finns. Sambanden kan sedan användas för att koppla ihop ny indata med rätt utdata. I vårt arbete har vi undersökt hur maskininlärning kan användas för att identifiera orsaken till varför tester går fel och för att gruppera liknande tester. Automatisering av denna process leder till att mindre ingenjörstid behöver läggas på arbetet, vilket i sin tur leder till lägre kostnader. Dessutom gör det verifieringsprocessen snabbare vilket leder till att komponenter kan börja produceras tidigare.

Vårt resultat visar att en dator kan lära sig att hitta orsaken till varför tester går fel. Med hjälp av maskininlärning lär sig datorn mönster i loggfiler och kan därefter i nio fall av tio identifiera orsaken till varför ett test går fel. Däremot har en dator problem med att kunna gruppera och hitta likheter mellan testerna.

Vårt arbete visar hur en dator hanterar information i loggfiler för att kunna hitta orsaken till varför tester går fel och kunna gruppera liknande tester. En metod för att kunna omvandla vanlig text till ett språk en dator förstår presenteras. Denna omvandling används sedan i en jämförelse av maskininlärningsalgoritmer. Vårt arbete visar även att resultatet beror på algoritmen som används för identifiering av orsaken till varför tester går fel. Resultatet beror också på hur mycket och vilken information som omvandlas.