Anomalidetektion hos molntjänster med hjälp av logmeddelanden och maskininlärning.

Jacob Gummesson Atroshi och Christian Le

May 4, 2021

I vårt arbete har vi visat att det går att skapa ett automatiskt system för att detektera fel hos datacenter som driver olika molntjänster. Detta görs genom att analysera logmeddelanden som ständigt genereras. Genom användning av sådant system kan man på ett mycket mer effektivt och sofistikerat sätt hitta avvikelser i jämförelse med manuella sökningar. Särskilt hos dagens enorma datacenter har detta blivit väldigt viktigt.

Idag används många tjänster som drivs i molnet. Med molnet kan en normal användare på ett snabbt och smidigt sätt få tillgång till stora filutrymmen, processorkraft med mera. Molnet används främst för att köra tjänster som vanligtvis körs på en lokal dator men idag drivs av stora datacenter. De består av många datorer som tillsammans fördelar det jobb som krävs för att hålla igång molnet.

Ifall det blir något fel i dessa datacenter kan det leda till stora problem för många användare. För att förhindra och hitta när avvikelser sker använder utvecklarna vad som kallas för logmeddelanden. För ett datacenter kan det generaras miljontals av loggar varje dag som beskriver vad systemet gör och systemets tillstånd. Antalet logmeddelanden ökar med storleken av datacenterna. Idag finns det massiva datacenter som är så stora att en manuell felsökning i logmeddelandena inte är effektivt. Lösningen är att skapa en AI som använder maskininlärning och automatiskt larmar ifall det uppstår något problem.

Maskininlärning har på senare tid utvecklats väldigt snabbt och används för att hitta mönster till exempel i bilder när man gör bildigenkänning. I vårt fall använder vi det för att identifiera strukturen bakom de logmeddelanden som datacenterna ger oss. På sådant sätt kan vi detektera när det finns logmeddelanden som avviker från denna struktur.

Vi skapade flera AI-modeller som använder maskininlärning. Dessa testades och utvärderades på olika feltyper som kan uppstå i ett datacenter, exempelvis att nätverket blir långsamt. AI-modellerna följer flera olika principer. En metod bygger på att vi samlar logmeddelanden som genereras under en viss tidsperiod och sedan jämför alla dessa grupper. Ifall en grupp är betydligt annorlunda än de andra, till exempel att ett visst logmeddelande förekommer mycket oftare, är det en stor chans att det finns ett fel under den tidsperiod gruppen kommer ifrån.

Vi undersökte även en metod som använder ett neuralt nätverk för att förutse vad nästa logmeddelande kommer att vara. Denna slags AI gör detta genom att studera vad som har hänt tidigare, och utifrån det beräknar sannolikheten att ett visst logmeddelande kommer att uppstå. Till sist tittade vi på en metod som endast använder siffrorna i loggarna och försöker detektera ovanliga värden. Dessa siffror kan exempelvis vara hur lång tid det tog för en användare att logga in vid ett specifikt ögonblick. Flera av de olika modellerna fungerade väl, men har olika användningsområden och bör användas tillsammans.

Ett flertal av felen kan hittas med nästan inga falska alarm. I snitt detekterades felen i 60% av fallen. Dessutom visade vi hur effektiviteten påverkas av olika villkor, exempelvis hur lång tid vi tillåter ett fel att vara aktivt innan AI modellerna detekterar det. Detta gav möjlighet till att göra ett övervägande mellan villkor som påverkar resultatet. Resultaten kan hjälpa eventuella molnleverantörer att identifiera vad man måste tänka på när man skapar liknande system för att detektera fel.

Det främst unika i denna studie var användningen av ett virtuellt datacenter där vi kunde styra och skapa fel exempelvis hur servarna kommunicerade med varandra. Genom detta tillvägagångsett kunde vi testa hur bra våra AI fungerade på varje typ av fel vilket inte hade varit möjligt utan ett virtuellt datacenter. Därför är det naturligt för kommande studier att använda liknande system fast på riktiga datacenter istället.