

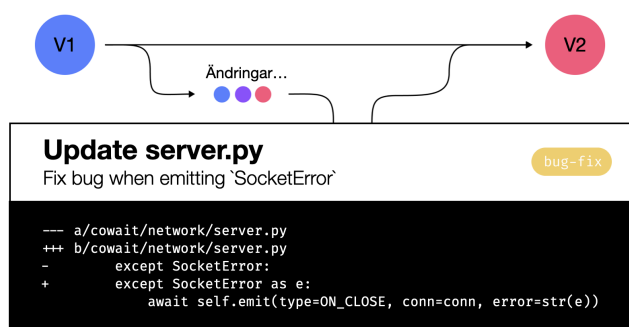
EXAMENSARBETE Classification of Pull Requests using Transformers**STUDENTER** Oscar Fridh, Szymon Stypa**HANDLEDARE** Pierre Nugues (LTH)**EXAMINATOR** Jacek Malec (LTH)

Klassificering av kodändringar som stöd för utvecklare

POPULÄRVETENSKAPLIG SAMMANFATTNING Oscar Fridh, Szymon Stypa

I större mjukvaruprojekt kategoriseras och beskrivs varje kodändring för att sedan enkelt kunna granskas av andra medarbetare. Detta arbete undersöker hur väl *Transformer*-arkitekturen presterar vid klassificering av sådana ändringar, och är ett försök att förbättra tidigare modeller med utökad data bestående av både text och kod.

Mjukvaruprojekt med många utvecklare använder ofta ett arbetssätt där ny funktionalitet utvecklas i separata miljöer för att eliminera konflikter mellan versioner. När en ändring är klar, kan en begäran om att slå ihop dess miljö med centrala projekversionen göras. Denna begäran kallas för *pull request*, och måste godkännas av andra medarbetare för att ändringen ska sammanfogas.



För att underlätta granskningen är det nödvändigt att beskriva alla begäranden, samt kategorisera dem med klasser som tydliggör ändringen. I detta arbete har vi utforskat hur väl det går att automatisera kategoriseringen med hjälp av modeller baserade på *Transformer*-arkitekturen. Målet är att främja nya verktyg som hjälper utvecklare att sköta administrativa uppgifter, och låter dem fokusera på att skriva kod.

I arbetet behandlar vi flera olika förtränade nätverk, och granskar hur de reagerar på olika typer av indata. Genom att använda olika sammanställningar av komponenter från *pull requests* som prediktorer (eng. *features*), studerar vi hur nätverkens förmåga att klassificera påverkas.

Vi demonstrerar att *Transformer*-arkitekturen överlag lämpar sig väl för uppgiften. Slutligen bygger vi en sammansatt modell som uppnår avsevärt bättre resultat än tidigare forskning på vår egeninsamlade data. Samtidigt finns det mycket rum för förbättringar, och framförallt ett behov av mer kvalitativt annoterad data.

Ett oväntat och intressant resultat är att modeller förtränade på vanlig text presterade nästan lika bra som modeller förtränade på kod vid klassificering av råa kodändringar. Detta tyder på att det finns ett behov av mer detaljerad forskning kring förträning av modeller, med fokus på hur träningsdata relaterar till modellernas inlärd förståelse.

Vi hoppas att vårt arbete kan agera byggsten för vidare forskning i domänen där kod och språk möter varandra. Automatiserande verktyg för att skriva och underhålla dokumentation är ett exempel där vi tror att arbetet kan komma till nytta.