

Evolvering av textklassificerare med hjälp av genetisk programmering

POPULÄRVETENSKAPLIG SAMMANFATTNING AV EMA EMINAGIC OCH HOANG NGO

Genetisk programmering (GP) är ett område inom AI som är inspirerat av den biologiska evolutionen. I detta arbete använder vi genetisk programmering för att evolvera fram textklassificerare. Denna metod ger många fördelar som dagens teknik saknar.

Textklassificering går ut på att kategorisera textdata efter fördefinierade klasser. Några vanliga tillämpningsområden är detektion av skräppost i mail-inkorgen, kategorisering av nyhetsartiklar, analys av produktrecensioner, och chattbotar.

Textklassificering kan uppnås med många olika metoder. Trots en del framgångar råder många brister med dagens teknik. De enklare metoderna kräver mycket manuellt arbete. Vad gäller de mer avancerade metoderna som baseras på maskininlärning krävs så kallade features; features är numeriska representationer av en text. För att ta fram högkvalitativa features krävs oftast mycket goda domän-specifika kunskaper. Det finns djup-inlärningsmodeller som fungerar utan manuellt framtagna features, men komplexiteten hos dessa modeller gör dem till mer eller mindre svarta lådor. Det går alltså inte att tolka på vilka grunder modellerna har gjort sina prediktioner.

Vi presenterar en GP-algoritm som kan generera text-klassificerare. Denna metod är helt automatiserad, kräver inga domän-specifika kunskaper och inga manuellt framtagna "features" och genererar klassificerare som är tolkningsbara. Algoritmen lyckas generera klassificerare

som kommer ända upp i 96% noggrannhet för ämnesdetektering. En förfinad version av algoritmen kommer integreras i ett automatiserat system hos företaget, Oxide, där arbetet utfördes. Detta system kallas Polychaos och är en själv-förbättrande sökmotor som kommer bedriva automatiserad forskning inom bland annat hälsa och life-science.

För att vår GP-algoritm ska kunna generera text-klassificerare definierade vi olika byggstenar. Dessa är logiska operatörer och n-gram. N-gram är en sekvens av n stycken bokstäver som kan uppträda i en text. Logiska operatörer kan vara "AND", "OR" eller "NOT", och används för att sammanfoga olika satser i en "string matching"-förfrågan. Nedan är en text-klassificerare som kan tolkas som förfrågan: "Innehåller texten inte 'pro' eller innehåller texten båda 'ret' och 'tro'?"

