

EXAMENSARBETE Improving the semantic accuracy and consistency of OpenStreetMap using modern machine learning techniques

STUDENTER Hannes Sandberg, Michal Stypa

HANDLEDARE Pierre Nugues (LTH)

EXAMINATOR Elin Anna Topp (LTH)

AI-assistent för kartläggning i OpenStreetMap

POPULÄRVETENSKAPLIG SAMMANFATTNING **Hannes Sandberg, Michal Stypa**

Över fyra miljoner bidragande användare och ett nära laglöst bidragssystem. Hur säkerhetsställs det att riktlinjer tolkas och implementeras rätt? I detta arbete används maskininlärning för att upplysa kartläggare om avvikelser, ofullkomligheter och direkta fel samt föreslå fler egenskaper.

Genom att förse maskininlärningsalgoritmer med data från över 600 000 vägar i Sverige har vi lyckats ta fram ett hjälpverktyg för kartläggning och felkontroll. Ta inmatning av ett nybyggt stycke av E4:an som exempel. Genom att kolla på en rad olika parametrar som liknande vägar, området i nära anslutning och vägens uppbyggnad föreslår systemet vägens mest sannolika egenskaper. Dessa skulle kunna vara en hastighetsbegränsning på 110 km/h, enkelriktning och två filer. Vid inmatning av orimliga egenskaper föreslår systemet lämpliga åtgärder. Skulle kartläggaren beskriva E4:an som ett objekt där fotgängare tillåts kommer systemet säga att det med största sannolikhet är fel.

Geografiska objekt märks i OpenStreetMap (OSM) med så kallade taggar. Dessa innehåller all information om objektet så som position, förhållande till andra objekt och en rad beskrivande egenskaper. Hur olika objekt ska beskrivas finns dokumenterat på OSM's hemsida men OSM låter användarna själva avgöra hur ett objekt ska taggas. Denna frihet gör att systemet är väldigt lättanvänt, särskilt för nya kartentusiaster, men medför skillnader beroende på kartläggarens tolkning av vad som är korrekt.

Detta brus gör det nästintill omöjligt att an-

vända datan i analytiska sammanhang då datorer behöver tydliga regler. För att underlätta kartläggningsarbetet och framför allt uppmåna konsekvent beskrivningsmetodik har vi tagit fram ett verktyg som bygger på befintliga mönster.

För att få fram dessa har vi använt maskininlärningsalgoritmer där datorn, utan vår hjälp, härleder mallar för hur olika objekt bör se ut. Detta åstadkoms genom att analysera samband mellan egenskaper och dess olika värden. Eftersom systemet bygger på statistisk analys kan även tidigare osedda objekt kategoriseras. Sambandsmallen mellan egenskaper ger även en procent på hur sannolikt ett befintligt värde är tillsammans med andra egenskaper samt hur sannolika eventuella förslag är.

Skulle samma problem vilja lösas utan artificiell intelligens skulle ett enormt regelverk behöva tas fram för hand där alla möjliga utfall täcks. Även då skulle systemet inte vara kapabelt att jobba med nyintroducerade objekt eftersom regelverket då bara skulle passa den ursprungliga datan.

Under arbetets gång har det framkommit att det enbart krävs 5% av alla vägar i Sverige för att systemet skulle kunna hantera vilken svensk väg som helst.