

EXAMENSARBETE Constructing a Multilingual Relation Extraction System

Using Neural Networks

STUDENT Erik Gärtner & Axel Larsson**HANDLEDARE** Pierre Nugues (LTH) & Håkan Jonsson (Sony Mobile)**EXAMINATOR** Jacek Malec (LTH)

Teknik bekämpar falska nyheter

POPULÄRVETENSKAPLIG SAMMANFATTNING Erik Gärtner & Axel Larsson

Falsa nyheter är något som cirkulerar på Internet i allt högre grad och orsakar stor skada mot samhället. Vi har byggt ett system som kan användas för att motverka denna trend genom att automatiskt extrahera påståenden från texter som sedan kan kontrolleras mot känd fakta.

Faktakontroll av nyheter är i sig inget nytt men idag konsumerar allt fler människor sina nyheter via sociala medier och undkommer härvid källkritik och faktakontroll som utförs av journalister. Den stora mängden nyheter gör det idag omöjligt att manuellt kontrollera allt. Självklart vore det bra om dessa nyheter automatiskt kunde kontrolleras med datorer men dessa har svårt att förstå vanliga texter. De behöver mer strukturerad data.

Vi har skapat ett system som kan extrahera och strukturera påståenden från vanlig text och som dessutom gör det mycket snabbt. Att gå igenom hela engelska Wikipedia, med mer än 5 miljoner artiklar tar bara 20 minuter med vårt system. Systemet kör på ett tiotal datorer samtidigt för att kunna uppnå denna prestanda.

Det bästa med systemet är att det är självlärande. Genom *maskininlärning* och *språkteknologi* kan vi lära systemet att känna igen olika typer av påståenden. Vi visar systemet exempel på hur påståenden brukar se ut och utifrån det lär sig systemet att känna igen nya påståenden. Internt använder systemet ett *neuralt nätverk* för att skilja mellan olika typer av påståenden, t.ex. om ett påstående beskriver var en person är född eller vem denna är gift med.

Vi lärde systemet genom att låta det bearbeta Wikipedia, vilket i detta fall är en lämplig källa då den är gratis och innehåller flera miljoner artiklar

inom alla ämnesområden. Till sin hjälp använde systemet Wikidata som är en öppen databas av fakta – fakta som gav systemet ledtrådar till vilka fakta som systemet kunde förvänta sig att hitta i Wikipedia.

Våra resultat visar att systemet är pålitligt för vissa typer av påståenden. Givet tid är det möjligt att expandera systemet för andra typer av påståenden. Som demonstration konstruerade vi ett plugin till webbläsare som automatiskt kontrollerar påståenden på en hemsida mot tidigare etablerade fakta.

Systemet är dessutom flerspråkigt vilket betyder att fakta och påståenden från olika språk kan kontrolleras och sammanfogas till en enhetlig databas. Detta är viktigt då det i vissa länder finns en stark inhemsk propaganda på det egna språket och inte engelska.

I nuläget är systemet på forskningsstadiet men i framtiden lär denna typ av system komma att användas i många olika sammanhang. Dels för att kontrollera fakta i texter, men även för att bygga databaser till virtuella assistenter och liknande system där informationen främst finns i vanliga texter. Kanske kommer din doktor om 50 år att vara ett system som lärt själv från medicinsk kurslitteratur.