Per Classon
pclasson@kth.se
DD143X

Oppositionsprotokoll

Rapportförfattare: Mattias Folke, Mitra Khorsand
Rapportenstitel: En jämförelse av matchningsmetoder för big data twitterbotar
Opponent: Per Classon

Var det lätt att förstå vad kexjobbet gick ut på? Kommentarer.
Det var lätt att förstå vad arbetet gick ut på: det handlade om att jämföra två olika typer av chatbotar. Det var en lättläst inledning där strukturen av rapporten förklarades tydligt. Definition av begreppen fanns i början av rapporten, vilket hjälpte för den oinsatta.

Hur tycker du att titeln avspeglar rapportensinnehåll?
Titeln speglar rapportens innehåll väl, den är dock ganska generell. Den säger att det är en jämförelse av *matchningsmetoder*, men kanske hade det varit bättre att specificera vilka typer av matchningsmetoder som används. För att förtydliga att det handlar om matchning med trigram jämfört mot semantisk klassificering.

Hur beskrev författaren bakgrunden till kexjobbet? Finns det en introduktion till och översikt av området?
Författarna har gjort en ordentlig och väl tilltagen bakgrund där först turingtest och en kort historia om chatbotar gås igenom. De beskriver även hur chatbotar används idag och vad big data är. Det finns kort en beskrivning av textklassificering och trigram också. Sammantaget är det en bra översikt av området och man får en förståelse för vad deras undersökning och arbete utgår från. Det som däremot saknas i bakgrunden är en översikt av liknande undersökningar och ifall andra har undersökt hur olika klassificeringar och liknande kan användas för att skapa chatbotar.

Hur väl har författaren motiverat sitt val av metod att angripa problemet?
Turingtest och relevanstest användes för att utvärdera de två matchningsmetoderna, och dessa två metoder var väl motiverade. Det förklarades varför dessa test passade just för utvärdering av chatbotar och hur de användes. Vad som däremot inte motiveras i metoden är varför algoritmen Levenshtein Distance användes. Man får inte heller veta varför just uClassify ramverket användes till att klassificera tweetsen.

Diskuterar författaren huruvida de förutsättningar som gäller för att metoden ska kunna användas är uppfyllda?
Det känns inte applicerbart på denna undersökning. Kanske hade de kunnat diskuterat begränsningen av tweetens längd och hur detta påverkar metoden.

Är metoden väl beskriven?
Metoden är för det mesta väl beskriven där man tydligt förstår *hur* undersökningen har gjorts. Det vill säga vilka verktyg som användes och hur deras matchningsmetoder fungerade. Det fanns

också väldigt fina figurer som ökade förståelsen mycket. Något som jag saknade var en förklaring om hur Levenshtein Distance-algoritmen fungerade. Det förklarades vad den gjorde, men rapporten gick inte in på hur algoritmen faktiskt räknade ut "avståndet" mellan olika tweets. Det förklarades inte heller hur klassificering med uClassify fungerade, utan det nämndes bara att ett ramverk användes för att utföra detta.

Har författaren presenterat sina resultat tydligt?
Resultatet av undersökningarna är presenterade tydligt i olika cirkeldiagram och stapeldiagram, som har förklaringar så att man förstår vad som presenterades. Det gavs också beräknar med statistisk säkerhet på resultaten. Även om mycket resultat finns i rapporten, hade det kunnat vara bra med en tabell där all data från utvärderingen presenterades.

Finner du författarensslutsatser trovärdiga?
Slutsaterna är för det mesta trovärdiga och känns välmotiverade. Det diskuterades om hur undersökningen hade kunnat förbättras och även en hel del om problemen med trigram. Vad som saknades var en diskussion om hur klassificeringen hade kunnat förbättras. Det var också väldigt svårt att förstå varför botarna var bättre än slumpmässiga tweets, det förklarades med en argumentation som jag inte förstod.

Vad tycker du om litteraturlistan? Vilken typ av litteratur finns med? Förefaller litteraturen relevant?
Litteraturförteckningen var bra. Många forskningsartiklar refererades och dessa förehaller relevanta för rapporten.

Vilka avsnitt i rapporten var svåra att förstå?
Det var svårt att förstå statistikberäkningar (oparat t-test etc.) i resultatdelen, det gavs ingen ingående förklaring till hur och varför det beräknades. Det var också svårt att förstå i slutsatsen varför botarna var bättre än slumpmässiga tweets.

Övriga kommentarer om rapporten och dess struktur.
Rapportens struktur var bra. Rätt sak på rätt plats, de hade inte blandat ihop olika delar.

Vilka är arbetets/rapportens starka sidor?
Rapportens starka sidor är att det är en väldigt intressant jämförelse som har gjorts, det känns verkligen som ett aktuellt ämne. Arbetet som har gjorts verkar också ha varit väldigt bra utfört. Rapporten är också väldigt välskriven, enkel att förstå och kul att läsa.

Vilka är arbetets/rapportenssvaga sidor?
Dess svaga sidor är hur utvärderingen har utförts, men detta diskuteras också i slutsatsen. I vissa delar (som metod och resultat) känns det också som att man skulle kunna ha gått in lite mer ingående och förklarat mer.

Per Classon
pclasson@kth.se
DD143X

Hur bedömer du arbetets nyhetsvärde?
Arbetets nyhetsvärde är svårt att bedömma, eftersom rapporten inte går in forskning om klassificering av tweets etc. Det känns dock som ett aktuellt område.

Sammanfatta arbetet på ett par rader.
Två olika botar har skapts, som svarar på tweets. En som använder sig av trigram och en som använder sig av klassificering. Det har jämförts vilken metod som är bäst genom att göra en utvärdering med turingtest och relevanstest. Inget statistiskt säkerstället resultat kunde fås, däremot lutade det mot att semantisk textklassificering av tweetsen var den bästa metoden.

Frågor till författaren:
1. Hur fungerar Levenshtein Distance-algoritmen?
2. Hur fungerar klassificering av tweets?
3. De flesta gick på magkänsla när de valde svar. Enligt rapporten tyder detta på att de “genererade svaren” hade varit helt orelevanta. Varför då?
4. Följdfråga på 3:an. Varför fanns det ingen “kontroll” med helt slumpmässiga tweets i utvärderingen? Hur vet ni ändå att era bottar är bättre än helt slumpmässiga tweets från databasen?

Vilket är ditt sammanfattande omdöme om kexjobbet?
Sammanfattande är det en väldigt bra rapport, som var lättläst och intressant. Det var en bra vetenskaplig struktur på uppsatsen. Arbetet som har gjorts verkar väldigt väl genomfört också. Det enda som drog ner rapporten lite, var att utvärderingen av resultatet verkade ha en del brister. Metoden hade också kunnat vara lite mer ingående. Avslutningsvis tycker jag det var kul rapport, och man får själv lust att göra en egen chatbot efter att ha läst rapporten!