

RECENZJA

rozprawy doktorskiej mgr. Wojciecha Włodarczyka pt. "Modele ewaluacji poprawności danych lingwistycznych pozyskanych metodą crowdsourcing"

1. Tematyka rozprawy.

Praca poświęcona jest problematyce anotowania danych lingwistycznych z wykorzystaniem crowdsourcingu, który polega na angażowaniu do zadania dużych grup ludzi pracujących na wspólnych danych przy użyciu aplikacji sieciowych. Doktorant skupił się na ocenie jakości takiej anotacji oraz sposobie jej poprawy przez wykorzystanie informacji zwrotnej służącej do edukacji anotatorów w trakcie wykonywania zadań. Autor zaproponował też własny model automatycznego generowania informacji zwrotnej. W pracy opisane są liczne eksperymenty przeprowadzone na pięciu zbiorach danych lingwistycznych, których wyniki Doktorant poddał szczegółowej analizie.

2. Zawartość merytoryczna

Praca składa się z czterech rozdziałów i podsumowania. W pierwszym rozdziale Doktorant wprowadził czytelnika w tematykę pracy, zdefiniował pojęcia wykorzystywane w rozprawie oraz przedstawił aktualny stan badań.

Rozdział drugi stanowi wprowadzenie w problematykę mechanizmu informacji zwrotnej oraz oceny różnych aspektów jego działania. W rozdziale tym Doktorant zreferował artykuł autorstwa Zhu, X., Singla, A., Zilles, S., Rafferty, A. N. „An Overview of Machine Teaching” poświęcony nauczaniu maszynowemu. Następnie omówił jak mechanizm nauczania maszynowego może zostać wykorzystany w kontroli i poprawie jakości zadań anotacyjnych realizowanych z wykorzystaniem outsourcingu.

W rozdziale trzecim Doktorant sformułował hipotezy oraz pytania badawcze, na które następnie udzielił odpowiedzi przeprowadzając eksperymenty na pięciu zbiorach danych. Hipotezy dotyczą wpływu informacji zwrotnej (i jej jakości) na jakość pozyskiwanych danych, zaś pytania badawcze dotyczą wpływu informacji zwrotnej na efekt edukacyjny, szybkość i zaangażowanie anotatorów. W przypadku ostatniego z pytań Doktorant zdefiniował zaangażowanie anotatora jako łączną liczbę wykonanych przez niego mikro-zadań wchodzących w skład danego zadania, co wydaje się wątpliwą miarą tej cechy. Skoro wszystkie zadania były opłacane tak samo, to wydaje się, że prostsze, szybciej wykonywane anotacje były chętniej wybierane, bo przekładały się na wyższe wynagrodzenie za godzinę pracy. Jest to więc raczej ocena atrakcyjności wykonywania zadania. W rozdziale przedstawiona została szczegółowa analiza eksperymentów pozwalająca na sformułowanie wniosków wskazujących między innymi na wpływ informacji zwrotnej na jakość anotacji w zależności od wiedzy potrzebnej do wykonania zadania. Eksperymenty wykazały większy pozytywny wpływ w przypadku zadań wymagających od anotatora wiedzy dziedzinowej. Istotnym wnioskiem jest to, że nawet niedoskonała informacja zwrotna ma zwykle pozytywny wpływ na jakość pozyskiwanych anotacji, co wskazuje na możliwość wykorzystywania w tym celu anotacji generowanej automatycznie.

W rozdziale czwartym Doktorant przedstawił autorski model automatycznego generowania Dynamicznej Informacji Zwrotnej. Postawił dwa pytania badawcze dotyczące opracowanego modelu w celu oceny jego działania w procesie anotacji. Następnie przeanalizował działanie modelu na danych rzeczywistych oraz symulowanych dla pięciu zadań. Uzyskane wyniki nie pozwoliły mu na jednoznacznie pozytywną rekomendację zaproponowanego modelu Dynamicznej Informacji Zwrotnej. Autor we wnioskach sugeruje wykorzystywanie tego mechanizmu po wcześniejszej analizie rozwiązywanego problemu.

3. Uwagi dotyczące języka matematycznego

Praca zawiera bardzo wiele błędów i usterek w formułach matematycznych i algorytmach. Istotnie utrudnia to zrozumienie wywodu, gdyż wymaga domyślania się, w którym miejscu został popełniony błąd.

W rozdziale 2.1 zamęt wprowadzony przez Doktoranta w oznaczeniach i symbolach matematycznych zmusza czytelnika do sięgnięcia do oryginalnego artykułu, po którego lekturze pojawiają się pytania, dlaczego we wzorze 2.3 (przytaczanym za oryginalną pracą) nastąpiła zmiana znaku z plus na minus i KosztNauczania jest odejmowany od RyzykaNauczania. Ponadto, we wzorze pojawia się \hat{g} zamiast g , oraz oznaczenie D_u , choć poszukując minimum rozważamy zmienne D i g (wskazane byłoby również zdefiniowanie odpowiednich zbiorów do których należą te zmienne). Podobne problemy z oznaczeniami pojawiają się we wzorach na stronie 69. Natomiast wzór na górze str. 65 (bez przypisanego numeru) pozostawiony jest bez jakiegokolwiek komentarza, choćby dotyczącego pojawiających się w nim symboli.

Poniżej podaję listę zauważonych błędów we wzorach i algorytmach:

- We wzorach 2.24, 2.26, 2.27 pojawia się niezdefiniowany symbol y_i^* zamiast wcześniej opisanego y_i .
- W algorytmie znajdującym się na stronie 85 i we wzorze 2.28 sumowanie powinno rozpoczynać się od wartości $t=1$. Ponadto, nie każdy anotator rozwiązuje wszystkie mikro-zadania, co powinno znaleźć odzwierciedlenie w algorytmie. Nie wiemy jakie wartości są przypisane, gdy mikro-zadanie nie jest rozwiązywane przez anotatora.
- W Procedurze 2 na str. 21 pojawiają się nieokreślone zmienne.
- W Przykładzie 5 zmienna j powinna się zmieniać od 1 do n , a nie od 0.
- W algorytmie na stronie 88 nie wiemy czym jest R ? Co oznacza, że zmienna n ma wartości od 1 do n ?
- Strona 90: parametry θ nie są zdefiniowane.
- Na stronie 93 parametry są oznaczone Φ , znów nie wiemy jakie.
- Strona 135, w nieoznaczonym wzorze, nie wiadomo czym jest n_a i n_b .
- W algorytmie na stronie 139 nie wiadomo czym jest $R_{i,1}$ do $R_{i,n}$, zdefiniowane jest R_i .
- W tabeli 3.10 są źle policzone wartości dodatnie i ujemne.
- W procedurze 12 na stronie 167 nie wiadomo czym jest D_u .
- Na stronie 170 źle spisane są z pracy Jurafsky&Martin wzory tf-idf.

Proponuję też wprowadzić rozróżnienie między pojęciem „dane” i „anotacje”. Obecnie są używane wymiennie co bywa mylące, patrz np. rozdział 1.2.4 lub na stronie 29 w rozdziale 1.3.2

Należałoby rozważyć uzupełnienie pracy o legendę oznaczeń, które następnie powinny być konsekwentnie stosowane.

4. Uwagi dotyczące formy i języka

Praca napisana jest językiem, który umożliwia jej zrozumienie, wymaga jednak znacznego dodatkowego wysiłku ze strony czytelnika. Zawiera bowiem bardzo wiele potknięć językowych, niepotrzebnych powtórzeń, niekonsekwencji, niewłaściwie użytych słów, niepoprawnych przymków i fraz rzeczownikowych w niewłaściwym przypadku. Wszystko to odwraca uwagę czytelnika od istoty przedstawianych treści. Błędy te powinny być usunięte na etapie redakcji, którego praca najwyraźniej nie przeszła.

W pierwszym przypisie Doktorant wyjaśnia potrzebę użycia angielskiego terminu „crowdsourcing” podlegającego polskiej deklinacji, po czym zarówno w tytułach całej rozprawy i pierwszego rozdziału termin ten pojawia się w mianowniku zamiast w dopełniaczu.

W rozdziałach 3 i 4, które zawierają istotnie mniej błędów językowych, Autor powieliła całe fragmenty tekstu, patrz rozdział 3.5.4. Rozdział 4.2.5 powstał na zasadzie pięciokrotnego kopiowania opisu z drobnymi zmianami na poziomie kilku słów. Taka metoda referowania wyników jest nużąca dla czytelnika.

W wielu miejscach Autor pisząc o zawartości nadchodzącego rozdziału używa czasu przeszłego, co myli czytelnika, bo te treści dopiero będzie czytał.

Poniżej przytaczam jedynie wybrane przykłady błędów językowych:

- Nie zaczynałabym rozdziału będącego wprowadzeniem do pracy od „Mimo że”.
- Autor w jednym akapicie pisze „W Rozdziale 2”, a już w następnym „W rozdziale trzecim” (strona 12). Nie jest utrzymana konwencja zapisu.
- Strona 15: „stanowią postawę dla rozważań”.
- Strona 21: „gdzie a Y_i”.
- Strona 24 „na osiągnięcie poprawę dystrybucji”.
- Strona 28 „Po dokonaniu oceny jakości danych wykonywane są konkretne akcje, mające na celu poprawę jakości.” Te „konkretne” akcje nie są opisane. Dalej w pracy też się pojawia to słowo bez dalszego podania szczegółów, np. str. 34 „mają postać konkretnych algorytmów”.
- Strona 32 „Na podstawie przeprowadzony przeze mnie przegląd literatury”.
- Strona 34 „w oparciu wybrany algorytm”.
- Strona 34 „Jakość anotatora obliczana poprzez porównanie jego odpowiedzi z odpowiedziami referencyjnymi.” brak czasownika.
- Strona 42 „informacja zwrotna ma formę punktową ocenę anotacji zdefiniowaną w wybranej skali”.
- Praca jest napisana po polsku, więc w przypadku wieloautorskich publikacji powinno się używać „i inni” a nie „et al.”
- Strona 49 „W ramach niniejszej rozprawy opracowałem eksperyment, w ramach którego zagłębiłem oba powyższe tematy.”
- Strona 54 „W niniejszym rozdziale opisana została definicja oraz charakterystyka nauczania maszynowego”. Definicję się „podaje” lub „wyjaśnia”, a nie „opisuje”.
- Strona 58 „W przypadku problemu jednowymiarowego, określenie zbioru D okazało się to bardzo proste.”
- Na str. 61, pojawia się odesłanie do „Protokołu 4”, który nie wiadomo gdzie się znajduje.
- W opisach niektórych algorytmów, zmienna pojawia się nie w tym miejscu zdania co trzeba, co powoduje przypisanie jej innego, niż zamierzone przez Autora, znaczenia. Przykład takiej sytuacji jest na str. 72: „System aktualizuje aktualny model wiedzy anotatora M_u^j w oparciu

o odpowiedź udzieloną przez anotatora y_i ." Ze zdania wynika, że y_i jest anotatorem, a nie odpowiedzią, na co wskazuje pozostała treść opisywanego procesu.

5. Konkluzja

Temat rozprawy wraz z przedstawionymi eksperymentami uważam za ciekawy i nowatorski. Autor wykazał się wszechstronnymi umiejętnościami: od implementacji interfejsu do systemu MTurk umożliwiającego wykorzystanie informacji zwrotnej, przez przygotowanie i przeprowadzenie serii eksperymentów badawczych, po wnikliwą ocenę ich wyników. Praca stanowi istotny wkład w dziedzinę badań nad metodami crowdsourcingu, w szczególności nad poprawą jakości uzyskiwanych przy jej zastosowaniu anotacji. Natomiast warsztat pisarski rozprawy oraz niepoprawne, nieprecyzyjne stosowanie języka matematycznego stanowią ogromny mankament przedstawionej do recenzji pracy, którą zamiast kierować do poprawy, postuluję uzupełnić o erratę.

Wartość merytoryczna pracy, mimo istotnych uwag do edycji tekstu, pozwala mi stwierdzić, że rozprawa pod tytułem: *Modele ewaluacji poprawności danych lingwistycznych pozyskanych metodą crowdsourcing* spełnia wymogi, jakie ustawa o stopniach i o tytule naukowym przewiduje dla rozpraw doktorskich. Na tej podstawie wnoszę o dopuszczenie Autora pracy – mgr. Wojciecha Włodarczyka – do publicznej obrony.

Matejona Marciniak