

mgr Wojciech Włodarczyk

Modele ewaluacji poprawności danych lingwistycznych pozyskanych metodą crowdsourcing

Streszczenie

Crowdsourcing pozwala na wykorzystanie zbiorowej inteligencji dużej grupy ludzi do rozwiązywania zadań z dziedzin takich jak sztuczna inteligencja, uczenie maszynowe i rozwój badań naukowych za pośrednictwem platform internetowych. Współcześnie badania nad rozwojem metody crowdsourcingu skupione są przede wszystkim w trzech obszarach: optymalizacji jakości pozyskanych danych, optymalizacji kosztu procesu oraz optymalizacji czasu trwania procesu.

Niniejsza rozprawa skupia się na zagadnieniach związanych z optymalizacją jakości procesu crowdsourcingu dla zadań dotyczących danych lingwistycznych. Praca opisuje autorski model Dynamicznej Informacji Zwrotnej (DIZ), którego zadaniem jest generowanie informacji zwrotnej w sposób automatyczny. Rozprawa weryfikuje skuteczność tego modelu dla danych empirycznych oraz danych symulacyjnych.

Analiza wyników przeprowadzonego eksperymentu wykazuje skuteczność modelu DIZ w poprawie jakości generowanej informacji zwrotnej, jednak jakość tego rozwiązania jest zależna od jakości oznaczeń tworzonych przez anotatorów.

Słowa kluczowe

crowdsourcing, obliczenia ludzkie, nauczanie maszynowe, przetwarzenie języka naturalnego

Wojciech Włodarczyk