

Features for the People: An Expert-Enabling User Interface for Feature Engineering in Spoken Language Identification

Within the field of Natural Language Processing, the spoken language identification problem (the task of “detect[ing] the presence of a hypothesized target language, given a segment of conversational speech”) is a clear instance of the process of feature engineering. Feature engineering is the task of determining the costliest errors made by a system and elucidating novel features to reduce errors without causing regressions. A primary obstacle to improved system performance is the disconnect dividing linguists and other domain experts from the inaccessible mechanics of the computer system. The Feature Engineering Console (FEC) addresses these problems by providing a user interface for expert involvement in feature engineering process. It acts as a GUI abstraction upon which non-coders or those with limited time can experiment with and contribute to the performance of the identification system. The FEC should prove a useful tool for any classification or detection task in NLP, signal processing, machine learning, etc.