EP11 Using machine learning to recover unrecorded prehospital data

Thilo Reich, Adam Bancroft, Marcin Budka

Abstract

Background The recording practices, of electronic patient records for ambulance crews, are continuously developing. South Central Ambulance Service (SCAS) adapted the common AVPU-scale (Alert, Voice, Pain, Unresponsive) in 2019 to include an option for 'New Confusion'. Progressing to this new AVCPU-scale made comparisons with older data impossible. We demonstrate a method to retrospectively classify patients into the alertness levels most influenced by this update.

Methods SCAS provided ~1.6 million Electronic Patient Records, including vital signs, demographics, and presenting complaint free-text, these were split into training, validation, and testing datasets (80%, 10%, 10% respectively), and under sampled to the minority class. These data were used to train and validate predictions of the classes most affected by the modification of the scale (Alert, New Confusion, Voice).

A transfer-learning natural language processing (NLP) classifier was used, using a language model described by Smerity et al. (2017) to classify the presenting complaint free-text.

A second approach used vital signs, demographics, conveyance, and assessments (30 metrics) for classification. Categorical data were binary encoded and continuous variables were normalised. 20 machine learning algorithms were empirically tested and the best 3 combined into a voting ensemble combining three vital-sign based algorithms (Random Forest, Extra Tree Classifier, Decision Tree) with the NLP classifier using a Random Forest output layer.

Results The ensemble method resulted in a weighted F1 of 0.78 for the test set. The sensitivities/specificities for each of the classes are: 84%/ 90% (Alert), 73%/ 89% (Newly Confused) and 68%/ 93% (Voice).

Conclusions The ensemble combining free text and vital signs resulted in high sensitivity and specificity when reclassifying the alertness levels of prehospital patients. This study demonstrates the capabilities of machine learning classifiers to recover missing data, allowing the comparison of data collected with different recording standards.