

2020 年度研究助成 研究実績報告書

代表研究者	井口 元三
研究テーマ	深層学習(ディープラーニング)を活用したストレスチェックに基づく労働災害リスク予測モデルの開発

<助成研究の要旨>

【背景】厚生労働省の「患者調査」によると、過労死・うつ病・自殺へと繋がる職場でのメンタルヘルス不調者総数は年々増加している。このような状況から「ストレスチェック」制度が開始され、全労働者に実施することが義務付けられた。本制度の実施率は約83%と高く、現在最も多くの事業場で行われているメンタルヘルス調査制度となる。しかしながら、ストレス反応を総合点で算出して評価するため、メンタルヘルス不調に対して識別性の高い項目が、総合点の中に埋没する可能性が指摘され、新たな予測モデル創出の社会的ニーズが高まっていた。

【目的】機械学習および深層学習(ディープラーニング)を用いた探索的データ解析を活用して、ストレスチェックに基づいたメンタルヘルス不調者を見逃さないための新規予測モデルの構築を目的とする。

【方法】2019 年度に本学全職員を対象としてストレスチェックを実施した。ストレスチェックデータとメンタルヘルス不調者との紐づけを行い、機械学習および深層学習(ディープラーニング)による予測モデルの作成を試みた。ストレスチェックの各項目を前処理してデータセットとし、メンタル相談の有無をターゲットとしてメンタルヘルス不調者の予測モデルを作成した。

【結果】基礎統計データとしてのストレスチェック対象者 5,315 人の内、回答者は 3,352 人(回答率 63.1%)であった。その中でメンタル相談受診者は 116 人(3.5%)であった。

得られたデータセットを教師データ: 予測データの8:2に分割し、教師データを用いて各種機械学習アルゴリズム(ランダムフォレスト、ロジスティック回帰、サポートベクターマシン(SVM)等)により複数の予測モデルを作成し、ハイパーパラメータチューニングおよびデータセットを 10 分割して交差検証を行い、正解率(Accuracy)が 95%を超える複数の予測モデルが作成された。

機械学習におけるベストモデルは Light Gradient Boosting Machine(Light GBM)(Accuracy: 0.9654, AUC: 0.7395)であり、Top5 のモデルはそれぞれ、Light GBM、Random Forest Classifier(0.9648, 0.7344)、Quadratic Discriminant Analysis(0.9648, 0.5000)、Extra Trees Classifier(0.9648, 0.7712)、CatBoost Classifier(0.9638, 0.7708)であった。

また、得られた各予測モデル(弱学習器)を集約したアンサンブル学習を行うことで、高精度な予測モデルを構築した。Top5 の予測モデルを用いたスタッキングでは(0.9503, 0.6399)とベストモデルの Light GBM に及ばなかったが、Light GBM を用いた Bagging (0.9648,0.7539)、Boosting (0.9643, 0.7741)では良好な予測モデルが構築できた。

さらに、深層学習(ディープラーニング)ソフト Neural Network Console を用いて、より効率の良い予測モデルを求めて構造探索を行った結果、予測精度の高いニューラルネットワークによる予測モデル(Accuracy: 0.9582~0.9716)を複数構築することができた。

今後、より大規模なストレスチェックデータを用いてより精度の高い予測モデルを構築することを目指す。

【結語】複雑な要因が多重的に関わる職業性ストレスにおけるメンタルヘルス不調者予測モデルの構築に成功した。将来的に疾病を予防する質の高い保健サービスの創出が期待される。