

共変量シフト下における医療費の予測モデリング

小暮厚之*

東京経済大学 経営学部

概要

世界中で長寿化が急速に進行する中で、医療費の高騰は多くの国々において喫緊の課題である。このような課題を扱う最初のステップは医療費の予測であろう。 t 年において翌年の医療費 Y_{t+1} を予測するために通常行われる方法は以下の通りである：

1. t 年における医療費 Y_t を目的変数、前年の共変量 X_{t-1} を説明変数とする予測モデルを構築
2. 構築した予測モデルの説明変数の値を t 年の共変量 X_t に置き換え、 Y_{t+1} を予測。

このような方法の背後には、 (X_{t-1}, Y_t) と (X_t, Y_{t+1}) の確率分布に変化がないという想定がある。しかし、現実のデータを用いる場合には、確率分布に何らかのシフトが生じている可能性を考慮すべきであろう。その場合、通常の予測法を用いると何らかのバイアスが生じると考えられるであろう。機械学習の世界では、このような問題をデータシフトという名称で扱い、その困難さを軽減するための適応手段を議論してきた。

本報告では、データシフトの重要なケースである共変量シフトを考察する。目的は、実際の健康保険データに対して Two-part モデルを適用した場合に、共変量シフトに対応する適応手段によって医療費の予測精度が向上するかどうかを検討することである。分析に用いるデータは、日本の複数の健康保険組合から無作為に抽出された 1 万人の利用者に関する 2010-2012 年の 3 年間のレセプトデータ及び健診データである。

1 序

世界中で長寿化が急速に進行する中で、医療費の高騰は多くの国々において喫緊の課題である。このような課題を扱う最初のステップは医療費の予測であろう。 t 年において翌年の医療費 Y_{t+1} を予測するために通常行われる方法は以下の通りである：

1. t 年における医療費 Y_t を目的変数、前年の共変量 X_{t-1} を説明変数とする予測モデルを構築
2. 構築した予測モデルの説明変数の値を t 年の共変量 X_t に置き換え、 Y_{t+1} を予測。

予測モデルは、前年の共変量 $\mathbf{X}_{t-1} = \mathbf{x}$ を所与とした今年の医療費 Y_t の条件つき期待値 $g(\mathbf{x})$ を推定することによって得られることが多い。 g の推定値を \hat{g} とするとき、翌年の医療費 Y_{t+1} は、 $\hat{g}(\mathbf{X}_t)$ によって予測される。このような方法の背後には、 (X_{t-1}, Y_t) と (X_t, Y_{t+1}) の確率分布に変化がないということを前提にしている。しかし、現実のいわゆる「リアルワールドデータ」を用いる場合には、確率分布に何らかの変化が生じている可能性がある。その場合、通常の予測法を

*本論文では、株式会社 JMDC から提供を受けたデータを用いました。同社に深く感謝申し上げます。