

## 生活習慣病ハイリスクグループ特定に向けた データマイニングシステムの開発

○中山 洋一<sup>\*1)</sup>、上島 優一郎<sup>\*1)</sup>、岸本 充弘<sup>\*1)</sup>

### 1. はじめに

生活習慣病は40歳以降で罹患者が増加しはじめ、年齢とともに増加する。現在、高齢化に伴い、これら生活習慣病の罹患者数が急激に増加しており、それとともに関連医療費が急増し、健保組合の財政を圧迫している。これを防ぐためには、そもそも生活習慣病にかからないよう、その予防のための啓蒙活動が重要である。

生活習慣病の原因は、主に20代・30代に確立するライフスタイルにあり、生活習慣病に直面していない世代に対して、予防の重要性について啓蒙することは、これまでも多数の試みがなされているが、非常に困難な状況にある。疾患の予防には、特にインフルエンザ流行時の対応に代表されるような、予防のための行動の啓蒙も重要ではあるが、一方で、遺伝子検査に代表されるような、個々人の疾患リスクの把握も疾患予防に効果的であることがすでに知られている。特に、生活習慣病の場合は、特定の生活習慣を変更するよう動機付けを行う必要があり、往々にしてその変更は対象者の大きな精神的負担を伴う。

以上のようなことから、生活習慣病の中長期的な罹患確率を算定できるような、簡便かつ安価で正確な方法が望まれている。今回、健保組合が法令に基づいて実施している一般検診にいくつかのデータを付加したもので生活習慣病の中長期的な罹患リスクの高いハイリスクグループを予測するシステムを開発した。

### 2. 方法・結果

本システムは、検査データを保持するデータベース部と、予測を行うためのデータマイニング部から成る。データマイニングには、主成分分析やニューラルネットを組み合わせた手法を用いている。本システムの供試データとしては、約3万人×3年分の一般検診と健保独自の追加検査及び問診のデータを使用した。

本システムから出力される結果は、6種類の生活習慣病（糖尿病、脂質異常症、高血圧、血栓症、慢性腎疾患、脂肪肝）の各疾患の10年以内の罹患リスクを、100点が最高点（最高リスク値）として算出するようにした。また、各被験者の状況に応じて、期待される余命の長さを「総合健康度数」として表現するようにした。

### 3. まとめ・考察

本システムは、まだデータ数が十分ではないために、予測性能には限界があるが、本手法でこのようなリスク値の予測が可能であることを示すことができた。また、本手法を応用することによって、内科系疾患のコンピュータによる自動診断も可能であることを示すことができた。



\*1)株式会社サイトライン・ソリューションズ