情報システム工学科　平成１９年度「自主課題研究」

研究テーマ：R と階層型クラスタリングを用いた食品の分類

学年：3　名列番号：036　氏名：辰巳　和也

１．研究課題

統計解析ソフト R を用いて、クラスター分析による対象の分類を行う。その際、複数の「クラスター間距離の定義」を利用してクラスター分析を行い、それぞれの特性を理解する。

2．研究方法

① 本研究では、解析用データに一般食品 100g における栄養成分表を用いた（http://www4.ocn.ne.jp/~katonet/）。

② 対象間距離の定義には標準ユークリッド距離を用いた。

③ クラスター間距離の定義には、最短距離法、最長距離法、群平均法の三種類を用いた。

3．結果・考察

①最短距離法：帯状の大きなクラスターが出力された。他の二つの定義法と比べて、クラスターの分割部がわかりにくい。

②最長距離法：同じような大きさのクラスターを形成する傾向にある。そのため、クラスター同士の分割部が非常にわかりやすい。その反面、極端に数値が大きい（小さい）データがあっても、樹形図に反映されにくい。

③群平均法：3 つの手法の中で、一番バランスがよい樹形図が出力された（図１）。クラスターごとに、それぞれの食品群に分割できる。糖類などの極端な栄養素のデータも、最長距離法に比べて高い位置に配置されているためわかりやすい。

4．結論

本研究で用いたデータについては、群平均法が最も適していた。対象とするデータや分類の目的によって、その都度最適な手法を用いることが大切である。

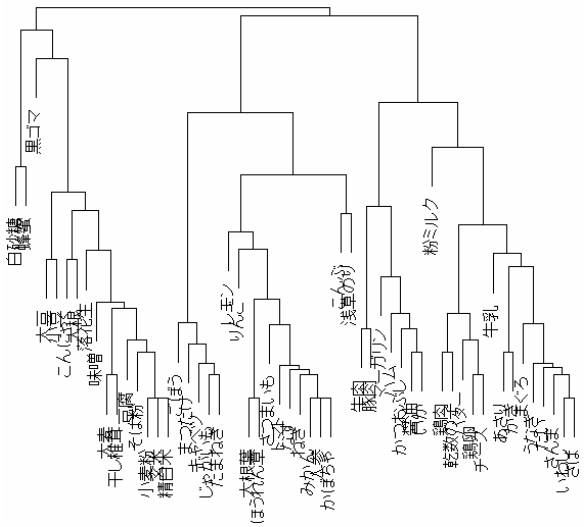

図１　群平均法を用いた場合の樹形図