

を抽出したところ、1127項目の表示名について103450レコードが抽出された。

表2 稼働額データ

診療年月日	患者ID	入外区分	診療区分名	表示名
2006-8-1	A001	外来	単純	胸部X-P
2006-8-2	A001	入院	調剤料	調剤料(入院)
2006-8-2	A001	入院	内服	内服薬
2006-8-2	A001	入院	その他注射	中心静脈注射

続いて、稼働額データには「室料差額料」など医療行為には直接関係がないと考えられるレコードや、注射薬剤名や検体検査項目など患者数に対してばらつきが大きいと考えられたレコードを分析対象から外した。また、残ったレコードの中で、異なる表示名ではあるが同じ医療行為として考えることができるレコードの表示名が同じになるようにそろえた結果、対象データは147項目の表示名について8291レコードとなった。

更に、8291レコードの中で出現頻度が5回以下の医療行為について対象データから外した結果、最終的に対象データは57項目の表示名について8123レコードとなった。この対象データから表1に示すような分析用データを作成した。対象データが57項目、対象入院期間のべ2561日(であることから、分析用データは患者ID、入院経過日数、および57項目の表示名のあわせて59列、2561レコードから成り立つデータである。

### 3.2 クラスタ分析

クラスタ分析の手法には様々な手法があるが、階層的な手法の中で一般的に最初に適用される場合が多いと考えられるWard法を用いることにした<sup>1)</sup>。Ward法では点がベクトルで記述されている必要があるため、表1における1(有)、0(無)をそれぞれ数値として扱うことによって適用した。クラスタ分析にあたっては、全て0(無)であるレコードは分析の対象から外した。その結果、最終的な点の数は2324となった。また、分析の結果得た樹形図より、クラスタ数は6として結果を検討することにした。Ward法による分析にあたっては、SAS Institute Japan株式会社 JMPジャパン事業部のJMP(TM)を使用した。

### 4. 結果

6に分けられたクラスタはそれぞれクラスタ1~クラスタ6と呼ぶことにする。それぞれのクラスタに分類された点の数、57項目の表示名のうちそれぞれのクラスタにおいて一番出現頻度の高かったものとその各クラスタにおける出現割合(表示名に該当する医療行為が行われていた点の数を、各クラスタに分類された点の数で割った値)を表3に示す。また、入院経過日数と各クラスタに分類された点の数の関係を図1、(入院経過日数-1)/(入院日数-1)と各クラスタに分類された点の数の関係を図2に示す。図2では、(入院経過日数-1)/(入院日数-1)=0が入院日、(入院経過日数-1)/(入院日数-1)=1が退院日を表す。

表3 クラスタ分析の結果

クラスタ	分類された点の数	各クラスタにおける出現割合					
		点滴注射	経管栄養	皮下、筋肉内注射(入院)	胃瘻栄養	理学療法(1)(個別)	中心静脈注射
クラスタ1	826	0.873	0.215	0.172	0.007	0.318	0.005
クラスタ2	163	0.086	0.926	0.061	0.000	0.393	0.074
クラスタ3	286	0.091	0.003	0.941	0.000	0.430	0.000
クラスタ4	132	0.318	0.000	0.811	0.992	0.273	0.000
クラスタ5	678	0.109	0.038	0.029	0.000	0.723	0.000
クラスタ6	239	0.000	0.318	0.017	0.000	0.473	0.757

### 5. 考察

表3より、各クラスタにおいて特に出現頻度が高かった表示名は、他のクラスタにおけるその出現頻度と比較しても高いものが多く、各クラスタは何らかの特徴を持つように分類されたと考えられる。また、図1、図2より、クラスタ1とクラスタ6は入院前半から中盤にかけて良く見られ、クラスタ4とクラスタ5は中盤から後半にかけて良く見られることが分かる。特に入院経過日数が100日を超えてくるような点については、クラスタ5に分類される点が多い。また、入院日の点はほとんどがクラスタ1に分類されるのに対し、退院に近くなるに従ってクラスタ5に分類される点が多くなるものの、他のクラスタに分類される点も多く見られる。このことより、入院時の医療行為に定型的なものがあるように考えられるのに対し、様々な病態で退院して行くことが推察される。

クラスタ分析の手法については、必ずしもWard法が良いとは限らず、この結果から得られる知見に基づいて他の手法を検討すべきである<sup>[1]</sup>。特に、本研究ではWard法を適用するために、0(有)または1(無)というカテゴリカルな属性であった表示名を数値属性として扱う必要があった。一方、カテゴリカルな属性を扱うための手法としてROCK<sup>[2]</sup>やSTIRR<sup>[3]</sup>などが提案されている。今後、このような手法の適用を検討する必要がある。

### 6. 結論

脳梗塞で入院した患者の各入院経過日を点と定義し、医事会計システムの稼働額データにWard法を用いてクラスタ分析することにより、各患者の各入院経過日を6クラスタに分類することができた。各クラスタは何らかの特徴を持っていると考えられ、2つのクラスタは入院前半から中盤にかけて、別の2つのクラスタは入院中盤から後半にかけて良く見られた。今回の知見に基づいて今後他の手法を検討して行く必要がある。

### 参考文献

- [1] 神尾敏弘.データマイニング分野のクラスタリング手法(1).人工知能学会誌 2003; 18(1): 59-65.
- [2] Guha, S., Rastogi, R., and Shim, K. ROCK: A Robust Clustering Algorithm for Categorical Attributes. Proc. of the 15th Int'l Conf. on Data Engineering 1999; 512-521.
- [3] Gibson, D., Kleinberg, J., and Raghavan, P. Clustering Categorical Data: An Approach Based on Dynamical Systems. Proc. of the 24th Very Large Database Conf.

P18-8 ポスターセッション/ポスターセッション:P18

1998; 311-322.

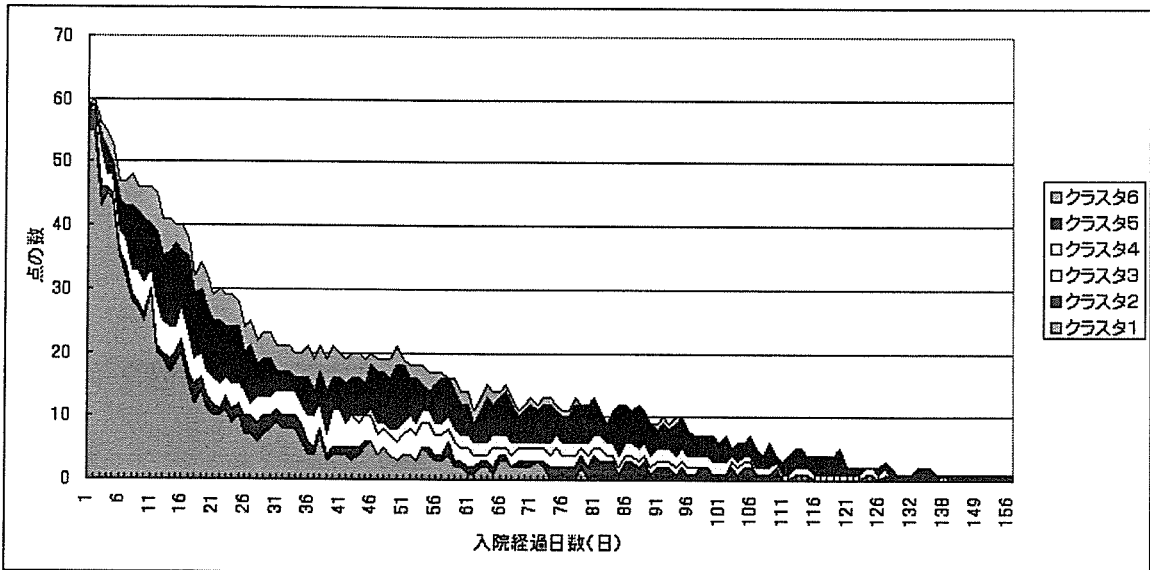


図1 入院経過日数と各クラスに分類された点の数の関係

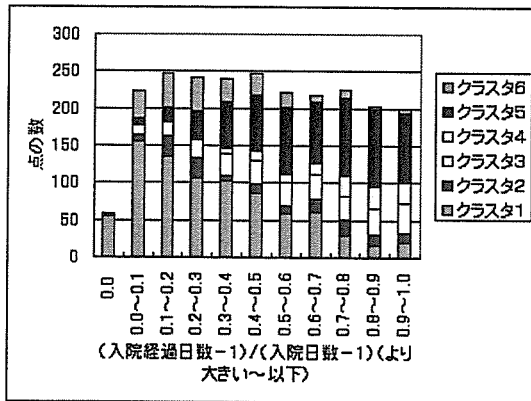


図2 (入院経過日数-1)/(入院日数-1)と各クラスに分類された点の数の関係