

国際情報処理学会 T C 4 第5回 Working Conference 報告
数学的理論、手法について
三尾 章彦 (日本医科大学)

IFIP, T C 4 主催の今回の Working Conference は, France の Dijon に於いて5月24日から5日間開かれた。第1日は総論, 第2, 第3日は Current Work, 第4, 第5日は理論, 手法という構成で, 研究報告と討論が行なわれた。

第1日午前 clinical medicine と public health に於ける decision making について5演者で哲学なり一般論なりが取扱われた。

第1日午後

1. Clinical Decision Making; B. Lusted (Chicago).

Decision Trees, Bayes Theorem, Decision Matrices, ROC (Receiver Operating Characteristic) Analysis, Cost and Benefit という概念によって階層的に構成される典型的な診断過程を想定し, 実地の立場から例を混じえて論じた。

2. Cost and Utilities; D. V. Lindley (London)

確率を用いて定義される Utilities の概念を医療の場に入用する事を試み, この Utilities の期待値による act の比較評価を試みた。

3. Multidimensional Data Analysis in Medical Decision; J. P. Nakache (Paris)

医療 Data の分類, 変換. 更に M. D. A. の有用性から説き起し, Medical Decision Making に適した M. D. A. の手法を網羅的に解説し, 更に実際の分析例を付け加えた意欲的な講演であった。

彼は殆んど総ての手法を尽くしていると呼び, 次の様な分類を行った。

★ DESCRIPTIVE METHODS

FACTORIAL ANALYSIS

Classical principal component analysis

Correspondence analysis

Proximity analysis

Factorial discriminant analysis

CLASSIFICATION METHODS

Hierarchical classification

Dynamic cluster analysis

★ PREDICATIVE AND DECISIONAL METHODS

y : response variable, x_1, \dots, x_{p-1} : explicative variable.

y, x_1, \dots, x_{p-1} ; continuous.

Linear multiple regression.

Regression analysis using a nearest neighbours procedure.

y ; ordinal discrete or continuous, x_1, \dots, x_{p-1} ; mixed variable.

First correspondence analysis

Second correspondence analysis.

y ; discrete (ordinal or nominal), x_1, \dots, x_{p-1} ; continuous.

Linear discriminant analysis.

Bayesian discriminant analysis.

- Quadratic discriminant analysis.
- Sebestyen allocation rule.
- Modified Sebestyen allocation rule.
- Discriminant analysis using a nearest neighbours procedure.
- y ; discrete (ordinal or nominal), x_1, \dots, x_{p-1} ; binary variable
- Bayesian discriminant analysis
- Binary discriminant analysis
- Nearest neighbours procedure

総体的に見て、Dataの範囲内で最適処理を目論むという記述的立場を高くして居らず、標本から母集団についての情報を抽出するという推測統計的立場はあまり見られない。

才 = 才三日の Current Work について、主題とその中で使われている論理的方法を略記しておく。

才 = 日前

- I-1 Three Example of Computer-Aided Medical Decision
 1. graph の最短 path を見出す OR 手法.
 2. Bayes Theorem に基づく tree.
- I-2 Computer-Aided Diagnosis; A practical proposition?
- I-3 Some Practical Problems in Transferring C.A.D. Systems from One Geographical Area to Another.
- I-4 C.A.D. of the Acute Abdomen: A system from Leeds used on Copenhagen patients.
 1. Bayes formula
- I-5 C.A.D. of Pancreatic Function Test in the Routine.
 1. Non-linear discriminant analysis.
 2. Parametric discriminant analysis for qualitative data.
 3. Linear discriminant analysis.
- I-6 Computer Diagnosis of Congenital Heart Disease by means of V.C.G. using the Likelihood Method.
 1. 尤度法, Shannon entropy
- I-7 Experience in Computer Classification of EEG'S
 1. Linear discriminant Analysis
 2. Euclidean Classifiers a) Weighted b) Unweighted.
 3. Minkowski - Manhattan Classifier a) Weighted b) Unweighted
 4. Likelihood.
- I-8 Cluster Analysis of Systemic Lupus Erythematosus - A case study of disease entities -
 1. Cluster analysis

才 = 日午後

- I-9 Decision-Making Aids used to determine the Content of Medical Teaching.
 1. Bayes process.
- I-10 Estimation of the Utilities of States of Health with different Visual Acutities using a Wagering Technique.
 1. Utility Function
- I-11 Review of Automated Decision Support at the Medical System Hannover.
 1. Bayesian model.
 2. Boolean logic.
- II-1 Processing of Chest X-Ray Images by Computer
- II-2 Effective and Cost-Effective Real-Time Picture Operators

II-3 Decision Theoretic Approach to White Cell Differentiation. For Medical Imagery.

1. Linear discriminant analysis

2. Cluster analysis.

才三日午前は、Public Health においてこの Current Work が四演者で報告された後、教理的問題について次の総括的な講演があった。

1. C.A.D. - An overview of some Theoretical Problems; M.J.R. Healy (U.K.)

C.A.D.への主たる approach として、yes-no decision, discriminant approach, Bayes theorem の三つを取り上げた。才一についていまだ実用性の存在する事を主張した。才二について allocation rule の performance の評価が困難であるが、それについては Lichenbruch の貢献について触れた。allocation rule が多種開発工中であるが、その複雑さに此して成果が期待出来ぬかの疑問を表明した。才三について医療 Data の不安定性の影響、変量間の独立性の不成立の影響について論じ、影響の少ない場合のある事も指摘した。最後に教理側の一層の理論展開を期待して講演を終えた。

才四日前は、高散変量による判別問題についての理論が討論された。

1. C.A.D. and the Atypical Case; J. Hilden (Copenhagen)

稀病患者乃至稀かな症状を示す患者を、C.A.D. に不適であるという理由で抽出する事を考へた。Warning index と称する型はずれ (atypical) の測度を生起確率から構成し、3.0 以上の値を取る患者は C.A.D. の適用外と判断する。

2. The Importance of Symptom Non-Independence in Diagnosis; J.A. Jaquez. (Michigan)

Symptoms の間に不当に独立性の仮定して、情報の損失を招く事は望ましくない事ではない。そこで Bayes decision を行うための確率計算に於いて、高次の相関 (= 次のオーダー) を考慮した。この際すべての項目の組合せについてではなく、cluster analysis により高次の相関のある項目の組合せを見付け出して、Bahadur の應用を利用した。

なお Hilden は、non-independence の補正を、independence の仮定の基で計算した確率を、1 より小さいある正数による累乗によって行っている。

3. Diagnosis on Qualitative Variables; C. Huber (Paris)

高散型の変量を互いに独立な group (= これを syndrome と称する) に分け、syndrome 内では高次相関を考慮して確率分布を推定し、それの積より全変量の同時確率分布を推定し、classification rule を構成する。

Syndrome の検出には χ^2 検定を使う。

4. Models for Diagnosis and Detection of Combinations of Diseases; J.D.F. Habema

複数の病気を持った患者を判別する手法を考へた。病気の (Netherland) すべての組合せを一つの群と見なしての多群判別は、推定すべき parameter 数が膨大になり過ぎ現実的でない。しかしある疾病にある symptom が存在しない確率が、他の疾病の有無と独立であるという仮定の基では大なる parameter 数の reduction が可能である。もう一つの方法として、従来通りの exactly-one-disease model の基で、複数の病気を持った患者を例外として検出する枠組の構成が考へられる。各疾病について出現し難い symptom の組合せを surprising observation として、ある symptom の組合せの事後確率の最大値と次に大きい値との比がある一定値以下の時を a case of doubt として、detection of combinations of diseases の指針とする。

5. Pseudoquestionnaires and Information ; M. Terrenoire (Lyon)

Decision tree の stopping rule を三種提案している。又 stop しない次の question を受けるに際して、情報量最大の question を選択する事を経理とし、その情報量として三種提案している。

以下は、= の例外はあるが大略連続変量の判別問題が扱われた。

1. Some Unsolved Practical Problems in Discriminant Analysis ; P.A. Lachenbruch (Caroline)

判別手法を (1) Linear discriminant function, (2) Quadratic dis. func. (3) Multiple group dis. func. (4) Eigen vector or Canonical vector (5) Density estimation techniques. の5通り分類し、問題点を (1) Unequal covariance matrices, (2) Non-Normality a) Dichotomous b) Continuous, (3) Contamination a) Scale b) Location, 4) Variable selection, 5) Initial Misclassification と指摘し、現在迄の成果を次表の様々整理した。

	(1)	(2) a) b)	(3) a) b)	(4)	(5)	
(1)	1	1 1	1 2	1.2	1	1: satisfactorily solved.
(2)	3	2 2	2 2	2.1	2	2: no conclusive results.
(3)	2	2 2	2 2	2.1	2	3: an appropriate solution for the problem indicated
(4)	2	2 2	2 2	4	2	4: essentially solved if the canonical vectors are ordered by the magnitude of the eigen-values.
(5)	3	2 3	2 2	2	2	

夕四日午後

2. Error Rate of Linear Discriminant Function ; A. Miyake (Tokyo)

非心F分布の Patnaik 近似を利用した Error Rate の信頼区間の導出法を提案した。

3. Tactical Aspects in the use of L.D.A. ; L. Dusserre (Dijon)

線型判別関数に於ける変数選択問題、標本数と変数数の関係、testing sample の問題、線型判別関数の前提となる仮説の検定の問題、分析結果の解釈の問題を総括的に論じた。

4. Dimension Reduction and Piecewise L.D.A. ; S. Blomer (München)

Pattern Recognition Procedure の Dimension Reduction と Classification Rule との二つの手法の有機的結合を構成せしめる事を主張し、前者として Sammon による Feature Selector を、後者として Chang による Piecewise Linear Classifier を利用する事を提案し、原則を示した。

5. Time Dependence and Bayesian Approach ; H.L. Christl (Germany)

一変量の時系列 Data に回帰構造を想定し、その回帰係数による判別分析を試みた。回帰係数が群毎に一定の場合と、回帰係数自身が確率変数で個人により変化する場合が考えられた。更に多変量への拡張が試みられた。

6. Optimal Training Sets ; S.J. Pöppel (München)

正規性の仮定の基で、線型乃至二次判別関数の error rate を、sample 数、変数数、Mahalanobis 距離等を考慮して、training sample 又は test sample による simulation 実験により算出し整理した精力的な報告である。

7. Hold-One-Out Rules ; R.J. Blomer (München)

正規性の仮定の基での判別手法について、Hold-One-Out Rules の計算 algorithm を求めた。

8. Non-Parametric Allocation Rules ; N. Victor (Germany)

先ず、kernel 法による密度関数を推定し、= による Allocation Rule を考えた。kernel としては rectangular, triangular, normal distribution 等

考えた。更に、標本数に応じて又は標本の散布の濃度に応じて concentration を高めて行く方法を考えている。次に、密度関数を直交関数族で表現する方法、許容限界を利用する方法等を論じている。

9. Ternary Algebra. R. Salamon (Paris)
Present, Absent に関する三値論理である Bool 代数に対して, Present, Absent, Don't Know に関する三値論理である代数を展開している。

廿五日午前

1. A Statistical Method for Quantification of Categorical Data and Its Application to Medical Science. T. Komajawa (Tokyo)
数量化理論のⅡ, Ⅲ, Ⅳの解説と実例。
2. Descriptive Methods in Multidimensional Data Analysis; A. Schroder (France)
Descriptive Methods に関する Nakache の報告の内容を少し省略し、幾何学的表現における種々の距離について、Factorial Methods の Principal component analysis, Correspondence analysis について、更に tree, cluster analysis について詳しく論じている。
3. Multivariate Methods and The Problem of Time Dependence in Clinical Trials
Cyclophosphamide (制癌剤) の肝臓と血中に関する Compartment Analysis を試みた。Time function について、処置群毎に一定の場合と、個人によって変動する場合とを考えている。又 long term studies における他の要因の影響の除去、例えば censored sample 等について論じている。
4. Recommendation for Future Work; Working Conference Committee.
Committee が作成した原案を基に参加者の討議が行われた。Pattern Recognition 及び Image analysis の研究継続が主張された。robust methods の必要性が強調された。

以上の講演以外に Hilden は次の採り内容の note を配布し注意を引いた。判別に関する結論は、各群への尤度乃至事後確率の vector で表現されるべきである。plug-in rules は optimal な allocation rule とは限らぬ。判別関数の種々の評価方法の優劣を論じている。

この conference を通して見出される傾向を私なりにまとめてみる。まず离散型多変量の non-independence の問題が多く議論された。Jacquez, Huber は多変量を互いに独立な group に分け、推定すべき parameter の数の節減を試み、Hilden は独立仮定で計算した確率を累乗により補正する事を試みた。次に判別分析に於いて、分析の過程に現われる尤度、事後確率から直接に情報を汲み取りようとする行き方が目立った。Habbema の surprising observation, a case of doubt, Hilden の warning index の概念、更に Hilden の如く判別の結論の尤度の確率の vector で与えるべきという主張が有力である。連続変量の理論のうご過半がなくなる形の線形判別関数を扱っているのが目についたし、White Cell の differentiation のように実績をあげている current work も存在した。時系列の問題が、多少歪曲された形ではあるが二つ演題に登場したが、これは今後重要な主題となる前触れの様にも思われた。利は、Victor の non-parametric な density estimation に興味を持ち、判別分析の究極の案を示す重要な主題と思われる。