

生物医学文献データベースを利用する医療データマイニング

トアンナム・チャン 沼尾 正行

東京工業大学 情報理工学研究所 計算工学専攻

tt-nam@nm.cs.titech.ac.jp numao@cs.titech.ac.jp

概要

アクティブマイニング技術において、情報収集とマイニング過程は共に重要であるが、今までの研究では殆ど独立に行われている。従来のデータマイニング技術は、データだけに注目し、属性と関係のある文献情報を考慮していないのが一般的である。我々は、属性に関係する文献検索結果を属性の外部重みとして導入することにより、従来のデータマイニングでは発見できないまたは発見しにくいパターンを抽出することを目的としている。実験では、髄膜脳炎データを用いて決定木学習の C4.5 手法と提案手法による予備実験を行い、提案手法の妥当性を検討する。

Mining medical data with the assistance of biomedical literature sources

TuanNam Tran Masayuki Numao

Graduate School of Information Science and Engineering,

Department of Computer Science, Tokyo Institute of Technology

tt-nam@nm.cs.titech.ac.jp numao@cs.titech.ac.jp

Abstract

Both information gathering and mining process play an important role in active mining technology, however they are often been conducted independently so far. The data mining techniques so far tend to focus on the given data itself, not considering the literature information pertaining to the attributes. We propose a new approach that makes use of the literature search results related to the attributes as the attributes' external weights. Preliminary experiments are conducted on C4.5, a standard decision-tree learning method and our proposed method, and the validity of the proposed method is discussed.

1 はじめに

データマイニング [10] は、巨大なデータベースから、専門家の持っている知識とは異なる新たな知識を抽出することが目的である。大量のデータが容易に利用できることにより、様々なデータマイニングシステムが開発された。しかし、データマイニングツールをそのまま適用するだけでは、当り前の知識しか抽出できなかつたり、興味深い知識を発見す

るまでに多大なコストを費している。そこで、新しいマイニングの枠組「アクティブマイニング技術」が要求されるようになった [2]。従来のデータマイニング技術と異なり、「アクティブマイニング技術」は情報収集とユーザー側からのシステムへの迅速なフィードバックを重視しているが、実際に情報収集技術と従来のデータマイニング手法を組み合わせる研究は殆どない。一方、データマイニング過程において、共通データとして医療データがよく用い

られる [5] [3]. 実際には、共通データを用いて様々な知識発見手法による生成された知識の比較と評価が行われた [9] [4] [8] [1] [6].

さらに、医学において、「Statistical Evidence」に基づいた EBM(Evidence-based medicine) と呼ばれる概念が注目されるようになった [7]. EBM も文献の検索およびデータ収集に密接に関係があるため、本研究は、文献検索と従来のデータマイニング技術を組み合わせた新しい手法を提案する。

2 医療データからのマイニング

2.1 問題点

前説に述べたように、現在医療データのマイニングをするにあたって、多くのデータマイニングシステムが利用できるが、殆どのシステムは支持度 (Support) または正解率に基づいてパターンを抽出している。[6] に指摘されたように、一般的には支持度 (Support) の高いルールは専門家にとって妥当であるが、新しい知識ではない。逆に、発見に貢献しそうなルールは、いわゆる妥当なルールの「辺縁」にでも存在するかのようである。効率的に専門家にとって新しい知識を抽出するには、与えられたデータだけでなく、専門家の知識および既存の文献データベース情報などが非常に役に立つと考えられる。データマイニング過程の流れとして、システムにより得られたパターンを専門家に評価してもらってからシステムにフィードバックするのが一般的である。マイニングをする前に専門家の知識をシステムに与えればより効果的なマイニングが実現できる、実際には手間やコストがかかることから、簡単には実現できないのが現状である。

2.2 提案手法

上記の理由により、本研究は有名なオンライン生物医学データベースである MEDLINE から情報収集・検索および従来のデータマイニング技術を組み合わせ、支持度がやや高くなくても専門家にとって新しいパターンを発見するのを目的とする。具体的

には、よく知られている決定木学習システム C4.5 のゲインを計算する際に属性の「外部の重み」という概念を導入する。つまり、提案手法は、内部のデータからの計算だけでなく、文献検索による属性の重要性を反映させることが重要だと考えている。

2.3 C4.5 のゲインの修正

属性 A_1, A_2, \dots, A_m からなる決定木学習において、 T をある訓練例の集合とする。属性 A_j に対応するテスト X によって、 T は T_1, T_2, \dots, T_n に分割されるとする。[11] より、

$$info_X(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} \times info(T_i) \quad (1)$$

$$gain(X) = info(T) - info_X(T) \quad (2)$$

$$split\ info(X) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \times \log_2\left(\frac{|T_i|}{|T|}\right) \quad (3)$$

$$gain\ ratio(X) = gain(X) / split\ info(X) \quad (4)$$

属性 A_j の文献検索情報を用いて、 A_j の「外部の重み」 $\omega(A_j)$ ($0 \leq \omega(A_j) \leq 1$) を定義し、 $gain$ 、 $gain\ ratio$ をそれぞれ下記の $gain'$ 、 $gain\ ratio'$ に修正する。

$$gain'(X) = gain(X) \times \omega(A_j) \quad (5)$$

$$gain\ ratio'(X) = gain\ ratio(X) \times \omega(A_j) \quad (6)$$

$$\omega(A_j) = \frac{F(|A_j|)}{\sum_{i=1}^m F(|A_i|)} \quad (7)$$

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ (int)(\log_k(x) + 1) & \text{if } x > 0 \end{cases} \quad (8)$$

なお、 $k > 0$ は係数であり、 $|A_i|$ は与えられたデータと属性 A_i に関連する MEDLINE 文献数とする。

3 実験

38 属性からなる髄膜脳炎データベースの更新版¹を用いて、C4.5 と提案手法により CULTURE、

¹Proc. Int'l Workshop of KDD Challenge on Real-World Data (KDD Challenge), 2000, <http://www.slab.dnj.ynu.ac.jp/challenge2000>

CULT_FIND, C_COURSE, COURSE (Grouped), DIAG, DIAG2 といったそれぞれのクラスについてパターンを生成し、評価を行う。表 1 はこのデータベースの属性構成、属性解釈および MEDLINE 検索条件、見つけた文献数を記載している。各属性の解釈は [5] から引用するものである。実際に、文献を検索する際に、日本語から英語の正確な専門用語を対応付けするには、「WebLSD Online Life Science Dictionary」² を用いた。

付録 A はそれぞれの生成されたパターンを示している。例えば、CULTURE クラスに対して、提案手法による 3 番目のパターンは、

「もし EEG_FOCUS(脳波上異常所見) に異常がありかつ白血球数が 5200 以下ならば、実際に発見されるウィルスはヘルペスである」ということを意味し、このルールの左辺は全体のデータベースに 3 回現れており、そのうち 2 回は上記のルールを満たしている。

4 考察・今後の課題

まず我々は、提案手法は C4.5 より優れているという証明することを目的としているのではなく、提案手法を用いることにより、従来のデータマイニング技術で発見できない、または発見できるとしても多大なコストがかかる場合、もし提案手法により専門家にとって新しいパターンを発見できたり、マイニングコストを削減できたりするのが本来の目的である。前節に得られた実験結果は、あくまでも提案手法を評価するための初期結果にすぎなく、時間の都合により、生成されたパターンは専門家の評価を受けていないことをご了承していただきたい。

[6] に示されたように、DIAG, DIAG2 のクラスに対して、性別や年齢と髄膜炎との関連性が今までの医学研究においては指摘されていないため、生成されたパターンが性別や年齢に関する情報を含めば専門家にとって新しい知識になる可能性がありうる。実際に、DIAG クラスにおいて提案手法により生成されたルール 2, 4, 5, 6 はいずれも性別または年齢に関する情報を含んでいる。さらに、ルール 5, 6 は

²<http://lsd.lab.nig.ac.jp/WebLSD/>

AGE に関するルールペアを表し、AGE の重要さを意味している。DIAG2 クラスにおいては、提案手法では、THERAPY2 に関する情報が得られるのが面白い。また、[6] にも記述しているが、C_COURSE のクラスに対しては、LOC に関するルールが得られるのが予後に関して髄膜炎の早期治療が重要である。提案手法は、

$$\begin{aligned} IF \quad LOC > 6 \\ THEN \quad C_COURSE = dead \end{aligned}$$

というパターンを発見できた。LOC だけでなく、AGE (ルール 1, 4) または THERAPY2 (ルール 3) に関する知識も獲得できた。さらにルール 3, 4 は SEIZURE に関するルールペアの「拡張」と考えられる。また、COURSE(Grouped) クラスに関して、提案手法により生成されたルール 2, 3 が AGE のルールペアの「拡張」と考えられ、AGE の重要さを意味している。本実験では、係数 k は自然対数で設定されたが、 k を変えることにより、システムが容易に生成ルールを調整でき、ユーザ (専門家) にとって新しいルールを抽出することが可能である。しかし、提案手法は、既存文献検索情報をデータマイニングアルゴリズムに反映できたが、属性に関連する文献の検索はまだ粗く、属性の「外部の重み」をより工夫する方法を考えるべきである。

5 結論

本研究では、従来のデータマイニング手法と異なり、MEDLINE の文献検索を用いて属性の外部重みを導入することにより、従来のデータマイニングで発見できない、または発見しにくいパターンを発見することができた。現時点では、初期結果のみを記載しているが、今後専門家の評価・コメントをいただき研究を続けていきたい。

参考文献

- [1] 鈴木 英之進. 共通データからの仮説駆動型例外ルール発見. 人工知能学会誌, Vol. 15, No. 5, pp. 782–789, 2000.
- [2] 元田 浩. 情報洪水時代におけるアクティブマイニングの実現. <http://www.ar.sanken.osaka>

表 1: 髄膜脳炎データと MEDLINE 検索結果. S0 は meningoencephalitis という MeSH term と「Limits = Human」という条件で検索された文献集合 (3376 件)

属性名	日本語の意味	検索条件	見つけた文献数	備考
AGE	年齢	S0 AND AGE	1049	
SEX	性別	S0 AND SEX	74	M OR F
DIAG	元々の診断			CLASS
DIAG2	DIAG をグループ化したもの			CLASS
COLD	何日前から風邪様症状があったか.	S0 AND COLD	16	
HEADACHE	何日前から頭痛があったか.	S0 AND HEADACHE	111	
FEVER	何日前から熱があったか.	S0 AND FEVER	171	
NAUSEA	何日前から吐き気があったか.	S0 AND NAUSEA	4	
LOC	何日前から意識障害があったか.	S0 AND LOC	51	
SEIZURE	何日前から痙攣・てんかんがあったか.	S0 AND SEIZURE	79	
ONSET	発症様式	S0 AND ONSET	129	ACUTE(急性) SUBACUTE(亜急性) CHRONIC(慢性) RECURR(反復性)
BT	来院時体温	S0 AND TEMPERATURE	17	
STIFF	来院時頂部硬直	S0 AND STIFF	4	
KERNIG	来院時 Kernig 徴候	S0 AND KERNIG	0	
LASEGUE	来院時 Lasegue 徴候	S0 AND LASEGUE	0	
GCS	来院時 Glasgow Coma Scale	S0 AND GCS	2	
LOC_DAT	意識障害の状態 (-:なし, +:あり)	S0 AND LOC	51	LOC と同様
FOCAL	巣症状	S0 AND FOCAL (-:なし, +:あり)	85	
WBC	白血球数	S0 AND WBC	153	
CRP	炎症性蛋白	S0 AND CRP	12	
ESR	血沈	S0 AND ESR	3	
CT_FIND	CT 所見	S0 AND CT AND (NORMAL OR ABNORMAL)	29	NORMAL OR ABNORMAL
EEG_WAVE	脳波所見	S0 AND EEG	154	
EEG_FOCUS	脳波の局所異常の有無	S0 AND EEG	154	
CSF_CELL	髄液細胞数	S0 AND spinal fluid AND CELL	63	
Cell_Poly	髄液中多核球数	S0 AND spinal fluid AND polymorphonuclear leukocyte	0	
Cell_Mono	髄液中単核球数	S0 AND spinal fluid AND mononuclear leukocyte	29	
CSF_PRO	髄液蛋白	S0 AND spinal fluid AND protein	53	
CSF_GLU	髄液ブドウ糖	S0 AND spinal fluid AND glucose	11	
CULT_FIND	培養・免疫検査で, 原因菌・ウイルスが発見できたかどうか.			CLASS
CULTURE	実際に発見できた菌・ウイルス名			CLASS
THERAPY2	実際の治療	S0 AND Therapy AND (ABPC OR CEX OR CTX OR CZX OR FMOX OR LMOX OR ARA OR Dara OR AMK OR INH OR RFP OR PCG OR PIPC OR ZOBIRAX OR GLOBULIN)	59	
CSF_CELL3	3 日後の髄液細胞数	S0 AND spinal fluid AND CELL	63	CSF_CELL と同様
CSF_CELL7	7 日後の髄液細胞数	S0 AND spinal fluid AND CELL	63	CSF_CELL と同様
C_COURSE	最終的な転帰			CLASS
COURSE(Grouped)	C_COURSE をグループ化したもの			CLASS
RISK	危険因子	S0 AND Risk AND (DM OR LC OR TB OR BECHET OR BRONCHO OR HEPATITIS OR MYELOMA OR SINUSITIS OR NO)	6	
RISK(Grouped)	RISK をグループ化したもの	S0 AND Risk AND (DM OR LC OR TB OR BECHET OR BRONCHO OR HEPATITIS OR MYELOMA OR SINUSITIS OR NO)	6	RISK と同様

u.ac.jp/activemining/, 2001.

- [3] 元田 浩. 総括班 活動報告. 情報洪水時代におけるアクティブマイニングの実現、研究成果報告書, 2002.
- [4] 山口 高平. 帰納アプリケーション自動構築環境に基づく共通データからの知識発見. 人工知能学会誌, Vol. 15, No. 5, pp. 768-773, 2000.
- [5] 津本 周作. 知識発見手法の比較と評価のための共通データ. 人工知能学会誌, Vol. 15, No. 5, pp. 751-758, 2000.
- [6] 津本 周作, 高林 克日己. 知識発見手法により生成された知識の比較と評価. 人工知能学会誌, Vol. 15, No. 5, pp. 790-797, 2000.
- [7] 津本 周作, 平野 章二, 高林 克日己, 柳樂 真佐美. アクティブマイニングと ebm. 情報洪水時代におけるアクティブマイニングの実現、研究成果報告書, 2002.
- [8] 鍾 寧. 確率的ラフ帰納学習システム gdt-rs による共通データからの知識発見. 人工知能学会誌, Vol. 15, No. 5, pp. 774-781, 2000.
- [9] 鷺尾 隆, 元田 浩. 構造データ及び数値データに対する相関ルールマイニングの拡張. 人工知能学会誌, Vol. 15, No. 5, pp. 759-767, 2000.
- [10] U. M. Fayyad. *Advances in Knowledge Discovery and Data Mining*. AAAI Press, 1996.
- [11] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

A C4.5 と提案手法によって得られたルール

各クラス CULTURE, CULT_FIND, C_COURSE, COURSE(Grouped), DIAG, DIAG2 に関して C4.5 と提案手法により生成されたパターンを示す. 実験では, 係数 k は自然対数 e とする.

* CULTURE

- C4.5

1. IF (COLD > 8) AND (NAUSEA > 5)
THEN (CULTURE = pi(B)) [2/0]
2. IF (SEX = F) AND (AGE > 29) AND (BT > 39.8)
THEN (CULTURE = varicella) [2/0]
3. IF (COLD ≤ 8) AND (ONSET = ACUTE) AND (BT ≤ 39.8) AND (GCS > 11) AND (CRP ≤ 11.4) AND (CSF_CELL > 12) AND (CSF_CELL7 ≤ 162)
THEN (CULTURE = -) [62/1]
4. IF (FOCAL = +) AND (ESR > 31)
THEN (CULTURE = neisseria) [3/1]
5. IF (CRP > 11.4)
THEN (CULTURE = strepto) [4/1]

- 提案手法

1. IF (ONSET = ACUTE) AND (AGE > 17) AND (COLD ≤ 8) AND (FEVER ≤ 24) AND (LOC ≤ 2) AND (CRP ≤ 15.9) AND (CSF_CELL > 12) AND (CSF_CELL7 ≤ 124)
THEN (CULTURE = -) [57/0]
2. IF (THERAPY2 = ope)
THEN (CULTURE = strepto) [2/0]
3. IF (EEG_FOCUS = +) AND (WBC ≤ 5200)
THEN (CULTURE = herpes) [3/1]

* CULT_FIND

- C4.5

1. IF (CRP ≤ 0.5) AND (CSF_PRO > 47) AND (CSF_GLU > 64)
THEN (CULT_FIND = T) [8/0]
2. IF (CRP > 7) AND (Cell_Poly ≤ 1220)
THEN (CULT_FIND = T) [4/0]
3. IF (COLD ≤ 8) AND (CRP > 0.5) AND (CRP ≤ 7)
THEN (CULT_FIND = F) [42/0]
4. IF (THERAPY2 = no_therapy) AND (NAUSEA ≤ 9) AND (ESR ≤ 11) AND (CSF_GLU ≤ 64)
THEN (CULT_FIND = F) [38/1]

- 提案手法

1. IF (CRP ≤ 7) AND (CSF_CELL > 12) AND (CSF_CELL ≤ 268)
THEN (CULT_FIND = F) [51/1]
2. IF (WBC > 5900) AND (CSF_CELL > 373) AND (CSF_GLU > 24) AND (CSF_GLU ≤ 64) AND (CSF_CELL7 > 162)
THEN (CULT_FIND = F) [22/0]
3. IF (ONSET = ACUTE) AND (COLD ≤ 8) AND (HEADACHE ≤ 15) AND (CRP ≤ 7) AND (CSF_CELL > 268) AND (CSF_CELL7 ≤ 162)
THEN (CULT_FIND = F) [27/1]
4. IF (ONSET = ACUTE) AND (CSF_GLU > 64) AND (CSF_CELL7 > 162)
THEN (CULT_FIND = T) [7/0]
5. IF (COLD > 8) AND (CSF_CELL > 268)
THEN (CULT_FIND = T) [6/1]
6. IF (CRP > 7)
THEN (CULT_FIND = T) [6/1]

* C_COURSE

- C4.5

1. IF (LOC_DAT = +) AND (BT ≤ 36) AND (Cell_Mono > 260) AND (CSF_GLU ≤ 72)
THEN (C_COURSE = aphasia) [3/0]
2. IF (BT ≤ 38.9) AND (CSF_GLU ≤ 17)
THEN (C_COURSE = dead) [5/0]
3. IF (COLD ≤ 10) AND (FEVER ≤ 24) AND (BT > 36) AND (CSF_GLU > 17)
THEN (C_COURSE = negative) [91/5]

- 提案手法

1. IF (AGE > 62) AND (WBC > 13900)
THEN (C_COURSE = dead) [3/0]
2. IF (LOC > 6)
THEN (C_COURSE = dead) [1/0]
3. IF (THERAPY2 = ARA_A) AND (SEIZURE > 1)
THEN (C_COURSE = aphasia) [3/1]

4. IF (AGE ≤ 62) AND (HEADACHE ≤ 21) AND (SEIZURE ≤ 1) AND (CSF_PRO ≤ 405)
THEN (COURSE = negative) [119/11]
- * COURSE(Grouped)**
- C4.5
1. IF (AGE > 21) AND (FEVER ≤ 8) AND (AGE ≤ 62) AND (FOCAL = -)
THEN (COURSE(Grouped) = n) [66/0]
 2. IF (LOC > 2) AND (CSF_GLU ≤ 61)
THEN (COURSE(Grouped) = p) [5/0]
 3. IF (COLD > 9) AND (FOCAL = +)
THEN (COURSE(Grouped) = p) [4/0]
 4. IF (AGE > 62) AND (STIFF > 2)
THEN (COURSE(Grouped) = p) [3/0]
 5. IF (BT > 39.6) AND (FOCAL = +)
THEN (COURSE(Grouped) = p) [2/0]
- 提案手法
1. IF (AGE > 23) AND (AGE ≤ 62) AND (FOCAL = +) AND (CSF_CELL > 39) AND (CSF_CELL ≤ 938)
THEN (COURSE(Grouped) = p) [11/1]
 2. IF (AGE > 62) AND (WBC > 13900)
THEN (COURSE(Grouped) = p) [3/0]
 3. IF (AGE > 21) AND (AGE ≤ 62) AND (HEADACHE ≤ 20) AND (FOCAL = -)
THEN (COURSE(Grouped) = n) [81/2]
- * DIAG**
- C4.5
1. IF (EEG_FOCUS = +) AND (CSF_CELL ≤ 1474) AND (CSF_PRO > 73)
THEN (DIAG = VIRUS(E)) [13/0]
 2. IF (LOC > 1) AND (Cell_Poly ≤ 221) AND (Cell_Mono > 3)
THEN (DIAG = VIRUS(E)) [3/0]
 3. IF (SEIZURE > 0) AND (Cell_Mono > 12)
THEN (DIAG = VIRUS(E)) [3/0]
 4. IF (LOC ≤ 0) AND (LOC_DAT = +) AND (EEG_FOCUS = -) AND (Cell_Poly ≤ 221) AND (Cell_Mono > 3)
THEN (DIAG = VIRUS(E)) [4/0]
 5. IF (CSF_CELL ≤ 29) AND (Cell_Mono > 12)
THEN (DIAG = VIRUS(E)) [3/0]
 6. IF (AGE ≤ 15) AND (Cell_Poly ≤ 221)
THEN (DIAG = VIRUS(E)) [2/0]
 7. IF (CT_FIND = abnormal) AND (Cell_Mono ≤ 3)
THEN (DIAG = ABSCESS) [7/1]
 8. IF (LOC_DAT = -) AND (EEG_FOCUS = -) AND (AGE > 15) AND (CRP ≤ 5.8) AND (CSF_CELL ≤ 29) AND (Cell_Poly ≤ 221)
THEN (DIAG = VIRUS) [59/1]
 9. IF (Cell_Poly > 221)
THEN (DIAG = BACTERIA) [32/10]
- 提案手法
1. IF (LOC_DAT = -) AND (HEADACHE ≤ 16) AND (EEG_FOCUS = -) AND (AGE > 15) AND (WBC ≤ 13600) AND (CRP ≤ 4.7) AND (CSF_CELL > 29) AND (CSF_CELL ≤ 1900)
THEN (DIAG = VIRUS) [55/0]
2. IF (SEX = F) AND (FEVER ≤ 15) AND (EEG_FOCUS = +) AND (CSF_CELL ≤ 2860)
THEN (DIAG = VIRUS(E)) [15/0]
3. IF (LOC ≤ 0) AND (LOC_DAT = +) AND (HEADACHE ≤ 16) AND (FEVER ≤ 24) AND (WBC ≤ 13600) AND (CSF_CELL ≤ 2220)
THEN (DIAG = VIRUS(E)) [6/0]
4. IF (AGE ≤ 30) AND (FEVER > 24)
THEN (DIAG = BACTE(E)) [2/0]
5. IF (AGE > 57) AND (CSF_CELL > 2860)
THEN (DIAG = BACTE(E)) [4/1]
6. IF (AGE ≤ 57) AND (CSF_CELL > 2860)
THEN (DIAG = BACTERIA) [11/1]
7. IF (WBC ≤ 13600) AND (CRP > 4.7) AND (CSF_CELL ≤ 2860)
THEN (DIAG = BACTERIA) [4/0]
8. IF (FEVER ≤ 24) AND (Cell_Mono ≤ 3)
THEN (DIAG = ABSCESS) [8/2]
- * DIAG2**
- C4.5
1. IF (Cell_Poly > 220)
THEN (DIAG2 = BACTERIA) [33/0]
 2. IF (CT_FIND = abnormal) AND (Cell_Mono ≤ 12)
THEN (DIAG2 = BACTERIA) [8/0]
 3. IF (Cell_Poly ≤ 220) AND (Cell_Mono > 12)
THEN (DIAG2 = VIRUS) [96/1]
 4. IF (CT_FIND = normal) AND (Cell_Poly ≤ 220)
THEN (DIAG2 = VIRUS) [3/0]
- 提案手法
1. IF (THERAPY2 = no_therapy) AND (WBC ≤ 13600)
THEN (DIAG2 = VIRUS) [55/0]
 2. IF (THERAPY2 = Zobirax)
THEN (DIAG2 = VIRUS) [25/0]
 3. IF (THERAPY2 = ARA_A)
THEN (DIAG2 = VIRUS) [11/0]
 4. IF (SEX = M) AND (THERAPY2 = ABPC)
THEN (DIAG2 = BACTERIA) [20/0]
 5. IF (CSF_CELL > 2860)
THEN (DIAG2 = BACTERIA) [10/0]
 6. IF (AGE > 24) AND (WBC > 13600)
THEN (DIAG2 = BACTERIA) [3/0]
 7. IF (THERAPY2 = multiple)
THEN (DIAG2 = BACTERIA) [4/1]
 8. IF (THERAPY2 = ope)
THEN (DIAG2 = BACTERIA) [1/0]