

平成 30 年 6 月 5 日現在

機関番号：17104

研究種目：挑戦的萌芽研究

研究期間：2015～2017

課題番号：15K12102

研究課題名(和文) 予兆探索と薬剤感受性解析の融合によるアンチバイオグラムからのパターン発見

研究課題名(英文) Pattern discovery from antibiograms incorporated propagation with drug susceptibility

研究代表者

平田 耕一 (Hirata, Kouichi)

九州工業大学・大学院情報工学研究院・教授

研究者番号：20274558

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、菌の薬剤感受性検査結果であるアンチバイオグラムからの予兆探索として、菌株の伝搬パターンを抽出する手法を開発した。そして、伝搬パターンにおいて異患者・同一病室・同一主治医、異患者・同一病室・異主治医、異患者・異病室・同一主治医の3通りとなる感染伝搬パターンが、他の菌と比較して肺炎球菌において割合が高かった。また、アンチバイオグラムからの薬剤感受性解析として、相互依存重集合を抽出する手法を開発した。その結果、医学的によく知られている、黄色ブドウ球菌の出現後に緑膿菌が出現する菌交代を裏付ける相互依存重集合を抽出することができた。

研究成果の概要(英文)：In this research, as discovery of propagation from antibiograms, we have designed the algorithm to extract propagation patterns. Then, we have observed that *Streptococcus pneumoniae* have the large ratio of the infectious propagation patterns, in which two patients are not identical, and they are in the same room and/or treated by the same physician. Also, as pattern discovery based on drug susceptibility for antibiograms, we have designed the algorithm to extract mutually dependent multisets. Then, we have succeeded in extracting the mutually dependent multisets suggesting microbial substitution such as the occurrence of *Pseudomonas aeruginosa* after the occurrence of *Staphylococcus aureus*.

研究分野：知能情報学

キーワード：パターン発見 アンチバイオグラム 伝搬パターン 相互依存重集合

1. 研究開始当初の背景

(1) 現在、医療施設における院内感染は、重大な社会問題の一つとなっている。このような院内感染の対策として、医療施設において電子化されたデータを用いて、院内感染の予測と予防を行う研究が進められている。これまでに研究代表者は、研究協力者の協力の下、菌の薬剤感受性検査結果である**アンチバイオグラム**から、院内感染の予防に有用な規則を抽出する研究を進めてきた。ここで、アンチバイオグラムは、患者情報、位置情報、診療情報、材料情報、検出菌、薬剤感受性から成る。一方で、このような規則の抽出に際して、さまざまな問題点も生じた。

(2) まず、アンチバイオグラムの検査結果の頻度は、菌の発生頻度よりも医療従事者の検査依頼頻度を表すことが多い点である。また、時間の経過と共に検出菌が別の検出菌に変化する菌交代やある薬剤に対して感受性が耐性に変化する菌の薬剤耐性変化では、DNA 情報が異なる菌株は同一菌名でも異なる菌株として扱わなければならない。

(3) さらに、アンチバイオグラムは、その定義から、一つのレコードに、重複する日付、患者情報、検出菌などの同一アイテムを含むデータとなっている。すなわち、アンチバイオグラムからのパターン発見では、アイテム集合ではなく、アイテム重集合を対象とする必要がある。一方、通常のデータマイニング手法では、このようなアイテムの重複出現は許されていない。

2. 研究の目的

(1) これらの問題を解決するために、本研究では、**アンチバイオグラムからの予兆探索と薬剤感受性解析**により、アンチバイオグラムから院内感染の要因を**パターンとして発見する基盤**を構築することが目的である。なお、本研究では、1999年から2007年に蓄積された大阪府立急性期・総合医療センターのアンチバイオグラムを対象とする。このアンチバイオグラムは、患者ID、日付、検査材料、検出菌、および、108種類の薬剤に関する感受性検査結果からなる。さらに、これに患者ID、病室No、ベッドNo、医師ID(主治医)を追加したデータを融合したアンチバイオグラムを利用する。

(2) 予兆探索のパターン発見では、アンチバイオグラムにおける各薬剤に対するS(感受性)、I(中間)、R(耐性)というSIR値から、菌株を同定することにより、菌株の**伝播パターンを発見する**手法を開発する。

(3) 一方、薬剤感受性解析によるパターン発見では、アイテム集合ではなく**アイテム重集合**を対象としたパターン発見に取り組む。特に、相互依存アイテム集合を重集合に拡張

した**相互依存重集合の抽出アルゴリズム**を開発し、それをアンチバイオグラムに適用することでパターンを発見する。

3. 研究の方法

(1) 予兆探索のパターン発見では、アンチバイオグラムからの同一菌株の近似推定法である**2次元キャリアマップ(2DCM)**に着目する。2DCMの原理は、薬剤感受性検査において、ある薬剤に対してSとRを持つ2つのレコードは異なる菌株と推定することである。しかし、この2DCMの原理では、本研究で扱うアンチバイオグラムから複雑な同一菌株の推定を導く。そこで本研究では、**2DCMの原理を以下のように拡張する**。

- A) すべての検査結果がSか空欄となっているレコードは対象から外す。
- B) 一方の検査結果が空欄でもう一方の検査結果がS、I、Rのいずれかとなっている薬剤数が**閾値以上**(この閾値を**共有閾値**という)となっているレコードは異なる菌株と判断する。

そして、同一菌株のコードrとsに対して、特定の期間においてrの出現後にsが出現する**伝播パターン r s**を**全探索**によってすべて抽出する。

(2) 薬剤感受性解析によるパターン発見に対して、アンチバイオグラムからの相互依存重集合の抽出に取り組む。

アイテム集合Xは、任意のY、Z、Xに対してZが出現するときにYが出現する条件付き確率が与えられた閾値以上となるとき、**相互依存集合**という。この閾値を**最小依存度**という。この定義は、任意のx、Xに対してxが出現するときにXが出現する条件付き確率が与えられた閾値以上となる、と言い換えることもできる。そこで本研究では、まず、相互依存集合を**相互依存重集合**に拡張する。ここで重集合とは、直観的には同一要素の重複を許したアイテム集合であり、形式的には要素としてのアイテムの集合から自然数への写像と定義できる。相互依存重集合においても、その定義を相互依存集合のように言い換えることができる。

また、既存の相互依存集合の抽出アルゴリズムは、頻出アイテム集合マイニングアルゴリズムであるAPRIORIに基づいて設計されている。そこで本研究では、1回のデータの走査で頻出アイテム集合を抽出するような、APRIORIをより高速にしたアルゴリズムであるAPRIORITIDに基づいて相互依存重集合を抽出するアルゴリズムを設計する。この中で本研究では、**重集合の末尾出現**という概念を定式化し、それによって重集合の出現検査が容易になる。

4. 研究成果

(1) 表1は、伝播パターンr sにおいてrの後にsが出現する期間を5日、10日、15

日、20日、25日、30日としたときの、すべての伝搬パターンを抽出するための計算時間(ミリ秒)である。計算機環境は、CPU Core i7 3770 3.40GHz, RAM 8GB であり、共有閾値は6である。

表 1. 計算時間(ミリ秒)

5日	10日	15日	20日	25日	30日
7017	8082	8624	9262	10287	11337

いずれの場合も高速に抽出できた。また、計算時間は期間の増加に伴い、指数的ではなく線形的に増加している。

このような伝搬パターンを、患者・病室・主治医の異同、および、材料の観点から解析する。

患者・病室・主治医の異同のうち、異患者・同一病室・同一主治医、異患者・同一病室・異主治医、異患者・異病室・同一主治医の3通りが院内感染において共慮すべきパターンだと考え、これを**感染伝搬パターン**とよぶ。

表2, 3, 4は、期間ごとの黄色ブドウ球菌、緑膿菌、肺炎球菌における伝搬パターンの数と感染伝搬パターンの割合(%)である。ここで、共有閾値は6である。

表 2 黄色ブドウ球菌

5日	10日	15日	20日	25日	30日
29507	58147	85000	107153	133012	157930
2.9%	2.91	2.90	2.89	2.91	2.93

表 3. 緑膿菌

5日	10日	15日	20日	25日	30日
6315	12342	18203	22849	28232	33432
2.91	3.16	3.36	3.45	3.56	3.61

表 4. 肺炎球菌

5日	10日	15日	20日	25日	30日
357	651	747	1177	1457	1738
14.71	15.21	15.42	15.12	14.82	14.84

表2, 3, 4の菌以外でも、肺炎球菌を除く上位12位までの菌では感染伝搬パターンの割合は概ね1%から5%程度である。一方、**肺炎球菌の感染伝搬パターンの割合は約15%と非常に多くなった**。肺炎球菌の感染伝搬パターンの材料はほとんど鼻腔であったこと、一方で肺炎球菌は空気感染しないことなどから、感染伝搬パターンの割合が高い理由を調べることで院内感染の防止に繋がることが予想される。

材料によるパターンの解析において、表5, 6, 7は、共有閾値7に対して、期間ごとに黄色ブドウ球菌、緑膿菌、肺炎菌に対して抽出された伝搬パターンの数(上)、その中の検査材料が異なるパターンの割合(中)、

および、中分類として検査材料が異なるパターンの割合(%) (下)である。ここで、検査材料の中分類は、血液、髄液、髄液、血管内留置カテーテル、ペースメーカー、カテーテル、ドレーン、穿刺液、胸水、腹水、関節液、灌流液、喀痰、呼吸器粘液、胃液、十二指腸液、胆汁、腸液、大便、尿、膿汁、付着物、羊水、搾母乳、リンパ節、組織、臓器、鼻腔である。なお、より詳細に分析するために表1, 2, 3, 4の共有閾値6を表5, 6, 7では7に増やしているため、抽出される伝搬パターンの数が多くなっている。

表 5. 黄色ブドウ球菌

5日	10日	15日	20日	25日	30日
59014	116294	170000	214306	266024	315860
90.64	90.47	90.64	90.90	91.06	91.20
68.44	68.64	68.72	69.03	69.06	69.21

表 6. 緑膿菌

5日	10日	15日	20日	25日	30日
12630	24684	36406	45698	56464	66864
88.28	87.81	88.30	88.84	89.18	89.50
64.61	64.31	65.03	65.67	65.79	66.24

表 7. 肺炎球菌

5日	10日	15日	20日	25日	30日
734	1302	1894	2354	2914	3476
72.21	71.43	70.12	69.58	70.49	70.89
49.05	49.77	49.63	50.04	51.00	52.13

黄色ブドウ球菌は抽出パターン数が最多となる菌であり、緑膿菌は二番目である。また上位12位までの菌のうち肺炎球菌以外は、異なる検査材料の割合が80%以上、中分類としての異なる検査材料の割合が概ね50%以上である。一方、抽出パターン数が11位だった肺炎球菌では、異なる検査材料の割合が約70%、中分類として異なる検査材料の割合が約50%である。期間を増やすと抽出パターン数は増加するが、(中分類も含めて)異なる検査材料の割合はあまり変化しない。

以下は、30日において、中分類の検査材料が異なるパターン数(菌名の後に記載)、および、頻出上位3位のパターン(材料の変化を で表記)の頻度とその割合である。

- A) 黄色ブドウ球菌(218610 . 頻度 2000 以上のパターン数は 11) :
 - 喀痰 膿汁(17282, 7.91%)
 - 膿汁 喀痰(6359, 2.91%)
 - 喀痰 呼吸器粘液(4500, 2.06%) .
- B) 緑膿菌(44288 . 頻度 500 以上のパターン数は 9) :
 - 喀痰 膿汁(4144, 9.36%) ,
 - 喀痰 尿(1823, 4.12%) ,
 - 膿汁 喀痰(1674, 3.78%) .
- C) 肺炎球菌(1812 . 頻度 50 以上のパターン数は 6) :

呼吸器粘液 喀痰(294,16.23%) ,
 喀痰 呼吸器粘液(162,8.94%) ,
 鼻腔 喀痰(77,4.25%) .

(2) 相互依存重集合の抽出では、アンチバイオグラムのうち、場所データを融合する前のアンチバイオグラムに適用する。ここで、日付を ID とした**日データ**、および、患者 ID を ID とした**患者データ**の2種類のアンチバイオグラムを利用する。レコード数は、日データが3338、患者データが15596である。また、計算機環境は、CPU Xeon CPU E5-1650 v3 (3.5GHz), RAM 1024MB である。

表 8 と 9 は、日データおよび患者データから、最小依存度を 100% から 10% まで 10% ずつ減少させたときに抽出される相互依存重集合の数(上)、および、そのときの相互依存重集合における最大要素数(中)と計算時間(秒)である。

表 8 日データからの相互依存重集合

100%	90%	80%	70%	60%
463	466	471	488	533
4	4	4	5	6
3.9166	3.9045	3.9253	3.9241	3.9719

50%	40%	30%	20%	10%
638	935	1918	6910	68519
7	9	11	13	17
4.1261	4.2289	4.8282	7.3796	71.0553

表 9 患者データからの相互依存重集合

100%	90%	80%	70%	60%
467	467	467	467	472
5	5	5	5	5
7.4423	7.4423	7.4451	7.4455	7.2321

50%	40%	30%	20%	10%
497	509	590	708	1142
5	5	8	8	6
7.2598	7.2434	7.1902	7.2680	7.3376

日データでは最小依存度が 10% になると極端に計算時間が増えているが、それでも 2 分以内であり、それ以外の依存度では、どちらのデータも 10 秒以内には抽出が終了している。したがって、本研究の相互依存重集合の抽出では、今回設計したアルゴリズムは、対象とするアンチバイオグラムに対して有効であると言える。

次に、抽出された相互依存重集合のうち、院内感染の要因の一つと言われている**菌交代**を示唆する重集合について議論する。表 10 は 要素数が 7 となる抽出された相互依存重集合、表 11 は要素数が少なくとも 3 となる抽出された相互依存重集合、および、その重集合がデータに出現する割合である支

持度(%)である。なお菌名の左の数字は出現回数である。

表 11 日データからの相互依存重集合

(黄色ブドウ球菌)6(緑膿菌)1	56.68
(黄色ブドウ球菌)5(緑膿菌)1(大腸菌)1	54.76
(黄色ブドウ球菌)6(大腸菌)1	52.67
(黄色ブドウ球菌)5(緑膿菌)2	51.56
(黄色ブドウ球菌)7	51.47
(黄色ブドウ球菌)5(緑膿菌)1(大腸連鎖球菌)1	50.36
(黄色ブドウ球菌)5(緑膿菌)1(酵母菌)1	50.21

表 12 患者データからの相互依存重集合

(黄色ブドウ球菌)4	11.17
(緑膿菌)4	4.13
(黄色ブドウ球菌)3	15.88
(緑膿菌)3	5.60
(酵母菌)1(緑膿菌)2	4.08
(溶血菌群)1(ナイセリア菌群)1(ヘモフィルス菌群)1	1.90
(腸内細菌群)1(溶血菌群)1(ナイセリア菌群)1	1.53
(溶血菌群)1(ナイセリア菌群)1(非溶血菌群)1	1.42
(S. maltophilia)3	1.15
(セラチア)3	0.97
(Enterobacter aerogenes)3	0.56

表 11 より、日データから抽出された相互依存重集合は、5 回以上重複している黄色ブドウ球菌を含んでいる。また、「(黄色ブドウ球菌)6(大腸菌)1」と「(黄色ブドウ球菌)7」以外は、緑膿菌が 1 回か 2 回重複している。しかも、これらの相互依存重集合の支持度は 50% 以上であるため、高頻度に出現していることが分かる。

医学的な観点から、黄色ブドウ球菌の出現後に緑膿菌が出現する菌交代が知られている。表 11 の抽出結果は、その**菌交代を裏付ける相互依存重集合が抽出できている**ことを表している。

表 12 により、患者データから抽出された相互依存重集合のうち、要素数が 4 のものは「(黄色ブドウ球菌)4」と「(緑膿菌)4」だけである。一方、要素数 3 のものには、黄色ブドウ球菌や緑膿菌を含まない相互依存重集合も出現している。一方、表 12 の支持度は表 11 の支持度よりもかなり低い。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 2件)

Natsuki Kiyota, Sho Shimamura, Kouichi Hirata: *Extracting Mutually Dependent Multisets*, Proc. 20th International Conference on Discovery Science (DS 2017), Lecture Notes in Artificial Intelligence 10558, 267-280, 2017 (査読有). DOI: 10.1007/978-3-319-67786-6_19

Kazuki Nagayama, Kouichi Hirata, Shigeki Yokoyama, Kimiko Matsuoka: *Extracting Propagation Patterns from Bacterial Culture Data in Medical Facility*, New Frontiers in Artificial Intelligence, Lecture Notes in Artificial Intelligence 10091, 409-417, 2017 (査読有). DOI 10.1007/978-3-319-50953-2_28

〔学会発表〕(計 8件)

Sho Shomamura, Kouichi Hirata: *On the Applications of Iterative Consistency-Based Feature Selection to Nucleotide Sequences for Influenza A Viruses*, Proc. 2017 International Workshop on Smart Info-Media Systems in Asia (SISA 2017), 234-237, 2017.

Sho Shimamura, Kouichi Hirata: *Iterative Feature Selection Based on Binary Consistency*, Proc. 8th International Conference on E-Service and Knowledge Management (ESKM 2017), 396-399, 2017.

清田菜月, 平田耕一: *アンチバイオグラムからの相互依存重集合の抽出*, 第103回人工知能学会基本問題研究会, 人工知能学会研究会資料 SIG-FPAI-B506, 95-100, 2017.

Sho Shimamura, Kouichi Hirata: *On Temporal and Regional Analysis for Nucleotide Sequences of Influenza A (H1N1) Viruses Based on Feature Selection*. Proc. 2016 International Workshop on Smart Info-Media Systems in Asia (SISA 2016), 38-42, 2016.

嶋村翔, 平田耕一: *反復特徴選択を用いた塩基配列の解析*, 第102回人工知能学会基本問題研究会, 人工知能学会研究会資料 SIG-FPAI-B505, 1-4, 2016.

嶋村翔, 平田耕一: *特徴選択に基づくA型H1N1亜型インフルエンザウイルス塩基配列の時間性および地域性の解析*, 第101回人工知能学会基本問題研究会, 人工知能学会研究会資料 SIG-FPAI-B504, 68-72, 2016.

平田 耕一, 長山 和樹, 横山 茂樹, 松岡 喜美子: *伝搬パターンにおける検査材料の解*

析, 第20回日本医療情報学会春季学術大会, 2016.

Kazuki Nagayama, Kouichi Hirata, Shigeki Yokoyama, Kimiko Matsuoka: *On Extracting Propagation Patterns from Bacterial Culture Data in Medical Facility*, Proc. Workshop on Time Series Data Analysis and Its Applications (TSDAA2015), 2015.

〔その他〕
ホームページ等
<http://www.dumbo.ai.kyutech.ac.jp/hirata>

6. 研究組織

(1) 研究代表者

平田 耕一 (HIRATA, Kouichi)
九州工業大学・大学院情報工学研究院・教授
研究者番号: 20274558

(2) 研究分担者

(3) 連携研究者

(4) 研究協力者

横山 茂樹 (YOKOYAMA, Shigeki)
ケーディーアイコンズ(株)
松岡 喜美子 (MATSUOKA, Kimiko)
株式会社西日本健康科学