

# 救急医療トリアージノートを用いた症候群サーベイランス

中菅 章浩<sup>†</sup>      関 和広<sup>‡</sup>      上原 邦昭<sup>§</sup>

<sup>†</sup> 神戸大学工学部

<sup>‡</sup> 神戸大学自然科学系先端融合研究環

<sup>§</sup> 神戸大学大学院システム情報学研究科

a.nakasuga@ai.cs.kobe-u.ac.jp

## 1 はじめに

近年、重症急性呼吸器症候群（SARS）や鳥インフルエンザ、新型インフルエンザなどの急性感染症により世界中で多くの人々が死亡したり、高度な医療サービスを必要としたことは記憶に新しい。このような急性感染症の爆発的な感染の拡大を最小限に抑えるためには、感染初期における症状の正確な分類、すなわち症候群罹患判定によるバイオサーベイランス（症候群サーベイランス）が重要である [1, 2]。バイオサーベイランスでは、トリアージナースによって自然言語で記述された主訴やトリアージノートといった患者の初期症状、救急車の出動記録、学校の欠席者数、薬局の売上データなどを利用して、感染症の発生を検知する。

主訴やトリアージノートといったテキスト情報を利用した既存手法には、人手によって生成された規則を用いた症状分類の方法や、機械学習の枠組みを取り入れた分類の方法がある。規則ベースによる症候群の分類方法では、「w/o」（without）, 「w/」（with）, 「pt.」（patient）といった医療分野特有の省略表現に対して、個別に規則を追加することで比較的簡単に対応することができる。しかしながら、多種多様な症候群に対応する網羅的な規則を作成することは容易ではない。一方、機械学習の手法においては、信頼性の高い分類器を学習するために必要十分な訓練データをいかにして生成するかといった問題が存在する。さらに、両手法とも地域及び言語依存性が非常に強いため、ある一地域において有効な手法であったとしても、それを他地域で利用するためには、規則・訓練データの再構築が必要となる。

本研究は、既存手法の一つである規則に基づく手法と診断コードを活用して訓練データを自動的に生成し、これに機械学習を適用することで症候群への罹患判定を行う。具体的には、症状定義を用いた規則に基づく

手法と確定診断に基づく診断コードを用い、確定診断データベースから注目する症候群への感染の可能性が高い（低い）患者を抽出することで、大量の訓練データを自動的に生成する。さらに、診断コードの信頼性を区別することで症候群罹患判定の精度向上を目指す。

バイオサーベイランスの対象として、本稿では、特に呼吸器症候群（respiratory）に注目する。以下、提案手法で利用する確定診断データベース、診断コード、症例定義を用いた規則に基づく既存手法について解説し、続いて、診断コードと症候群の対応関係の妥当性を検証する。最後に、評価実験を通して提案手法の有効性を示す。

## 2 提案手法

### 2.1 概要

急性感染症の早期検知のためには、医師の確定診断を待たずに患者の判定を行う必要がある。そのため、退院時に付与される診断コードをバイオサーベイランスに使用することはできない。本研究では、診断コードを訓練データ構築のために利用する。訓練データ構築の流れを Fig.1 に示す。

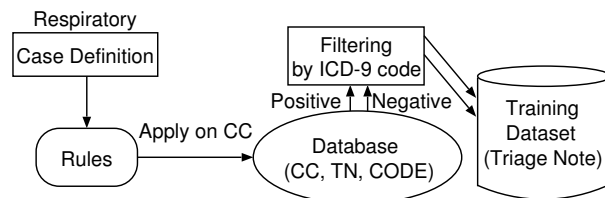


Fig. 1 訓練データ構築の流れ。

まず、信頼性の高いデータを抽出するために、症例定義から生成される規則と診断コードの両方を用い

て自動的に患者データを絞り込む。このデータを訓練データとし、分類器の学習に利用する。なお、症状定義を用いた規則を適用する際は、先行研究の知見 [3] に基づき、主訴のみを対象とすることで正例の純度向上を図る。

## 2.2 確定診断データベース

米国ノースカロライナ州では、バイオサーベイランスのため、州内の主要病院の救急外来に訪れた患者の診療情報を共有・提供するシステムが運用されている。本研究では、2006 年に救急外来を訪れた患者のうち、258,365 件のデータを使用する。このデータの一部を Table.1 に示す。ここで、1 行が 1 レコード（つまり 1 患者のデータ）に対応する。

Table. 1 確定診断データベースの一部。

CC	Triage Note	Diag1
Cold	He has a fever and [...]	464.4
Cough	have a harsh and cough [...]	376.1
FEVER	pt c/o fever and [...]	36.7

CC (Chief Complaint) は主訴と呼ばれ、患者の症状を数語で簡潔に表現したものである。トリアージノート (Triage Note) は患者の訴える症状が詳細に記述された文章である。トリアージノートは、ナースが直接コンピュータに入力するため、「w/o」(without)、「w/」(with)、「pt.」(patient) といった略語やスペルミスが非常に多く含まれている。Diag は診断コード (ICD-9-CM) であり、確定診断の際、医師によって重要な (患者の容態をよく表す) コードから順に Diag1 ~ Diag11 が付与される。ICD-9-CM は、世界保健機関 (WHO) によって策定された分類コード ICD-9 を基にしており、医療機関における診療記録の管理に利用されている。

ICD-9-CM と症候群の対応は、米国の疾病対策予防センター (Centers for Disease Control and Prevention; CDC) の専門家によって検討が行われており、その対応関係の信頼性が高い順に コンセンサス コード 1 ~ 3 が付与されている [4]。なお、この対応関係は症候群と一対一に定義されているわけではなく、また、対応関係が存在しても診断コードから症候群の罹患を正確に判断することはできないことに注意を要する。症候群と診断コードの対応関係の一部を Table.2 に示す。

例えば、呼吸器症候群 (respiratory) に定義されている診断コードのうち、コンセンサス 1 の 464.4 (croup)

Table. 2 症候群と診断コードの対応関係の例。

Syndrome	Code	Consensus
Fever	780.6 Fever (発熱)	1
	024 Glanders (鼻疽)	3
	⋮	⋮
Respiratory	464.4 Croup (喉頭炎)	1
	786.2 Cough (咳)	1
	786.07 Wheezing (喘鳴)	2
	⋮	⋮
GI	787.91 Diarrhea (下痢)	1
	787.2 Dysphagia (嚥下障害)	2
	⋮	⋮

や 786.2 (cough) が付与された患者は特に呼吸器症候群である可能性が高い。ただし、この診断コードは、医療保険の支払いに際し利用されるコードでもあるため、患者あるいは病院にとって有益となるように付与されている可能性がある。このため、診断コードを研究に利用する際には、その信頼性に留意する必要がある。

## 2.3 症例定義を用いた規則に基づく手法

症候群罹患の判定のため、公衆衛生の専門家によって各国で症例定義が作成されている。この症例定義から、症候群の罹患判定を行う規則を容易に作成することができる。SQL によって表現された規則の例を Fig.2 に示す。

```

WHERE (
  Case When
    CC like '%fvr%'
  OR CC like '%croup%'
  OR CC like '%cough%'
  OR CC like 'flu'
  OR CC like 'virus%'
  ...

```

Fig. 2 SQL による罹患判定規則の一部。

既存研究における規則に基づく手法では、この規則を新たな来院患者のデータに適用することで、症候群の罹患を判定していた。本研究では、規則を確定診断データベースへ適用し、訓練データの自動生成に用いる。

### 3 評価実験

#### 3.1 診断コードの有用性評価

診断コードは医療保険の支払いに利用されることから、付与の際に何らかのバイアスがかかっている可能性がある。よって、診断コードと症候群の対応関係 [4] がどの程度妥当であるかは明らかではない。このため、今回実験に使用するデータについて、診断コードと症候群の対応関係が妥当であるか否かを間接的に調査した。まず、確定診断データベースの中の 258,365 件のデータに関して、フィールド Diag1 に呼吸器症候群と対応する コンセンサス 1 の診断コードが付与されているデータを仮に陽 (positive) クラス、Diag1 ~ Diag11 のいずれにも呼吸器症候群に対応するコードが付与されていないデータを仮に陰 (negative) クラスとした。これにより、44,760 件のレコードが陽クラス、214,521 件のレコードが陰クラスとなった。これらのレコードから抽出したトリアージノートに関して、各単語のカイ二乗値を計算することで、陽クラスに特徴的な単語、陰クラスに特徴的な単語を同定した。Table.3 に、最頻出の 10,000 語を対象にカイ二乗値 ( $\chi^2$ ) を計算したときの結果を上位のものから順に示す。

Table. 3 単語のカイ二乗値.

Positive		Negative	
Term	$\chi^2$	Term	$\chi^2$
cough	138,817	pain	11,957
throat	48,856	to	6,289
fever	45,385	abd	3,940
coughing	42,430	left	3,886
sore	41,275	lower	3,341
congestion	40,697	right	3,281
productive	20,181	back	2,966
cold	13,643	bleeding	2,675
⋮	⋮	⋮	⋮

Table.3 をみると、陽クラスに特徴的な単語は、「cough」、「throat」、「fever」など呼吸器症候群と特に関係の強い単語であり、一方陰クラスに特徴的な単語は、「pain」、「abd」(abdominal の略)、「left」など呼吸器症候群との関係が弱い単語であることが観察出来る。この結果から、今回対象とするデータにおいては、診断コードと症候群の対応関係の妥当性が認められる。これは、確定診断データベースからの特定症候

群に関するデータ抽出に、診断コードの利用が有効であることを示唆する。

#### 3.2 罹患判定実験

本実験では、まず Fig.2 に示したような症例定義に基づく規則を用いて、確定診断データベースから呼吸器症候群に対応するレコードを抽出した。このレコードのうち、フィールド Diag1 に呼吸器症候群に対応する コンセンサス 1 の診断コードが付与されているレコードのトリアージノートを正例として 1,000 件抽出した (これを PosCons1 と呼ぶ)。負例としては、コンセンサス コードに関わらず、呼吸器症候群に対応する診断コードが Diag1 ~ Diag11 のいずれにも付与されていないレコードを同定し、このうち 4,000 件を無作為に抽出した (これを Neg4000 と呼ぶ)。テストデータとしては、医師が人手でラベルを付与したデータセットを用意した。このテストデータは、正例 171 件、負例 750 件のレコードから成る。

機械学習には、ベース分類器として決定木 (J48) を用い、さらにバギング (bootstrap aggregating; bagging) を適用して精度の向上を試みた。バギングは、与えられた訓練データを重複を許して複数回サンプリングし、それぞれを学習に用いた分類器を生成したのち、それらの分類器の多数決によって最終的な分類結果を出力するアルゴリズムであり、通常分類器より精度の高い結果を得られることが多い。ベースライン (Baseline) には、規則に基づく既存手法を用いた結果を使用する。評価指標として、精度 ( $P$ )、再現率 ( $R$ )、 $F_1$  値 ( $F$ ) を用いる。 $F_1$  値は分類性能を示す一般的な指標で、精度と再現率の調和平均として定義される。

$$F = \frac{2PR}{P + R}$$

Table.4 に、PosCons1 と Neg4000 を訓練データとして使用した際の罹患判定実験の結果を示す。太字の値は、比較した手法中での最も高い性能を示し、括弧内の数字はベースライン (既存手法) と比較した時の性能向上・低下を示す。

Table. 4 自動構築した訓練データ PosCons1 と Neg4000 を用いた罹患判定実験の結果.

	Baseline	J48	Bagging
$P$	<b>0.569</b>	0.450 (-21%)	0.485 (-15%)
$R$	0.325	<b>0.743</b> (+128%)	<b>0.743</b> (+128%)
$F$	0.413	0.561 (+36%)	<b>0.587</b> (+42%)

従来手法と比較し、再現率、 $F_1$  値において性能が大幅に改善された一方、精度に関して分類性能が低下した。なお、バイオサーベイランスにおいては陽性の患者を検出する感度、すなわち再現率が重要であり、 $F_1$  値も向上していることから、総合的には提案手法による分類器の挙動は好ましい。

次に、コンセンサスコードを有用性を評価するため、Diag1 フィールドがコンセンサス 1, 2, 3 のいずれかの診断コードであるレコードを正例と見なし、1,000 件のレコードを抽出した（これを PosCons1-3 と呼ぶ）。この際、確定診断データベース全体でのコンセンサスコードの割合に適合するように層化抽出法を適用し、コンセンサス 1, 2, 3 の診断コードを持つレコードの件数がそれぞれ 703, 258, 39 件となるようにした。この操作により利用する診断コードの範囲が広がるため、罹患判定の精度は低下しつつも、さらに再現率が向上するものと期待される。Table.5 に、PosCons1-3 と Neg4000 を訓練データとして使用した際の罹患判定実験の結果を示す。

Table. 5 PosCons1-3 と Neg4000 を訓練データとしたときの罹患判定実験の結果。

	Baseline	J48	Bagging
$P$	<b>0.569</b>	0.444 (−22%)	0.480 (−16%)
$R$	0.325	<b>0.772</b> (+138%)	<b>0.772</b> (+138%)
$F$	0.413	0.564 (+37%)	<b>0.592</b> (+43%)

予想した通り、Table.4 と比較して提案手法の再現率が改善していることが分かる。一方、精度への影響は非常に小さく、コンセンサスコード 2 および 3 の診断コードを訓練データの作成に用いても、データの質の低下は極めて限定的であることが分かった。

上述の実験では、異なるコンセンサスコードに対応する事例を学習時には全く同一に扱った。しかし、コンセンサスコード 1 に対応する事例と 2 や 3 に対応する事例では、その重要性は異なる。そこで、PosCons1-3 中の各事例へコンセンサスコードに対応した重みを付与し、再実験を行った。具体的には、コンセンサス 1, 2, 3 の事例に対して、それぞれ 1.0, 0.5, 0.33 の重みを付与した（これを WPosCons1-3 と呼ぶ）。Table.6 に実験の結果を示す。

期待に反し、Table.4, Table.5 の結果と比較して、ほとんどの場合に分類性能が低下する結果となった。今後、コンセンサスコードのより有効な利用方法を検討する必要がある。

Table. 6 WPosCons1-3 と Neg4000 を訓練データとしたときの罹患判定実験の結果。

	Baseline	J48	Bagging
$P$	<b>0.569</b>	0.481 (−15%)	0.477 (−16%)
$R$	0.325	0.684 (+110%)	<b>0.725</b> (+123%)
$F$	0.413	0.565 (+37%)	<b>0.575</b> (+39%)

## 4 おわりに

本研究では、症例定義に基づく規則ベースの分類法と診断コードに着目し、機械学習の手法を用いて急性感染症の罹患判定を行った。実データを用いた評価実験の結果、規則に基づく既存手法と比較し、提案手法では再現率（感度）、 $F_1$  値が大幅に向上した。また、異なるコンセンサスコードを訓練データ生成に利用した結果、さらに再現率が向上した。これらの結果は、感染症罹患者の早期検出につながるものと期待される。今後の課題として、コンセンサスコードのより有効な利用方法の検討、大規模な訓練データを用いた実験と評価、他言語への応用などが考えられる。

## 参考文献

- [1] 重茂浩美. 症候群サーベイランス-感染症流行の早期探知に向けて. 科学技術動向, Apr 2010.
- [2] H. Lu, D. Zeng, L. Trujillo, K. Komatsu, and H. Chen. Ontology-enhanced automatic chief complaint classification for syndromic surveillance. *Journal of biomedical informatics*, Vol. 41, No. 2, pp. 340–356, Apr 2008.
- [3] M. Scholer, G. Ghneim, S. Wu, M. Westlake, D. Travers, A. Waller, A. McCalla, and S. Wetterhall. Defining and applying a method for improving the sensitivity and specificity of an emergency department early event detection system. *AMIA Annu Symp Proc*, pp. 651–655, 2007.
- [4] Center for Disease Control and Prevention. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. U.S. Department of Health and Human Services, 2003. Retrieved August 4, 2010, from [www.bt.cdc.gov/surveillance/syndromedef/](http://www.bt.cdc.gov/surveillance/syndromedef/)