

医薬品の取り違いミスを防止するための薬名類似度の定量的指標の構築

大谷壽一,^a 竹田正幸,^{b,c} 今田結城,^b 澤田康文^{*,a,d}

Development of the Measures to Evaluate the Similarity of Drug Brand Names

Hisakazu OHTANI,^a Masayuki TAKEDA,^{b,c} Yuki IMADA,^b and Yasufumi SAWADA^{*,a,d}

^aLaboratory of Drug Informatics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, ^bGraduate School of Information Science and Electrical Engineering, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka City 812-8581, Japan, ^cSolution Oriented Research for Science and Technology (SORST), Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi City 332-0012, Japan, and ^dInterfaculty Initiative in Information Studies, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

(Received January 18, 2006; Accepted February 20, 2006)

Brand name confusion is one of the most common causes of drug-related errors. The aim of this study was to develop quantitative measures of similarity among brand names of drugs. We modified the fragmentary pattern-based measure, a measure of similarity for character strings based on the string resemblance system, to develop three novel measures of similarity, *i.e.*, the head and tail-weighted fragmentary pattern-based measure (htfrag), visually weighted htfrag (vwhtfrag), and auditorily weighted htfrag (awhtfrag). The 227 pairs of brand names for which confusion errors have been reported were used as a positive control group. Ten sets of 2270 random pairs of brand names were generated as negative controls. Then we evaluated the measures developed by using the geometric mean of sensitivity and selectivity as an objective function, in comparison with two conventional measures of similarity based on the vector space model (cos1 and htco). The measures developed, htfrag, vwhtfrag, and awhtfrag, provided better discrimination with mean objective function values of 0.953, 0.962, and 0.940, respectively, which were higher than those for the conventional measures cos1 and htco (0.922 and 0.892, respectively). The rates of false-positives and false-negatives were 3.3–10.7% and 5.3–11.9% for cos1, respectively, while the rates for vwhtfrag were 4.8–5.9% and 2.2%, respectively. The measures of similarity developed may provide significant information to avoid drug-related errors associated with brand name confusion.

Key words—brand name; medical error; similarity measure; similarity index; fragmentary pattern; character string

緒 言

近年頻発している医療ミスの中でも、投薬ミスは大きな問題となっている。その原因の1つに、類似した医薬品名同士の取り違いに起因する処方ミス、調剤ミスが挙げられている。例えば2000年12月には、β遮断薬であるアルマール（一般名塩酸アロチノロール）とアマリール（一般名グリメピリド）を取り違えて後者を誤調剤した結果、患者が死亡する

という調剤ミスが起こっている。¹⁾同様の取り違いは、医師による処方作成の段階においても数多く報告されている。¹⁾日本薬剤師会においても、調剤事故防止マニュアルの中で間違い易い医薬品の例を提示するなどして、注意を喚起している。²⁾

このような、薬名類似に起因する医療ミスを低減するためには、今後新たに上市される医薬品について、既存の医薬品に類似した名称を付さないとともに、処方作成の段階で薬名が類似した医薬品を誤って処方した場合に警告を発するコンピュータシステムを導入するといった対策が必要になると考えられる。しかし、いずれの対策においても、2 医薬品間の薬名の類似度を定量的に評価するための指標が必要となる。土屋らは現在までに、医薬品の薬名の類

^a東京大学大学院薬学系研究科医薬品情報学講座, ^b九州大学大学院システム情報科学研究院, ^c独科学技術振興機構 SORST, ^d東京大学大学院情報学環・学際情報学府

*e-mail: sawada@mol.f.u-tokyo.ac.jp

似度を定量的に評価するための指標を複数設定し、これを製薬企業が医薬品名を新たに命名する際の資料として提供している。^{3,4)} すなわち、薬名における各カタカナ文字の生起回数を成分とすることで薬名を特徴ベクトル化したときの、ベクトルのなす角の余弦 ($\cos\theta$)、薬名の先頭 2 文字と末尾 2 文字に着目した htco (head and tail cosine)、及び一方の薬名を他の薬名に変換するために必要な編集操作の最小適用回数である編集距離 (edit)、という 3 種の指標である。しかし彼らは、これらのどの指標を用いるべきかは明確にしておらず、医薬品の取り違えが起り易くなる閾値の値についても言及していないため、あくまで類似度の参考となり得る指標を複数提示しているに過ぎない。指標の評価についても、被験者を用いた少数の試験は行っているものの、⁴⁾ 実際に取り違えが生じた医薬品名を用いた大規模な評価は行われていない。また土屋らは、視覚的形狀の類似したカタカナが含まれている場合には取り違えが起り易いことから、それらの影響を指標に取り入れる必要性について考察しているが、⁴⁾ この点を考慮した類似度の評価は行われていない。

そこで本研究においては、薬名類似度を定量化するために、実際に医薬品の取り違えが起こった類似薬名の組み合わせを positive control (正例)、無作為に発生させた薬名の組み合わせを negative control (負例) として用い、より有用な類似度指標を構築することを目的とした。その際、カタカナ文字の視覚的類似性なども考慮に入れて、より判別性の高い指標を検索した。さらに、ある薬名の組み合わせ間で取り違えが起り易いか起り難いかを判別するための閾値を設定することが可能かについても検討を行った。

方 法

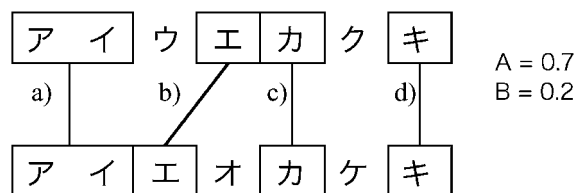
1. 新規類似度指標の構築

1-1. 指標 1 (Htfrag; Head and Tail-weighted Fragmentary Pattern Based Measure) まず、2 薬名間の共通文字列部分が長ければより類似度が高いと考え、断片パターン指標⁵⁾を応用した指標を導入した。土屋らの研究により、薬名類似度には先頭及び末尾部分の一致が重要であるとされていることから、それ以外 (中間) の部分での一致のスコアを軽くするための係数 A ($0 < A \leq 1$) を導入した。また、

文字の一致が断続的である場合よりも連続している場合のスコアを重くするために、係数 B ($0 < B \leq 1$) を導入した。実際の指標の計算手順としてはまず、両薬名の中から共通文字列の組を抽出し、各組の一致文字列長から係数 B を減じた値に、さらに先頭・末尾以外の一致では係数 A を乗じた。これをすべての共通文字列の組について加算し、両薬名の文字列長の平均値で除して正規化することで htfrag の値とした。薬名間の共通文字列抽出にあたっては、1 つの共通文字列と別の共通文字列とが同一薬名中において重なりを持つことを許さない。そのような共通文字列の組の選び方は一意に決まらないため、類似度の値が最も高くなるものを用いることにした。Figure 1 (A) には、係数 A 及び係数 B がそれぞれ 0.7, 0.2 のときの、「アイウエカクキ」と「アイエオカケキ」の組における htfrag 値の計算例を示す。

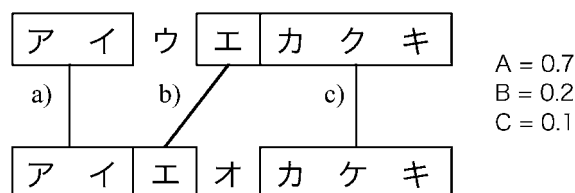
1-2. 指標 2 (Vwhtfrag; Visually Weighted Htfrag) カタカナの中には、視覚的に形状が類似した組み合わせも多いために、文字としては一致していなくても、類似度に影響を及ぼしている可能性が考えられる。そこで、濁点及び半濁点はその有無に関わらず同一文字とみなすとともに、田中の文字パターン認知に関する研究⁶⁾を基に、視覚的に類似したカタカナの 22 対の組み合わせ、すなわち「ナ, メ」「ソ, ン」「ツ, シ」「ム, マ」「ク, ケ」「ユ, ヲ」「リ, レ」「ア, ワ」「ラ, テ」「チ, ヌ」「イ, フ」「ヘ, ト」「ハ, ソ」「ハ, ン」「コ, ユ」「ヒ, セ」「シ, ミ」「ヌ, マ」「ト, メ」「フ, レ」「コ, エ」については、文字が一致したものとみなして共通文字列を抽出し htfrag と同様の計算を行い、これを vwhtfrag とした。ただし、一致文字列長を求める際には、これら形状の類似した文字に対して C 文字分の一致 ($0 \leq C \leq 1$) がみられたものとして計算を行っている。この指標の特徴は、htfrag では文字の不一致により複数の共通文字列に分断されていたものを、不一致を無視して長い共通文字列が存在するものとみなす点にある。すなわち、「パントシン」と「パソトミン」を例にとると、htfrag では「パ」「ト」「ン」の 3 つの共通文字列に分断されるが、vwhtfrag では「ン」と「ソ」、「シ」と「ミ」の形状類似を考慮することにより、共通文字列は 1 つに纏まる。したがって、 $C=0$ とした場合でも、htfrag に比べて類似度は高くなる。Figure 1

(A) calculation of measure "htfrag"



a) length of "アイ" - factor B = 2 - 0.2 =	1.8
b) (length of "エ" - factor B) × factor A = (1 - 0.2) × 0.7 =	0.56
c) (length of "カ" - factor B) × factor A = (1 - 0.2) × 0.7 =	0.56
d) length of "キ" - factor B = 1 - 0.2 =	0.8
sum	3.72
	$3.72 / 7 = \mathbf{0.53}$

(B) calculation of measure "vwhtfrag"



a) length of "アイ" - factor B = 2 - 0.2 =	1.8
b) (length of "エ" - factor B) × factor A = (1 - 0.2) × 0.7 =	0.56
c) length of "カ*キ" - factor B = 2.1 - 0.2 =	1.9
sum	4.26
	$4.26 / 7 = \mathbf{0.61}$

Fig. 1. Calculation of the Similarity Measures, (A) "Htfrag" and (B) "Vwhtfrag"

Factor A ($0 < A \leq 1$) was introduced to make light the significance of coincidence on the middle of words. Factor B ($0 < B \leq 1$) was introduced to emphasize the coincidence of adjacent letters of two or more. As for vwhtfrag, factor C was additionally introduced to treat the incomplete coincidence for a pair of visually similar characters.

(B)には、係数 A, B 及び C がそれぞれ 0.7, 0.2 及び 0.1 のときの、「アイウエカクキ」と「アイエオカケキ」の組における vwhtfrag 値の計算例を示す。

1-3. 指標 3 (Awhtfrag; Auditorily Weighted Htfrag) 文字の音感が類似している場合、文字としては一致していなくても類似度に影響を及ぼしている可能性が考えられる。そこで、カタカナを母音と子音に分解し、母音・子音のいずれかが一致している場合は文字が一致したものとみなして共通文字列を抽出し、htfrag と同様の計算を行い、これを awhtfrag とした。ただし、「フェ」「ヴィ」等については、2 文字ではなく 1 文字として扱い「fe」「vi」のように分解して処理を行っている。一致文字列長の算出に際しては、母音及び子音の一致に対してそれぞれ C, 1-C という値を付与している ($0 \leq C \leq 1$)。

2. 指標の評価法 実際に医薬品名の取り違えが生じた正例としては、日本薬剤師会が収集した事例より抜粋した資料²⁾に記された全 227 組を用いた。また、負例として、医薬品名の語幹 5817 個を抽出し、ここから無作為に作成した 2270 組を用いた。正例及び負例のすべての組について、その類似度を各指標により定量化した。このとき、それぞれの指標について閾値 t を設定すると、正例は、真陽性 (true positive; TP) と偽陰性 (false negative; FN) に、負例は偽陽性 (false positive; FP) と真陰性 (TN; true negative) にそれぞれ分類される。ここで、検出の感度 (sensitivity) と特異度 (specificity) はそれぞれ $TP / (TP + FN)$ ($= TP / 227$) と $TN / (TN + FP)$ ($= TN / 2270$) と定義される。感度と特異度の相乗平均を目的関数 $Obj(t)$ とし、 t を

Positive cases			Negative cases			
index			index			
TP 222	0.8556	パロチン	メパロチン	0.6727	レストマート	レスラート
	0.8556	ストミン	アストミン	0.6465	オビソート	ビタソート
	0.8556	カドラル	ドラール	0.6333	ブロン	モリブロン
	0.8545	サイトテック	サイトック	0.6167	アップノール	アセタノール
	⋮			⋮		
FN 5	0.2833	バナ	ベネン	0.2804	グリブラチン	グリフィリン
	0.2804	エクセグラン	エブランチル	0.2804	パラブラチン	ハルブラッセ
	0.2692	エンピナス	エラスチーム	0.2775	アルビラック	ボルビックス
	0.2465	ユリノーム	ユーロジン	0.2739	ロイアン	チクピロン
	0.2465	ニボラジン	ニバジール	⋮		
0.2241	ユリノーム	ユーエフティ	0	アクアチム	デキストファン	
0.2125	ミケラン	メブチン	0	アキリデン	フィジオゾール	

$$\text{Obj}(t) = \sqrt{\frac{\text{TP} \cdot \text{TN}}{\text{TP} + \text{FN} \cdot \text{TN} + \text{FP}}} = 0.9649$$

Fig. 2. Evaluation of Measures

The objective function (Obj(t)) is defined as the geometric mean of sensitivity (TP/total number of positive cases) and specificity (TN/total number of negative cases). The t value (threshold) was chosen to yield the maximum value of objective function. (TP: true positive, FN: false negative, FP: false positive, TN: true negative).

変化させたときの、目的関数の最大値を指標の評価尺度とした (Fig. 2)。なお、薬名の組み合わせが有限で、それらに対する指標の値は離散分布となるので、最良の評価関数を与える閾値 t は、一定の幅を持った値となる。

負例 2270 組は無作為に 10 セット (2270 組 × 10 セット) を作成して評価を行い、各目的関数の 10 セットにおける平均値を指標の評価に用いた。評価対象の指標としては、今回新たに構築した htfrag, vwhtfrag, awhtfrag 及び土屋らの報告による cos1, htco を選択した。

結 果

1. 指標の評価 2270 組の負例を無作為に 10 セット作成し、各指標の評価関数値の最大値、最小値、並びに平均値を算出した結果を Table 1 に記す。なお、htfrag, vwhtfrag 並びに awhtfrag に関しては、パラメータ A, B, C についてそれぞれ 0.05 刻みで変化させた場合についてすべて計算し、最良の評価関数値が得られた 5 組について記した。Htfrag, vwhtfrag 並びに awhtfrag に関してはいずれも既報の指標よりも優れた評価関数値を示した。中でも特に、文字の視覚的な類似性を考慮した指標 vwhtfrag の評価関数値が最良であった。偽陽性率

及び偽陰性率は生成した負例により変動したが、検討した 10 セットの負例において、cos1 では、偽陽性率及び偽陰性率はそれぞれ 3.3—10.7%、5.3—11.9% の範囲であった。これに対して、vwhtfrag (A=0.45, B=0.15, C=0.00) では、偽陽性率及び偽陰性率はそれぞれ 4.8—5.9% の範囲、及び 2.2% と、特に偽陰性率が著しく改善された。

また、cos1 では false positive 及び false negative と判定されたが、htfrag, vwhtfrag 並びに awhtfrag ではそれぞれ true negative 及び true positive と正しく判定された薬名の組み合わせ例の一部と、そのときの指標値及び閾値を Table 2 に示す。

2. 閾値の安定性 薬名が類似しているか否かを絶対的に判断するためには、その指標について、用いる負例に依存せず目的関数の最大値を与える閾値 t が安定している必要がある。そこで、各指標について、10 セットの負例の間で閾値の値がどのように変化するかを検討した結果を Table 3 に示す。Cos1 については、使用した負例の組み合わせによって最良の評価関数値を与える閾値の値が異なっていたが、htfrag, vwhtfrag 並びに awhtfrag に関してはいずれも、負例の組み合わせによらず一定の閾値となった。

3. 正例及び負例における指標値の分布 Cos1,

Table 1. Evaluation of the Measures to Quantify the Similarity of Brand Name

Measure	Parameters	Mean	Min	Max
Cos1		0.92209	0.9196	0.9260
Htco		0.89243	0.8883	0.8947
Htfrag	(A=0.35, B=0.30)	0.95279	0.9506	0.9569
	(A=0.35, B=0.25)	0.95279	0.9506	0.9569
	(A=0.35, B=0.35)	0.95272	0.9506	0.9567
	(A=0.35, B=0.40)	0.95247	0.9506	0.9565
	(A=0.35, B=0.45)	0.95245	0.9506	0.9565
Vwhtfrag	(A=0.45, B=0.15, C=0.00)	0.96199	0.9595	0.9649
	(A=0.45, B=0.15, C=0.05)	0.96192	0.9595	0.9647
	(A=0.45, B=0.10, C=0.00)	0.96191	0.9595	0.9649
	(A=0.45, B=0.20, C=0.00)	0.96187	0.9593	0.9647
	(A=0.45, B=0.15, C=0.15)	0.96183	0.9593	0.9647
Awhtfrag	(A=0.60, B=0.20, C=0.60)	0.93995	0.9363	0.9418
	(A=0.55, B=0.20, C=0.60)	0.93828	0.9337	0.9399
	(A=0.45, B=0.00, C=0.50)	0.93810	0.9330	0.9413
	(A=0.65, B=0.20, C=0.60)	0.93805	0.9341	0.9414
	(A=0.60, B=0.25, C=0.60)	0.93788	0.9344	0.9394

Each value represents the geometric mean of sensitivity (true positive/total true case) and specificity (true negative/total negative case).

Table 2. Examples of Discrimination of Similarity by Cos1, Htfrag, Vwhtfrag and Awhtfrag

Measure	Cos1	Htfrag	Vwhtfrag	Awhtfrag
Threshold	0.3873	0.1718	0.2775	0.3644
	-0.4000	-0.1750	-0.2804	-0.3657
Positive cases				
ブラダロン-ブレディニン	0.365×	0.254○	0.309○	0.440○
フランドル-フラベリック	0.365×	0.309○	0.336○	0.371○
バナナ-ベネン	0.333×	0.233○	0.283○	0.667○
Negative cases				
ランドセン-アンフラマイド	0.571○	0.122×	0.122×	0.280×
ワンアルファー-リファンピシン	0.544○	0.129×	0.129×	0.240×
レベトール-ナルコルト	0.507○	0.098×	0.153×	0.288×

○: judged to be similar, ×: judged to be not similar.

htco, htfrag, vwhtfrag 及び awhtfrag により 227 組の正例及び 2270 組の負例の薬名類似度を評価した場合の指標値の頻度分布を Fig. 3 に示す. Cos1, htco では, 正例と負例の分布に重なりがあるのに対し, htfrag, vwhtfrag 及び awhtfrag では, その重なりは小さいものであった.

重なりを表す指標として, Fig. 3 中に目的関数 $Obj(t)$ の最大値 Obj を併せて示した. 分布が完全に分離している場合, 並びに完全に一致している場合, Obj の値はそれぞれ 1.0 及び 0.5 を与える.

考 察

本研究においては, 従来薬名類似度の指標として用いられてきたベクトル空間指標に基づく評価 (cos1, htco など) や編集距離 (edit) ではなく, 断片パターン指標による類似度評価を導入し, 薬名類似度をより高精度に評価することが可能となった. すなわち, 今回構築した 3 種の指標はいずれも, 従来提案されてきた cos1, htco, 並びに edit と比較して, 実際に取り違えが生じた薬名と, 無作為に発生

Table 3. The Threshold (t) Values of Each Measure for 10 Sets of 2270 Random Negative Pairs

Cos1	Htco	Htfrag (a=0.35, b=0.30)	Vwhtfrag (a=0.45, b=0.15, c=0.00)	Awhtfrag (a=0.60, b=0.20, c=0.60)
0.3873—0.4000	0.2500—0.5000	0.1718—0.1750	0.2775—0.2804	0.3644—0.3657
0.3873—0.4000	0.2500—0.5000	0.1718—0.1750	0.2739—0.2804	0.3644—0.3657
0.3873—0.4000	0.2500—0.5000	0.1718—0.1750	0.2791—0.2804	0.3644—0.3657
0.4880—0.5000	0.2500—0.5000	0.1718—0.1750	0.2739—0.2804	0.3644—0.3657
0.4950—0.5000	0.2500—0.5000	0.1727—0.1750	0.2782—0.2804	0.3644—0.3657
0.3873—0.4000	0.2500—0.5000	0.1718—0.1750	0.2739—0.2804	0.3644—0.3657
0.4743—0.5000	0.2500—0.5000	0.1718—0.1750	0.2739—0.2804	0.3644—0.3657
0.4924—0.5000	0.2500—0.5000	0.1718—0.1750	0.2782—0.2804	0.3644—0.3657
0.4811—0.5000	0.2500—0.5000	0.1718—0.1750	0.2782—0.2804	0.3644—0.3657
0.3873—0.4000	0.2500—0.5000	0.1727—0.1750	0.2791—0.2804	0.3644—0.3657

させた負例とを良好に判別することができた。特に、vwhtfrag においては偽陰性率が著しく低下しており、医薬品の命名において類似薬名を付さないようにするためには有用な指標と考えられる。このような観点から、今回構築した指標は、医薬品の類似性を評価する指標として極めて有用と考えられる。

情報科学分野では、文字列間の類似性を測る尺度として古くから編集距離及びその変種が考案され、テキスト編集における綴り訂正を始め、DNA の塩基配列や蛋白質のアミノ酸配列の相同性評価などにも広く用いられてきた。一方、筆者らは和歌文学研究者との共同研究において和歌の類似性を定量化する必要に迫られ、編集距離やその変種を適用するのでは限界があることから、類似性指標を汎化した理論体系である SRS (String Resemblance System)⁷⁾を新たに構築し、その基でいくつかの類似性指標を開発した。断片パターン指標はその1つであり、和歌文学研究への応用において、従来知られていなかった本歌取りの例を見出すなどその有用性が示されている。⁸⁾ 医薬品名の類似性を定量化する目的においても、断片パターン指標による評価は、編集距離やベクトル空間指標を用いた評価と比較して、人間が認識する類似性を最も良好に反映した指標であると考えられる。

今回導入した指標には、A、B 及び C という複数のパラメータが導入されている。これは、それぞれ「共通文字列の連続性」「先頭・末尾部分の重要性」「文字形状や音韻の類似性」という因子に対応したパラメータである。本論文の指標を使用するためにはこれら複数のパラメータを選定しなければならな

いという側面もあるが、実験で得た最良パラメータの値を基に、どの因子がどの程度重要かを評価することができるというメリットもある。今回検討した指標のパラメータのうち、vwhtfrag についてはパラメータ C の値が 0 のときに最良の評価関数値を与えた。この条件では、htfrag との違いは、濁点・半濁点を含む文字を同一視したことに加え、形状の類似した文字を介してその前後の文字列断片が接続され、ひとまとまりの長い共通文字列とみなされる点である。すなわち、形状的に類似した文字は、C = 0 の場合それ自身はスコアに寄与しないが、前後の共通文字列を繋いでひとまとまりの長い共通文字列が存在するように見せかけることで視覚的類似性を増加させる働きを担っているのかもしれない。

今回導入した htfrag, vwhtfrag 及び awhtfrag を比較すると、いずれも既存の cos1, htco と比較して高い評価関数値を与えたものの、awhtfrag は他の 2 者に比較して評価関数値が比較的 low だった。これは、医薬品名のやり取りが、口頭ではなく処方せんとして書かれた情報を中心として行われるためかもしれない。このため、薬名の類似度を判定するためには、vwhtfrag が現時点では最も良好な指標と考えられる。今後は、取り違えのミスが生じた事例を詳細に分析し、口頭での取り違えに関する正例だけを選択的に蓄積すれば、awhtfrag において高い評価関数値となる可能性はあるかもしれない。

ある類似薬名の組み合わせが与えられたときに、その組み合わせが取り違えを誘発する可能性があるか否かを安定的に判別するためには、各指標毎に一定の閾値を設定する必要がある。今回、10 セット

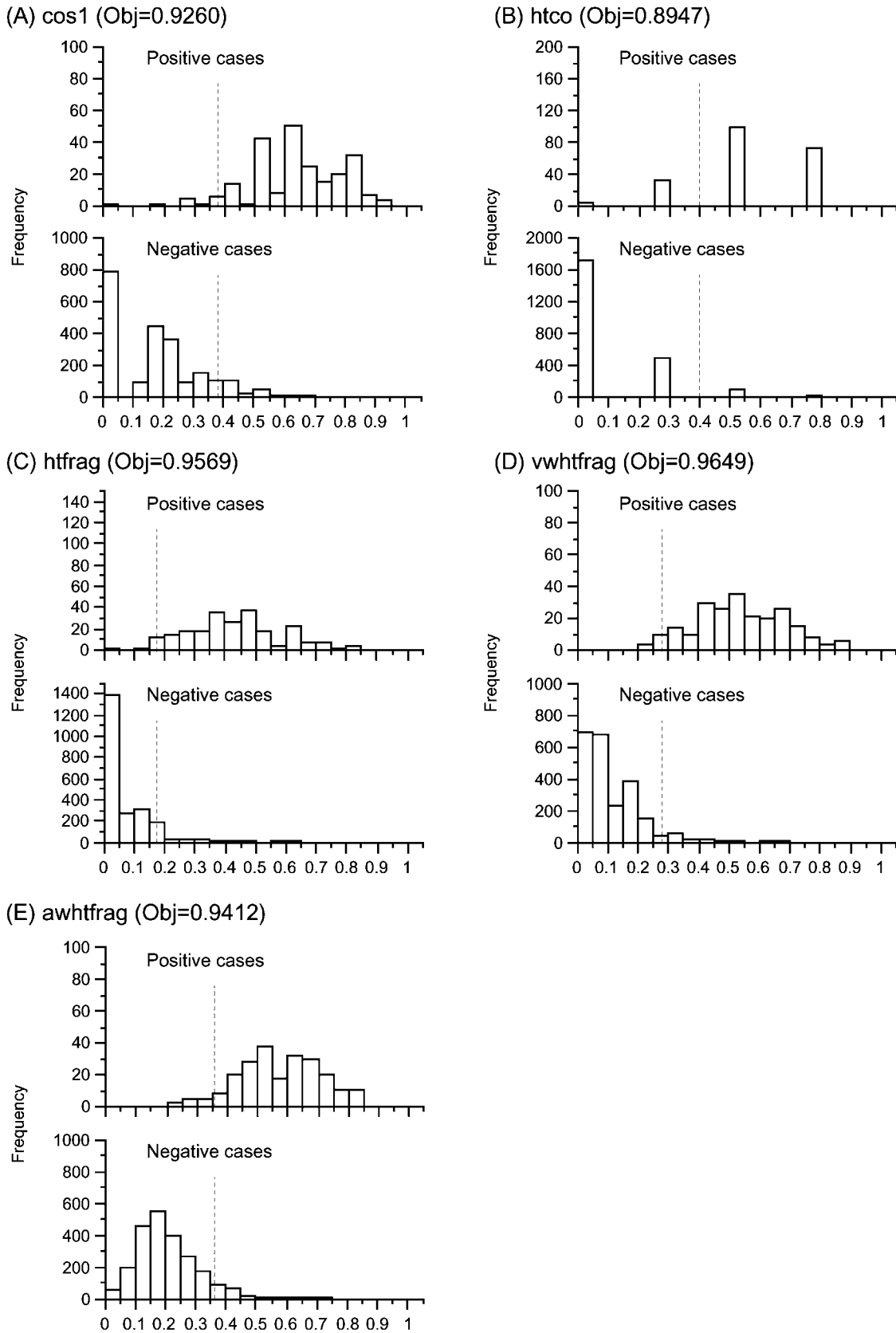


Fig. 3. Histograms of the Measures, Cos1, Htco (Head and Tail-weighted Cosine), Htfrag (Head and Tail-weighted Fragmentary Pattern Based Measure), Vwhtfrag (Visually Weighted Htfrag) and Awhtfrag (Auditorily Weighted Htfrag). Each vertical broken line represents the threshold that gives the maximum value of objective function (Obj).

の負例を用いて、それぞれの指標において最良の評価関数値を与える閾値を評価した結果、今回われわれが構築した *htfrag*, *vwhtfrag* 及び *awhtfrag* については、負例を変えてもその閾値は変化しなかった。これに対して、*cos1* については、負例により最良の評価関数値を与える閾値が変化していた。したがって、*cos1* については、閾値を一定値に固定した場合、負例によっては評価関数値が低下することになる。この点においても、今回構築した指標は薬名類似度を定量的に評価する上で有用であると考えられた。

今回構築した指標により、薬名類似度の判別力は向上したが、FP (偽陽性) や FN (偽陰性) を完全にゼロにすることはできなかった。これは以下の2つの理由による。1つは、負例は5817の医薬品名から無作為に2つを取り出し、これを2270回繰り返すことにより発生させている。このため、無作為に取り出した中にも一定の割合で類似した薬名が含まれてしまうと考えられる。特に、後発医薬品には、同一成分の先発医薬品と類似した薬名を付すことが多いため、一定の割合で類似した薬名の組み合わせが負例として作り出されるのはやむを得ないと考えられる。事実、負例として作り出された中にもかなり類似した薬名が散見された (Fig. 2)。もう1つの理由として、今回正例として用いた227組の中にも、かならずしも薬名類似だけに起因する取り違えとは言い切れない組み合わせが含まれていたことが挙げられる。例えばミケランとメプチンの成分はそれぞれ β 遮断薬の塩酸カルテオロールと β 刺激薬の塩酸プロカテロールであり、いずれも β 受容体に作用する点や、一般名が類似している点、さらには、同じ製薬企業から市販されているために包装や商品の書体、デザインなどが類似している点、などが複合的に関与して取り違えを生じている可能性は否定できない。このような条件を考慮すると、今回構築した指標は、陽性適中率、陰性適中率いずれも95%を超えており、十分な判別能力があると考

えてよいだろう。

今回構築した指標は、今後製薬企業が新医薬品を命名する際に有力な指標となるであろう。また、処方作成時に、類似した薬名による取り違えを警告することも可能になると思われる。すなわち、今回の薬名類似度とは別に、2医薬品間の処方関連度を定量的に指標化できれば、両指標を的確に組み合わせることで、取り違えによる誤処方に対する的確に警告を発することができるようになると考えられる。

以上、実際に薬名の類似により取り違えが起こった例を用いて、医薬品の薬名類似度を定量的に評価するための指標を構築することができた。本指標は、薬名の取り違えに基づく医療ミスを防止する上で重要な情報を提供するものとなるであろう。

謝辞 本研究の一部は、平成16—17年度文部科学省科学研究費(萌芽研究：研究代表者竹田正幸)の助成を受けて行われたものである。

REFERENCES

- 1) Sawada Y., "Yakugaku To Shakai", Jiho, Tokyo, 2001.
- 2) Japan Pharmaceutical Association, *J. Jpn. Pharm. Assoc.*, **53**(4) Sup, 60–62 (2001).
- 3) Tsuchiya F., Kawamura N., Oh C., Hara A., *Jpn. J. Med. Informatics*, **21**, 59–67 (2001).
- 4) Yamada Y., Shin H.-S., Haga S., Tsuchiya F., *Rikkyo Psychol. Res.*, **45**, 21–30 (2003).
- 5) Hori H., Shimozono S., Takeda M., Shinohara A., *Lecture Notes Comput. Sci.*, **2223**, 719–730 (2001).
- 6) Tanaka H., *J. Soc. Sci. Hum.*, **133**, 53–74 (1979).
- 7) Takeda M., Fukuda T., Nanri I., Yamasaki M., Tamari K., *Theor. Comput. Sci.*, **292**(2), 497–524 (2003).
- 8) Takeda M., Fukuda T., *Inform. Process Soc. Jpn. Magazine*, **43**(9), 941–949 (2002).