論　文

情報検索における信頼性評価基準について

石岡　恒憲[†]

Evaluation of Criteria on the Information Retrieval

Tsunenori ISHIOKA[†]

あらまし　情報検索においてその性能評価に用いられる精度と再現率を一次元のスカラ量に変換した指標である Van Rijsbergen の F 尺度，ブレークイーブン・ポイント，11 点平均精度の 3 つのそれぞれを，統計学で用いる 2×2 分割表における関連度の指標である 4 分相関係数，および ϕ 係数と比較した．その結果，フォールアウトが 0.1 以下であるという条件で，(1) F_1 尺度がほぼ同じ ϕ 係数の値を示すこと (2) ブレークイーブン・ポイントが ϕ 係数とほぼ同等の尺度であること (3) 11 点平均精度は ϕ 係数より大きく，4 分相関係数よりは小さい値を持つ尺度であること，がわかった．

キーワード　Van Rijsbergen の F 尺度，ブレークイーブン・ポイント，11 点平均精度，2×2 分割表，4 分相関係数，ϕ 係数

1. はじめに

インターネットの普及に伴い，goo (http://www.goo.ne.jp) や google (http://www.google.co.jp) などの検索エンジンの利用が一般のオフィスワーカーの日常となりつつある．このような状況において，情報検索における検索結果の信頼性という観点からの評価は，もっと論じられてよいと思われる．

情報検索において，現在もっともよく用いられている評価尺度としては，評価結果にどれだけ「漏れ」がないかを示す再現率 (recall) と評価結果にどれだけ「ゴミ」がないかを示す精度 (precision) であろう．再現率と精度の要約値として，Van Rijsbergen の F 尺度 (F-measure) [11] もよく知られている．また検索エンジン，あるいは検索システムの相互の比較には，後述するブレークイーブン・ポイントや 11 点平均精度も用いられ，これらの指標が情報検索における評価基準の事実上の標準（ディファクト・スタンダード）になっているように思われる（[10] など）．

しかしながら，これら評価基準そのものについての統計的性質について調査した文献は，文書検索における最も著名な国際集会 TREC(Text REtrieval Conference) の 10 年間にわたる論文集においてもただの一つもなく，またコンピュータ科学で最も著名な ACM(Association for Computing Machinery) の情報検索における国際集会 SIGIR の 1998 年から 2001 年にわたる全ての採択論文においても関連文献を発見できなかった.（評価基準そのものについて論じた文献に Ingwersen [4] の aboutness について言及した Fujita [2] や文書要約における評価基準を提案している Goldstein [3] があるのみである.）

そこで本稿では，これらの情報検索で用いられている評価尺度が，通常，統計学で用いられる 2×2 分割表における評価尺度とどのような関係にあるかを明らかにすることを試みる．

2 節では，情報検索で用いられる評価尺度について要約しておく．3 節では，統計学における 2×2 分割表における評価尺度について解説し，情報検索で用いる評価尺度との比較検討を行なう．4 節はまとめである．

2. 情報検索における評価尺度

2.1 再現率と精度

ある文書集合と検索質問集合について，検索質問に対する文書の適合性が与えられたとする．このとき，検索質問に対して，表 1 のような交差行列を作ることができる．ここで行は検索質問に対して適合するか否かを表わし，列は検索質問によって検索されたか否か

[†] 独立行政法人 大学入試センター 研究開発部, 東京都
The National Center for University Entrance Examinations, Komaba 2-19-23, Meguro-ku, Tokyo, 153-8501

を表わす．行列の要素は文書数である．

表 1 交差行列
Table 1. Cross matrix

	検索	非検索	計
適合（正解）	f_{11}	f_{12}	$f_{1\cdot}$
非適合（不正解）	f_{21}	f_{22}	$f_{2\cdot}$
計	$f_{\cdot 1}$	$f_{\cdot 2}$	$f_{\cdot\cdot} = n$

表 1 が与えられるとき，再現率 r(recall) と精度 p(precision) は，以下のように定義される:

$$再現率：r = \frac{f_{11}}{f_{1\cdot}} \tag{1}$$

$$精度：\ p = \frac{f_{11}}{f_{\cdot 1}} \tag{2}$$

すなわち，再現率は適合文書 $f_{1\cdot}$ のうちどれだけ検索できたかの比であり，検索における「漏れ」のなさを示している．一方，精度は，検索された文書 $f_{\cdot 1}$ のうちどれだけ適合文書を含むかの比であり，検索における「ゴミ」のなさを示している．

したがって，再現率と精度はともに大きい値ほど良いが，再現率と精度はトレード・オフの関係にあり，一般に再現率を高くしようとすると精度は低くなり，逆に精度を高くしようとすると再現率は低くなる．

さて，再現率の代わりに，フォールアウト (fallout) という指標が用いられる場合がある．フォールアウトは次式で定義される:

$$フォールアウト：a = \frac{f_{21}}{f_{2\cdot}} \tag{3}$$

フォールアウトは再現率と双対の尺度であり，不要な文書（非適合文書）をどれだけ排除できなかったかを示す尺度である．再現率は適合文書をどれだけ漏れなく検索できたかを示すという意味でユーザ指向の尺度といえるが，フォールアウトはシステムのエラーを表しており，その意味でシステム指向の尺度と考えることができる [7]．

ここで再現率，精度，およびフォールアウトの実際的な値の範囲について言及しておく．Belkin [1] によれば，現実のシステムにおける再現率と精度は最良のもので，再現率 40%に対して精度が 60%程度であるという．また最近の TREC-9 [9] の Web 検索の性能を見るに，最良のもの (ric9dpn) で完全に自動のとき再現率 40%に対して精度が 27%である（ユーザのフィードバックを入れて精度を 60%程度にまで上げることができる）．いま NTT が提供する「最新 Web 検索実験サービス (http://infobee.ne.jp/)」で "情報検索" という語を検索すると，4 千万ある Web ページの中から 117,240 件を検索することができる．再現率 40%，精度 27%とすると，$f_{11} = 117,240$，$f_{12} = 175,860$，$f_{21} = 316,982$，$f_{22} = 39,389,918$ であり，このときフォールアウト a は

$$a = 316,982/39,706,900 = 0.008$$

である．"情報検索 & 信頼性" という語を検索すると 1,364 件がヒットし，

$$a = 3,688/39,996,590 = 0.000\ 09$$

である．

情報検索では適合文書に比べ非適合文書の方が圧倒的に多いのが普通（そうでないとそもそも情報検索になり得ない）であり，フォールアウトはその膨大な非適合文書から（誤って）検索する割合を示しているから，$a = 0.01$ であっても十分な絞り込みになっていないことがこの例からもわかる．通常は $a < 0.01$ としてよいと考えられる．

2.2 再現率と精度の要約値

再現率と精度を 1 次元のスカラに変換する方法として Van Rijsbergen の F 尺度が知られる．

$$F_\beta = \frac{(1+\beta^2)pr}{\beta^2 p + r} = \frac{(1+\beta^2)f_{11}}{\beta^2 f_{1\cdot} + f_{\cdot 1}} \tag{4}$$

ここで β は再現率に対して精度を何倍重要視するかの比を示す．たとえば $\beta = 1$ ならば再現率と精度は同程度重要であることを示し，$\beta = 2$ ならば精度は再現率に比べ 2 倍重要であることを示す．特に $\beta = 1$ のとき，(4) 式は

$$F_1 = \frac{2pr}{p + r} \tag{5}$$

となり，再現率 p と精度 r の調和平均と一致する．F 尺度は値が大きいほど良い検索結果であることを示す．

一方，F 尺度の代わりに $1 - F$ を E 尺度と呼び，これが用いられる場合もある．

$$E_\beta = 1 - \frac{(1+\beta^2)pr}{\beta^2 p + r} \tag{6}$$

また E 尺度，F 尺度の如何にかかわらず，β の代わりに

$$\alpha = \frac{1}{\beta^2 + 1} \tag{7}$$

の関係を示す $\alpha(0 < \alpha \leqq 1)$ を用いる場合がある．$\alpha = 1/2$ ならば $\beta = 1$ であり，再現率と精度を同程度に扱うことを意味する．

Van Rijsbergen の F 尺度の他に TREC(Text REtrieval Conference) などでしばしば用いられる評価尺度には，以下の 2 つがある．

- ブレークイーブン·ポイント (break-even point)

再現率-精度グラフを描いたときに，傾き 1 の直線と交わる点，すなわち再現率と精度が一致する点．しかしながら実際のグラフはなめらかな曲線にならないので，適当な補間を必要とする．

- 11 点平均精度 (11-point averaged precision)

再現率が $(0.0, 0.1, 0.2, \ldots, 1.0)$ の 11 点における精度を平均したもの．再現率が 0 の点は理論的には求まらないが，実際は適合文書が最初に検索できたときの精度を用いて近似する．ブレークイーブン・ポイントが再現率と精度が一致する点しか考慮しないのに対し，11 点平均精度では大域的な精度を考慮している．

これら評価基準の詳細については TREC-8 [8] の付録にある Evaluation Techniques and Measures を参照されたい．

3. 関連性の指標

表 1 のような 2×2 分割表は，$k \times \ell$ 分割表の特殊な場合として古くから特別な扱いがなされてきており，2 値変量間の関連性指標としてたとえば統計学辞典 [6] には 21 もの指標が示されている．

このうち最も一般的な関連性指標として，4 分相関係数 (tetrachoric correlation coefficient または four-fold correlation coefficient) と ϕ 係数 (phi correlation coefficient または four-fold point correlation coefficient) の 2 つが挙げられる．

4 分相関係数は x と y の双方がともに 2 分割されたとみなし，分割前の 2 変量正規分布の相関 ρ を考えるものである．ρ を求めるには計算機プログラムが必要で，たとえば StatLib(Applied Statistics algorithms) http://phase.etl.go.jp/stat/apstat/ の 116.f がそれに該当する．(実行オブジェクトを作成するには，正規分布のパーセント点を求める 111.f および正規分布の下側確率を求める 66.f のオブジェクトをリンクさせる必要がある.)

一方，ϕ 係数は，変数 x と y が 2 値しか取りえないときにピアソンの積率相関係数の定義式を適用したもので，

$$\widehat{\phi} = \frac{f_{11}f_{22} - f_{21}f_{12}}{\sqrt{f_{1\cdot} f_{2\cdot} f_{\cdot 1} f_{\cdot 2}}} \tag{8}$$

で与えられる．

本節では，2×2 分割表における関連性指標として最も代表的な 4 分相関係数および ϕ 係数と，情報検索の評価尺度である F 尺度，ブレークイーブン・ポイント，11 点平均精度の関係について調査する．

3.1 4 分相関係数および ϕ 係数の F 尺度との比較

表 1 で与えられる分割表において総データ数 n を固定しても一般性を失うことはない．しかしながら情報検索の場合は，周辺度数（$f_{\cdot 1}, f_{\cdot 2}$）を固定することはできないことに注意する必要がある．たとえばベクトル空間モデル (vector space model) を用いた文書検索では，問い合わせベクトルと似た文書ベクトルをもつ文書を類似文書として検索するわけだが，実際には類似度を示すある閾値を越えた文書を類似文書と判定する．この閾値の設定次第で検索/非検索のそれぞれの総数は変わってしまう．そもそも適合/非適合の文書の判定からして絶対的なものでありえない．したがって例えば f_{11} に関して超幾何分布を想定した統計的解析は不適である．

いま，仮に f_{11} を既知とすれば，$f_{\cdot 1} = f_{11}/p$ より

$$f_{21} = f_{\cdot 1} - f_{11} = f_{11}\left(\frac{1}{p} - 1\right) \tag{9}$$

となる．同様に

$$f_{12} = f_{11}\left(\frac{1}{r} - 1\right) \tag{10}$$

を得る．また $f_{2\cdot} = f_{21}/a$ であるから

$$f_{22} = f_{2\cdot} - f_{21} = f_{11}\left(\frac{1}{p} - 1\right)\left(\frac{1}{a} - 1\right) \tag{11}$$

となる．ここで

$$n = f_{11} + f_{12} + f_{21} + f_{22} \tag{12}$$

であるから，式 (9) – (12) より始めに既知と仮定した f_{11} は，

$$f_{11} = n/\left\{\frac{1}{p} + \frac{1}{r} - 1 + \left(\frac{1}{p} - 1\right)\left(\frac{1}{a} - 1\right)\right\} \tag{13}$$

で示される．ここで f_{11} は r のみの関数であると仮定すれば，f_{11} は定数 c を用いて

$$f_{11} = \frac{n}{\dfrac{1}{r} + c} \tag{14}$$

と書ける．これを r で偏微分すれば，

$$\frac{\partial f_{11}}{\partial r} = \frac{n}{(1+cr)^2} > 0 \tag{15}$$

であるから f_{11} は r に関して単調増加であることがわかる．同様に p に関しても単調増加となる．

さて，再現率 r，精度 p，フォールアウト a の 3 つを定めれば，式 (13), (9)–(11) より $f_{11}, f_{12}, f_{21}, f_{22}$ が一意に定まり，4 分相関係数 ρ，および ϕ 係数を計算することができる．いま $a = 0.01, 0.1, 0.2$ に対して x 軸に再現率 r，y 軸に精度 p をとり，z 軸に 4 分相関係数 ρ，および ϕ 係数を示したのが図 1 および図 2 である．一般に 4 分相関係数と ϕ 係数は類似した指標であろうが，$a < 0.1$ の場合（すなわち通常の情報検索の範囲では）必ずしもそうではなく，これらの図より 4 分相関係数の方がかなり大きめの値が示されていることがわかる．実際，たとえば $r = 0.4$ における ρ と ϕ 係数の差分を計算すると, $p : 0.05 \to 0.50 \to 0.95$ のとき, $(\rho - \phi) : 0.503 \to 0.381 \to 0.346$ である．ρ および ϕ のいずれも，関連度を示す最大で 1 までの値をとる指標であることを考えれば，これらの差分値は全体の 1/3 から半分を占める量であり, かなり大きな値といってよいであろう．

比較のために同じ a に対して z 軸に F 尺度を示したのが図 3 である．(a), (b), (c) は，再現率 r に対する精度 p の重み係数 $\beta = 0.5, 1.0.2.0$ を示す．F 尺度は r と p に関して双対であるので，(a) と (c) はそれぞれ r と p を入れ換えたものに一致する．これより，$a = 0.1$ 以下のときは，図 2 と図 3 が似た形状を示していること，すなわち ϕ 係数と F 尺度（特に $\beta = 1.0$ のとき）とが似た形状を示していることがわかる．たとえば $r = 0.4$ において $p : 0.05 \to 0.50 \to 0.95$ のとき, $(\phi - F_1) : -0.050 \to 0.009 \to 0.031$ である．しかしながら，a が大きくなるにつれ，その近似の度合いは悪くなり，$a = 0.2$ 程度以上ではもはや似ているとはいえないであろう．たとえば $r = 0.4$ において $p : 0.05 \to 0.5 \to 0.95$ のとき, $(\phi - F_2) : 0.159 \to -0.094 \to -0.276$ である．

なお図中に描かれた曲面の中央部と周辺部と色の濃淡が違うのは，曲面の形状が複雑になったとき，すなわち曲面の凸性が喪失されたときにも，その形状がわかるようにとの配慮であり，色の濃淡自体に意味はない．

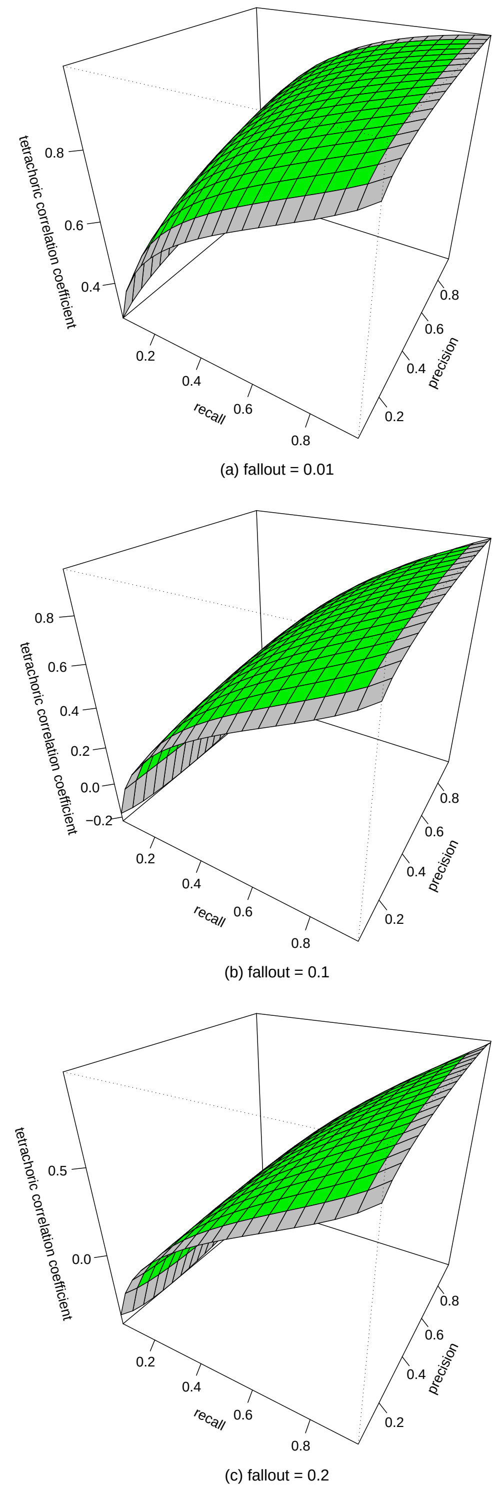

図 1　4 分相関係数 ρ
Fig. 1 Tetrachoric correlation coefficient ρ

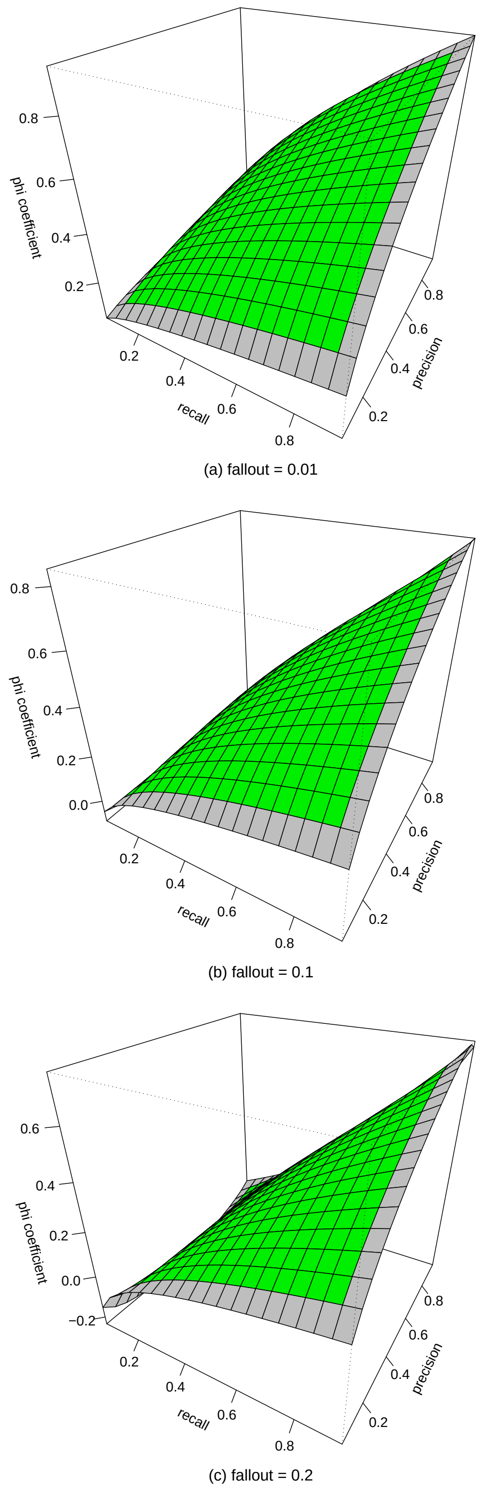

図 2 ϕ 係 数
Fig. 2 Phi correlation coefficient

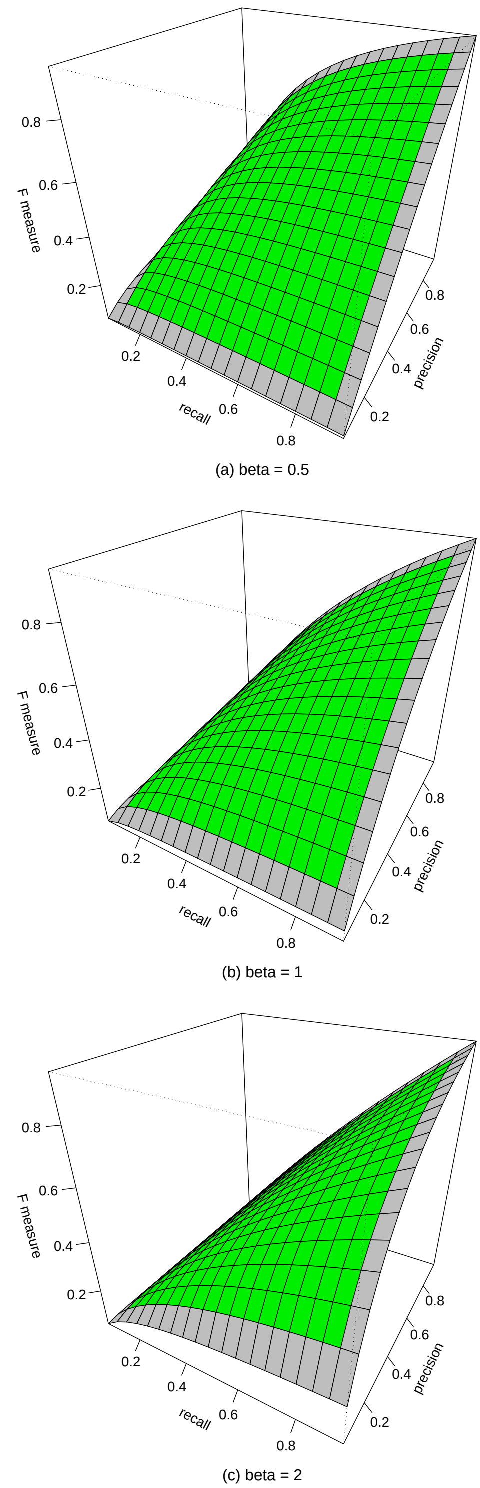

図 3 F 尺 度
Fig. 3 Van Rijsbergen's F-measure

3.2 4分相関係数および ϕ 係数のブレークイーブン・ポイントとの比較

ブレークイーブン・ポイントは再現率 r と精度 p の一致した点のことであるから，ブレークイーブン・ポイントにおける4分相関係数および ϕ 係数は，図1および図2において $r = p$ の平面で切ったときの ρ および ϕ を示せばよいことがわかる．

x 軸にブレークイーブン・ポイントにおける再現率 r または精度 p をとり，y 軸にフォールアウト a，z 軸に4分相関係数 ρ，および ϕ 係数を示したのが図4および図5である．

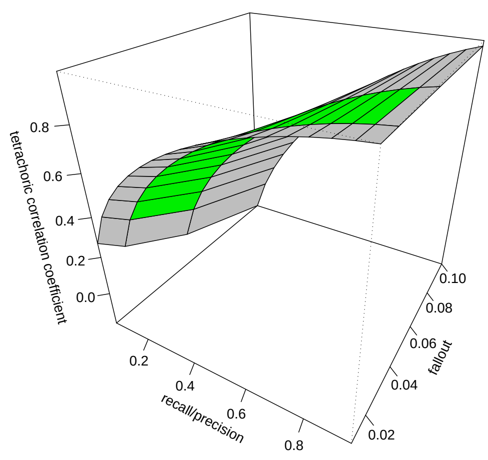

図4 ブレークイーブン・ポイントと4分相関係数
Fig. 4 Break-even point and tetrachoric correlation coefficient

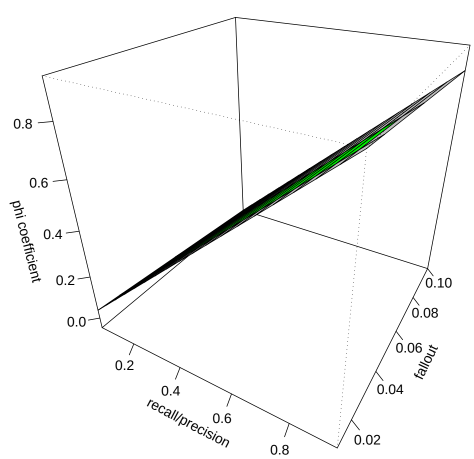

図5 ブレークイーブン・ポイントと ϕ 係数
Fig. 5. Break-even point and phi correlation coefficient

もしブレークイーブン・ポイントが4分相関係数ときわめて似た尺度であるならば，描かれる曲面は $x = z$ なる平面，すなわち $(x, z) = (0, 0)$ と $(x, z) = (1, 1)$ を結ぶ線分を y 軸方向に平行移動してできる平面になるであろう．一般に鳥瞰図において描いた曲面が $x = z$ なる平面より上方にあるか下方にあるかを判別することは難しい．しかしながら視点の方位角 (azimuth) と余緯度 (colatitude) が等しければ，$x = z$ なる平面は $(x, z) = (0, 0)$ と $(x, z) = (1, 1)$ を結ぶ線分にしか見えないはずであるから，この視点においては描いた曲面の $x = z$ なる平面との位置関係を判別することはきわめて容易である．そして本稿における図は全て方位角と余緯度を 30° で同じにしてある．その点に注意して図4および図5を改めてみると，図4における曲面は $x = z$ なる平面より全ての x-y 座標において上方にある，すなわち4分相関係数がブレークイーブン・ポイントより大きな値を得ていることがわかる．一方，図5で示される曲面はまさに $x = z$ なる平面そのものであるといってよいほど類似しており，ϕ 係数はブレークイーブン・ポイントとほぼ同等の指標であることがわかる．

実際に ϕ 係数は r, p, a を用いて表現することができ，式 (1)–(3), (8)–(11) より

$$\phi = (r - a)\sqrt{\frac{1 - p}{r\left\{a + \left(\frac{1}{p} - 1\right)(1 - a)\right\}}} \qquad (16)$$

となる．ここでブレークイーブン・ポイントでは $r = p$ であるから，これを用いて p を消去すれば，

$$\phi = (r - a)\sqrt{\frac{1 - r}{ra + (1 - r)(1 - a)}} \qquad (17)$$

を得る．情報検索において一般に a は小さいので，$a \ll 1$ とすれば，式 (17) において $ra \approx 0,\ 1 - a \approx 1$ より根号の中はほぼ1となり，また $r - a \approx r$ となるから，これより $\phi \approx r$ を導くことができる．

3.3 4分相関係数および ϕ 係数の11点平均精度との比較

11点平均精度は再現率 r が $(0.0, 0, 1. \cdots, 1.0)$ の11点における精度を平均したものであるから，4分相関係数 ρ，および ϕ 係数と比較するためには，ρ および ϕ と a を与えて，$r = 0.0, 0, 1. \cdots, 1.0$ に変えたときのそれぞれの p を計算する必要がある．p を解析的に求めることは難しいが，与えた ρ および ϕ にできる限

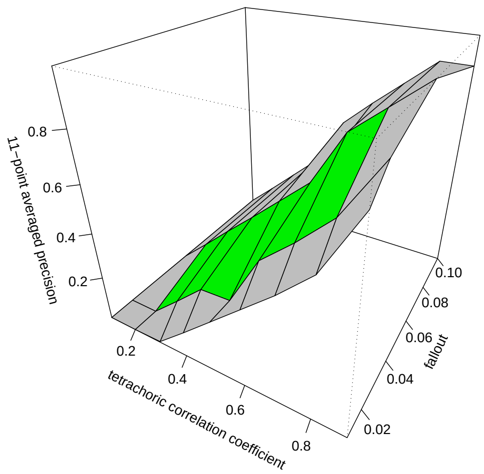

図 6　11 点平均精度と 4 分相関係数
Fig. 6 11-point averaged precision and tetrachoric correlation coefficient

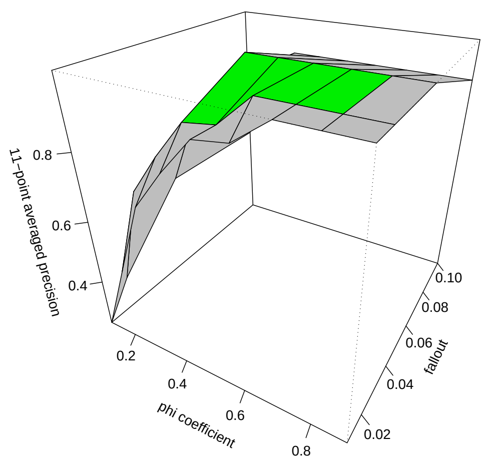

図 7　11 点平均精度と ϕ 係数
Fig. 7 11-point averaged precision and phi correlation coefficient

り近くなる p をその存在区間，たとえば [0.01, 0.99] に対して 2 分割法 [5] などで探索することで，p を数値的に求めることができる．しかしながら図 1–2 からもわかるように，ρ および ϕ は p に関して必ずしも単調でないので，2 分割法で得た p は近似解であるかもしれないが最適解とは限らない．

いずれにせよ数値的に求めた p の 11 点における平均をとることで平均精度 $\bar{p}$ を得ることでき，これを z 軸にとり，x 軸に ρ および ϕ を，y 軸に a を目盛ったのが図 6 および図 7 である．3.2 節と同様に描かれた曲面の $x = z$ 平面との位置的な上下関係を観察するに，図 6 から $a < 0.1$ のどのフォールアウトの範囲に対しても 11 点平均精度 $\bar{p}$ は 4 分相関係数 ρ よりはやや小さく，$a : 0.01 \rightarrow 0.02 \rightarrow 0.05 \rightarrow 0.1$ のとき $\mathrm{Ave}(\bar{p} - \rho) : -0.21 \rightarrow -0.14 \rightarrow -0.014 \rightarrow -0.003$ となる．図 7 から $\bar{p}$ は ϕ 係数よりは大きい値をとり，$a : 0.01 \rightarrow 0.02 \rightarrow 0.05 \rightarrow 0.1$ のとき $\mathrm{Ave}(\bar{p} - \phi) : 0.29 \rightarrow 0.35 \rightarrow 0.42 \rightarrow 0.34$ となる．なお曲面が滑らかでない，すなわち全 x-y 平面座標に対して z が凸でないのは，p の計算に 2 分割法を用いたことによる．

4. おわりに

表 1 に示す交差行列が与えられた場合，一般に良好な検索，すなわち再現率と精度の良い検索が行なわれたとすると，主対角線上にある f_{11} および f_{22} の頻度は大きく，それ以外の f_{12} および f_{21} の頻度は小さくなるであろう．しかし注意しなければならないのは，通常の情報検索では検索されるデータの数は，検索されないデータの数に比べ圧倒的に少ないから，同じ主対角線上にある要素であっても $f_{11} \ll f_{22}$ の関係が成り立つことが多いことである．たとえば前述の “情報検索” という語で検索した場合の例で $f_{11}/f_{22} = 3.0 \times 10^{-3}$ であり，“情報検索 & 信頼性” という語で検索した場合の例で $f_{11}/f_{22} = 3.4 \times 10^{-5}$ である．したがって情報検索でよく用いられる評価基準が，統計学で用いる 2×2 分割表における関連指標と必ずしも類似しないことは容易に予想されていた．本稿は情報検索が通常行なわれる範囲で，すなわちフォールアウト a が 0.1 以下の場合において，情報検索で用いる評価基準と統計学で用いる関連指標とその相互の関係を明らかにするもので，これら指標の大小関係を把握することができたことに価値があるものと考えている．

なお本稿で用いた 3 次元の鳥瞰図を含む全てのプログラムは，統計言語 S のクローンである R により作成されている．R は http://www.r-project.org/より入手でき，Windows および Linux に対してはバイナリが供給されている．

謝辞　本稿執筆の動機は，2000 年 2 月から 12 月にわたる文部省長期在外客員研究員として滞在したカーネギーメロン大学コンピュータサイエンス学部言語技術研究所において，Yiming Yang 博士と討議したことによるものである．ここに記して謝意を表したい．また編集委員ならびに 2 名の匿名の査読者には，論文

の誤りをご指摘いただき，修正すべき事項について多くのご教示をいただいた．心よりお礼申し上げる次第である．

文　献

[1] Belkin, N.J., The Problem of "matching" in Information Retrieval, Theory and Application of Information Research, 187–197, 1980.

[2] Fujita, S., More Reflections on "Aboutness" TREC-2001 Evaluation Experiments at Justsystem, NIST Special Publication 500-250: The Tenth Text REtrieval Conference (TREC 2001), Gaithersburg, Maryland, November 13-16, 2001.

[3] Goldstein, J., Kantrowitz, M., Mittal, M., Carbonell, J., Summarizing Text Documents: Sentence Selection and Evaluation Metrics, 22nd International Conference on Research and Development in Information Retrieval, SIGIR 99, University of California, Berkeley, August 15–19, 1999.

[4] Ingwersen, W., Information Retrieval Interaction, Taylor Graham Publishing, London, 1993.

[5] Rao, S.S., *Optimization, theory and applications*, A Holsted Press Book, John Wiley & Sons, 1979.

[6] 竹内 啓（編），「統計学辞典」, 東洋経済新報社, 1989.

[7] 徳永 健伸, 「情報検索と言語処理」，言語と計算 5，東京大学出版会, 1999.

[8] TREC-8, Evaluation Techniques and Measures, TREC-8 Results, page A-1, NIST Special Publication 500-246: The Eighth Text REtrieval Conference (TREC-8), Gaithersburg, Maryland, November 17-19, 1999. Available online. http://trec.nist.gov/pubs/trec8/t8_proceedings.html

[9] TREC-9, http://trec.nist.gov/pubs/trec9/appendices/A/web_results.html

[10] Quenot, G.M., TREC-10 Shot Boundary Detection Task: CLIPS System Description and Evaluation, 2001. NIST Special Publication 500-250: The Tenth Text REtrieval Conference (TREC 2001), Gaithersburg, Maryland, November 13-16, 2001.

[11] Van Rijsburgen, C.J., *Information Retrieval* (2 ed.) Butterworths, 1979.

（平成 x 年 xx 月 xx 日受付）

石岡　恒憲

1985 年東京理科大学 大学院 修士課程 工学研究科 経営工学専攻 修了．同年，株式会社リコー ソフトウェア研究所（現，研究開発本部 同研究所）．1998 年文部省 大学入試センター（現，独立行政法人 同センター）研究開発部 助教授（現職）．2000 年–2001 年　カーネギーメロン大学コンピュータ・サイエンス学部 言語技術研究所 客員研究員．工学博士．応用統計学会（編集委員），日本信頼性学会（理事，論文審査委員），日本計算機統計学会，日本行動計量学会，日本品質管理学会，各会員．IEEE Trans. on Reliability 誌 論文審査員．