社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－NL－179 (14)
2007－SLP－ 66 (14)
2007／5／25

音声理解を指向した音声認識のための単語重要度の自動推定

七里 崇　　南條 浩輝　　吉見 毅彦

龍谷大学 理工学研究科
〒 520-2194 大津市瀬田大江町横谷 1-5
e-mail: shichiri@nlp.i.ryukoku.ac.jp

内容梗概　情報検索を指向した音声認識のための単語の重要度の自動推定手法を提案する．検索要求中には検索に影響の大きい単語とそうでない単語が存在する．このため，検索要求の音声認識においては，検索に影響の大きい単語の誤りが少ないことが望ましく，近年 ASR の評価尺度として単語の重要度が重みづけされた誤り率である「重みつき単語誤り率（WWER）」が提案されている．しかし，単語の重要度の検討はこれまで十分になされていない．このような背景に基づき，本研究では，各単語の音声認識誤りが検索に与える影響の大きさを用いて単語の重要度を決定する手法を提案する．具体的には，検索要求の音声認識結果の WWER とその認識結果で検索を行った場合の検索精度の低下率の誤差が最小となるように重要度の決定を行う．NTCIR-3 の WEB 検索タスクで評価を行ったところ，WWER と検索スコア低下率の相関係数 0.969 （推定前 0.224 ）が得られ，単語の重要度が適切に求められることがわかった．各検索要求を用いた検索の結果が正しいかどうかの教師信号を与えない場合でも，相関係数 0.712 を得ることができ，本手法が広範囲に応用可能であることを示した．

キーワード　音声理解，音声認識，単語重要度，ベイズリスク最小化デコーディング

Automatic Estimation of Word Significance for Speech Understanding-oriented ASR

Takashi Shichiri　　Hiroaki Nanjo　　Takehiko Yoshimi

Graduate School of Science and Technology,
Ryukoku University
1-5 Yokotani, Oe-cho, Seta, Otsu, 520–2194

Abstract　This paper addresses automatic estimation of word significance for speech understanding-oriented automatic speech recognition (ASR). Since the significance of words differs in speech understanding (SU), ASR performance has been evaluated based on weighted word error rate (WWER), which gives a weight on errors from the viewpoint of SU, instead of word error rate (WER), which treats all words uniformly. A decoding strategy that minimizes WWER based on a Minimum Bayes-Risk framework has been shown, and the reduction of errors on both ASR and SU has been reported. In this paper, we propose an automatic estimation method for word significance (weights) based on its influence on SU. Specifically, weights are estimated so that evaluation measures of ASR and SU are equivalent. We apply the proposed method to a speech-based information retrieval system, which is a typical speech understanding system, and show that the method works well.

Key words　speech understanding, speech recognition, word significance, minimum Bayes-risk decoding

1 はじめに

近年の自然言語処理技術の発展により，音声メディアを処理する研究分野では，音声をテキスト化する音声認識にとどまらず，音声理解のための研究が行われるようになってきた．音声を用いた情報検索では，ユーザの発話から検索要求（意図）の解釈を行う必要があるため，情報検索は音声理解のファーストステップとしてとらえることができる．本研究では，このような音声入力型の情報検索 [1][2] に焦点をあてる．

これまでの情報検索システムは，主として検索対象のドメインが限定されたものであり，代表的なものとしてフライト検索（ATIS）[3] があげられる．このようなシステムでは，基本的に SQL ベースの検索が行われており，検索キーとなるキーワード（「地名」や「日付」など）が明確に定義されている．このため音声認識では，そのようなキーワードのみを認識すれば検索要求を解釈すること（音声理解）ができた．これに対し，本研究ではドメインを限定しない，すなわちオープンドメインの情報検索を考える．オープンドメインの検索タスクとしては，自然言語で記述されたマニュアル [4] や WEB ページ検索などがある．これらのタスクでは，キーワードを明確に定義することができないため，音声認識においてキーワードのみを認識するアプローチは適用できない．文節の係り受けの情報が検索要求の解釈に有効であることが示されており [4]，音声認識では，音声全体を自然言語文として認識する必要がある．

ただし，オープンドメインの検索システムにおいても，明確に定義はされていないものの，検索に決定的な影響を与える単語とそうでない単語の区別が存在する．このため，音声認識の評価は“認識できなかった単語の数（割合）がどれだけ少ないか”ではなく“重要な単語の認識誤りがどれくらい少ないか”という観点から行う必要があり，認識戦略として“重要な単語を優先的に認識する戦略”が必要となる [5]．[5] は，単語の重要度（重み）を単語の認識誤りが検索に与える情報損失の観点から定義し，この重みを用いて，音声認識の評価尺度「重みつき単語誤り率（WWER）」を定義した上で，WWER の最小化を目指す音声認識を行うことで情報損失の最小化，すなわち検索の向上を行うものである．実際に，重要度が高い単語の認識誤りを抑えることができ，重要文抽出及び情報検索の向上に効果があることが示されている．

このように，情報損失の観点から各単語に適切な重みを定義することができれば，情報検索に有効な音声認識が行えることが示されているものの，単語の重み（重要度）の定義方法についての十分な検討はなされていない．

このような背景に基づき，本研究では，情報検索を目的とした音声認識のための単語重みの自動推定手法を提案する．

2 音声認識の評価尺度

2.1 重みつき単語誤り率

音声認識の一般的な評価尺度は単語誤り率（Word Error Rate : WER）であり，式（1）で定義される．

$$WER = \frac{I + D + S}{N} \cdot 100 \tag{1}$$

ここで N は実際に発話された単語の数，S は置換誤りの単語数，D は削除誤りの単語数，I は挿入誤りの単語数を表す．実際の音声認識の評価では，発話された内容を人手で書き起こしたもの（正解文）と音声認識結果を DP マッチングにより対応づけ，WER を算出している．式（1）からもわかるように，WER は，単語の誤りを全て同一に扱う，すなわち重要な単語とそうでない単語を同等に扱う尺度である．これに対し，単語の重要度を考慮した評価尺度として重みつき単語誤り率（Weighted Word Error Rate : WWER）がある [5]．WWER は式（2）で定式化される．

$$\begin{aligned} WWER &= \frac{V_I + V_D + V_S}{V_N} \cdot 100 \qquad (2)\\ V_N &= \sum_{w_i} w_i \\ V_I &= \sum_{\hat{w}_i \in I} \hat{w}_i \qquad V_D = \sum_{w_i \in D} w_i \\ V_S &= \sum_{seg_j \in S} v_{seg_j} \\ v_{seg_j} &= \max\Big(\sum_{\hat{w}_i \in seg_j} v_{\hat{w}_i}, \sum_{w_i \in seg_j} v_{w_i} \Big) \end{aligned}$$

ここで w_i は正解文における i 番目の単語の重みを表し，$\hat{w}_i$ は認識結果における i 番目の単語の重みを表している．また，seg_j は置換誤りの区間を示す．この区間の重みは，認識結果中の置換区間に含まれる単語の重みの合計値と，正解文中の置換区間に含まれる単語の重みの合計値を計算し，合計値の大きい方の値とする．重みつき単語誤り率の計算例を図 1 に示す．全ての単語の重みを等しく（例えば 1）に設定すると，この WWER は WER と一致する．すなわち WWER は，WER の拡張となっている．

ASR result	:	a	b	c	d	e	f	
Correct transcript	:	a		c	d'		f	g
DP result	:	*C*	*I*	*C*	*S*		*C*	*D*

$$\mathrm{WWER} = (V_I + V_D + V_S)/V_N$$

$$V_N = v_a + v_c + v_{d'} + v_f + v_g,\ V_I = v_b$$
$$V_D = v_g,\ V_S = \max(v_d + v_e, v_{d'})$$

v_i: weight of word i

図 1: 重みつき単語誤り率の計算例

2.2 ベイズリスク最小化デコーディング

統計的な音声認識は一般的に，与えられた入力音声信号 X を最もよく説明する単語系列 $\hat{W}$ を求めるプロセスとして式 (3) により定式化される．

$$\hat{W} = \underset{W}{\mathrm{argmax}}\ P(W|X) \tag{3}$$

ベイズ決定理論に基づくと，音声認識は決定規則 $(\delta(X) : X \rightarrow \hat{W})$ と記述できる．ここで，損失関数を $l(W', \delta(X)) = l(W', W)$ とすると，音声認識は以下のベイズリスク最小化の枠組みで記述できる（式 (4)）．

$$\delta(X) = \underset{W'}{\mathrm{argmax}} \sum_{W} l(W', W) \cdot P(W|X) \tag{4}$$

ここで，W，W' はともに単語列である．なお，式 (3) で示されている音声認識のプロセスは，式 (4) において 0/1 損失関数を用いる場合と等価であり，文誤り率を最小化するプロセスであるといえる．

重みつき単語誤り率の最小化を目的とした場合は，式 (4) の損失関数 $l(W', W)$ に重みつき単語誤り率 (WWER) を用いことで，WWER 最小化音声認識ができることが示されている [5]．

3 単語重みの自動推定

本章では，情報検索のための音声認識における単語重みの自動推定手法について述べる．単語重みは，その単語が音声認識で誤認識されたときの，検索に及ぼす影響に基づいて決定する必要がある．本手法では，誤認識が検索に与える影響として検索スコア低下率 (IRDR) を定義し，音声認識の誤り率 (WWER) と検索スコア低下率が等しくなるように行う．検索スコア低下率については，4.2 節で後述する．推定の手順を以下に示す．

1. 検索要求の書き起こしと誤りを含む書き起こし（音声認識結果）のペアを N 個用意する．各検索要求 m ($m = 1 \ldots N$) に対して 2〜5 を行う．
2. 書き起こし文 W_m (誤り率 : 0%) を用いて検索を行い，検索スコア R_m を求める．
3. 誤りを含む書き起こし文 W'_m に対して重みつき単語誤り率 ($WWER_m$) を計算する．
4. W'_m を用いて検索を行い，検索スコア H_m を求める．
5. R_m と H_m から検索スコア低下率 $IRDR_m = 1 - \dfrac{H_m}{R_m}$ を求める．
6. 各検索要求の音声認識誤り率 (WWER) と検索スコア低下率が等しくなるように各単語の重みを決定する．

ここで，手順 6 は WWER と検索スコア低下率の誤差を最小化する問題として定式化できる (式 (5))．

$$F(\boldsymbol{x}) = \sum_{m} \left\{ \left(\frac{E_m(\boldsymbol{x})}{C_m(\boldsymbol{x})} - IRDR_m \right)^2 \right\} \rightarrow \min \tag{5}$$

ここで

$$\frac{E_m(\boldsymbol{x})}{C_m(\boldsymbol{x})} = WWER_m$$

$\boldsymbol{x}$ は各単語の重みを要素とするベクトルであり，m は手順 1 で用意した検索要求の ID である．$E_m(\boldsymbol{x})$ は音声認識単語列の誤り単語の重みの和 (式 (2) の分子) を求める関数であり，$C_m(\boldsymbol{x})$ は，発話の正解単語列の重みの和 (式 (2) の分母) を求める関数である．具体的には，検索要求ペア m の誤り単語に該当する要素のみを 1 とするベクトル $\boldsymbol{e}_m$ と検索要求ペア m の正解単語列に含まれる単語に該当する要素のみを 1 とするベクトル $\boldsymbol{c}_m$ を用意し，$\boldsymbol{x}$ との内積で重みの和を求めている．

本研究では式 (5) を最小化するパラメータ (単語重み) を最急降下法を用いて推定した．具体的には，以下の手順に従いパラメータの推定を行った．

1. 目的関数 (式 (5)) を各変数 x_k (単語重み) について偏微分する (式 (6))．
2. 求まった偏微分係数 $\dfrac{\partial F}{\partial x_k}$ から式 (7) に基づき単語重み x_k を更新する．式 (7) における α は単語重みの更新幅である．本研究では，α を 0.01 とした．
3. 更新後の $F(\boldsymbol{x})$ を計算し，更新前と比較することで重みの更新による変化量を求める．変化量が閾値以下の場合，または学習回数が一定回数に達した場合，処理を終了する．そうでない場合は更新処理を繰り返す．

$$\frac{\partial F}{\partial x_k} = \sum_m 2\left(\frac{E_m}{C_m} - \mathrm{IRDR}_m\right)\cdot\left(\frac{E_m}{C_m} - \mathrm{IRDR}_m\right)'$$
$$= \sum_m 2\left(\frac{E_m}{C_m} - \mathrm{IRDR}_m\right)\cdot\frac{E'_m\cdot C_m - E_m\cdot C'_m}{C_m^2}$$
$$= \sum_m 2\left(\frac{E_m}{C_m} - \mathrm{IRDR}_m\right)\cdot\frac{1}{C_m}\left(E'_m - C'_m\cdot\frac{E_m}{C_m}\right)$$
$$= \sum_m \frac{2}{C_m}\left(\mathrm{WWER}_m - \mathrm{IRDR}_m\right)\left(E'_m - C'_m\cdot\mathrm{WWER}_m\right) \quad (6)$$

$$x_k = \begin{cases} x_k - \alpha & \text{if } \dfrac{\partial F}{\partial x_k} > 0 \\ x_k + \alpha & \text{else if } \dfrac{\partial F}{\partial x_k} < 0 \\ x_k & \text{otherwise} \end{cases} \quad (7)$$

4 情報検索タスクにおける単語重みの自動推定

本章では，単語重みの推定に用いる情報検索システムについて述べ，その後，情報検索タスクであるNTCIR-3 WEB検索タスクを用いて行った評価実験について述べる．

4.1 情報検索システム

情報検索システムには，索引語の頻度分布に基づくシステム[6]を採用する．これは，検索要求Qが与えられると，索引語の頻度分布に基づいてQと検索対象文書iの類似度を計算し，類似度が高い文書から順に出力するものである．類似度には，文書の特徴ベクトル$\boldsymbol{D}_i$と検索要求の特徴ベクトル$\boldsymbol{Q}$の内積を用いた．文書iの特徴ベクトル$\boldsymbol{D}_i$は，索引語のtf*idfの値を要素とするベクトルであり，検索要求の特徴ベクトル$\boldsymbol{Q}$は，検索要求に含まれる索引語に対応する要素の値が1，それ以外が0であるベクトルである．文書iにおける索引語tのtf*idf値は，索引語tの文書iでの出現頻度（Term Frequency:TF）と，全文書中で該当単語tが出現する文書の頻度（Document Frequency:DF）の逆数の積によって定義する（式（8））．

$$tf*idf(t,i) = tf_{t,i}\cdot\log\frac{N}{df_i} \quad (8)$$

t：索引語
$tf_{t,i}$：文書iにおける索引語tの出現頻度
df_t：全検索対象において索引語tを含む文書数
N：文書総数

特定の文書にのみ出現し，かつ文書中での出現頻度が高い単語に対して，このtf*idfの値は大きな値となる．すなわち，tf*idfの値が大きい単語は，その文書を特徴づける単語といえる．ここでは，さらに，tf*idfの値を文書長で正規化する（式（9））．これは，長い文書ほど単語数が多く（tfの値が大），tf*idfの値が大きめに算出されるので，これを補正するためである．

$$tf*idf(t,i) = \frac{tf_{t,i}}{\frac{DL_i}{avglen} + tf_{t,i}}\cdot\log\frac{N}{df_t} \quad (9)$$

t：索引語
$tf_{t,i}$：文書iにおける索引語tの出現頻度
df_t：全検索対象において索引語tを含む文書数
N：テキスト総数
DL_i：文書Diの文書長
$avglen$：全文書の平均文書長

4.2 検索結果の評価尺度

検索結果の評価には，多値適合レベルに適した評価尺度の1つであるDCG（Discount Cumulative Gain）[7]を用いる．DCGは式（10）によって求められる．

$$dcg(i) = \begin{cases} g(1) & \text{if } i = 1 \\ dcg(i-1) + g(i)/log_b(i) & otherwise \end{cases} \quad (10)$$

$$g(i) = \begin{cases} h & \text{if d}(i) \in H \\ a & \text{else if d}(i) \in A \\ b & \text{else if d}(i) \in B \\ c & \text{otherwise} \end{cases} \quad (11)$$

ここでiは検索結果の出力順位を表し，$d(i)$は検索結果の上位i番目の文書を示す．H，A，Bはそれぞれ高適合，適合，部分的適合と判定された文書集合（正解データ）を示す．H，A，Bのいずれにも該当しない検索結果は不適合とする．DCGの値は，高適合の検索結果が上位に多くなるほど大きくなる．なお，本研究では式（11）で定義される利得の度合い，（h，a，b，c）はそれぞれ（3，2，1，0）とした．また，式（10）における対数関数の底bは2とした．

4.3 検索スコア低下率

次に，単語重みの推定に用いる検索スコア低下率について述べる．検索スコア低下率Sは式（12）によって定義する．

$$IRDR = 1 - \frac{H}{R} \quad (12)$$

Rは誤りを含まない検索要求（テキスト）での検索結果のDCGスコアである．Hは誤りを含む検索要求（音声認識結果）での検索結果のDCGスコアである．

4.4 評価実験

4.4.1 NTCIR3 WEB検索タスク

評価実験には，NTCIR3-WEB検索タスクの音声入力タスクで提供されているデータセット（テストコレクション）を用いる．データセットは以下の要素で構成される．

- 検索対象文書：jpドメイン上のWEBページ（100G）
- 利用者の検索要求を記述した「検索課題」：47課題の音声データ（10人分：計470発話）
- 検索課題を満たす「正解文書のリスト」

4.4.2 一般的なASRの評価尺度と検索スコア低下率の相関

検索要求ペアの作成（3章，手順1）のために以下の処理を行った．まず，47課題の書き起こしを用いて検索を行い，次に，検索要求470発話（47課題×10名分）の認識結果を用いて検索を行った．得られた470件のペアから，書き起こしの検索要求で検索結果が全く得られなかったもの，音声認識誤りがないもの，10人分の音声認識結果のうち，認識結果が同じものを除いた．最終的に検索要求ペアの数は，107件となった．

検索要求107ペアの音声認識結果を従来のASRの評価尺度である単語誤り率で評価し，検索スコア低下率との関係を調べた．結果を図3に示す．両者の相関係数は0.119であり相関は弱いことがわかる．これは，実験に用いた検索システムが名詞のみを索引語（キーワード）とする検索システムであるためである．従来，このような検索では，キーワード誤り率を評価尺度とする音声認識が行われてきた．キーワード誤り率は，キーワードのみを対象として算出される誤り率であり，これは重みつき単語誤り率の計算において，キーワードにすべて同じ重み（例えば1）を与え，そうでない単語には0を与えた場合と同じである．キーワード誤り率と検索スコア低下率との関係を図4に示す．単語誤り率と比べると，キーワード誤り率と検索スコア低下率の相関は高いものの相関係数自体は0.224と小さい．これらの結果は，情報検索を対象とした場合，単語の重要度を考慮しない従来の音声認識の評価尺度は適切でないことを示している．

4.4.3 重みつき単語誤り率と検索スコア低下率の相関

次に，107件の検索要求ペアに対し提案手法を用いて単語重みの推定を行った．推定後の単語重みを用いて算出した音声認識の誤り率（WWER）と検索スコア低下率との関係を図5に示す．相関係数は0.969であり，強い相関が得られるようになっていることがわかる．このことは，重みの推定が適切に機能したことを示している．推定結果の例を図2に示す．

正解文：キリストの復活を祝うイースターの祭りについて書かれている文書を探したい
認識文：キリストの復活大巌Pさんの祭りについて書かれている文書を探したい

誤り単語（置換誤り）：を祝うイースター→大巌Pさん
正解文での検索スコア：13.9
認識文での検索スコア：3.2　　検索スコア低下率：76.98%
推定後の単語重み："イースター"(1.00) → (11.00)

図2: 単語重みの推定例

最後に教師なし重み推定について述べる．単語重みの推定のためには検索スコア低下率$IRDR$，すなわち書き起こし及び音声認識結果を用いて検索を行った結果の検索スコアR_mとH_mが必要である．このR_mとH_mを求めるためには，検索結果が正しいかどうかの教師信号が必要である．この教師信号の作成（検索結果の正誤判定）は非常にコストが高いため，大規模システムへ本手法を適用する際には，この教師信号を人手で与えない方法（教師なし重み推定）が必要となる．教師なし重み推定の手順を以下に示す．

1. 書き起こしW_mを用いて検索を行い，疑似正解セットを作成する．このとき全ての検索結果を部分適合とする．
2. 疑似正解セットを用いて，書き起こしW_mを用いた検索結果のDCGスコアR'_mを求める．
3. 疑似正解セットを用いて，音声認識結果W'_mを用いた検索結果のDCGスコアH'_mを求める．
4. R'_mとH'_mから検索スコア低下率$IRDR'_m = 1 - \frac{H'_m}{R'_m}$を求める．

5. $IRDR'_m$ と $WWER_m$ の誤差が最小になるように単語重みを決定する．

教師なし重み推定の結果を図 6 に示す．教師なし学習により推定した単語重みを用いた場合でも相関係数 0.712 が得られた．このことは，本手法が大規模システムにも応用可能であることを示している．

5 おわりに

情報検索を指向した音声認識のための単語重要度の自動推定手法を提案した．様々な検索要求の音声認識結果とそれを用いた検索結果から音声認識の評価尺度である重みつき単語誤り率（WWER）と検索スコア低下率を計算し，これらの誤差を最小とする重みを最急降下法により求める方法を提案した．

NTCIR-3 WEB 検索タスクにおいて，WWER と検索スコア低下率の相関を高くする単語重みの推定が行えることを示した．また，検索結果が正しいかどうかの教師信号（正解データ）が与えられない場合でも WWER と検索スコア低下率の相関を高くする単語重みの推定が行えることを示し，全ての検索要求に対する正解データの用意が困難な大規模な検索タスクにおいても WWER 最小化音声認識に有効な単語重みの推定が行えることを示した．

参考文献

[1] 翠輝久, 駒谷和範, 清田陽司, 河原達也. 音声対話によるソフトウェアサポートのための効率的な確認戦略. 信学論, Vol.J88-DII, No.3, pp.499–508, 2005.

[2] 西崎博光. 音声文書を対象とした音声入力型情報検索システムに関する研究. 豊橋科学技術大学, 博士論文, 2003.

[3] E.Levin, S.Narayanan, R.Pieraccini, K.Biatov, E.Bocchieri, G.DiFabbrizio, W.Eckert, S.Lee, A.Pokrovsky, M.Rahim, P.Ruscitti, and M.Walker. THE AT & T-DARPA communicator mixed-initiative spoken dialog system. In *Proc. ICSLP*, vol.2, pp.122-125, 2000.

[4] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答—ダイアログナビ. 自然言語処理, Vol. 10, No. 4, pp. 145-175, 2003.

[5] 南條浩輝, 翠輝久, 河原達也. 単語重要度を考慮したベイズリスク最小化音声認識とそれに基づく情報検索. 信学技報, NLC2005-66, pp55-60, 2005.

[6] 伊藤克亘, 藤井敦. NTCIR-3 ワークショップにおける音声入力型ウェブ検索タスク. 情処学研報, 2002-SLP-43, pp.25-32, 2002.

[7] 江口浩二, 大山敬三, 石田栄美, 神門紀子, 栗山和子. NTCIR-3 WEB: Web 検索のための評価ワークショップ. In *NII Journal*, No.6, pp.31-56, 2003.

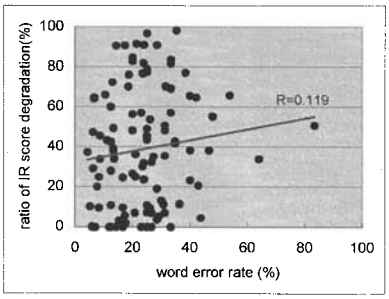

図 3: 単語誤り率と検索スコア低下率の相関

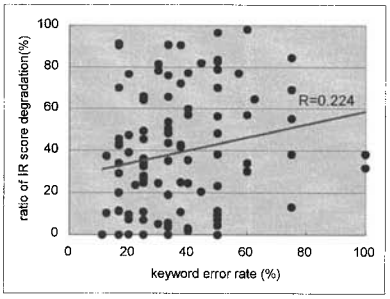

図 4: キーワード誤り率と検索スコア低下率の相関

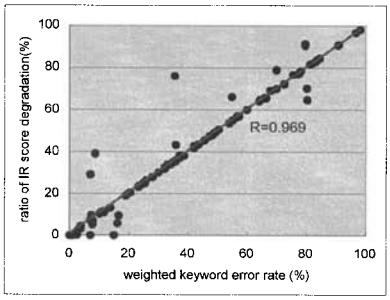

図 5: 重みつき単語誤り率（教師あり推定）と検索スコア低下率の相関

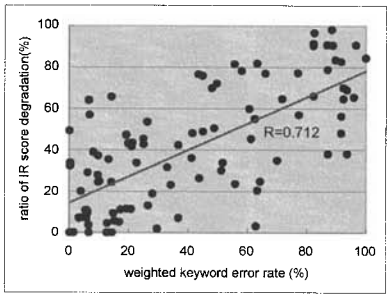

図 6: 重みつき単語誤り率（教師なし推定）と検索スコア低下率の相関