Vol.2009-SLP-77 No.14
2009/7/18

音声情報案内システムにおける SVM を用いたタスク外発話の検出

藤田洋子[†1] 竹内翔大[†1] 川波弘道[†1]
松井知子[†2] 猿渡洋[†1] 鹿野清宏[†1]

多様なユーザ発話に簡便かつ頑健に応答できる音声対話システムの実現手法として用例ベースの応答手法がある．質問応答データベース（Question and Answer Database: QADB）を用いる用例ベースの応答手法であれば，応答文を追加することによって応答内容が容易に拡張でき，質問例の拡張収集によって言語的なゆらぎに頑健なシステムとすることができる．ただし，このようなシステムは想定内の発話（タスク内発話）に対して柔軟な応答が可能であるが，想定されていない発話（タスク外発話）に対しては応答誤りが発生する．そこで，タスク外発話が生じた場合は，QADB による応答ではなく，Web 検索などの別処理によって応答を代替することを考える．本稿ではその前処理として入力発話の中からタスク外発話を検出する手法を検討した．実環境で稼動している音声情報案内システムから得られたデータを用い，Bag Of Words（BOW）を特徴量とした Support Vector Machines（SVM）によるタスク内発話とタスク外発話の分類を試みた結果，従来手法よりも高いタスク外発話の判定精度が得られた．

Detection of Out-of-Task Utterances Using SVM for Speech-Oriented Guidance Systems

Yoko FUJITA,[†1] Shota TAKEUCHI,[†1]
Hiromichi KAWANAMI,[†1] Tomoko MATSUI,[†2]
Hiroshi SARUWATARI[†1] and Kiyohiro SHIKANO[†1]

Example-based response selection approach is applied to a spoken dialogue system. This method can effectively cope with a variety of users' utterances. There are two advantages in this method using question and answer database (QADB). First one is that it's easy to extend the response categories of the system. Second one is that the system can respond to a variety of utterance expressions by using example collection. This method can respond to expected utterances (in-task utterances). However, this method cannot respond to unexpected utterances (out-of-task utterances), which cause inappropriate response. To deal with unexpected utterances, we introduce a response generation not using QADB, but for example Web retrieval. In this paper, we propose out-of-task utterances detection using Support Vector Machine (SVM) which bag-of-words (BOW) is employed as the SVM input feature. The experimental result employing the real-environmental data shows that the proposed method has better detection performance than that of the conventional one.

1. はじめに

近年，音声認識技術を用いた製品や音声検索などの技術に注目が集まっている．カーナビや携帯電話を用いたサービスに音声認識を用いる利点としては，入力に手を使う必要がないこと，音声がほとんどの人間にとって自然に扱えるインターフェースであることなどが挙げられる．これらの理由から音声情報案内システムなどをはじめとする音声対話システムに大きな期待が寄せられている．

実用的な音声情報案内システムの実現のためには，ユーザ発話の多様性への対応やシステム応答を拡張する際に必要なコストを検討する必要がある．この 2 つの問題に対応した音声対話システムの応答手法の 1 つに，質問応答データベース（QADB）を用いる手法がある．この手法では質問例と適切な応答のペアをデータベース化した QADB を用いたデータベース検索によって，入力発話に対しもっとも類似した質問例を選択し，応答を行う．我々が開発・運用を行っている音声情報案内システム「たけまるくん」[1] も QADB を用いた音声対話システムの 1 つである．このような用例ベースの音声情報案内システムは，QADB 中にある想定内の発話（タスク内発話）に対しては応答可能だが，QADB にはない想定外の発話（タスク外発話）には対処できないという問題がある．しかし，このようなタスク外発話に対しては，後段に Web 検索などの処理を入れることにより対処できる可能性がある．

そこで本研究では，タスク外発話に対しては後段において Web 検索などの処理を用いて応答することを想定し，その前段として入力発話の中からタスク外発話の検出を行う手法を検討した．以下，2 章では音声情報案内システム「たけまるくん」と「たけまるくん」から

†1 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of science and Technology
†2 統計数理研究所
The Institute of Statistical Mathematics

Vol.2009-SLP-77 No.14
2009/7/18

得られたデータの説明を行い，3 章では従来のタスク外発話検出法について説明する．4 章では本稿で提案するタスク外発話検出の手法を述べ，5 章にてその性能評価実験の結果を示し，6 章で結果をまとめる．

2. 音声情報案内システム「たけまるくん」

2.1 システムの概要

音声情報案内システム「たけまるくん」（図 1）は，2002 年 11 月より奈良県生駒市にある生駒市北コミュニティセンター ISTA はばたきに常設しているシステムであり，施設の利用者に対して施設・観光情報案内を行っている．対話戦略としては一問一答形式を取っており，情報案内の他に時間や天気，エージェント自身に対する質問などに応答できる．また「検索開始」というキーワード入力により，音声検索モードに切り替わり，ユーザの任意の発話に対する Web 検索も可能である[2)]．

「たけまるくん」の処理の流れを図 2 に示す．まず，マイクロホンより入力された音声には，Gaussian Mixture Model（GMM）による雑音棄却処理と音響尤度による年齢層識別が行われる．次に年齢層別に用意された音響モデルと言語モデルを用いて音声認識が行われる．この時，得られた音声認識結果と QADB の質問例を用いて式 (1) により類似度スコアが算出される[3)]．式 (1) は音声認識結果と質問例との形態素単位での一致数を求め，それを質問例の形態素数か音声認識結果の平均形態素数の最大値で除算した値である．システムの応答としては最近傍法（Nearest Neighbor Method: NNM）により類似度スコアがもっとも高い質問例に対応した応答が選択される．

$$\text{類似度スコア} = \frac{\text{形態素単位での一致数}}{\max(\text{質問例の形態素数}, \text{音声認識結果の平均形態素数})} \quad (1)$$

2.2 データベース

「たけまるくん」は 2002 年 11 月の運用開始以後，すべての入力データを収録している．この内，2002 年 11 月から 2004 年 12 月の間に収集された約 2 年分の発話データに対しては，聴取によるマニュアルでの書き起こし作業及び年齢性別ラベル，雑音ラベルなどのラベル付与作業が終了している．

現在の「たけまるくん」で使用されている言語モデル，音響モデル，QADB は，2002 年 11 月〜2004 年 10 月分の発話データを用いて構築されている．この時，大人の発話と子どもの発話とでは発話内容の傾向が異なるため, モデル及び，データベースは年齢層別に構築さている[1)]．

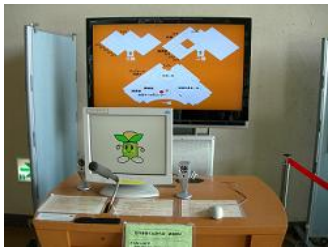

図 1 音声情報案内システム「たけまるくん」
Fig. 1 Speech-oriented guidance system “Takemaru-kun.”

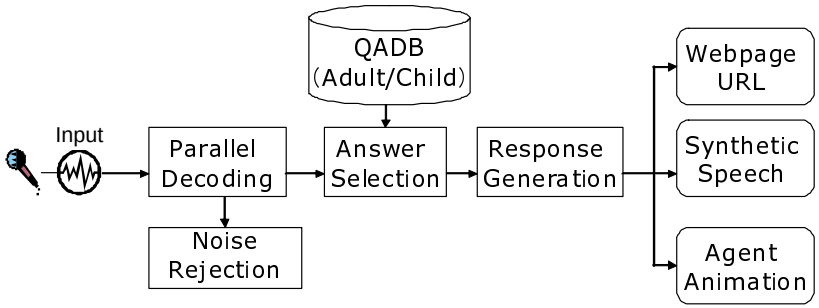

図 2 「たけまるくん」の応答処理の流れ
Fig. 2 Processing flow of “Takemaru-kun.”

2.3 タスク外発話

「たけまるくん」の運用においては，開発者側でタスクを限定せずに，実際のユーザ発話に基づき，応答できる内容を拡張していくという方針を取っている．そのため本稿におけるタスク外発話とは，その時点の QADB に対応する応答情報がないためシステムが処理できない発話を指すものとする．これまでの運用では，タスク外発話であってもユーザの意向を反映した応答を行わせるため，得られた音声データの書き起こし文にもっとも適していると考えられる応答文を追加してきた．具体的には，コミュニティセンターの案内システムにおいて明らかに不要と考えられる質問に対しては,「分かりません」という応答を行うことで QADB の枠組みを壊すことなくタスク外発話に対処してきた．なお現時点におけるタスク

Vol.2009-SLP-77 No.14
2009/7/18

表 1 タスク内発話とタスク外発話の一例
Table 1 Example of in-task utterances and out-of-task utterances

タスク内発話	タスク外発話
こんにちは	今流行っている映画は？
お名前は何ですか？	デービットベッカムはどこにいますか？
近くのバス停はどこ？	ボーダフォンショップはどこにありますか？
図書館の案内をお願いします	大和信用金庫はどこですか？

表 2 タスク外発話の種類
Table 2 The type of out-of-task utterances

質問の種類	発話数	発話例
流行情報	78	ポケモン知っていますか？
地域情報	91	真弓中央公園を見せて下さい
人名・組織名	109	ベッカム
一般名詞	224	肥満対策
上記以外の質問	159	ゴルフのルールを教えて
「たけまるくん」に対する質問	183	野球はできますか？
それ以外（発話の一部）	50	えーと
総発話数	894	-

外発話として判断されている発話データの例を表 1，2 に示す．

3. 従来のタスク外発話の検出方法

タスク外発話を検出する従来法として 2 つの手法を説明する．

3.1 タスク外質問例を用いる方法

現在運用している「たけまるくん」において，タスク内外の判定は明示的に行っていない．表 2 で示しているようなタスク外発話に対しては，2.3 節で述べたように応答不可能な質問例として「分かりません」という応答を付与してきた．このような質問例すべてがタスク外発話を定義しているクラスと考えると，現在の応答手法によりタスク外発話を判定することができる．

タスク外質問例を用いる方法： 現在の応答選択法では，入力発話の音声認識結果と QADB 中の質問例の類似度スコアを式 (1) により算出し，その値のもっとも高いものをユーザ質問として捉える．この際に類似度が高い質問例がタスク外発話の質問例であれば，それをタスク外発話と判定する．なお，この手法には，QADB に収集されたタスク外発話は正しく検出できるが，まったく未知のタスク外発話に対する検出精度は低いという問題点がある．

3.2 類似度スコアによる閾値棄却に基づく方法

これまでの類似度スコアの観察から，タスク外発話には適切な質問例が QADB に存在しないため，類似度スコアが低い傾向にあることが分かっている．また，同じ類似度スコアであっても，認識結果の入力形態素数が長い場合には誤った応答が多くなるという観測結果も得られている．そこで，入力形態素数と類似度スコアの分布から類似度スコアに対する入力形態素数の閾値を求め，誤った応答が選ばれたかどうかを判定する手法が早川らによって提案されている[3]．

類似度スコアによる閾値棄却に基づく方法： 今回は早川らの手法をタスク外発話検出に用いる．本稿においてタスク外発話は QADB にはない発話として定義しているため，類似度スコアに対する入力形態素数の閾値を式 (2) と学習データから求め，タスク内発話とタスク外発話の分類を行う．なお，この手法における QADB は学習データの書き起こし文の中からタスク外発話（今まで応答を「分かりません」としていた発話）をのぞき，学習データ中のタスク内発話だけで作成した．この方法は，式 (2) の y 以上の入力形態素数を持つ発話をタスク外発話として判定する．

$$y = ax + b \tag{2}$$

ここで x は類似度スコア，a と b は任意の数である．

4. SVM を用いたタスク外発話検出方法

タスク外発話を検出する問題は，入力発話がタスク内発話とタスク外発話のどちらのクラスに属するのかを分類する二値分類問題と見なせる．そこで，この 2 クラスの分類手法として構文解析や文章の抽出手法[4][5] として有効である SVM を用いる手法を提案する．

4.1 SVM

SVM は，二値分類問題を解くための統計学習理論に基づく教師あり学習アルゴリズムである[6]．与えられた n 次元の特徴量ベクトル $\mathbf{x_i} \in R^n, i = 1, ..., l$ とラベル $y_i \in \{1, -1\}$ のペア集合より，この 2 クラスを分類できる識別境界を求める．この時，SVM では両クラスに対するマージンが最大となる識別境界を求める．その際にある程度の誤分類を許可し，スラック変数 ξ_i を加えてソフトマージンを導入することができる．またここでは，正例と負例の数が顕著に異なるアンバランスなデータの問題を扱える SVM を用いる（式（3）参照）．

Vol.2009-SLP-77 No.14
2009/7/18

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C_+\sum_{y_i=1}\xi_i + C_-\sum_{y_i=-1}\xi_i \tag{3}$$

$$\text{Subject to} \quad y_i(\mathbf{w}^T\phi(\mathbf{x_i}) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, i = 1, ..., l.$$

C_+ : 正例のコストパラメータ

C_- : 負例のコストパラメータ

コストパラメータ C と C_+, C_- の間には式 (4) のような関係が成り立っている．本実験では，C に関しては予備実験から事後的に設定し，C_+, C_- は式 (5),(6) に従ってデータにより算出した．

$$C = C_+ + C_- \tag{4}$$

$$C_+ = \frac{y_i = -1 \text{ のデータ数}}{\text{全データ数}} \times C \tag{5}$$

$$C_- = \frac{y_i = 1 \text{ のデータ数}}{\text{全データ数}} \times C \tag{6}$$

4.2 特 徴 量

SVM において精度をもっとも左右する要素は特徴量の選択である．本論文中では，各データを表す特徴量ベクトル $\mathbf{x_i}$ として以下に述べる特徴量を検討した．

- 1 仮説あたりの平均形態素数（以下，平均形態素数）
  10-Best までの音声認識結果を用いた場合における 1 仮説あたりの平均形態素数.
- 類似度スコア
  質問例と，入力文との類似度を式 (1) により計算したスコア.
- Bag Of Words（BOW）
  入力文に含まれている単語の出現頻度を表したベクトル.（出現頻度を数えるのは Wordlist 中の単語のみである．Wordlist は学習データ中のタスク内発話に含まれている単語から作られている.）
- 未知語の出現頻度（以下，未知語クラス）
  入力文に含まれている単語中において，Wordlist の中にはない単語の出現回数.

4.3 SVM を用いたタスク外発話の検出方法

SVM を用いる方法：まず，学習・テストデータから 4.2 節で示した特徴量ベクトルを抽出する．次に，学習用の特徴量ベクトルを用いて，タスク内発話とタスク外発話を分類するための SVM のパラメータ（4.1 節）を推定する．この SVM を使用してテストデータの分類を行う.

表 3 学習データとテストデータ（大人データ）
Table 3 Training data and test data (for adult)

	タスク内発話数	タスク外発話数	総発話数
学習データ	18516	847	19363
テストデータ	1026	47	1073

5. 評 価 実 験

本実験における目的は次の 2 つである．

- 従来手法と提案手法の分類精度を比較する．
- タスク外発話を検出するために有効な SVM の特徴量を検討する．

この 2 つを踏まえた上で，タスク外発話を検出するために適した SVM のパラメータを調査する必要がある．

なお今回の実験においては C_+, C_- のパラメータは式 (5), (6) から算出した値を用い，C は 10〜10000 の値を 10 倍刻みで与え，その中でもっとも良い結果を示す値を事後的に設定した．

5.1 使用データ

本実験において使用したデータを表 3 に示す．このデータは聴取により大人と判別されたデータである．

5.2 実 験 内 容

まず従来手法と提案手法におけるタスク外発話検出の精度比較を行う．従来手法は NNM により音声認識結果ともっとも類似した質問例を選択する手法であるが，タスク外発話例を用いる手法と類似度スコアによる閾値を用いる手法がある．これに対し提案手法としては，類似度スコアと平均形態素数を特徴量として SVM による分類を行う手法と BOW を特徴量として SVM による分類を行う手法がある．今回はこれらの手法による分類精度を比較する．その他の実験条件については表 4 に示す.

5.3 評 価 方 法

本実験において，評価尺度には次の 2 種類を用いている．

- 平均精度（Average Precision: AP）
  AP は，Web 検索などのシステム性能評価の際によく使用されている指標の 1 つであ

表 4 実験条件
Table 4 Experimental condition

音声認識エンジン	Julius3.5.3
形態素解析器	Chasen2.3.3
SVM ツール	LIBSVM[6]
カーネル関数	Radial Basis Function（RBF）
比較手法	従来手法 1：NNM（タスク外発話質問例） 従来手法 2：NNM（類似度スコア閾値） 提案手法 1：SVM（類似度スコア，形態素数） 提案手法 2：SVM（BOW） 提案手法 3：SVM（BOW，未知語クラス）
パラメータ C	10〜10000（10 倍刻み）

る[7]．検出したいデータがどの程度の正しさで検出できているのかを示している（式 (7) 参照）．

$$AP(\rho) = \frac{1}{|R|}\sum_{k=1}^{A}\frac{|R \cap \rho^k|}{k}\psi(i_k) \tag{7}$$

R : タスク外発話の集合
A : 全発話数
ρ^k : SVM スコアを昇順に並べた時の，k 番目までの発話の集合
$|*|$: | 集合の要素数
i_k : 発話番号

$$\psi(i_k) = \begin{cases} \psi(i_k) = 1 & i_k \in R \\ \psi(i_k) = 0 & otherwise \end{cases}$$

- 等誤り率（Equal Error Rate: EER）
今回の二値分類の問題においては，タスク外発話をタスク内発話と誤分類する誤り率（False Acceptance Rate: FAR）とタスク内発話をタスク外発話と誤分類する誤り率（False Rejection Rate: FRR）がある．この 2 つの誤り率が等しくなる値が EER である．FAR は式 (8)，FRR は式 (9) より求める．

$$FAR = \frac{\text{タスク外発話をタスク内発話と誤分類した数}}{\text{タスク外発話数}} \tag{8}$$

$$FRR = \frac{\text{タスク内発話をタスク外発話と誤分類した数}}{\text{タスク内発話数}} \tag{9}$$

表 5 タスク外発話検出の実験評価
Table 5 Result of out-of-task utterances detection

		AP		EER	
	パラメータ C	学習データ	テストデータ	学習データ	テストデータ
NNM（タスク外質問例）	—	0.07	0.04	51.45%	47.50%
NNM（類似度スコア閾値）	—	0.14	0.15	22.77%	26.83%
SVM（平均形態素数，類似度スコア）	100	0.22	0.22	21.13%	22.90%
SVM（BOW）	100	0.57	0.36	4.01%	16.86%
SVM（BOW，未知語クラス）	1000	0.85	0.35	0.83%	15.69%

5.4 実験結果

実験結果を AP により評価したものを図 3 に，EER により評価したものを図 4 に示す．なお AP，EER の正確な値及び SVM による分類時に適していた C の値を表 5 に示す．SVM を用いた本提案手法はどちらの評価尺度においても従来手法より改善した結果を得た．

提案手法である SVM を用いる手法のうち，平均形態素数と類似度スコアを特徴量とした場合と BOW を特徴量とした場合の結果を比較する．図 3，図 4 から，BOW は平均形態素数と類似度スコアを特徴量とした結果よりも高い分類性能を持ち，特徴量として有効であることが分かる．つまり，音声情報案内システムに対するユーザ発話のような短い文であってもタスク内発話とタスク外発話とでは，単語の組み合わせに異なる傾向が見られることが判明した．しかし，BOW を用いた手法は学習データに対しテストデータの AP 値が低く，全体的に AP は向上しているものの学習データとテストデータの精度に落差がある．

また同じ BOW の結果を比較した場合であっても，未知語クラスの有無により学習データとテストデータの分類精度に差異が見られる．図 3 を見ると，BOW に未知語クラスを加えた場合，学習データに対しては AP に改善が見られるが，テストデータに対する AP はほとんど変わらない．これは学習データにおける未知語クラスの傾向とテストデータにおける未知語クラスの傾向が異なっていることが原因だと考えられる．学習データに含まれていた未知語の大半がタスク外発話に含まれる単語であったのに対し，テストデータ中の未知語クラスには学習データにはなかったタスク内発話の言い回しなどが含まれていた．しかし，図 4 を見ると，未知語クラスを含めることにより EER の値は 1 %程度改善している．これはタスク外発話の検出数こそあまり変わらないが，FRR が減少していたためであった．

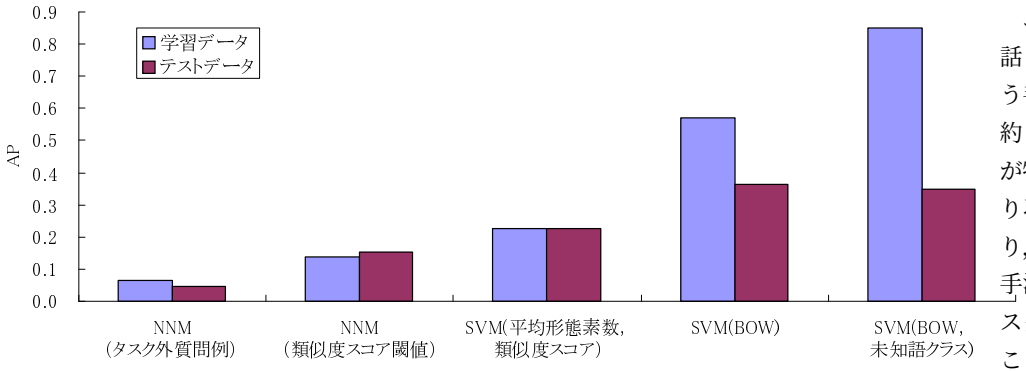

図 3 タスク外発話検出の実験評価（AP 評価）
Fig. 3 Average precision evaluation result for out-of-task utterances detection.

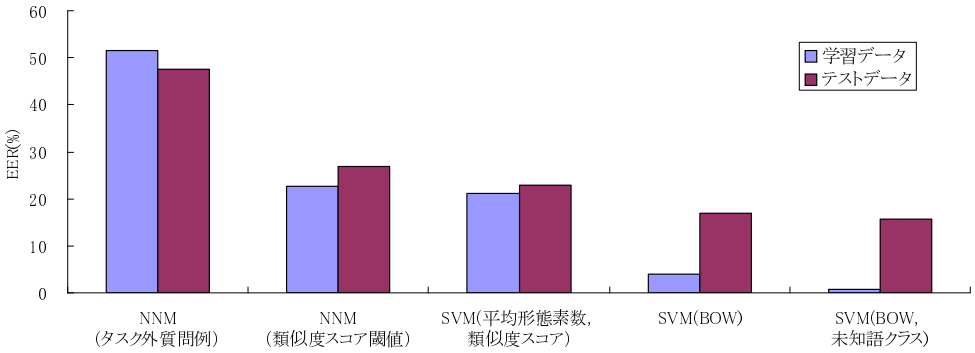

図 4 タスク外発話検出の実験評価（EER 評価）
Fig. 4 Equal error rate evaluation result for out-of-task utterances detection.

6. ま と め

ユーザ発話の中からタスク外発話を検出するための手法として，SVM によるタスク内発話とタスク外発話の分類を行った．提案手法（BOW を特徴量として SVM による分類を行う手法）は従来手法（タスク外発話質問例を用いる手法）と比べ，AP 値で約 0.3，EER で約 30 ポイントの改善が見られた．この結果から質問のような比較的短い文であっても BOW が特徴量として有効であることが示された．しかし，BOW のような特徴量を扱うことにより次元数が膨大な数に膨れ上がり，学習効率が減少している．また学習データの不足により，うまく分類できないデータができている．そこで，今後は複数の特徴量を組み合わせる手法について検討を進めるつもりである．類似度スコアや平均形態素数とタスク内発話，タスク外発話の関係はすでに明らかとなっている部分もある．これらの特徴量を組み合わせることにより，BOW だけでは分類することのできなかったタスク外発話も分類できる可能性がある．

参 考 文 献

1) R. Nisimura, A. Lee, H. Saruwatari, K. Shikano: Public Speech-oriented Guidance System with Adult and Child Discrimination Capability, *In Proc. ICASSP* 2004, pp.433-436, 2004.
2) 三宅純平，竹内翔大，川波弘道，猿渡洋，鹿野清宏：括弧表現に基づく Web テキストマイニングを用いた流行語への自動読み付与の提案，電子情報通信学会技術研究報告，SP2008-126，Vol. 108，No. 422，pp. 1-6，2009.
3) 早川直樹，ツィンツァレク・トビアス，川波弘道，猿渡洋，鹿野清宏：音声情報案内システムの応答文選択におけるスコア閾値を用いた棄却処理の導入，日本音響学会講演論文集，1-P-26，pp. 175-176，2007.
4) 山田寛康，工藤拓，松本裕治：Support Vector Machine を用いた日本語固有表現抽出，情報処理学会論文誌，Vol. 43，No. 1， pp. 43-53，2002.
5) 平尾努，前田英作，松本裕治：Support Vector Machine による重要文抽出，情報処理学会論文誌，vol. 44，pp. 2230-2243，2003.
6) Chih-Chung Chang, Chih-Jen Lin: LIBSVM : a library for support vector machines, http://www.csie.ntu.edu.tw/˜cjlin/libsvm.
7) Cees G. M. Snoek, Marcel Worring, Jan C. van Gemert, Jan-Mark Geusebroek, Arnold W. M. Smeulders:The Challenge Problem for Automated Detection of 101 Semantic Concepts in Multimedia, *ACM International Conference on Multimedia*, page 421-430, 2006.