

# Context-Aware SVM に基づく状況依存型情報推薦方式の提案

Context-Aware SVM for Context-Dependent Information Recommendation

奥 健太<sup>◇</sup> 中島 伸介<sup>△</sup>  
宮崎 純<sup>△</sup> 植村 俊亮<sup>△</sup>

Kenta OKU Shinsuke NAKAJIMA  
Jun MIYAZAKI Shunsuke UEMURA

状況依存型情報推薦システムを実現するための基礎技術として、Context-Aware Support Vector Machine (C-SVM) を提案する。ユーザの嗜好はそのときのコンテキスト（日時や同伴者、天気、体調など）により変化するため、情報推薦においてユーザのコンテキストを考慮することは重要である。しかし、これまでにユーザのコンテキストを考慮した情報推薦技術はほとんどない。そこで、我々は2クラス分類手法である Support Vector Machine (SVM) を、その特徴空間にコンテキスト軸を追加することにより拡張した C-SVM を提案する。また、その C-SVM に基づく協調フィルタリング (C-SVM-CF) を提案する。さらに、C-SVM 及び C-SVM-CF を飲食店情報推薦システムに適用した実験を行い、各手法の有効性を検証する。

The purpose of this study is to propose Context-Aware Support Vector Machine (C-SVM) for application in a context-dependent recommendation system. It is important to consider users' contexts (e.g. time of day, companions, weather, and physical conditions) in information recommendation, as users' preferences change with their contexts. However, currently there are few methods which take users' contexts into account. Thus, we propose C-SVM extended the functionality of a Support Vector Machine (SVM), a popular classifier method used for two classes, by adding axes of contexts to the feature space in order to consider the users' contexts. We also propose Collaborative Filtering based on C-SVM (C-SVM-CF). We then apply C-SVM and C-SVM-CF to a restaurants recommendation system, and then examine the effectiveness of each approach.

## 1. はじめに

膨大な情報源の中からユーザの嗜好に合った情報を見つけ出し、これをユーザに能動的に提供する情報推薦技術が注目されている。人の嗜好は、その人の置かれている状況に応じて変化するものであるため、情報推薦においてユーザの状況を考慮することは重要である。しかしながら、従来の情報

推薦技術には、情報推薦においてユーザの状況を考慮したものはいくつかある。

我々は、人の情報選択の判断に対して影響を及ぼす外的及び内的状況（日時や同伴者、天気、体調など）を「コンテキスト」と定義する。本研究では、コンテキスト依存型情報推薦システムを実現するための基礎技術として、Context-Aware Support Vector Machine (C-SVM) 及び Context-Aware Support Vector Machine に基づく協調フィルタリング (C-SVM-CF) を提案する。

C-SVMは、2クラス分類手法である Support Vector Machine (SVM) の特徴空間にユーザの情報選択の判断に影響するコンテキスト軸を追加することにより拡張した手法である。C-SVMは、学習データ集合に対するユーザの嗜好をあらゆるコンテキストも含めて分類することができ、ユーザのコンテキストに依存した推薦情報を判定することができる。

C-SVM-CFは、C-SVMをベースにしたコンテキスト依存型の協調フィルタリングである。C-SVM-CFにおけるユーザ類似度は、C-SVMに基づいて計算され、コンテキストも含めたユーザの嗜好の類似度から類似ユーザを決定することができる。

本研究では、C-SVM及びC-SVM-CFを飲食店情報推薦システムに適用した実験を行い、各手法の有効性を検証する。

## 2. 関連研究および関連手法

従来の情報推薦技術として、a) ユーザプロフィールに基づく情報フィルタリング、b) 協調フィルタリングなどがある。a) はユーザの過去の行動履歴などからユーザプロフィールを作成し、それに基づいて情報推薦を行う。Google [1] はユーザの検索履歴に基づいて、よりユーザの興味のあるコンテンツを提示する“Personalized Search”というサービスを提供している。b) は複数のユーザが協調して利用することを前提とした情報推薦手法であり、“嗜好が類似している他のユーザ”が気に入っている情報を推薦するという手法である。Amazon.com [2] や movielens [3] は、協調フィルタリングを用いて書籍や映画などをユーザに推薦する。a) に比べ b) は、一度人間の判断に基づいて評価された情報を対象とするため、推薦情報に対する信頼度は比較的高いといえる。しかしながら、いずれの手法もユーザのコンテキストを考慮していないため、推薦された情報に対して常に満足するとは限らない。それに対して、C-SVM及びC-SVM-CFはコンテキストを考慮することにより、より有効な情報推薦を可能にしようとするものである。

Support Vector Machine (SVM) は2クラスの分類手法の一つであり、1995年にVapnikら [4] によって提案された。クラス分類手法には、他にニューラルネットワークや $k$ 最近傍法、決定木、ベイジアンフィルタなどがあるが、SVMは、

- 少ない学習データから識別関数を構成できる
- 膨大な次元数を持つ学習データにも対応可能
- 解が一意に定まる
- 最適化すべき学習パラメータが少ない

などの利点を持つ。SVMは主に画像認識や自然言語処理などの分野で利用されている。

図1のように学習データが分布する空間をSVMの特徴空間という。図1中の“○”、“●”及び“□”、“■”は学習データを表現し、それぞれの形状は正データ及び負データを表す。SVMにより、ユーザの学習データ集合を満足、不満足 of 2クラスに分類する識別モデルを構築する。その識別モデルに基づいて、

<sup>◇</sup> 学生会員 奈良先端科学技術大学院大学情報科学研究科 博士後期課程 [kenta.o@is.naist.jp](mailto:kenta.o@is.naist.jp)

<sup>△</sup> 正会員 奈良先端科学技術大学院大学情報科学研究科 [shin.miyazaki, uemura}@is.naist.jp](mailto:{shin.miyazaki, uemura}@is.naist.jp)

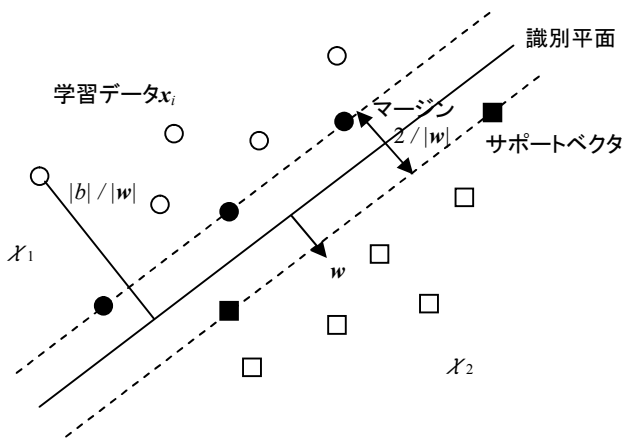


図1 SVMの特徴空間

Fig.1 SVM Feature Space

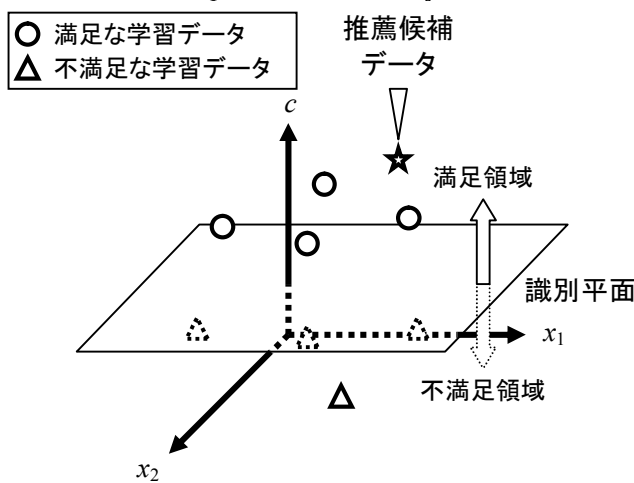


図2 C-SVMの特徴空間

Fig.2 C-SVM Feature Space

未知のデータがユーザにとって満足されるものか否かを予測する。ここで、満足されると予測されたデータがユーザに対する推薦候補として判定される。なお本研究ではSVMにおける識別モデルをSVMモデルと呼ぶ。

### 3. 提案手法

C-SVMは、図2のように、SVMの特徴空間にユーザの情報選択の判断に影響するコンテキスト軸を追加することにより拡張した手法である。図2の $x_1, x_2$ は学習データの特徴を表現するパラメータ軸を、 $c$ はコンテキスト軸を表す。このようにコンテキスト軸を追加することにより、コンテキストごとに異なる、学習データ集合に対するユーザの満足、不満足进行分类することができるようになる。

C-SVMでは、学習データの特徴及びコンテキストをパラメータ化することにより、あらゆるコンテキストにおけるユーザの嗜好を一つの特徴空間上で記述することを可能とする。このため、ユーザの情報選択に影響を及ぼすあらゆるコンテキストを総合的に考慮したうえで、推薦候補を判定すること、さらには未知のコンテキスト(未学習コンテキスト)においても推薦候補を判定することが可能となる。なお、本来なら学習データのパラメータ軸及びコンテキスト軸は多次元で表現されるものであるが、ここでは説明のため便宜上それぞれ2次元及び1次元にて表現している。本研究ではC-SVMにおけるSVMモデルをC-SVMモデルと呼ぶ。

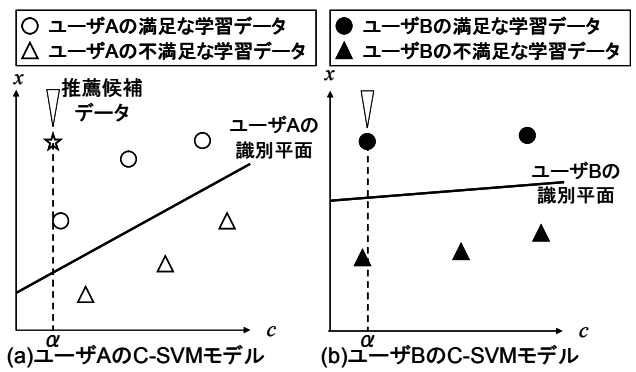


図3 ユーザA及びユーザBのC-SVMモデル

Fig.3 C-SVM Model of User A and User B

$$\text{類似度} = ((0.800 + 0.833) / 2) \times 100 = 81.7\%$$

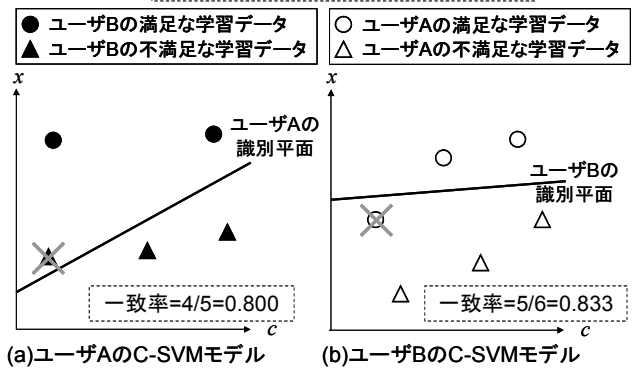


図4 ユーザ類似度の計算

Fig.4 Calculating User Similarity

ここで、ある未知データがユーザのあるコンテキストにおいて、識別平面から満足領域に存在する場合、そのデータがユーザに対して推薦候補となる情報として判定される。図2の例では、データ“☆”が推薦候補データとして扱われる。

C-SVM-CFは、C-SVMをベースにしたコンテキスト依存型の協調フィルタリングである。従来の協調フィルタリングでは、ユーザ類似度は単に嗜好の類似のみから計算される。それに対して、C-SVM-CFでは、ユーザ類似度はC-SVMの識別平面の類似度に基づいて計算されるため、コンテキストも含めたユーザの嗜好に基づいて、類似ユーザを判定することができる。これは、単に同じ嗜好を持つというだけでなく、“同じコンテキストにおいて、かつ同じ嗜好を持つユーザ”を類似ユーザとして扱うことにより、より正確に類似ユーザを判定することができるという考えに基づく。

C-SVM-CFにおけるユーザ $u$ とユーザ $v$ との間のユーザ類似度 $\text{sim}(u, v)$ は式(1)により求められる。

$$\text{sim}(u, v) = \frac{1}{2} \left( \frac{M_{v \rightarrow u}}{N_u} + \frac{M_{u \rightarrow v}}{N_v} \right) \times 100(\%) \quad (1)$$

ここで、 $N_u$ はユーザ $u$ の学習データ数、 $M_{v \rightarrow u}$ はユーザ $v$ のC-SVMモデルによりユーザ $u$ の学習データ集合を分類したときのクラスが一致したデータ数である。ユーザ類似度が一定の閾値を超えたユーザが類似ユーザとして扱われる。

図3及び図4は、ユーザA及びユーザBのC-SVMモデルである。図中の“○”、“△”及び“●”、“▲”は学習データを表現し、それぞれ、ユーザAにとって満足、不満足なデータ、ユーザBにとって満足、不満足なデータを表す。図4のように、ユーザAとユーザBとの間のユーザ類似度 $\text{sim}(A, B)$ は、 $((5/6 + 4/5) / 2) \times 100 = 81.7(\%)$ となる。これが、類似ユーザ

表1 飲食店パラメータ  
Table 1 Restaurants Parameters

駐車場あり	ランチサービスあり	パーティにおすすめ
禁煙席あり	食べ放題プランあり	デートにおすすめ
個室あり	テイクアウト可	家族で食事におすすめ
カラオケあり	クーポンサービスあり	夜景がきれい
ライブあり	接待におすすめ	海がみえる
深夜営業 OK	宴会におすすめ	

表2 コンテキストパラメータ  
Table 2 Contexts Parameters

カテゴリ	パラメータ	設定値
時間	月	1~12月
	時刻	0~23時
	曜日	月=0~日=6
スケジュール	エリアタイプ	なし, 歓楽街, 駅周辺, 観光地
	予算	0~10000円
	休日	なし, 一日休み, 休憩時間, 休日前
同伴者	人数(男性)	0~10人
	人数(女性)	0~10人
	最低年齢	0~100歳
	最高年齢	0~100歳
	関係	なし, 家族, 恋人, 友人, 上司, 部下
	身分	なし, 学生, 社会人
外的要因	天気	晴れ, 曇り, 雨
	気温	-5~40°C

判定の閾値を越えたとすると、ユーザ A とユーザ B は類似ユーザとして扱われる。

図 3 に示すように、いま、ユーザ A がコンテキスト  $a$  に置かれているとすると、ユーザ B がコンテキスト  $a$  において満足と評価している学習データ (図 3 中の☆) がユーザ A への推薦候補データとして扱われる。

## 4. 検証実験と考察

C-SVM 及び C-SVM-CF の有効性を検証するために、C-SVM 及び C-SVM-CF を飲食店情報推薦システムに適用した検証実験を行った。

### 4.1 実験環境

本実験では『Yahoo!グルメ』[5]を参考に、実験用の飲食店データベースを作成した。飲食店データの登録対象エリアとして、大阪の「心斎橋、なんば」エリアを選択した。登録飲食店件数は 938 件である。飲食店の特徴を表現するパラメータ (以下、飲食店パラメータ) として、『Yahoo!グルメ』を参考に、表 1 に示す 17 個のパラメータを用いた。それぞれ、設備やサービスなどの有無を {0, 1} の 2 値で設定することにより、飲食店の個性を表現する。本実験の被験者は 8 名とした。いずれの被験者も「心斎橋、なんば」エリアについて土地勘はあるものの、当然ながらすべての飲食店について把握しているわけではない。本実験では表 2 に示す 14 個のコンテキストパラメータを用いた。各パラメータの設定値は表 2 のとおりである。ただし、コンテキストパラメータの項目及び設定値は、今回の実験用に暫定的に設定しただけであるため、コンテキストパラメータの選定及びその直交性の保証については別途検討課題とする。なお、SVM の学習データとして与えるパラメータ値の範囲は {0~1} をとりうる。そのため、表 2 に示したコンテキストパラメータの設定値を {0~1}

の範囲に正規化する。

### 4.2 実験手順

本実験では以下の二つ観点から C-SVM 及び C-SVM-CF の有効性を検証する。1) 従来型 SVM モデルと C-SVM モデルの識別性能、2) C-SVM 及び C-SVM-CF を用いた情報推薦に対する被験者満足度。

まず、第 1 段階の実験手順を示す。1) 被験者は、あるコンテキストにおける各飲食店データに対する満足、不満足の評価を与える。2) 同様に、他のコンテキストにおける各飲食店データに対する満足、不満足の評価を与える。ここで、被験者は 20 個の飲食店データに対して評価を行い、それを 5 個のコンテキストにおいて行う。したがって、被験者 1 人あたり合計 100 個の学習データが得られる。3) 学習データ集合に基づいて、従来型 SVM モデル及び C-SVM モデルを構築する。4) 各モデルの識別性能を得る。識別性能とは、構築された SVM モデルがどの程度正確に構築されたかを示す指標であり、式(2)で表される。

$$\text{識別性能} = \frac{\text{正確に分類されたデータ数}}{\text{総学習データ数}} \times 100(\%) \quad (2)$$

つづいて、第 2 段階の実験手順を示す。1) 被験者は、あるコンテキストにおいてランダムに推薦された飲食店データに対して満足、不満足の評価を与える。2) 同様に、従来型 SVM、C-SVM 及び C-SVM-CF を用いて推薦された飲食店データに対して満足、不満足の評価を与える。ここで被験者は各推薦手法において 50 個の飲食店データに対して評価を行う。したがって、合計 200 回の推薦に対する評価を行うことになる。3) 被験者の評価に基づいて被験者満足度を得る。被験者満足度とは、被験者が各手法による情報推薦に対してどの程度満足したかを示す指標であり、式(3)で表される。

$$\text{満足度} = \frac{\text{被験者が満足した回数}}{\text{推薦回数}} \times 100(\%) \quad (3)$$

### 4.3 結果及び考察

#### 4.3.1 従来型 SVM モデルと C-SVM モデルの識別性能

構築した従来型 SVM モデル及び C-SVM モデルの識別性能を比較し、識別モデルの構築にコンテキストを考慮することによる識別性能を評価した。ここでは、識別モデルの構築に用いた学習データをそのままテストデータとして識別性能を得た。したがって、既知データに対する識別性能であるため、識別性能が 100% となることが望ましい。

図 5 は被験者別の従来型 SVM モデル及び C-SVM モデルの既知データに対する識別性能を比較したものである。C-SVM モデルではいずれの被験者も 100% に近い識別性能が得られたのに対し、SVM モデルでは平均 70% にも満たなかった。

この実験結果は、識別モデルの構築にユーザのコンテキストを考慮した C-SVM モデルでは、コンテキストごとに変化するユーザの好みも正確に分類することができることを示唆している。一方、ユーザのコンテキストを考慮していない SVM モデルでは、ユーザの好みもコンテキストごとに変化するということに対応できていないため、識別性能が平均 70% という結果となったと考える。

#### 4.3.2 情報推薦に対する被験者満足度

本実験では、a) Random 提示、b) 従来型 SVM、c) C-SVM、d) C-SVM-CF の 4 種の推薦手法により被験者に対して情報推薦を行った。これらの 4 種の推薦手法による情報推薦に対する被験者の満足度を比較することによって、C-SVM 及び C-SVM-CF による情報推薦の有効性を検証した。なお、今回

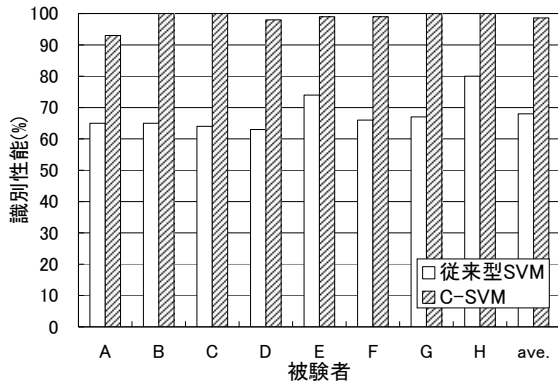


図5 従来型SVMモデル及びC-SVMモデルの識別性能

Fig.5 Accuracy of Simple SVM Model and C-SVM Model

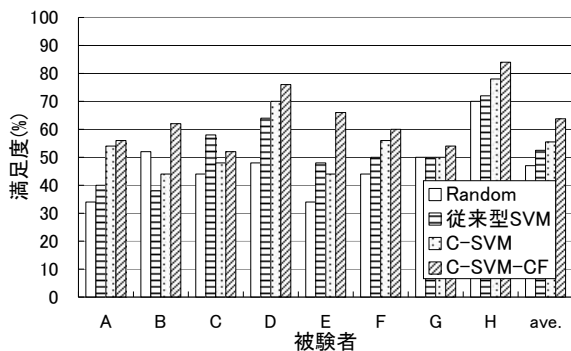


図6 情報推薦に対する被験者満足度

Fig.6 Users' Satisfaction Rate for Recommendation

の被験者におけるユーザ類似度は、42.5%~62.5%となり、極端に類似度が高い、若しくは低いということはなかった。このうち、ユーザ類似度 50%以上を類似ユーザとして、C-SVM-CFによる情報推薦を行った。

図6は4種の推薦手法による情報推薦に対する被験者の満足度を比較したものである。平均的にみると、C-SVM-CFに対する満足度が最も高く、次いで、C-SVM、SVM、Random提示の順に高かった。

今回の実験は、実験対象エリアを「心斎橋、なんば」エリアとし、被験者は8名とも「心斎橋、なんば」エリアについて土地勘がある、というような条件下において C-SVM-CFによる情報推薦を行った。この実験結果は、“類似ユーザが推薦対象エリアにおけるいくつかの飲食店データに対して評価済みである”，という条件下では、C-SVM-CFが有効に働くことを示した。

今回の実験結果では、C-SVM-CFに対する満足度は平均して64%であったが、被験者数を増やし、類似ユーザが増えることによって、より満足度を向上させることができると期待できる。

## 5. まとめと今後の課題

本研究では、コンテキスト依存型の情報推薦システムを実現するための技術として、C-SVM及びC-SVM-CFを提案した。本研究では、C-SVM及びC-SVM-CFを飲食店情報推薦システムに適用した実験を行い、提案手法の有効性を検証した。実験結果より、以下の事項を示した。

- 識別モデルの構築にユーザのコンテキストも含めたC-SVMモデルでは、コンテキストごとに変化するユー

ザの好みも正確に分類することができる

- “類似ユーザが推薦対象エリアにおけるいくつかの飲食店データに対して評価済みである”，という条件下では、C-SVM-CFが有効に働く

なお、今回はC-SVM及びC-SVM-CFを飲食店情報推薦システムに適用したが、他にも音楽情報推薦や旅行情報推薦などのように、ユーザのコンテキストが情報選択の判断基準に影響を及ぼすような分野に応用可能であると考えられる。

今後の検討課題として、以下の事項が挙げられる。

- より詳細なコンテキストパラメータ（過去の行動履歴、ユーザの目的、現在地からの距離、周辺環境との相対性、同伴者の嗜好など）
- 飲食店パラメータ及びコンテキストパラメータの直交性の検証
- 被験者数の増加や実験対象エリアの多様化などによる、多様な条件下におけるC-SVM及びC-SVM-CFの挙動の検証
- 多クラス分類への対応、複数飲食店データベースへの対応、飲食店パラメータの検討

## 【謝辞】

本研究の一部は、科学技術振興機構戦略的創造研究推進事業(CREST)「高度メディア社会の生活情報技術」、科学研究費補助金基盤研究(A)(2)(課題番号:15200010)、若手研究(B)(課題番号:17700132)の支援による。ここに記して謝意を表す。

## 【文献】

- [1] Google, <http://www.google.co.jp>
- [2] Amazon, “Amazon.com”, <http://www.amazon.com>
- [3] GroupLens Research, “movielens”, <http://movielens.umn.edu>
- [4] C. Cortes and V. Vapnik: “Support-Vector Networks”, Machine Learning Vol. 20, No. 3, pp. 273-297 (1995).
- [5] Yahoo Japan, “Yahoo!グルメ”, <http://gourmet.yahoo.co.jp/gourmet>

## 奥 健太 Kenta OKU

奈良先端科学技術大学院大学情報科学研究科博士後期課程在学中。2006 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。情報処理学会、日本データベース学会、電子情報通信学会 各学生会員。

## 中島 伸介 Shinsuke NAKAJIMA

奈良先端科学技術大学院大学情報科学研究科助手。博士(情報学)。日本データベース学会、情報処理学会、IEEE CS 各会員。

## 宮崎 純 Jun MIYAZAKI

奈良先端科学技術大学院大学情報科学研究科助教授。データベースシステムの研究に従事。情報処理学会、電子情報通信学会、日本データベース学会、ACM SIGMOD, IEEE CS 各会員。

## 植村 俊亮 Shunsuke UEMURA

奈良先端科学技術大学院大学情報科学研究科教授。データベースシステムの研究に従事。情報処理学会フェロー。電子情報通信学会フェロー、IEEE Fellow、日本データベース学会正会員。著書に「データベースシステムの基礎」(オーム社)など。