サポートベクターマシンによるe-Learningの適正コンテンツの決定

甲南高等学校・中学校　非常勤講師　中上　香代子
甲南高等学校・中学校　吉田　賢史
甲南大学理工学部情報システム工学科　中山　弘隆

1　はじめに

近年，e-Learningは教育機関において特に注目されており，さまざまな試みが行われている．e-Learningシステムには，webページに教材を記述するWBT(Web Based Training)に代表されるような非同期型システムと呼ばれるものがあり，様々な分野で使用されている．

しかし，従来の非同期型のe-Learningシステムにおけるコンテンツの配信は，学習者の都合に合わせて学習を進めて行くことができる利点はあるが，学習者は単調で受身となるため長続きしない．そこで，能動的に学習させるために実験的に数学を取り入れ，発見的学習させることを試みた．

そこで本論文では，学習者の状況に合わせて適切な配信をするためにパターン分類の手法の1つであるサポートベクターマシン(Support Vector Machine; SVM)を用いてコンテンツの配信を決定する手法を提案する．また，教師側，つまりe-Learningのコンテンツを作成する側にとって，作成する作業は莫大な時間と労力を要するため，日々時間に追われている教師にとっては作成することが困難である．そこで，SVMを用いてコンテンツの決定を行うことにより，その作業に要する労力の軽減にもつながると思われる．コンテンツの配信を決定する際，多様なコンテンツの中から学習者の状況に適したコンテンツを選択する必要があるため，多値分類を用いる必要がある．しかし，SVMによる多値分類では，どの集合にも分類しがたいあいまいな領域が現れる．その領域に対する分類を適切にするために，メンバーシップ関数を導入して判別を行う．ここで用いるメンバーシップ関数は[0,1]以外の値も許容しており，これを拡大メンバーシップ関数と呼ぶことにする．問題やデータにあわせて拡大メンバーシップ関数を変えることにより，データの特徴を考慮した判別が可能になり，的確な分類を行うことが可能になると考えられる．

2　サポートベクターマシンについて

SVMは，Vapnikによって提案されたパターン分類の手法で，マージンを最大化することにより分離超平面を求め，2つの集合の分類を実現する．サポートベクターマシンは，一般的には2次計画問題(QP)として定式化されるが，計算時間がかかるなどの問題がある．そこで本論文では線形計画問題(LP)を用いる．

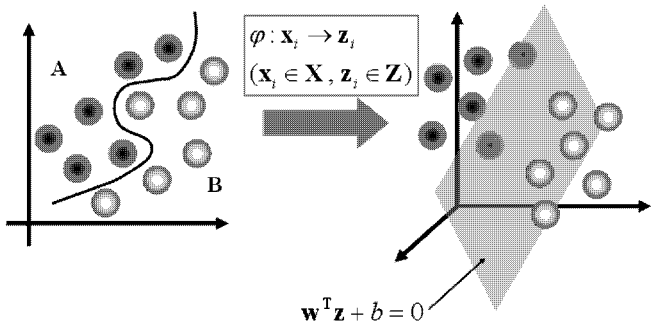

図 1: 高次元空間への非線形写像

2.1　サポートベクターマシンの定式化

n次元実空間上に2つのクラスA, Bがあり，このいずれかに属するようなデータx_i ($i = 1, \cdots, m$)があるとする．クラスAの要素x_iに対して$y_i = +1$を与え，クラスBの要素x_jに対して$y_j = -1$を与えることにする．ここで，クラスAとBを超平面で分離することを考える．この2つのクラスが線形分離不可能な場合は，原空間Xからある高次元特徴空間Zへの非線形写像$\varphi : \mathsf{x} \to \mathsf{z}$ ($\mathsf{x} \in X, \mathsf{z} \in Z$)によって図1のように線形分離可能な状態にする．このとき，Zにおける分離超平面を$D(\mathsf{z}) = \mathsf{w}^{\mathrm{T}}\mathsf{z} + b = 0$とする．$A$と$B$を分離する超平面は一般に数多く存在するが，SVMではマージン(分離超平面と各クラスとの最短距離)を最大化することを目的とする．クラスA, Bはそれぞれ，

$$\mathsf{w}^{\mathrm{T}}\mathsf{z}_i + b \geq 1 \qquad (\mathsf{z}_i \in \varphi(A)) \tag{1}$$

$$\mathsf{w}^{\mathrm{T}}\mathsf{z}_j + b \leq -1 \qquad (\mathsf{z}_j \in \varphi(B)) \tag{2}$$

を満たすとすると，SVMは次のような数理計画問題として定式化できる．

$$\begin{aligned} \text{[SVM]} \quad & \text{Minimize} && ||\mathsf{w}||_q \\ & \text{subject to} && y_i D(\mathsf{z}_i) \geq 1 \\ & && (i = 1, \cdots, m) \end{aligned} \tag{3}$$

ここで，問題(3)の制約条件において，等号を成立させるようなz_iをサポートベクターという．また，マージンの大きさを測る距離として，ℓ_2ノルムを用いると

問題 (3) は2次計画問題 (4) として定式化され，ℓ_∞ ノルムを用いると線形計画問題 (5) として定式化できる．

$$\begin{array}{lll} \text{[QPSVM]} & \text{Minimize} & \frac{1}{2}\mathbf{w}^{\mathsf{T}}\mathbf{w} \\ & \text{subject to} & y_i D(\mathbf{z}_i) \geq 1 \\ & & (i = 1, \cdots, m) \end{array} \tag{4}$$

$$\begin{array}{lll} \text{[LPSVM]} & \text{Minimize} & \sum_{i=1}^{} |w_i| \\ & \text{subject to} & y_i D(\mathbf{z}_i) \geq 1 \\ & & (i = 1, \cdots, m) \end{array} \tag{5}$$

2.2 サポートベクターマシンによる多値分類について

SVM における多値分類では，1対他分類を組み合わせることによって実現することができる．しかしそれだけでは分類不可能な領域がでてくる．その分類不可能な領域を分類可能にするために，メンバーシップ関数を導入し，判別を行う．

2.2.1 拡大メンバーシップ関数の導入

n クラス分類の場合，あるクラス C_k に属するデータと，クラス C_k 以外のクラス ($n-1$ 個) に属するデータで SVM による二値分類を行う．そして n 回の二値分類を行うことで n クラスを分類することができる．

ここで，クラス C_k に注目したときに求めることができる最適な分離超平面を $D_k(\mathbf{z}) = \mathbf{w}_k^{\mathsf{T}}\mathbf{z} + b_k = 0$ とする．このとき，例えば3クラスの分類の場合，図2のようにそれぞれのクラスに対して分離超平面を求めることができるが，図中のグレーの部分が分類不可能な領域となる．この領域を分類するために文献 [2] と同様に，メンバーシップ関数を導入する．ここで用いるメンバーシップ関数は [0,1] 以外の値も許容しており，これを拡大メンバーシップ関数と呼ぶことにする．

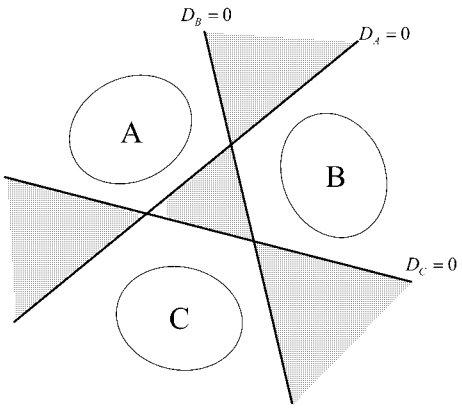

図 2: 分離不可能な領域

任意のデータ $\mathbf{z}_i$ $(i = 1, \cdots, m)$ についてメンバーシップ値を拡大メンバーシップ関数 f_k によって式 (6) のように定義する．

$$m_k(\mathbf{z}_i) = f_k\left(D_k(\mathbf{z}_i)\right) \qquad (k = 1, \cdots, n) \tag{6}$$

これより，n 個のクラスぞれぞれに対するメンバーシップ値を求め，その最大値を $m_p(\mathbf{z}_i)$ とするとき，式 (7) より，z_i は C_p に分類する．

$$m_p(\mathbf{z}_i) = \max_{1 \leq k \leq n} m_k(\mathbf{z}_i) \Rightarrow \mathbf{z}_i \in C_p \quad (p = 1, \cdots, n) \tag{7}$$

3 e-Learningにおけるコンテンツの決定

3.1 ねらい

従来の e-Learning においては，学習者の状況を考えずに，画一的な教材を一方的に配信するだけのものが多く，それでは学習が長続きしない．そこで，SVM を用いたコンテンツ決定のねらいとして，SVM を用いて分類を行うことで学習者の状態を推定し，それによって適切なコンテンツを決定することをねらいとしている．これによって，学習者それぞれに合わせて適切は教材を配信することをを目的としている．また，学習者へのコンテンツの配信をある程度自動化することができると考えている．

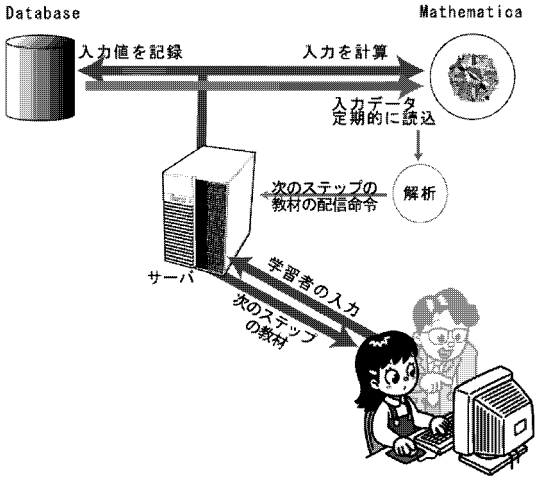

図 3: システムの構想

3.2 コンテンツについて

本論文で用いた e-Learning のコンテンツは,「2次関数のグラフの平行移動」のコンテンツである．ここで学習者に対して「$y = ax^2 + bx + c$ のグラフと a, b, c の値の関係を見て下さい.」という問いかけをして，学習者は a, b, c の値を入力する (図4).

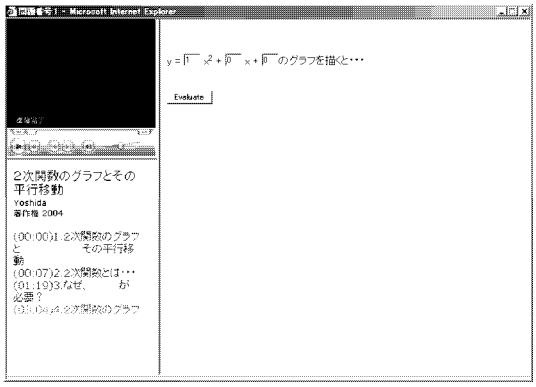

図 4: 入力画面

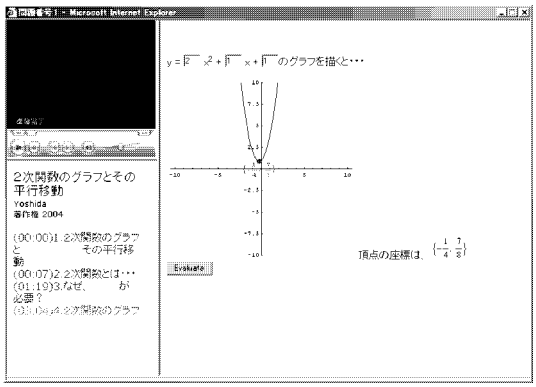

図 5: グラフの表示

そして，学習者が入力した値（a,b,c の値）に対してグラフ (図 5) を表示し，その入力値の変化を見ることにより学習者がどこに注目しているのかを推定し，次のステップを決定する．そして，それに対するコンテンツを配信する．

次のステップのコンテンツは下記の５つのコンテンツである．

コンテンツ A： a に注目している場合
コンテンツ B： b に注目している場合
コンテンツ C： c に注目している場合
コンテンツ D： どれにも注目していない場合
コンテンツ E： 入力方法がわかっていない場合

3.3 判別方法

判別方法として，学習者が入力した値の変化に注目することで，次のコンテンツを決定することを考える．その方法として，学習者が５回入力したデータの入力の変化をみて図 6 のような方法でビット列（学習者一人につき 1 × 12 のデータ）を生成し，これを用いて SVM によって判別を行う．(0:入力値に変化なし，1:入力値に変化あり)

	実際の入力値	ビット列の生成			
		変化	a	b	c
1回目	$y=\boxed{1}x^2+\boxed{1}x+\boxed{2}$				
2回目	$y=\boxed{1}x^2+\boxed{1}x+\boxed{5}$	1回目から2回目	0	0	1
3回目	$y=\boxed{1}x^2+\boxed{1}x+\boxed{6}$	2回目から3回目	0	0	1
4回目	$y=\boxed{1}x^2+\boxed{1}x+\boxed{8}$	3回目から4回目	0	0	1
5回目	$y=\boxed{1}x^2+\boxed{1}x+\boxed{10}$	4回目から5回目	0	0	1

図 6: 入力値とそれに対するビット列の生成の例

3.4 データ詳細

本論文で用いたデータは，Test data として実際に生徒が入力したデータを用い，それに対する理想の判別結果は教師側が作成している．Training data として，教師が作成した予想される入力データ及びそれに対する教師値を用いている．これらのデータ数を表 1 に示す．

表 1: データ数の詳細

コンテンツ	Training data	Test data
A	8	13
B	8	9
C	8	1
D	13	45
E	79	46
合計	116	114

3.5 判別結果

前節のデータを使用し，SVM を用いて Test data を判別した結果は表 2 である．

表 2: 判別結果

コンテンツ	データ数	正答数	正答率 (%)
A	13	13	100.0
B	9	9	100.0
C	1	1	100.0
D	45	36	80.0
E	46	45	97.8
全体	114	104	91.2

この結果のうち，判別結果に誤りがあったものに注目した．そうすると，誤った判別をしているデータの中で，コンテンツ D を理想値としているデータが多いことがわかった．そこで，コンテンツ D への判別の影響度を強めることにより，正答率が上がるのではない

かと予想される．

3.5.1 メンバーシップ関数の導入

コンテンツ D の影響度を強めるために，コンテンツ D のメンバーシップ関数を変更する．本研究ではコンテンツ D のメンバーシップ関数として，式 (8) を用いた．

$$m_{\mathrm{D}}(z) = D_{\mathrm{D}}(z) + a \qquad (8)$$

そこで，式 (8) の a の値を変化させて，正答率の変化を見た．その様子をあらわしたグラフは図 7 である．また，クラス別の正答率の変化をあらわしたグラフは図 8 である．

a の値を大きく設定するとコンテンツ D の正答率が上がっていることがわかる．しかし，a の値をあまり大きく設定しすぎると全体の正答率が下がってしまう．ここで，全体の正答率が最大値を示したものは a = 0.2 のときであった．

表 3 に示すように，a = 0.2 とすることにより，全体の正答率も上がり，コンテンツ D のクラスの正答率も上がった．これより，今回の場合において，a = 0.2 が最適な値であると考える．ただしこれは，今回の場合にのみ言えることであり，取り扱う問題にあわせてメンバーシップ関数を設定する必要がある．

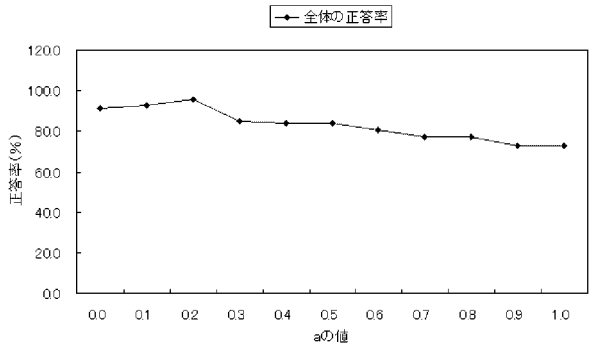

図 7: a の値による正答率の変化 (全体)

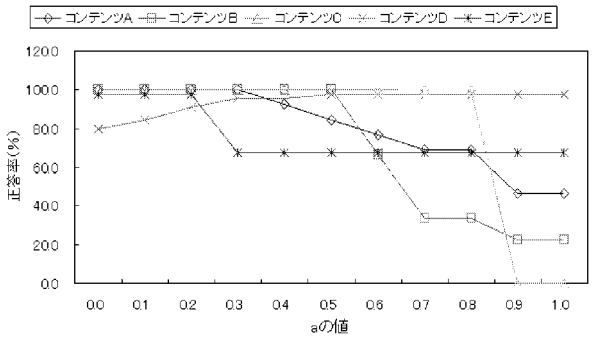

図 8: a の値による正答率の変化 (クラス別)

表 3: 判別結果の比較

		a=0.0		a=0.2	
	データ数	正答数	正答率	正答数	正答率
A	13	13	100.0	13	100.0
B	9	9	100.0	9	100.0
C	1	1	100.0	1	100.0
D	45	36	80.0	41	91.1
E	46	45	97.8	45	97.8
全体	114	104	91.2	109	95.6

4 終わりに

従来の e-Learning システムにおけるコンテンツの配信は，どの学習者に対しても同じストーリーであるため，学習が長続きしないという問題点があった．

そこで本論文では，サポートベクターマシン (Support Vector Machine; SVM) を用いたコンテンツの選択する手法を紹介した．この e-Learning のコンテンツ選択に SVM を用いることで，学習者それぞれにあったコンテンツの配信を行うことが可能になる．

また，従来のサポートベクターマシンによる多値分類では，分類不可能な領域があった．そこで，その領域を拡大メンバーシップ関数を用いることで分類可能にした．そして，メンバーシップ関数を問題ごとに設定することにより，より的確な分類を行うことが可能になる．

今後の課題としては，さらに的確な判別を行うために，問題それぞれに適したメンバーシップ関数の設定が重要となる．

謝辞

本研究の一部は日本文部科学省のオープン・リサーチ・センター整備事業による私学助成を得て行われた．（平成 16 年度-20 年度）

参考文献

[1] 浅田武史・中山弘隆：多目的計画法を用いたサポートベクターマシン，システム制御情報学会論文誌，Vol.166, No.2，pp.70/pp.76（2003）

[2] 井上拓也・阿部重夫：パターン認識用ファジィサポートベクトルマシンのアーキテクチャ，システム制御情報学会論文誌，Vol.15，No.2，pp.92/pp.98（2002）

[3] 吉田賢史：ヒューマンライク e-ラーニングシステムに関する研究，甲南大学大学院　博士論文（2006）