

Pendiskretan Data Set Kasar Menggunakan Ta'akulan Boolean

ROKIAH @ ROZITA AHMAD, MASLINA DARUS, SITI MARIYAM SHAMSUDDIN
& AZURALIZA ABU BAKAR

ABSTRAK

Pendiskretan data set kasar terhadap nilai nyata atribut merupakan salah satu aspek yang paling penting di dalam konsep pelombongan data, terutamanya yang melibatkan masalah pengkelasan masalah. Hasil kajian secara empirik menunjukkan bahawa kualiti pengkelasan bergantung kepada algoritma pendiskretan yang digunakan di dalam fasa pra-pemprosesan input data. Oleh yang sedemikian, proses pendiskretan boleh dinyatakan sebagai satu operasi gelidahan partisi terhadap domain atribut ke dalam bentuk selang dan menggabungkan nilai tersebut ke atas setiap selang yang berkaitan. Proses pendiskretan melibatkan gelidahan terhadap potongan yang menentukan selang berkaitan. Kesemua nilai yang berada pada setiap selang tersebut dipetakan terhadap nilai yang sama, iaitu menukarkan nilai atribut berkaitan kepada satu nilai berangka, yang boleh dijadikan sebagai atribut simbolik. Gelidahan terhadap potongan dilaksanakan terhadap perwakilan dalaman integer bagi input sistem keputusan. Kertas kerja ini bertujuan mengupas peranan set kasar dalam pendiskretan data menggunakan ta'akulan Boolean yang melakukan proses penukaran nilai nyata simbol matematik bercetak ke dalam perwakilan yang lebih mudah untuk menghasilkan kadar pencaman yang lebih baik menggunakan rangkaian nural.

ABSTRACT

Data discretization of rough set towards real attribute values is one of the important aspect in the data mining concepts, particularly the ones which involved classification problems. Empirical results showed that the quality of the classification depends on the discretization algorithm used in the input data pre-processing phase. In general, discretization is a process of searching for partition of attribute domains into intervals and unifying the values over each interval. Discretization process involves searching of cuts which determine the intervals acquired. All values which lie within each interval are mapped to the same values, in effect converting numerical attributes that

can be treated as being symbolic. The search for cuts is performed on the internal integer representation of the input decision system. This paper describes the role of rough set with Boolean reasoning discretization process in converting the real values of printed mathematical symbol that leads to the better recognition rates using neural network.

ASAS SET KASAR

Teori Set Kasar telah diperkenalkan oleh Zdzislaw Pawlak pada awal tahun 1980an (Pawlak 1991), dan merupakan suatu fenomena matematik yang baru berkaitan dengan kekaburan dan ketidakpastian. Falsafah set kasar ditemui berdasarkan kepada andaian bahawa untuk setiap objek bagi *universe of discourse*, objek tersebut disekutukan dengan maklumat yang berkaitan sama ada dalam bentuk data atau pengetahuan. Sebagai contoh, jika objek adalah pesakit yang menderita penyakit tertentu, maka punca penyakit berkenaan merupakan maklumat bagi pesakit berkaitan. Objek yang mempunyai maklumat yang sama dikatakan mempunyai kesamaan terhadap koleksi maklumat yang sudah sedia ada. *Hubungan Kesetaraan* inilah yang merupakan asas matematik kepada konsep set kasar.

KONSEP ASAS TA'AKULAN BOOLEAN

Jika d ialah suatu nilai keputusan dan $\hat{A} = (U, A \cup \{d\})$ adalah satu jadual keputusan yang besar, dengan bilangan nilai a yang banyak terhadap objek daripada satu set semesta U bagi sebarang nilai $a \in A$ yang A suatu set pungutan objek-objek, maka wujud kemungkinan peluang yang tipis iaitu satu objek baru akan dikenalpasti berdasarkan padanan nilai vektor atribut dengan baris bagi jadual berkaitan. Oleh yang sedemikian, bagi suatu jadual keputusan dengan nilai atribut nyata, maka terdapat beberapa strategi pendiskretan yang bertujuan untuk memperolehi kualiti pengkelasan yang tinggi [Nguyen et al. 1995,1998].

Katalah $\hat{A} = (U, A \cup \{d\})$ adalah jadual keputusan dengan $U = \{x_1, x_2, \dots, x_n\}$. Untuk sebarang $a \in A$, andaikan $V_a = [l_a, r_a] \subset \mathbb{V}$, yang \mathbb{V} adalah set nombor nyata. Di sini diandaikan bahawa \hat{A} adalah jadual keputusan yang konsisten.

Katalah P_a partisi pada V_a (bagi $a \in A$) di dalam subselang i.e.,

$$P_a = \{[c_0^a, c_1^a), [c_1^a, c_2^a), \dots, [c_k^a, c_{k+1}^a)\}$$

untuk beberapa integer k , yang mana:

$$l_a = c_0^a < c_1^a < c_2^a < \dots < c_k^a < c_{k+1}^a = r_a \quad \text{dan} \\ V_a = ([c_0^a, c_1^a]) ([c_1^a, c_2^a]) (\dots ([c_k^a, c_{k+1}^a])).$$

P_a ditakrif secara unik oleh set $Ca = \{c_1^a, c_2^a, \dots, c_k^a\}$ yang dikenali sebagai set potongan pada V_a (set potongan adalah sifar jika kardinal $(P_a) = 1$). Acapkali P_a dengan set potongan pada V_a tertakrif oleh C_a . Sebarang famili $\{P_a : a \in A\}$ yang P_a adalah partisi pada V_a dikenali sebagai partisi pada A . Maka sebarang partisi bagi famili $P = \{P_a : a \in A\}$, boleh diwakilkan sebagai $P = \cup \{a\} \times C_a$. Sebarang pasangan $(a, c) \in P$, digelar potongan pada V_a .

Partisi P bagi sebarang famili $P = \{P_a : a \in A\}$ pada A adalah konsisten dengan $\hat{A} = (U, A \cup \{d\})$, satu jadual keputusan baru $\hat{A}^p = (U, A^p \cup \{d\})$, yang $A^p = \{a^p : a \in A\}$ dan $a^p(x) = i \Leftrightarrow a(x) \in [c_i^a, c_{i+1}^a)$ untuk selang $x \in U$ dan $i \in \{0, \dots, k\}$. Jadual \hat{A}^p dikenali sebagai Pendiskretan- P pada A .

Katahlah $\hat{A} = (U, A \cup \{d\})$ satu jadual keputusan dengan $U = \{x_1, \dots, x_n\}$. Sebarang atribut $a \in A$ menakrif satu jujukan:

$$v_1^a < v_2^a < \dots < v_{n_a}^a, \quad \text{dengan} \quad \{v_1^a, v_2^a, \dots, v_{n_a}^a\} = \{a(x) : x \in U\} \quad \text{dan} \quad n_a \leq n.$$

Katahlah p_k^a adalah suatu pembolehkan usulan terhadap selang (v_k^a, v_{k+1}^a) untuk sebarang $k \in \{1, \dots, n_a - 1\}$ dan $a \in A$. Dengan $BV(A)$, kita nyatakan set bagi semua pemboleh ubah usulan dalam bentuk di atas. Sebarang partisi $P \subseteq \cup_{a \in A} \{a\} \times V_a$ mentakrifkan suatu penilaian val_p bagi pemboleh ubah usulan p_k^a dengan $val_p(p_k^a)$ adalah benar jika dan hanya jika wujud satu potongan $(a, c) \in P$ yang memuaskan ungkapan $v_k^a \leq c_a < v_{k+1}^a$.

Disjunki bagi kesemua pemboleh ubah Boolean adalah daripada set:

$$\{p_k^a : [v_k^a, v_{k+1}^a]\} \subseteq \left[\min(a(x_i), a(x_j)); \max(a(x_i), a(x_j)) \right]$$

Rumus Usulan Boolean Ketaksamaan diberikan oleh:

$$\Phi^A = \Lambda \{ \Psi(i, j) : d(x_i) \neq d(x_j) \}$$

Sebarang set tak sifar $S = \{p_{k_1}^{a_1}, p_{k_2}^{a_2}, \dots, p_{k_r}^{a_r}\}$ bagi pemboleh ubah usulan Boolean daripada $BV(A)$ mentakrifkan suatu famili partisi $P(S)$ seperti berikut:

$$P(S) = \left\{ \left(a_1, \frac{v_{k_1}^{a_1} + v_{k_1+1}^{a_1}}{2} \right) \left(a_2, \frac{v_{k_2}^{a_2} + v_{k_2+1}^{a_2}}{2} \right), \dots, \left(a_r, \frac{v_{k_r}^{a_r} + v_{k_r+1}^{a_r}}{2} \right) \right\}$$

Di sini dinyatakan kaedah pendiskretan dengan menggunakan contoh di bawah. Pertimbangkan jadual keputusan (Jadual 1) dengan 2 atribut a , b dan 7 objek u_1, \dots, u_7 . Nilai bagi setiap atribut pada objek ini dan nilai keputusan d diwakilkan oleh Jadual 1a. Nilai jangkaan akhir yang diperoleh bagi pendiskretan-P pada A ialah $P = \{(a,0.9), (a,1.5), (b,0.75), (b,1.5)\}$. Gambaran secara geometri pada objek dan kelas keputusan ditunjukkan oleh Rajah 1.

JADUAL 1. Proses pendiskretan

A	a	b	d
u1	0.8	2	1
u2	1	0.5	0
u3	1.3	3	0
u4	1.4	1	1
u5	1.4	2	0
u6	1.6	3	1
u7	1.3	1	1

 \Rightarrow

Ap	ap	bp	d
u1	0	2	1
u2	1	0	0
u3	1	2	0
u4	1	1	1
u5	1	2	0
u6	2	2	1
u7	1	1	1

PROSES PENDISKRETAN

Di bawah dinyatakan langkah-langkah kaedah ta'akulan Boolean untuk pendiskretan data bagi contoh di atas.

- Nilai kemungkinan bagi set pada a dan b ditakrifkan oleh

$$V_a = [0, 2)$$

$$V_b = [0, 4)$$

- Nilai set a dan b pada objek daripada U diberi oleh,

$$a(U) = \{0.8, 1, 1.3, 1.4, 1.6\}$$

$$b(U) = \{0.5, 1, 2, 3\}$$

- Kardinal bagi $a(U)$ dan $b(U)$

$$n_a = 5$$

$$n_b = 4$$

- Set bagi pembolehubah boolean tertakrif oleh A adalah,

$$BV(A) = \{p_1^a, p_2^a, p_3^a, p_4^a, p_1^b, p_2^b, p_3^b\}$$

dengan

$$p_1^a \sim [0.8, 1) \quad p_2^a \sim [1, 1.3) \quad p_3^a \sim [1.3, 1.4) \quad p_4^a \sim [1.4, 1.6) \\ p_{1b} \sim [0.5, 1) \quad p_2^b \sim [1, 2) \quad p_3^b \sim [2, 3)$$

- Rumus ketaksamaan $\psi(i, j)$ bagi pasangan objek yang berbeza (u_i, u_j) daripada $U \times U$ mempunyai bentuk seperti berikut berdasarkan kepada atribut a dan b dengan keputusan yang berbeza :

$$BV(A) = \{p_1^a, p_2^a, p_3^a, p_4^a, p_1^b, p_2^b, p_3^b\}$$

$$\psi(2, 1) = p_1^a \vee p_1^b \vee p_2^b$$

$$\psi(2, 4) = p_2^a \vee p_3^a \vee p_1^b$$

$$\psi(2, 6) = p_2^a \vee p_3^a \vee p_4^a \vee p_1^b \vee p_2^b \vee p_3^b$$

$$\psi(2, 7) = p_2^a \vee p_1^b$$

$$\psi(3, 1) = p_1^a \vee p_2^a \vee p_3^b$$

$$\psi(3, 4) = p_2^a \vee p_2^b \vee p_3^b$$

$$\psi(3, 6) = p_3^a \vee p_4^a$$

$$\psi(3, 7) = p_2^b \vee p_3^b$$

$$\psi(5, 1) = p_1^a \vee p_2^a \vee p_3^a$$

$$\psi(5, 4) = p_2^b$$

$$\psi(5, 6) = p_4^a \vee p_3^b$$

$$\psi(5, 7) = p_3^a \vee p_2^b$$

Sebagai contoh : rumus $\psi(5, 6)$ adalah benar pada set potongan jika wujud satu potongan $p_1 = (a, c)$ pada V_a dalam set ini supaya $c \in [1.4, 1.6)$ atau satu potongan $p_2 = (b, c)$ pada V_b iaitu $c \in [2, 3)$.

Rumus ketaksamaan Φ^A dalam bentuk CNF diberi oleh,

$$\Phi^A = (p_1^a \vee p_1^b \vee p_2^b) \wedge (p_2^a \vee p_3^a \vee p_1^b) \wedge (p_2^a \vee p_3^a \vee p_4^a \vee p_1^b \vee p_2^b \vee p_3^b) \\ \wedge (p_2^a \vee p_1^b) \wedge (p_1^a \vee (p_2^a \vee p_3^b) \wedge (p_2^a \vee p_2^b \vee p_3^b) \wedge (p_3^a \vee p_4^a) \wedge (p_2^b \vee p_3^b) \\ \wedge (p_1^a \vee p_2^a \vee p_3^a) \wedge (p_2^b) \wedge (p_4^a \vee p_3^b) \wedge (p_3^a \vee p_2^b)$$

Menggunakan Hukum Kalis Agihan maka ungkapan Φ^A di atas dapat diungkapkan dalam bentuk DNF. Bagi memudahkan penghasilan ungkapan yang dicari,

- (i) pernyataan berikut akan digunakan:

$$f_1 : (a \vee b \vee c) \wedge (a \vee b) = (a \vee b)$$

$$f_2 : (a \vee b \vee c) \wedge (b \vee d) = b \vee (a \wedge d) \vee (c \wedge d)$$

- (ii) pemboleh ubah Boolean diwakilkan dengan:

$$\begin{array}{llll}
 p_1^a = \alpha, & p_1^b = \beta, & p_2^b = \delta, & p_2^a = \eta, \\
 p_3^a = \xi, & p_4^a = \gamma, & p_{3b} = \kappa. &
 \end{array}$$

Maka, ungkapan (A) dalam bentuk CNF di atas boleh ditulis sebagai:

$$\begin{aligned}
 \Phi^1 = & (\alpha \vee \beta \vee \gamma) \wedge (\eta \vee \xi \vee \beta) \wedge (\eta \vee \xi \vee \gamma \vee \beta \vee \delta \vee \kappa) \wedge (\eta \vee \beta) \wedge \\
 & (\alpha \vee \eta \vee \kappa) \wedge (\eta \vee \delta \vee \kappa) \wedge (\xi \vee \gamma) \wedge (\delta \vee \kappa) \wedge (\alpha \vee \eta \vee \xi) \wedge (\delta) \wedge \\
 & (\gamma \vee \kappa) \wedge (\xi \vee \delta)
 \end{aligned}$$

Menggunakan pernyataan f_1 , maka sebutan:

$$\begin{aligned}
 & (\eta \vee \xi \vee \beta) \wedge (\eta \vee \xi \vee \gamma \vee \beta \vee \delta \vee \kappa) \wedge (\eta \vee \beta) = \eta \vee \beta \text{ dan} \\
 & (\eta \vee \delta \vee \kappa) \wedge (\delta \vee \kappa) (\delta) \wedge (\xi \vee \delta) = \delta
 \end{aligned}$$

Oleh itu,

$$\Phi^1 = (\alpha \vee \beta \vee \delta) \wedge (\eta \vee \beta) \wedge (\alpha \vee \eta \vee \kappa) \wedge (\xi \vee \gamma) \wedge (\alpha \vee \eta \vee \xi) \wedge (\delta) \wedge (\gamma \vee \kappa).$$

Dengan Hukum Kalis Tukar tertib dan pernyataan f_2 di atas,

$$\begin{aligned}
 \Phi^1 = & [(\alpha \vee \beta \vee \delta) \wedge (\beta \vee \eta)] \wedge [(\alpha \vee \kappa \vee \eta) \wedge (\kappa \vee \gamma)] \wedge [(\alpha \vee \xi \vee \eta) \wedge (\xi \vee \gamma)] \wedge \delta \\
 = & [\beta \vee (\alpha \vee \eta) \vee (\delta \wedge \eta)] \wedge [\kappa \vee (\alpha \wedge \gamma) \vee (\eta \wedge \gamma)] \wedge [\xi \vee (\alpha \wedge \gamma) \vee (\eta \wedge \gamma)] \wedge \delta
 \end{aligned}$$

Andaikan $(\alpha \wedge \gamma) \vee (\eta \wedge \gamma) = Q$ maka sebutan ke-2 dan ke-3 memberikan:

$$\begin{aligned}
 [\kappa \vee Q] \wedge [\xi \vee Q] &= [(\kappa \vee Q) \wedge \xi] \vee [(\kappa \vee Q) \wedge Q] \\
 &= (\kappa \wedge \xi) \vee (Q \wedge \xi) \vee Q \\
 &= (\kappa \wedge \xi) \vee Q
 \end{aligned}$$

$$\begin{aligned}
 \Phi^1 = & [\beta \vee (\alpha \wedge \eta) \vee (\delta \wedge \eta)] \wedge [(\kappa \wedge \xi) \vee (\alpha \wedge \gamma) \vee (\eta \wedge \gamma)] \wedge (\delta) \\
 = & [\beta \vee (\alpha \wedge \eta) \vee (\delta \wedge \eta)] \wedge [(\kappa \wedge \xi \wedge \delta) \vee (\alpha \wedge \gamma \wedge \delta) \wedge (\eta \wedge \gamma \wedge \delta)] \\
 = & (\beta \wedge \kappa \wedge \xi \wedge \delta) \vee (\beta \wedge \alpha \wedge \gamma \wedge \delta) \vee (\beta \wedge \eta \wedge \gamma \wedge \delta) \vee (\alpha \wedge \eta \wedge \kappa \wedge \xi \wedge \delta) \vee (\alpha \wedge \eta \wedge \gamma \wedge \delta) \vee (\alpha \wedge \eta \wedge \gamma \wedge \delta) \vee (\delta \wedge \eta \wedge \kappa \wedge \xi) \vee (\delta \wedge \alpha \wedge \gamma \wedge \eta) \vee (\eta \wedge \gamma \wedge \delta) \\
 = & (\beta \wedge \delta \wedge \xi \wedge \kappa) \vee (\alpha \wedge \beta \wedge \delta \wedge \gamma) \vee (\beta \wedge \delta \wedge \eta \wedge \gamma) \vee (\alpha \wedge \delta \wedge \eta \wedge \xi \wedge \kappa) \vee (\alpha \wedge \eta \wedge \gamma \wedge \delta) \vee (\eta \wedge \gamma \wedge \delta) \vee (\delta \wedge \eta \wedge \xi \wedge \kappa)
 \end{aligned}$$

tetapi

$$\begin{aligned}(\alpha \wedge \eta \wedge \gamma \wedge \delta) \vee (\eta \wedge \gamma \wedge \delta) &= \eta \wedge \gamma \wedge \delta \\(\beta \wedge \delta \wedge \eta \wedge \gamma) \vee (\eta \wedge \gamma \wedge \delta) &= \eta \wedge \gamma \wedge \delta\end{aligned}$$

dan

$$(\alpha \wedge \delta \wedge \eta \wedge \xi \wedge \kappa) \vee (\delta \wedge \eta \wedge \xi \wedge \kappa) = \delta \wedge \eta \wedge \xi \wedge \kappa$$

maka,

$$\Phi^A = (\eta \wedge \gamma \wedge \delta) \vee (\eta \wedge \xi \wedge \delta \wedge \kappa) \vee (\xi \wedge \beta \wedge \delta \wedge \kappa) \vee (\alpha \wedge \gamma \wedge \beta \wedge \delta).$$

Ungkapan dalam bentuk DNF diperoleh dan dengan menggantikan kembali perwakilan di atas dengan pemboleh ubah Boolean terhasil 4 'penunjuk perdana' iaitu:

$$\Phi^A = (p_2^a \wedge p_4^a \wedge p_{2b}) \vee (p_2^a \wedge p_3^a \wedge p_2^b \wedge p_3^b) \vee (p_3^a \wedge p_1^b \wedge p_2^b \wedge p_3^b) \vee (p_1^a \wedge p_4^a \wedge p_1^b \wedge p_2^b)$$

Oleh yang demikian, kita mempunyai 4 set penunjuk perdana yang boleh dipilih untuk dijadikan sebagai set potongan seperti berikut:

$$\begin{aligned}S_1 &= (p_2^a, p_4^a, p_2^b) \\S_2 &= (p_2^a, p_3^a, p_2^b, p_3^b) \\S_3 &= (p_3^a, p_1^b, p_2^b, p_3^b) \\S_4 &= (p_1^a, p_4^a, p_1^b, p_2^b)\end{aligned}$$

Pertimbangkan penunjuk perdana $S_4 = (p_1^a, p_4^a, p_1^b, p_2^b)$, yang dipilih sebagai set potongan. Dengan menggunakan takrifan famili partisi:

$$P(S) = \left\{ \left(a_1, \frac{v_{k1}^{a1} + v_{k1+1}^{a1}}{2} \right) \left(a_2, \frac{v_{k2}^{a2} + v_{k2+1}^{a2}}{2} \right) \dots \left(a_r, \frac{v_{kr}^{ar} + v_{kr+1}^{ar}}{2} \right) \right\}$$

maka akan diperoleh:

$$P(S) = \left\{ \left(a, \frac{0.8+1}{2} \right) \left(a, \frac{1.4+1.6}{2} \right) \left(b, \frac{0.5+1}{2} \right) \left(b, \frac{1+2}{2} \right) \right\}$$

menghasilkan potongan set berikut:

$$P(S) = \{(a, 0.9), (a, 1.5), (b, 0.75), (b, 1.5)\}$$

Jadual keputusan yang baru $A^{P(S)}$ diwakili oleh Jadual 1b.

Rajah 1 menunjukkan perwakilan geometri selang dan potongan yang diperolehi menerusi algoritma pendiskretan. Paksi melintang x mewakili selang dan potongan yang dikira untuk atribut a , manakala paksi menegak y mewakili selang dan potongan untuk atribut b . Garisan tebal pada Rajah 1 mewakili nilai potongan yang telah diperolehi. Contohnya nilai pasangan $(u1, a) = 0.8$ berada di dalam selang $[0, 0.9]$ yang telah didiskretkan kepada nilai 0 dan nilai pasangan $(u1, b) = 2$ berada di dalam selang $[1.5, \infty)$ dengan nilai diskret 2. Oleh itu objek $u1$ berada di dalam kawasan G. Jadual 2 menunjukkan kawasan nilai diskret di Rajah 1 yang mewakili setiap objek.

JADUAL 2. Kawasan pendiskretan objek

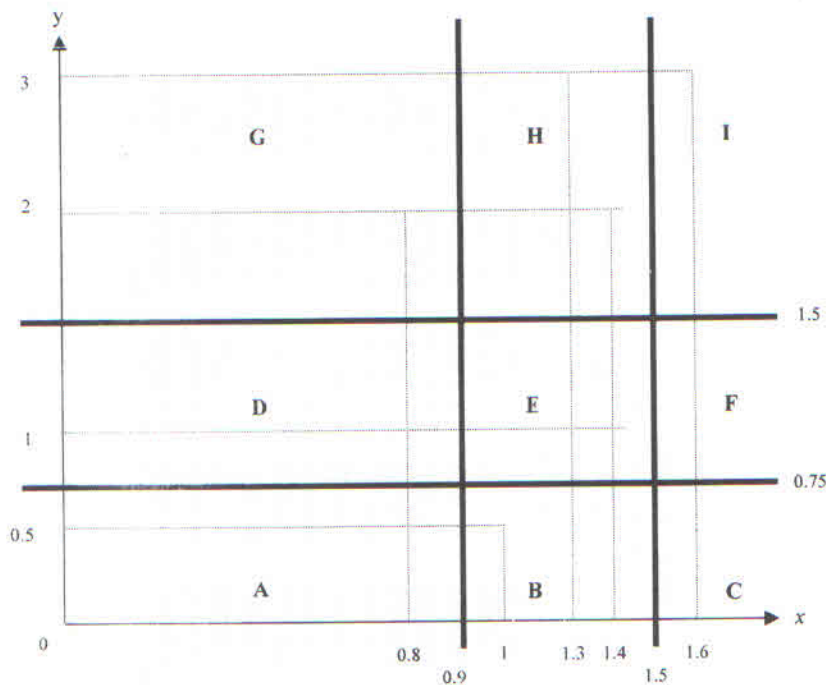
Objek	u1	u2	u3	u4	u5	u6	u7
Diskret	G	B	H	E	H	I	E

APLIKASI TERHADAP SIMBOL MATEMATIK BERCETAK

Sebanyak 40 contoh simbol matematik bercetak dilampirkan dengan pelbagai orientasi dan saiz (Rajah 2). Geometri momen tak ubah digunakan untuk mengesktrak fitur bagi simbol matematik berkaitan (Siti Mariyam et al. 1999).

Jadual 3 menunjukkan sebahagian daripada ciri momen tak ubah yang telah diekstrak daripada simbol bercetak matematik dengan pelbagai orientasi, dan Jadual 4 adalah hasil pendiskretan simbol matematik menggunakan set kasar dengan ta'akulan Boolean berdasarkan kepada algoritma yang telah dibincangkan.

Kajian ini dilaksanakan dengan menggunakan momen geometri tak ubah daripada tertib ke-2 sehingga kepada tertib ke-4. Parameter pembelajaran dan momentum ditetapkan sebagai 0.2 dan 0.7, serta menggunakan fungsi sigmoid sebagai fungsi keaktifan. Kadar pencaman bagi simbol matematik adalah 95% dengan masa pemprosesan adalah 100 saat iaitu pencapaian yang baik jika dibandingkan dengan kajian sebelum ini oleh Siti Mariyam et al. (1999) tanpa menggunakan konsep set kasar (Rajah 3).



RAJAH 1. Perwakilan secara geometri



RAJAH 2. Sampel simbol matematik bercetak pelbagai orientasi

JADUAL 3. Fitur takubah bagi simbol matematik

Symbol	02	03	11	12	13	21	22	30	31
χ	0.86711	0.18849	0.08184	0.16839	0.12728	0.01923	0.24873	0.12638	0.04125
	0.54536	0.02198	0.02583	0.0241	0.01231	0.01844	0.1193	0.00087	0.00535
δ	0.58806	0.05518	0.08122	0.00895	0.07504	0.01626	0.18318	0.03664	0.05776
	0.61814	0.00880	0.05408	0.01927	0.05894	0.00178	0.07934	0.01363	0.02165
ε	0.88477	0.14812	0.01660	0.13137	0.06236	0.02861	0.21195	0.04551	0.00528
	0.80491	0.05006	0.03593	0.01596	0.04019	0.00195	0.12116	0.01324	0.01841
α	0.73293	0.05052	0.16291	0.05135	0.11263	0.02107	0.1385	0.00799	0.07375
	0.66253	0.08034	0.03918	0.01415	0.10883	0.01978	0.11662	0.0049	0.01161
β	0.91948	0.02059	0.01081	0.06653	0.00924	0.01543	0.15602	0.00388	0.00697
	0.82281	0.06182	0.02135	0.03221	0.03237	0.01006	0.12365	0.00398	0.00606
μ	2.213	0.71402	0.059	0.22918	0.00903	0.01181	0.63556	0.05279	0.08960
	2.15402	0.18761	0.08548	0.33771	0.81689	0.11741	0.70659	0.03468	0.13071
π	0.15565	0.00002	0.00662	0.00547	0.00182	0.00775	0.03896	0.02263	0.00017
	0.16081	0.01299	0.01091	0.00812	0.00205	0.01267	0.04902	0.04908	0.01069

JADUAL 4. Pendiskretan simbol matematik

Orientasi	02	03	11	12	13	21	22	30	31	Keputusan
Orientasi #1	1	2	1	2	2	2	2	1	2	(
Orientasi #2	0	1	0	1	1	1	1	0	0	(
Orientasi #1	0	1	1	0	2	1	2	1	2	(
Orientasi #2	0	0	1	1	1	0	0	1	1	(
Orientasi #1	2	2	0	2	1	2	2	1	0	(
Orientasi #2	1	1	0	1	1	0	1	1	1	(
Orientasi #1	0	1	1	1	2	2	1	0	2	(
Orientasi #2	0	2	1	0	2	2	0	0	1	(
Orientasi #1	2	0	0	2	0	1	1	0	1	(
Orientasi #2	1	1	0	1	1	0	1	0	0	(
Orientasi #1	2	2	1	2	0	1	2	1	2	(
Orientasi #2	2	2	1	2	2	2	2	1	2	(
Orientasi #1	0	0	0	0	0	0	0	1	0	(
Orientasi #2	0	0	0	0	0	1	0	1	1	(

KESIMPULAN

Konsep set kasar dengan teknik pendiskretan data menggunakan ta'akulan Boolean merupakan pendekatan yang baik untuk dipertimbangkan sekiranya kajian yang berkaitan melibatkan proses pengkelasan dan pencaman. Ini bertujuan untuk mendapatkan satu kadar pengkelasan dan pencaman yang tinggi di samping mempercepatkan masa penumpuan bagi teknik pengkelasan dan pencaman yang digunakan. Oleh yang sedemikian, kajian berkenaan pengkelasan dan pencaman simbol matematik sedang dilakukan menggunakan teknik rangkaian neural.

RUJUKAN

- Nguyen, H. S. 1998. Discretization Problem for Rough Sets Methods. In Polkowski, et al. (Eds.) *Proceedings of the First International Conference on Rough Sets and Current Trend in Computing (RSCTC'98)*. Warsaw, Poland, June 1998. pp 545-552.
- Nguyen, H. S. and Skowron, A. 1995. Quantization of Real-Valued Attribute. *Proceedings of the 2nd International Joint Conference on Information Science*, Wrightsville Beach, NC, September 1995. pp 34-37.
- Pawlak, Z. 1991. *Rough Set: Theoretical Aspect of Reasoning About Data*. New York: Kluwer Publications.
- Siti Mariyam Shamsuddin, Md Nasir Sulaiman dan Maslina Darus. 1999. Classification and Clustering of Printed Mathematical Symbols with Improved Backpropagation and Self Organizing Map. *International Bulletin of Malaysian Mathematical Society* 22(2): 57-167.

Rokiah@Rozita Ahmad
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor D.E.
rozy@pkrisc.cc.ukm.my

Maslina Darus
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor D.E.
maslina@pkrisc.cc.ukm.my

Siti Mariyam Hj. Shamsuddin
Fakulti Sains Komputer dan Teknologi Maklumat,
Universiti Teknologi Malaysia,
81310 UTM Skudai
Johor D.T.
mariyam@fsksm.utm.my

Azuraliza Abu Bakar
Fakulti Teknologi dan Sains Maklumat
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor D.E.
aab@ftsm.ukm.my