

PENGGUNAAN FUZZY SUBTRACTIVE CLUSTERING PADA PENGELOMPOKAN SISWA

Arifin

Univeristas Muhammadiyah Pare-Pare,

Info:

Jurnal MSA Vol. 4 No. 1
Edisi: Januari – Juni 2016
Artikel No.: 1
Halaman: 1 - 4
ISSN: 2355-083X
Prodi Matematika UINAM

ABSTRAK

Algoritma *clustering* telah digunakan secara luas dalam menangani permasalahan pengelompokan data mining, dengan meminimumkan jarak antar objek didalam *cluster* (*intra-cluster*) dan memaksimalkan jarak antar *cluster* (*inter-cluster*) sehingga mendapatkan kualitas *clustering* yang baik. Pengelompokan siswa dalam suatu kelas biasanya dilakukan berdasarkan nomor induk siswa dan urutan nilai rata-rata semua mata pelajaran. Makalah ini mengusulkan sistem pengelompokan siswa yang didasarkan atas kesamaan setiap mata pelajaran utama sehingga siswa yang memiliki kemampuan diatas atau dibawah rata-rata dapat dilayani dengan baik. Pengelompokan siswa tersebut akan dianalisis dengan menggunakan metode *fuzzy subtractive clustering*.

Kata Kunci: clustering, fuzzy

1. PENDAHULUAN

Analisis *Clustering* merupakan suatu proses pengelompokan objek-objek yang didasarkan pada ukuran kesamaan atau ketidaksamaan (Johnson dan Wichern, 2002). Analisis *Clustering* merupakan salah satu analisis data eksploratori yang bertujuan untuk menentukan kelompok atau grup dari sekelompok data. Prinsip yang digunakan dalam *clustering* adalah meminimumkan jarak antar objek didalam *cluster* (*intra-cluster*) dan memaksimalkan jarak antar *cluster* (*inter-cluster*) sehingga mendapatkan kualitas *cluster* yang baik (Jang dkk, 2003).

Pengelompokan yang mempertimbangkan tingkat keanggotaan yang mencakup himpunan *fuzzy* sebagai dasar pembobotan bagi pengelompokan disebut dengan *fuzzy clustering*, termasuk diantaranya *fuzzy subtractive clustering* (Castellano dkk, 2003). Metode ini merupakan pengembangan dari metode *partitional* dengan pembobotan *fuzzy* yang memungkinkan pengelompokan dimana kelompok data tidak terdistribusi secara jelas. Pada bidang pendidikan, analisis *clustering* dapat diaplikasikan dalam pengelompokan siswa.

2. PENGELOMPOKAN SISWA

Kebanyakan sekolah memberikan perlakuan yang standar (rata-rata), bersifat klasikal dan massal, terhadap semua siswa. Akibatnya, siswa

yang di bawah rata-rata, akan selalu tertinggal dalam mengikuti kegiatan belajar-mengajar yang berlangsung, sebaliknya siswa yang di atas rata-rata, akan merasa jenuh karena harus menyesuaikan diri dengan kecepatan belajar siswa-siswa lainnya (Widyastono, 2001).

Pengelompokan siswa perlu didasarkan pada tingkat kecerdasannya sehingga siswa dapat terlayani secara baik dan potensi yang dimiliki dapat tersalurkan atau berkembang secara optimal. Selama ini pengelompokan siswa didasarkan pada nilai rata-rata sehingga tidak menginterpretasikan semua nilai mata pelajaran. Sehingga diperlukan metode pengelompokan yang didasarkan pada kesamaan/ketidaksamaan antar mata pelajaran, misalnya Matematika, Bahasa Indonesia dan Bahasa Inggris. Model Matematis

3. Himpunan Fuzzy

Teori himpunan *fuzzy* dikembangkan oleh Lotfi Zadeh (Zimmermann, 2000) yang bertujuan untuk menerangkan masalah *imprecision* atau ketidaktepatan secara matematis. Masalah ketidaktepatan ini disesuaikan dengan pola pikir manusia yang bersifat ambigu/relatif. Konsep dari himpunan *fuzzy* sejalan dengan himpunan tegas, hanya saja derajat atau tingkat keanggotaan dari himpunan *fuzzy* bersifat kontinyu $[0,1]$.

Himpunan tegas atau himpunan klasik adalah himpunan dengan batas yang jelas. Jika X adalah koleksi objek-objek yang dinyatakan secara generik (umum) oleh x, maka himpunan fuzzy \tilde{A} di X didefinisikan sebagai himpunan pasangan terurut. $\tilde{A} = \{(x; \mu_{\tilde{A}}(x)) | x \in X\}$. Disini $\mu_{\tilde{A}}(x)$ disebut fungsi keanggotaan (membership function) atau disingkat (MF) untuk himpunan fuzzy \tilde{A} . MF memetakan setiap elemen x ke tingkat keanggotaan (membership grade) atau nilai keanggotaan (membership value) antara 0 dan 1. Sangat jelas bahwa definisi himpunan fuzzy adalah perluasan sederhana dari definisi himpunan tegas yang membolehkan fungsi karakteristik mempunyai nilai-nilai antara 0 dan 1.

4. Fuzzy Subtractive Clustering (FS)

Fuzzy subtractive clustering pertama kali dikemukakan oleh Chiu (1994). Metode ini pada awalnya tidak perlu ditentukan jumlah kelompok atau clustering yang akan dibentuk, tetapi akan terbentuk berdasarkan kriteria rasio (Castellano dkk, 2003). Konsep dasar dari fuzzy subtractive clustering adalah menentukan daerah-daerah dalam suatu variabel yang memiliki densitas tinggi terhadap titik-titik disekitarnya (Jang dkk, 1997). Titik yang telah terpilih sebagai pusat cluster akan dikurangi densitasnya.

Pada penerapannya, fuzzy subtractive clustering menggunakan faktor pembanding, yaitu accept ratio dan reject ratio yang merupakan bilangan pecahan 0 sampai 1. Accept ratio merupakan batas bawah dimana suatu titik data yang menjadi kandidat (calon) pusat cluster diperbolehkan untuk menjadi pusat cluster. Sedangkan reject ratio merupakan batas atas dimana suatu titik data yang menjadi kandidat (calon) pusat cluster tidak diperbolehkan untuk menjadi pusat cluster. Algoritma fuzzy subtractive clustering adalah sebagai berikut:

- 1) Input data yang akan dicluster $X_{m \times n}$ (dimana m adalah banyaknya data dan n adalah banyaknya variabel).
- 2) Tetapkan nilai:
 - a. Jari-jari setiap atribut data (r_j)
 - b. Faktor kuadrat (q)
 - c. Accept ratio (ar) dan Reject ratio (rr)
 - d. Minimum (\min_j) dan maximum (\max_j) data yang diperbolehkan.
 dimana $i = 1, 2, \dots, m$

- 3) Lakukan normalisasi:

$$x_{ij} = \frac{x_{ij} - \min_j}{\max_j - \min_j}$$

dimana $j = 1, 2, \dots, n$

- 4) Tentukan potensi awal tiap-tiap titik data (sebanyak n)

$$D_i = \sum_{k=1}^m e^{-4 \left(\sum_{j=1}^n \left(\frac{r_j - x_{kj}}{r_j} \right)^2 \right)}$$

- 5) Pilih titik awal dengan potensi tertinggi

$$P_t = \max [D_i]$$

$$h = i \text{ sehingga } D_i = P_t$$

$$z = P_t$$

- 6) Tentukan kondisi berdasarkan rasio.

- a. Rasio = z/P_t

- b. Jika rasio > accept, maka kondisi = 1

- c. Jika rasio < accept dan rasio > reject maka

- i. $S_i = \sum_{j=i}^n \left(\frac{V_j - Center_{ij}}{r} \right)^2$

- ii. $SS = \sqrt{\min(S_i)}$

- iii. Jika (Rasio + SS) ≥ 1 , maka kondisi = 1 dan Jika (Rasio + SS) < 1, maka kondisi = 2.

- d. Jika rasio < accept dan rasio < reject maka kondisi = 2

- 7) Tentukan pusat cluster dan kurangi potensinya terhadap titik-titik disekitarnya.

- a. Jika kondisi = 1 maka Kurangi potensi dari titik-titik didekat pusat cluster

- i. $c = c + 1$

- ii. $P_{stc} = v$

- iii. $D_c = P_t * e^{-4 \left[\sum_{j=1}^n \left(\frac{v_j - x_{ij}}{r_j * q} \right)^2 \right]}$

- iv. $D = D - D_c$

- v. Jika $D_i < 0$, maka potensi $i = 0$

- b. Jika kondisi = 2 maka potensi $i = 0$

$$D_h = 0$$

- 8) Update titik dengan potensi tertinggi

- i. $z = \max[D_i]$

- ii. $h = i$ sehingga $D_i = z$

- iii. Jika $z \neq 0$, maka ulangi langkah 6

- 9) Kembalikan pusat cluster dari bentuk ternormalisasi ke bentuk semula.

$$Pst_{kj} = Pst_{kj} * (\max_j - \min_j) + \min_j$$

10) Hitung nilai sigma cluster

$$\sigma_j = \frac{r_j * (\max_j - \min_j)}{\sqrt{8}}$$

11) Hitung derajat keanggotaannya

$$\mu_{ki} = e^{-\sum_{j=1}^n \frac{(x_{ij} - Pst_{kj})^2}{2\sigma_j^2}}$$

5. Kualitas Cluster

Kualitas cluster diperoleh berdasarkan hasil perbandingan antara rata-rata jarak intra cluster (*intra-similarity*) dan rata-rata jarak antar cluster (*inter-similarity*). Suatu cluster dikatakan baik jika mempunyai nilai kualitas cluster yang besar, atau *intra-similarity* yang tinggi dan *inter-similarity* yang rendah. *Intra-similarity* dan *inter-similarity* dihitung berdasarkan rumus sebagai berikut (Castelano dkk, 2003):

$$Intra - similarity = \frac{2 * \sum_{i=1}^{\frac{1}{2}((c_i-1)*c_i)} S(A, B)}{(c_i - 1) * c_i}$$

dimana S(A,B) adalah ukuran kesamaan, A,B adalah cluster ke-i dan A < B, dan c_i adalah banyaknya anggota dari cluster ke-i.

$$Inter - similarity = \frac{\sum_{k=1}^{c_i * c_j} S(A, B)}{c_i * c_j}$$

dimana S(A,B) adalah ukuran kesamaan, A ∈ cluster ke-i; B ∈ cluster ke j, c_i adalah banyaknya anggota dari cluster ke-i, dan c_j adalah banyaknya anggota dari cluster ke-j.

Nilai S(A,B) dihitung berdasarkan rumus jarak Euclidean (Bernstein dkk, 2006):

$$S(A, B) = \frac{1}{1 + \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}}$$

Jadi kualitas cluster (KC) dapat dihitung berdasarkan rumus

$$KC = \frac{(n-1) * \sum_{i=1}^n (Intra - similarity)_i}{2 * \sum_{j=1}^n (Inter - similarity)_j}$$

dimana n adalah banyaknya cluster.

6. Simulasi

Data nilai rapor dari 100 siswa dengan mata pelajaran Matematika, Bahasa Indonesia, dan

Bahasa Inggris yang dibangkitkan secara acak (terlampir) dan dikelompokkan dengan menggunakan *fuzzy subtractive clustering* serta menghitung kualitas masing-masing cluster diperoleh hasil sebagai berikut (Program Matlab):

- a. *fuzzy subtractive clustering* untuk data dengan variansi nilai Matematika, Bahasa Indonesia, dan Bahasa Inggris masing-masing adalah 0,53; 0,33; dan 0,51

i. Pusat cluster

Cluster	Nilai Mata Pelajaran		
	Mat	B. Ind	B. Ing
1	7	7	7
2	6	7	6
3	7	9	6

ii. Hasil Pengelompokan

	Cluster1	Cluster 2	Cluster 3
Siswa	1,3,8,9,10,12,13,15,20,29,30,31,32,34,35,36,37,38,42,45,47,48,52,56,59,63,65,66,67,69,71,72,74,75,80,81,84,85,95,96,98	2,4,5,6,7,11,14,16,17,18,19,21,22,23,24,25,26,27,28,33,39,40,41,43,44,49,51,53,55,57,61,64,68,70,76,77,78,82,83,86,88,90,91,92,93,94,99,100	50,54,58,60,73,79,87,89,97

- b. *fuzzy subtractive clustering* untuk data dengan variansi nilai Matematika, Bahasa Indonesia, dan Bahasa Inggris masing-masing adalah 0,17; 0,05; dan 0,08

i. Pusat cluster

Cluster	Nilai Mata Pelajaran		
	Mat	B. Ind	B. Ing
1	7	7	7
2	6	7	6
3	7	9	6

ii. Hasil Pengelompokan

	Clust1	Clust. 2	Clust. 3
Siswa	1,2,6,12,13,17,20,23,26,27,29,30,31,38,39,41,43,44,47,50,56,57,59,60,65,68,71,72,73,78,80,81,83,87,89,92,97,98,99	3,5,7,8,9,11,14,24,25,28,33,34,36,40,42,45,48,52,54,58,63,64,66,67,70,74,82,84,86,88,91	4,10,15,16,18,19,21,22,32,35,37,46,49,51,53,55,61,62,69,75,76,77,79,85,90,94,95,96,100

c. Kualitas *cluster*

Fuzzy Subtractive Clustering	Kualitas Cluster
Variansi 0,53; 0,33; dan 0,51 (a)	0,4194
Variansi 0,17; 0,05; dan 0,08 (b)	0,4776

7. Kesimpulan

Pengelompokan yang mempertimbangkan tingkat keanggotaan yang mencakup himpunan *fuzzy* sebagai dasar pembobotan bagi pengelompokan disebut dengan *fuzzy clustering*, termasuk diantaranya *fuzzy subtractive clustering*. Pengelompokan dengan metode *fuzzy subtractive clustering* adalah menentukan daerah-daerah dalam suatu variabel yang memiliki densitas tinggi terhadap titik-titik disekitarnya.

Berdasarkan hasil simulasi diatas, dapat dibentuk *cluster* siswa dengan menggunakan metode *Fuzzy Subtractive Clustering*. Pada simulasi tersebut, pengelompokan dengan *Fuzzy Subtractive Clustering* sulit digunakan apabila jumlah *cluster* (rombongan belajar) telah ditetapkan terlebih dahulu. Variansi dari data yang dibangkitkan berpengaruh terhadap hasil pengelompokan.

DAFTAR PUSTAKA

Bernstein, A., Kaufmann, E., Kiefer, C., dan Burki, C. (2006), *SimPack: A Generic Java Library for Similarity Measures in Ontologies*, University of Zurich Switzerland.

Castelano, G., Castiello, C., Fanelli, A.M., dan Sidella, M. (2003), *Symbolic Fuzzy Clustering for Mining User Profile*, Italy.

Dumitrescu, D., Lazzerini, B., dan Jain, L.C. (2000), *Fuzzy Sets and Their Application to Clustering and Training*, CRC Press, United States of America.

Jang, J.S., Sun, C.T., dan Mizutani, E. (1997), *Neuro-Fuzzy and Soft Computing*, Prentice Hall, New York.

Johnson, R.A., dan Wichern, D.W. (2002), *Applied Multivariate Statistical Analysis*, Prentice Hall, New York.

Klir, G.J. dan Yuan, B. (2001), *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Inc., New Delhi.

Kusumadewi, S. dan Purnomo, H. (2004), *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*, Graha Ilmu, Yogyakarta.

Pop, H.F. (2004), *Fuzzy Sets in Data Analysis: Between Theory and Applications*, Babeş-Bolyai University.

Susanto, S., Suharto, I.G.N., dan Sukapto, P. (2002), "Using Fuzzy Clustering Algorithm for Allocation of Student", *World Transaction on Engineering and Technology Education*. Vol. 1, No. 2, hal. 245-248.

Widyastono, H. (2001), *Sistem Percepatan Kelas (Akselerasi) bagi Siswa yang Memiliki Kemampuan dan Kecerdasan Luar Biasa*, Jakarta.

Yen, J. & Langari, R. (1999), *Fuzzy Logic: Intelligence, Control, and Information*, Prentice-Hall, Inc., United States of America.

Zimmermann, H.J. (2000), *Fuzzy Set Theory and Its Applications*. Kluwer Academic Publishers, London.