

LECTURE NOTES

COMP8014

Knowledge Data Discovery

Classification: Advanced Methods 1

LEARNING OUTCOMES

LO4: Mahasiswa diharapkan mampu untuk menentukan metode data mining yang sesuai dengan sifat kealamiahannya permasalahan yang dihadapi.

LO5: Mahasiswa diharapkan mampu untuk mengimplementasikan metode data mining.

OUTLINE MATERI :

1. Bayes classification methods
2. Rule-based methods
3. Bayesian Belief networks

ISI MATERI

Kenapa Klasifikasi Bayesian?

- Pengelompokan statistik: melakukan prediksi probabilistik, yaitu memprediksi probabilitas keanggotaan kelas
- Berdasarkan Teorema Bayes.
- Kinerja: Pengklasifikasi Bayesian sederhana, pembaur Bayesian naif, memiliki kinerja yang sebanding dengan pohon keputusan dan pengenalan jaringan syaraf terpilih
Inkremental: Setiap contoh pelatihan dapat secara bertahap meningkatkan / menurunkan probabilitas bahwa hipotesis benar - pengetahuan sebelumnya dapat dikombinasikan dengan data yang diamati.
- Standar: Bahkan ketika metode Bayesian secara komputasi sulit dilakukan, mereka dapat memberikan standar pengambilan keputusan yang optimal yang dengannya metode lain dapat diukur.

Teorema Bayes:

$$P(B) = \sum_{i=1}^M P(B|A_i)P(A_i)$$

$$P(H|\mathbf{X}) = \frac{P(\mathbf{X}|H)P(H)}{P(\mathbf{X})} = P(\mathbf{X}|H) \times P(H) / P(\mathbf{X})$$

(Note: Keterangan notasi bisa dilihat di slide)

Naïve Bayes Classifier:

Contoh penggunaan bisa dilihat di slide.

Diberikan data sebagai berikut, dengan dua kelas

C1:buys_computer = 'yes'

C2:buys_computer = 'no'

age	income	student	credit_rating	comp
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Problem: dengan metode Naïve Bayes Classifier, data X = (age <=30, Income = medium, Student = yes, Credit_rating = Fair) akan masuk ke kelas C1 atau C2?

Keunggulan dan kelemahan metode naïve bayes:

Keuntungan:

- Mudah diimplementasikan
- Hasil bagus didapat pada sebagian besar kasus
- Robust ke titik kebisingan terisolasi
- Tangani nilai yang hilang dengan mengabaikan contoh selama perhitungan perkiraan probabilitas
- Robust terhadap atribut yang tidak relevan
- Asumsi independensi mungkin tidak berlaku untuk beberapa atribut
- Gunakan teknik lain seperti Bayesian Belief Networks (BBN) untuk meningkatkan akurasi.

Kekurangan:

Asumsi: independensi bersyarat kelas, oleh karena itu kehilangan keakuratan

- Praktis, ketergantungan ada antar variabel
 - Misalnya, rumah sakit: pasien: Profil: usia, riwayat keluarga, dll.
Gejalanya: demam, batuk dll. Penyakit: kanker paru-paru, diabetes, dll.
 - Ketergantungan di antara ini tidak dapat dimodelkan oleh Naïve Bayes Classifier

Bayesian Belief Network (BNN):

Kelebihannya dibandingkan dengan naïve bayes classifier adalah BNN memungkinkan independensi bersyarat kelas antara himpunan bagian dari variabel.

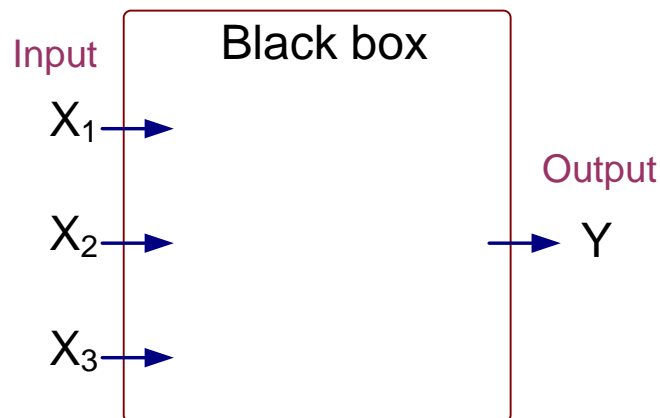
Untuk contoh: lihat di slide ppt yang sudah disediakan.

Backpropagation Neural Network (BPNN)

- Backpropagation: Algoritma pembelajaran jaringan syaraf tiruan
- Dimulai oleh psikolog dan neurobiologists untuk mengembangkan dan menguji analog komputasi neuron
- Jaringan syaraf tiruan: Satu set unit input / output terhubung dimana masing-masing sambungan memiliki bobot yang terkait dengannya
- Selama tahap pembelajaran, jaringan belajar dengan menyesuaikan bobot sehingga bisa memprediksi label kelas yang benar dari input tupel.
- Juga disebut sebagai pembelajaran koneksionis karena adanya hubungan antar unit

Contoh BPNN:

X_1	X_2	X_3	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0



Dari contoh di atas, diperoleh rule bahwa output Y adalah 1 jika paling sedikit dua dari tiga input sama dengan 1.

SIMPULAN

1. Telah dijelaskan konsep dari Bayes classifiers
2. Telah diperkenalkan teknik dan algoritma dari Naive Bayes classifiers
3. Telah diperkenalkan teknik dan algoritma dari Bayesian Belief networks
4. BPNN

DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, Y. (2006). "Data Mining: Concepts and Technique". Edisi 3. Morgan Kaufman. San Francisco
2. Tan, P.N., Steinbach, M., & Kumar, V. (2006). "Introduction to Data Mining". Addison-Wesley. Michigan
3. Witten, I. H., & Frank, E. (2005). "Data Mining : Practical Machine Learning Tools and Techniques". Second edition. Morgan Kaufmann. San Francisco