Lazy Associative Classification

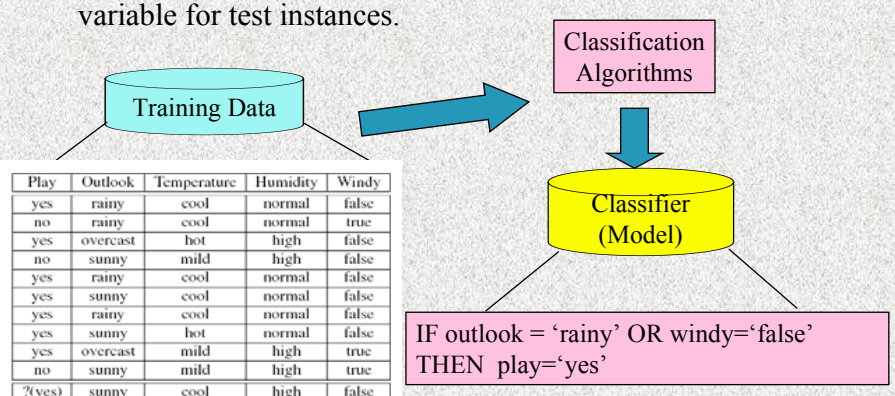
By Adriano Veloso,Wagner Meira Jr. , Mohammad J. Zaki

Presented by:
Fariba Mahdavifard
Department of Computing Science
University of Alberta

Contents:

Classification: Model Construction and Prediction

- **Learning Step:** The training data is used to construct a model which relates the feature variables.
- **Test Step:** The training model is used to predict the class variable for test instances.



Play	Outlook	Temperature	Humidity	Windy
yes	rainy	cool	normal	false
no	rainy	cool	normal	true
yes	overcast	hot	high	false
no	sunny	mild	high	false
yes	rainy	cool	normal	false
yes	sunny	cool	normal	false
yes	rainy	cool	normal	false
yes	sunny	hot	normal	false
yes	overcast	mild	high	true
no	sunny	mild	high	true
?(yes)	sunny	cool	high	false

Classification Models

- Several models have been proposed over the years, such as neural network, statistical model, decision trees (DT), genetic algorithms, etc.
- The most suitable one for data mining is DT.
  - DT could be constructed relatively fast
  - DT models are simple and easy to be understood.

Contents:

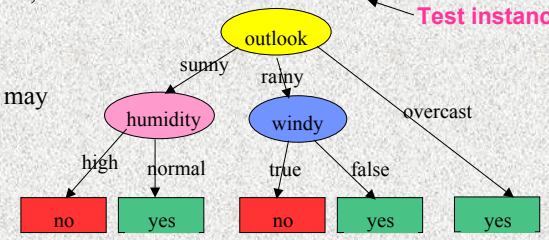
- Classification
- **Decision Tree Classifier**
- (Eager) Associative Classifier
- Comparison between Decision Tree and Associative Classifier
- Lazy Associative Classifier
- Comparison between Lazy and Eager Associative Classifier
- Shortcomings of Lazy Associative Classifier
- Conclusion

Decision Tree Classifier

- At each internal node, the best split is chosen according to the information gain criterion.
- A DT is built using a greedy recursive splitting strategy
- Decision tree can be considered as a set of disjoint decision rules, with one rule per leaf.
- Such greedy (local) search may prune important rules!

Play	Outlook	Temperature	Humidity	Windy
yes	rainy	cool	normal	false
no	rainy	cool	normal	true
yes	overcast	hot	high	false
no	sunny	mild	high	false
yes	rainy	cool	normal	false
yes	sunny	cool	normal	false
yes	rainy	cool	normal	false
yes	sunny	hot	normal	false
yes	overcast	mild	high	true
no	sunny	mild	high	true
?(yes)	sunny	cool	high	false

Test instanc



Contents:

- Classification
- Decision Tree Classifier
- **(Eager) Associative Classifier**
- Comparison between Decision Tree and Associative Classifier
- Lazy Associative Classifier
- Comparison between Lazy and Eager Associative Classifier
- Shortcomings of Lazy Associative Classifier
- Conclusion

Eager Associative Classifier

- Class association rules (CARs) : $X \rightarrow c$
- CARs are essentially decision rules
- They are r

Antecedent is composed of feature variables

Consequent is class

- During the testing phase, Associative classifier checks weather each CAR matches the test instance.
- The class associated with the first match is chosen.

Note:

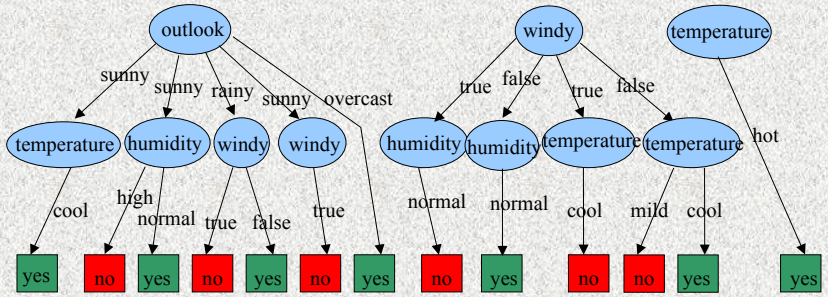
- **Decision tree is a greedy search for CARs that only expands the current best rule.**
- **Eager Associative Classifier mines all possible CARs with a given minimum support.**

Eager Associative Classifier Steps:

1. Algorithm mines all frequent CARs
2. Sort them in descending order of information gain.
3. For each test instance, the first CAR matching that, is used to predict the class.

let $\mathcal{D}$ be the set of all n training instances
let $\mathcal{T}$ be the set of all m test instances
1. let $\mathcal{C}^e$ be the set of all rules $\{\mathcal{X} \rightarrow c\}$ mined from $\mathcal{D}$
2. sort $\mathcal{C}^e$ according to information gain
3. for each $t_i \in \mathcal{T}$ do
4. pick the first rule $\{\mathcal{X} \rightarrow c\} \in \mathcal{C}^e | \mathcal{X} \subseteq t_i$
5. predict class c

Eager Associative Classifier



- Three CARs match the test instance are:

  outlook=sunny, temperature=cool, humidity=high -> play???

  1. {windy=false and temperature=cool -> play=yes}

The first rule would be selected, since it is the best ranked CAR.

Contents:

Comparison between Decision Tree and Associative Classifier

- The test instance is recognized by only on rule in decision tree.
- The same test instance is recognized by three CARs in associative classifier.
- **Intuitively associative classifiers perform better than decision trees because it allows several CARs to cover the same portion of the training data**.

- **Theorem1:** The rules derived from a decision tree are subset of the CARs mined using an eager associative classifier based on information gain.
- **Theorem 2:** CARs perform no worse than decision tree rules, according to the information gain principle.

Contents:

Lazy Learning Algorithms

- Eager learning methods create the classification model during the learning phase using training data
- But lazy learning methods postpone generalization and building the classification model until a query is given.

Lazy Associative Classifier

Lazy Associative Classifier induces CARs specific to each test instance.

1. Lazy Associative Classifier projects the training data only on features in the test instance (from all training instances, only the instances sharing at least one feature with test instance are used)
2. From this projected training data, CARs are induced and ranked, and the best CAR is used.

let $\mathcal{D}$ be the set of all n training instances
let $\mathcal{T}$ be the set of all m test instances
1. for each $t_i \in \mathcal{T}$ do
2. let $\mathcal{D}_{t_i}$ be the projection of $\mathcal{D}$ on features only from t_i
3. let $\mathcal{C}^l_{t_i}$ be the set of all rules $\{\mathcal{X} \rightarrow c\}$ mined from $\mathcal{D}_{t_i}$
4. sort $\mathcal{C}^l_{t_i}$ according to information gain
5. pick the first rule $\{\mathcal{X} \rightarrow c\} \in \mathcal{C}^l_{t_i}$, and predict class c

Contents:

Comparison between Lazy and Eager Associative Classifier

Test Instance:

Outlook=overcast, Temperature=hot and Humidity=low -> play?

- The set of CARs found by eager classifier (minsup=40%) is composed of:
  1. {windy=false and humidity=normal -> play=yes}
  2. {windy=false and humidity=cool -> play=yes}

None of the two CARs matches the test instance!

Comparison between Lazy and Eager Associative Classifier

Test Instance:

Outlook=overcast, Temperature=hot and Humidity=low -> play?

- Lazy Associative Classifier projects the training data (D) by the features in the test instance A
- The projected training data (D_A) has less instances, therefore CARs not frequent in D may be frequent in D_A .
- The Lazy Associative Classifier found two CARs in D_A:
  1. {Outlook=overcast -> play=yes}
  2. {Temperature=hot -> play=yes}

Play	Outlook	Temperature	Humidity	Windy
no	–	–	–	true
yes	overcast	hot	–	–
yes	–	hot	–	–
yes	overcast	–	–	true
no	–	–	–	true
?(yes)	overcast	hot	low	true

- The Lazy CARs predict the correct class and they are also simpler compaerd to the eager ones.

Comparison between Lazy and Eager Associative Classifier

- **Intuitively, lazy classifiers perform better than eager classifiers because of two characteristic:**

1. **Missing CARs:**

- Eager classifiers search for CARs in a large search space.
- This strategy generates a large rule-set, but CARs that are important for some specific test instances may be missed!

- Lazy classifiers focus the search for CARs in a much smaller search space, which is induced by the features of the test instance.

Comparison between Lazy and Eager Associative Classifier

- **Intuitively, lazy classifiers perform better than eager classifiers because of two characteristic:**

2. Highly Disjunctive Spaces:

- Eager classifiers often combine small disjuncts to generate more general predictions. It will reduce classification performance in highly disjunctive spaces where single disjunct may be important to classify specific instances.

- Lazy classifiers generalize their training examples exactly as needed to cover the test instance. More appropriate in complex search spaces!

Contents:

Shortcomings of Lazy Associative Classifier

First Problem:

- The more CARs are generated, the better is the classifier??!
- NO! it sometimes leads to overfitting, reducing the generalization and affecting the classification accuracy.
- **Overfitting and high sensitivity to irrelevant features are shortcoming of lazy classifier.**
- Features should be selected carefully.

Shortcomings of Lazy Associative Classifier

Second Problem:

- **Lazy classifier typically requires more work to classify all test instances.**
- Caching mechanism is used to decrease this workload.
- The basic idea of caching: different test instances may induce different rule-sets, but different rule-sets may share common CARs.

Contents:

Conclusion

- Decision tree classifiers perform a greedy search that may discard important rules.
- Associative classifiers perform a global search for rules, however it may generate a large number of rules. (many of them may be useless during classification and even worse important rules may never be mined)

- Lazy associative classifier overcome these problems by focusing on the features of the given test instance.
  - Lazy classifier is suitable in highly disjunctive spaces.
  - The most important problem of lazy classifier is its overfitting.

Thanks for you attention!

Question?

Reference

- A. Veloso,W. Meira Jr. , M. J. Zaki. "Lazy Associative Classification". In ICDM '06: *Proceedings of the Sixth International Conference on Data Mining,* pages 645-654, IEEE Computer Society, 2006.

- Y. Sun, A. K.C.Wong, and Y. Wang. An overview of associative classifiers. In *Proceedings of the 2006 International Conference on Data Mining, DMIN 2006*, pages 138–143. CSREA Press, 2006.