Naive Credal Classifier 2: an extension of Naive Bayes for delivering robust classifications

G. Corani M. Zaffalon

IDSIA
Switzerland
giorgio{zaffalon}@idsia.ch

SIPTA '08

Outline

1. Introducing NCC2
   - Background
   - Credal classifiers
   - NCC2
2. Experimental Results
   - Setup and indicators
   - Indeterminate classifications vs posterior probabilities
3. Software demonstration

Naive Bayes Classifier (NBC)

- *Naive* assumption (statistical indep. of the features given the class):

$$\theta_{c|f_1,f_2,\ldots f_k} \propto \theta_c \prod_{i=1}^{i=k} \theta_{f_i|c}$$

Probability computation

$$\theta_{POST} \propto \theta_{LIKELIHOOD}\,\theta_{PRIOR}$$

- *Maximum likelihood* estimators are for instance $\hat{\theta}_c = n(c)/N$ and $\hat{\theta}_{f_i|c} = n(f_i|c)/n(c)$.
- *The choice of any specific prior introduces necessarily some subjectivity, even if it is a non-informative one.*

Naive Bayes Classifier (NBC)

- *Naive* assumption (statistical indep. of the features given the class):

$$\theta_{c|f_1,f_2,\ldots f_k} \propto \theta_c \prod_{i=1}^{i=k} \theta_{f_i|c}$$

Probability computation

$$\theta_{POST} \propto \theta_{LIKELIHOOD}\,\theta_{PRIOR}$$

- *Maximum likelihood* estimators are for instance $\hat{\theta}_c = n(c)/N$ and $\hat{\theta}_{f_i|c} = n(f_i|c)/n(c)$.
- *The choice of any specific prior introduces necessarily some subjectivity, even if it is a non-informative one.*

NBC and prior sensitivity

- NBC computes a single posterior distribution.
- However, the most probable class might depend on the chosen prior, especially on *small data sets*.
- Prior-dependent classifications might be fragile.
- Solution via set of probabilities:
  - Robust Bayes Classifier (Ramoni and Sebastiani, 2001)
  - Naive Credal Classifier (Zaffalon, 2001)

Naive Credal Classifier (NCC) (Zaffalon, 2001)

- Extends Naive Bayes to imprecise probabilities; it specifies a set of priors by adopting the *Imprecise Dirichlet Model*.
- The set of priors is turned into a set of posteriors set via Bayes' rule.
- NCC returns the classes that are *non-dominated* within the set of posteriors.

Test of dominance and indeterminate classifications

Definition

Class c_1 dominates c_2 if $P(c_1) > P(c_2)$ in any distribution of the posterior credal set.
If no class dominates c_1, then c_1 is non-dominated.

- If there are more non-dominated classes, NCC returns all of them: the classification is *indeterminate.*
- NCC becomes indeterminate on the instances whose classification would be prior-dependent with NBC.
- Indeterminate classifications proofed to be viable in real world case studies (e.g., dementia diagnosis).

Incomplete data sets

- Most classifiers (including NBC) ignore missing data.
- This is correct only if data are missing-at-random (MAR).
- It is not possible to test the MAR hypothesis on the incomplete data.
- However, ignoring Non-MAR missing data can lead to unreliable conclusions.
- Missing data can be MAR for some features but not for some others; or can be MAR only in training and not in testing (or vice versa).

NCC2: NCC with conservative treatment of missing data (I)

- NCC2 receives the declaration of which features have Non-MAR missing data and at which stage (learning, testing or both).
- NCC2 ignores MAR missing data.
- NCC2 deals *conservatively* with Non-MAR missing data.

Conservative treatment of missing data (learning set)

- All possible completions of missing data are seen as possible.
- A set of likelihoods is computed.
- A set of posteriors is computed from a set of priors and a set of likelihoods.
- The conservative treatment of missing data can generate additional indeterminacy.

NCC2: NCC with conservative treatment of missing data (I)

- NCC2 receives the declaration of which features have Non-MAR missing data and at which stage (learning, testing or both).
- NCC2 ignores MAR missing data.
- NCC2 deals *conservatively* with Non-MAR missing data.

Conservative treatment of missing data (learning set)

- All possible completions of missing data are seen as possible.
- A set of likelihoods is computed.
- A set of posteriors is computed from a set of priors and a set of likelihoods.
- The conservative treatment of missing data can generate additional indeterminacy.

NCC2: NCC with conservative treatment of missing data (II)

Conservative treatment of missing data in the instance to be classified

- All possible completions of missing data are seen as possible, thus giving rise to several *virtual instances*.
- Test of dominance: c_1 should dominate c_2 on *all* the virtual instances.
- A procedure allows to find out the dominance relationships without actually building the virtual instances.
- Conservative treatment of missing data in the instance to classify can generate additional indeterminacy.

NCC2: NCC with conservative treatment of missing data (II)

Conservative treatment of missing data in the instance to be classified

- All possible completions of missing data are seen as possible, thus giving rise to several *virtual instances*.
- Test of dominance: c_1 should dominate c_2 on *all* the virtual instances.
- A procedure allows to find out the dominance relationships without actually building the virtual instances.
- Conservative treatment of missing data in the instance to classify can generate additional indeterminacy.

What to expect from NCC2

By adopting imprecise probabilities, NCC2 is designed to be robust to:

- prior specification, especially critical on small data sets;
- Non-MAR missing data, critical on incomplete data sets.
- However, excessive indeterminacy is undesirable.

What to expect from indeterminate classifications

- To preserve NCC2 reliability, avoiding too strong conclusions (a single class) on doubtful instances.
- To convey sensible information, dropping unlikely classes.

What to expect from NCC2

By adopting imprecise probabilities, NCC2 is designed to be robust to:

- prior specification, especially critical on small data sets;
- Non-MAR missing data, critical on incomplete data sets.
- However, excessive indeterminacy is undesirable.

What to expect from indeterminate classifications

- To preserve NCC2 reliability, avoiding too strong conclusions (a single class) on doubtful instances.
- To convey sensible information, dropping unlikely classes.

Indicators of performance for NCC2

NCC2

- *determinacy* (% of determinate classifications);
- *single-accuracy* (% of determ. classification that are accurate);
- *set-accuracy* (% of indeterm. classifications that contain the true class);
- *size of indeterminate output*, i.e., avg. number of classes returned when indeterminate.

Indicators for comparing NBC and NCC2

NCC2 vs NBC

- NBC(NCC2 D): accuracy of NBC on instances **determinately** classified by NCC2.
- NBC(NCC2 I): accuracy of NBC on instances **indeterminately** classified by NCC2.
- We expect NBC(NCC2 D) > NBC(NCC I).

Experiments on 18 UCI data sets

- MAR setup: 5% missing data generated via a MAR mechanism; all features declared as MAR to NCC2.
- Non-MAR setup: 5% missing data generated via a Non-MAR mechanism; all features declared as Non-MAR to NCC2.
- Average NBC accuracy under both settings: 82%.

NBC vs NCC2

- NBC (NCC2 D): 85%(95%)
- NBC (NCC2 I): 36%(69%)

On each data set and setup: NBC(NCC2 D)>NBC(NCC2 I)

NCC2

- determinacy: 95%(52%)
- single accuracy: 85%(95%)
- set-accuracy: 85%(96%)
- imprecise output size: ≅33% of the classes

- Indeterminate classifications do preserve the reliability of NCC2!

Experiments on 18 UCI data sets

- MAR setup: 5% missing data generated via a MAR mechanism; all features declared as MAR to NCC2.
- Non-MAR setup: 5% missing data generated via a Non-MAR mechanism; all features declared as Non-MAR to NCC2.
- Average NBC accuracy under both settings: 82%.

NBC vs NCC2

- NBC (NCC2 D): 85%(95%)
- NBC (NCC2 I): 36%(69%)

On each data set and setup: NBC(NCC2 D)>NBC(NCC2 I)

NCC2

- determinacy: 95%(52%)
- single accuracy: 85%(95%)
- set-accuracy: 85%(96%)
- imprecise output size: ≅33% of the classes

- Indeterminate classifications do preserve the reliability of NCC2!

Experiments on 18 UCI data sets

- MAR setup: 5% missing data generated via a MAR mechanism; all features declared as MAR to NCC2.
- Non-MAR setup: 5% missing data generated via a Non-MAR mechanism; all features declared as Non-MAR to NCC2.
- Average NBC accuracy under both settings: 82%.

NBC vs NCC2

- NBC (NCC2 D): 85%(95%)
- NBC (NCC2 I): 36%(69%)

On each data set and setup: NBC(NCC2 D)>NBC(NCC2 I)

NCC2

- determinacy: 95%(52%)
- single accuracy: 85%(95%)
- set-accuracy: 85%(96%)
- imprecise output size: ≅33% of the classes

- Indeterminate classifications do preserve the reliability of NCC2!

Experiments on 18 UCI data sets

- MAR setup: 5% missing data generated via a MAR mechanism; all features declared as MAR to NCC2.
- Non-MAR setup: 5% missing data generated via a Non-MAR mechanism; all features declared as Non-MAR to NCC2.
- Average NBC accuracy under both settings: 82%.

NBC vs NCC2

- NBC (NCC2 D): 85%(95%)
- NBC (NCC2 I): 36%(69%)

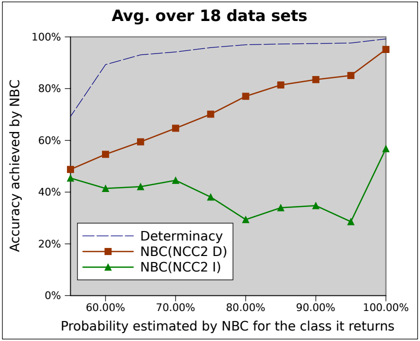
On each data set and setup: NBC(NCC2 D)>NBC(NCC2 I)

NCC2

- determinacy: 95%(52%)
- single accuracy: 85%(95%)
- set-accuracy: 85%(96%)
- imprecise output size: ≅33% of the classes

- Indeterminate classifications do preserve the reliability of NCC2!

NBC Probabilities vs indeterminate classifications (MAR setup, average over all data sets)



- Higher posterior probability of NBC → higher NCC2 determinacy.
- At any level of posterior probability, NBC(NCC2 D)>NBC(NCC2 I).
- Striking drop on the instances classified confidently by NBC.

Summary

- NCC2 extends Naive Bayes to imprecise probabilities, to robustly deal with small data sets and missing data.
- NCC2 becomes indeterminate on instances whose classification is doubtful indeed.
- Indeterminate classifications preserve the classifier' reliability while conveying sensible information.
- Bibliography, software and manuals: see `www.idsia.ch/~giorgio/jncc2.html`
- Software with GUI to arrive soon!