Clustering II

David Arroyo Guardeño

Initialization problem in K-means

- ✓ Extremely ☠ sensitive to cluster center initialization
- ✓ Bad initialization
  1. Poor convergence rate
  2. Bad overall clustering
- ✓ Safeguarding measures
  1. Choose first center, sencond which is the farthest form the first, third which is the farthest from both, so on
  2. Choose all centers randomly
  3. Try multiple initializations and choose the best result
  4. User other initialization procedures (Pena et al. 1999) $\Rightarrow$ In sklearn $\rightarrow$ **kmeans++**

kmeans++

Input: $\mathcal{S}$ set of all possible objects (with N attributes, and $\#\mathcal{S} = M$)
Input: K number of clusters (user-defined parameter)
Output: K centroids, T_i, with $T = \bigcup_i T_i$ and $i \in [0, K]$

INITIALIZATION: select $x \in \mathcal{S}$ randomly, and do $T \leftarrow \{x\}$
while $|T| < K$ **do**
  pick $x \in \mathcal{S}$ at random, with probability proportional to $cost(x, T) = \min_{z \in T} ||x - z||^2$
  $T \leftarrow T \bigcup \{x\}$
end

Examples

kmeans_initialization.ipynb

Comparing clusterings

- ✓ Is a clustering algorithm sensitive to small perturbations?
- ✓ Is the algorithm sensitive to the order of the data?
- ✓ How similar are the solutions of two different algorithms?
- ✓ In case there exists optimal solution, how far are we from that solution?

Why comparison?

Robustness: To combine and improve the results of different clustering algorithms

Re-use: Old clusterings that cannot be reconstructed but can be useful

Distribution computation/integration: Databases geographically split and centralization leads to $\Uparrow$ computational, bandwidth, and storage costs

Legal compliance: Legal restrictions impose several copies of data, each with a different feature set (think about anonymized data) $\Rightarrow$ feature distributed clustering + integration into one *mean value* clustering

Integration of different optimization criteria: In some scenarios, as in social sciences, we could have different clusterings obtained using several optimization criteria, e.g., different distance/similarity measures. A procedure to proper integration is required (Li et al. 2004)

Measures to compare clusters

As underlined in (S. Wagner and D. Wagner 2007):

1. Measures based on counting pairs
2. Measures based on set overlaps
3. Measures based on mutual information

Some definitions and notations

- ✓ Let X be a set of finite set with cardinality $|X| = n$
- ✓ Clustering $\mathcal{C}$ is a set $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k\}$ with $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset \;\; \forall i \neq j$, $X = \bigcup_{i=1}^{k} \mathcal{C}_i$, and assuming $|\mathcal{C}_i| \neq 0 \;\; \forall i \in \{1, 2, \ldots, k\}$
- ✓ $\mathscr{C}(X) \equiv$ the set of all clusterings of X
- ✓ Let $\mathcal{C}' = \{\mathcal{C}'_1, \ldots, \mathcal{C}'_l\} \in \mathscr{C}(X)$ be a second cluster of X

Definition (The confusion matrix $M = (m_{ij})$ or contigency table of the pair $\mathcal{C}, \mathcal{C}'$)

It is a $k \times l$ matrix whose ij-the entry is the number of elements in the intersection of the clusters $\mathcal{C}_i$ and $\mathcal{C}'_j$, i.e.,

$$m_{ij} = |\mathcal{C}_i \cap \mathcal{C}'_j|, \;\; i \in \{1, \ldots, k\}, j \in \{1, \ldots, l\}$$

Measures based on counting pairs

The set of all pairs of elements of X is the disjoint union of the following sets:

$$\mathcal{S}_{11} = \{\text{pairs included in the same cluster under } \mathcal{C} \text{ and } \mathcal{C}'\}$$

$$\mathcal{S}_{00} = \{\text{pairs included in different clusters under } \mathcal{C} \text{ and } \mathcal{C}'\}$$

$$\mathcal{S}_{10} = \{\text{pairs included in the same cluster under } \mathcal{C} \text{ but in different ones under } \mathcal{C}'\}$$

$$\mathcal{S}_{01} = \{\text{pairs included in different clusters under } \mathcal{C} \text{ but in but in the same one under } \mathcal{C}'\}$$

If $n_{ab} \triangleq |S_{ab}|, a, b \in 0, 1$ is the size of the respective set, then

$$n_{11} + n_{00} + n_{10} + n_{01} = \binom{n}{2}$$

Rand index I

Definition (General rand index)

It defines the ratio between the number of elements correct and uncorrectly classified and the total number of elements

$$\mathcal{R}(\mathcal{C}, \mathcal{C}') = \frac{2(n_{11} + n_{00})}{n(n-1)}$$

This measure is very dependent upon the number of clusters and, in the inlikely case of independent clusterings, it converges to 1 as the number of clusters increases (no desirable at all)

Rand index II

Definition (Adjusted Rand Index)

Assuming a generalized hypergeometric distribution as null hypothesis, it is given by

$$\mathcal{R}_{adj}(\mathcal{C},\mathcal{C}') = \frac{\sum_{i=1}^{k}\sum_{j=1}^{l}\binom{m_{ij}}{2} - t_3}{\frac{1}{2}(t_1+t_2) - t_3}$$

where $t_1 = \sum_{i=1}^{k}\binom{|\mathcal{C}_i|}{2}$, $t_2 = \sum_{i=1}^{j}\binom{|\mathcal{C}'_j|}{2}$, *and* $t_3 = \frac{2t_1t_2}{n(n-1)}$

PROBLEMS

- ✓ Strong assumptions on the distribution
- ✓ Sensitivity to the number of clusters

Rand index III

- ✓ The adjusted version can take negative values, but it should be in the interval $[0, 1]$
- ✓ For $n/k > 3$, the base-line of R_{adj} varies too much (Meilă 2007)
- ✓ High scores for clusterings with large number of clusters, since eventually all instances end up alone in a cluster

Measures to compare clusters

As underlined in (S. Wagner and D. Wagner 2007):

1. Measures based on counting pairs
2. Measures based on set overlaps
3. Measures based on mutual information

Now, clusterings that have a maximum absolute or relative overlap. Let us consider just one example:

Definition ($\mathscr{F}$-measure)

The $\mathscr{F}$-measure for a cluster $\mathcal{C}'_j$ with respect to a certain class $\mathcal{C}_i$ indicates how good the cluster $\mathcal{C}'_j$ describes the class $\mathcal{C}_i$. To do so, first it is calculated the harmonic mean of precision

$$p_{ij} = \frac{m_{ij}}{|C'_j|}$$

and recall

$$r_{ij} = \frac{m_{ij}}{|C_i|}$$

which leads to

$$\mathscr{F}(\mathcal{C}_i, \mathcal{C}'_j) = \frac{2 \cdot r_{ij} \cdot p_{ij}}{r_{ij} + p_{ij}}$$

The overall $\mathscr{F}$-measure is

$$\mathscr{F}(\mathcal{C}, \mathcal{C}') = \mathscr{F}(\mathcal{C}') = \frac{1}{n} \sum_{i=1}^{n} |C_i| \max_{j=1}^{l} \mathscr{F}(\mathcal{C}_i, \mathcal{C}'_j)$$

Problems of the measures based on overlaps

This kind of measures do not take into account unmatched parts of the clusters. For example

- ✓ Consider $\mathscr{C}'$ obtained from $\mathscr{C}$ just by shifting a fraction α of the elements of each cluster C_i to the *next* cluster $\mathcal{C}_{(i+1) \bmod k}$
- ✓ Let $\mathscr{C}''$ be a clustering derived from $\mathscr{C}$ by a reassigning a fraction α of the elements in each cluster $\mathcal{C}_i$ evenly between the others clusters
- ✓ If $\alpha < 0.5 \Rightarrow \mathscr{F}(\mathcal{C}, \mathcal{C}') = \mathscr{F}(\mathcal{C}, \mathcal{C}'')$... but $\mathcal{C}'$ is a less modified version of $\mathcal{C}$ than $\mathcal{C}''$!!!

Measures to compare clusters

As underlined in (S. Wagner and D. Wagner 2007):

1. Measures based on counting pairs
2. Measures based on set overlaps
3. Measures based on mutual information

Measures based on mutual information I

Definition (The entropy associated with clustering $\mathcal{C}$)

$$\mathcal{H}(\mathcal{C}) = -\sum_{i=1}^{k} P(i)\log_2(P(i))$$

with $P(i) = \frac{|\mathcal{C}_i|}{n}$

Measures based on mutual information II

Definition (The mutual information between two clusterings $\mathcal{C}$, $\mathcal{C}'$)

$$\mathcal{I}(\mathcal{C},\mathcal{C}') = \sum_{i=1}^{k}\sum_{j=1}^{k} P(i,j)\log_2 \frac{P(i,j)}{P(i)P(j)}$$

where $P(i,j)$ is the probability of an element belonging to cluster $\mathcal{C}_i$ in $\mathcal{C}$ and to cluster $\mathcal{C}'_j$ in $\mathcal{C}'$

$$P(i,j) = \frac{|\mathcal{C}_i \cap \mathcal{C}'_j|}{n}$$

The mutual information is a metric on the space of all clusterings

- ✓ It is not bounded by a constant value $\Rightarrow$ difficult to interpret
- ✓ $\mathcal{I}(\mathcal{C},\mathcal{C}') \leq min\{\mathcal{H}(\mathcal{C}),\mathcal{H}(\mathcal{C}')\}$

Definition (Normalized mutual information by Strehl & Ghosh)

$$NMI_{SG}(\mathcal{C}, \mathcal{C}') = \frac{\mathcal{I}(\mathcal{C}, \mathcal{C}')}{\sqrt{\mathcal{H}(\mathcal{C})\mathcal{H}(\mathcal{C}')}}$$

1. $0 \leq NMI_{SG}(\mathcal{C}, \mathcal{C}') \leq 1$
2. $NMI_{SG}(\mathcal{C}, \mathcal{C}') = 1 \Rightarrow \mathcal{C} = \mathcal{C}'$
3. $NMI_{SG}(\mathcal{C}, \mathcal{C}') = 0 \Rightarrow \quad \forall i \in \{1, \ldots, k\}$, and $\forall j \in \{1, \ldots, l\}$, $P(i, j) = 0$ <u>OR</u> $P(i, j) = P(i) \cdot P(j)$

Determining the number of clusters

- ✓ Most criteria (as SSE -Smallest Square Error-) are monotonically decreasing in K (i.e., the number of clusters)$\Rightarrow$ leads to the trivial cluster (one item per cluster)
- ✓ Alternatives mainly based on heuritic methodologies

Methods based on intra-cluster scatter

For example:

$$W_K = \sum_{k=1}^{K} \frac{1}{2N_k} D_k$$

with D_K as the sum of the pairwise distances for all instances in cluster k

$$D_k = \sum_{x_i, x_j \in \mathcal{C}_k} ||x_i - x_j||$$

Usually as the number of clusters increases, the within-cluster first decay. From a certain value of K, the curve flattens$\Rightarrow$ this gives the proper value of K

Methods based on inter- and intra-cluster scatter

Definition (Mean Intra-Cluster Distance for the k-th cluster)

$$MICD_k = \sum_{x_j \in \mathcal{C}_k} \frac{||x_i - \mu_k||}{n_k}$$

- ✓ Data under-partitioned $K < K^*$, at least one cluster has large $MICD$
- ✓ As the partition state moves towards over-partitioned $(K > K^*)$, the large $MICD$ abruptly decreased

Methods based on inter- and intra-cluster scatter

Definition (Inter-Cluster Minimum Distance)

$$ICMD = \min_{i \neq j} ||\mu_i - \mu_j||$$

- ✓ The $ICMD$ is large when hte data are under-partitioned or optimally partitioned
- ✓ $\Downarrow\Downarrow$ when the data enters the over-partioned state

Methods based on inter- and intra-cluster scatter

- ⇒ Several measures are derived from the previous ones to better capture the under-/over-partioned nature of a given clustering
- ⇒ Criteria based on probabilistic measures: Bayessian Information Criterion (BIC), Minimum Message Length (Minimum Message Length), Minimum Description Length (MDL)

- ✓ example_pca2.ipynb
- ✓ Time series Anomaly detection example: http://amid.fish/anomaly-detection-with-k-means-clustering

Some useful examples from sklearn

- ✓ The Silhouette coefficiente: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html
- ✓ http://scikit-learn.org/stable/modules/clustering.html#silhouette-coefficient
- ✓ http://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

EM trivial example

Let us consider the grades in a class:

- ✓ a students get an $A \Rightarrow P(A) = \frac{1}{2}$
- ✓ b students get a $B \Rightarrow P(B) = \mu$
- ✓ c students get a $C \Rightarrow P(C) = 2\mu$
- ✓ d students get a $D \Rightarrow P(D) = \frac{1}{2} - 3\mu$

with $\mu \in [0, 1/6]$

GOAL: get an estimation of μ from data b answering

$\Rightarrow$ What's the maximum likelihood estimate of μ given a, b, c, and d?

Trivial example: solution

1. $P(a,b,c,d|\mu) = (\frac{1}{2})^a \cdot \mu^b \cdot (2\mu)^c \cdot (\frac{1}{2} - 3\mu)^d$
2. $\log(P(a,b,c,d|\mu)) =$
   $a \cdot \log(1/2) + b \cdot \log\mu + c\log(2\mu) + d\log(1/2 - 3\mu)$
3. Max w.r.t. $\mu \Rightarrow \frac{\partial logP}{\partial\mu} = 0$
4. $\frac{\partial logP}{\partial\mu} = \frac{b}{\mu} + \frac{2c}{2\mu} - \frac{3d}{1/2-3\mu} = \frac{(b+c)(1-6\mu)-6d\mu}{\mu(1-6\mu)} = 0$
5. $\mu = \frac{b+c}{6(b+c+d)}$

In case $a = 14,\ b = 6,\ c = 9,\ d = 10 \Rightarrow \mu = \frac{1}{10}$

In case hidden information?

Let us consider we just know:

1. Number of high grades (A's and B's) is h
2. Number of C's is c
3. Number of D's is d

The ratio $a : b$ should be the same as the ratio $P(A) : P(B) \equiv 1/2 : \mu$. In the **expectation phase**:

$$a = \frac{1/2}{1/2 + \mu} h$$
$$b = \frac{\mu}{1/2 + \mu} h$$

Hidden information: maximization stage

If we know the values of a and b we can get μ through the maximum likelihood

$$\mu = \frac{b+c}{6(b+c+d)}$$

EM for our trivial problem

1. A first guess for μ
2. Iterate between expectation and maximization to improve the estimates of μ, a, and b

Definitions:

- ✓ $\mu(t)$ the estimate of μ in the t-th iteration
- ✓ $b(t)$ the estimate of b in the t-th iteration
- ✓ $\mu(0) \equiv$ initial guess
- ✓ **E−step**

$$b(t) = \frac{\mu(t)h}{1/2 + \mu(t)} = E[b|\mu(t)]$$

- ✓ **M−step**

$$\mu(t+1) = \frac{b(t) + c}{6(b(t) + c + d)}$$

EM convergence

It can be proven converge to a local optimum

In our example, let us consider $h = 20$, $c = 10$, $d = 10$, $\mu(0) = 0$:

t	$\mu(t)$	$b(t)$
0	0	0
1	0.08333	2.85714
2	0.09375	3.15789
3	0.09469	3.18452
4	0.09478	3.18706
5	0.09479	3.18734
6	0.09479	3.18734

Normal Sample I

- ✓ $X \sim N(\mu, \sigma^2)$
- ✓ Given n samples $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$
- ✓ And assuming

$$f(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)}{2\sigma^2}}$$

- ✓ $\hat{\mu}$, $\hat{\sigma}^2$??
- ✓ Another assumption: x_i are i.i.d. (independent and identically distributed)

$$f(\mathbf{x}|\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i|\mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n/2)} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

- ✓ We want to maximize

$$\mathcal{L}(\mu, \sigma^2|\mathbf{x}) = f(\mathbf{x}|\mu, \sigma^2)$$

Log-Likelihood function

Taken into account that:

$$x < y \Rightarrow \log(x) < log(y)$$

instead of the likelihood, we are maximizing

$$I(\mu, \sigma^2|\mathbf{x}) = \log(\mathcal{L}(\mu, \sigma^2|\mathbf{x})) =$$
$$= -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\sigma^2}{2\sigma^2}$$

and thus μ and σ^2 are derived by imposing

$$\frac{\partial I(\mu, \sigma^2|\mathbf{x})}{\partial \mu} = 0, \qquad \frac{\partial I(\mu, \sigma^2|\mathbf{x})}{\partial \sigma^2} = 0$$

Max. the Log-Likelihood function

$$\left.\frac{\partial I(\mu,\sigma^2|\mathbf{x})}{\partial \mu}\right|_{\mu=\hat{\mu}} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\hat{\mu}}{\sigma^2} = 0 \boxed{\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

$$\left.\frac{\partial I(\mu,\sigma^2|\mathbf{x})}{\partial \sigma^2}\right|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2} = -\frac{n}{2\hat{\sigma}^2} - \frac{1}{\hat{\sigma}^4} + \frac{\hat{\mu}}{\hat{\sigma}^4}\sum_{i=1}^{n} x_i - \frac{n\hat{\mu}^2}{2\hat{\sigma}^4} = 0$$

$$n\hat{\sigma}^2 = \sum_{i=1}^{n} x_i^2 - 2\hat{\mu}\sum_{i=1}^{n} x_i + n\hat{\mu}^2$$

$$= \sum_{i=1}^{n} x_i^2 - \frac{2}{n}\left(\sum_{i=1}^{n} x_i\right)^2 + \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 \boxed{\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}^2}$$

Consider EM for classification

✓ Given a training data set

$$\mathbf{x} = \{x_1, x_2, \dots, x_n\}$$

✓ Class labels

$$\mathbf{z} = \{z_i, z_2, \dots, z_n\}$$

✓ Data is modelled by a joint distribution

$$p(x_i, z_i) = p(x_i|z_i)p(z_i)$$

✓ Assumption: $z_i \sim multinomial(\boldsymbol{\theta})$

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_k]^T,\ \theta_j \geq 0,\ \sum_{j=1}^{k} \theta_j = 1,\ \theta_j = p(z_i = j)$$

$$x_i|z_i = j \sim \mathcal{N}(\mu_j, \sum_j)$$

✓ Known data: $\mathbf{x}$, $\mathbf{z}$; Unknown parameters: $\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$

EM clustering algorithm

- ✓ Given a training data set

$$\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$$

- ✓ Class labels

$$\mathbf{z} = \{z_i, z_2, \ldots, z_n\}$$

- ✓ In clustering $\mathbf{x}$ is given and $\mathbf{z}$ is unknown
- ✓ Expectation
  - ✗ If the expected values of $\mathbf{z}$ are known, it is possible to compute the maximum likelihood value of $\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$
- ✓ Maximization
  - ✗ If the values of $\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$ are known, it is possible to compute the expected values of $\mathbf{z}$
- ✓ We start with a guess for $\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$. Then iterate between EXPECTATION and MAXIMIZATION. . . until converge

EM clustering: the procedure I

The starting point, the log-likelihood

$$I(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(x_i | z_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \log p(z_i; \boldsymbol{\theta})$$

Second point, maximization of the log-likelihood with respect to $\boldsymbol{\theta}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$. The result:

- ✓ $\theta_j = \frac{1}{n} \sum_{i=1}^{n} 1\{z_i = j\}$
- ✓ $\mu_j = \frac{\sum_{i=1}^{n} 1\{z_i = j\} x_i}{\sum_{i=1}^{n} 1\{z_i = j\}}$
- ✓ $\Sigma_j = \frac{\sum_{i=1}^{n} 1\{z_i = j\} (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{n} 1\{z_i = j\}}$

EM clustering: the procedure II

Third step, repeat EXPECTATION and MAXIMIZATION until convergence:

expectation For each i, j set

$$w_{j,i} \triangleq p(z_i = j | x_i; \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

maximization Update de parameters

$$\theta_j \triangleq \frac{1}{n} \sum_{i=1}^{n} w_{j,i}, \;\; \mu_j \triangleq \frac{\sum_{i=1}^{n} w_{j,i} x_i}{\sum_{i=1}^{n} w_{j,i}}$$

$$\Sigma_j \triangleq \frac{\sum_{i=1}^{n} w_{j,i} (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{n} w_{j,i}}$$

Comparison with K-means

- ✓ Given a training data set

$$\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$$

- ✓ Class labels

$$\mathbf{x} = \{z_i, z_2, \ldots, z_n\}$$

- ✓ Assumption: $z_i \sim multinomial(\boldsymbol{\theta})$

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_k]^T,\ \theta_j \geq 0,\ \sum_{j=1}^{k} \theta_j = 1,\ \theta_j = p(z_i = j)$$

$$x_i | z_i = j \sim \mathcal{N}(\mu_j, \sum_j)$$

- ✓ K-means is a simplified EM
  - ✗ $\theta_i = \frac{1}{k},\ \forall i,\ \Sigma_i = \Sigma_j\ \forall i \neq j,\ i \in \{1, 2, \ldots, k\}$
  - ✗ k is given by user
  - ✗ $\mu_1,\ \mu_2,\ \ldots,\ \mu_k \Rightarrow$ the means of clusters, are the only unknown parameters of the model

Assignment

kmeans_vs_gmm-1617.ipynb

Some references. . . I

- Bishop, Christopher M (2006). *Pattern recognition and machine learning*. springer.
- Guttag, John V (2013). *Introduction to Computation and Programming Using Python*. Mit Press.
- Li, Tao, Mitsunori Ogihara, and Sheng Ma (2004). "On combining multiple clusterings". In: *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. ACM, pp. 294–303.
- Meilă, Marina (2007). "Comparing clusterings—an information based distance". In: *Journal of multivariate analysis* 98.5, pp. 873–895.
- Pena, José Manuel, Jose Antonio Lozano, and Pedro Larranaga (1999). "An empirical comparison of four initialization methods for the k-means algorithm". In: *Pattern recognition letters* 20.10, pp. 1027–1040.

Some references. . . II

Wagner, Silke and Dorothea Wagner (2007). *Comparing clusterings: an overview.* Universität Karlsruhe, Fakultät für Informatik Karlsruhe.