

A Possibilistic c -means Clustering Model with Cluster Size Estimation

László Szilágyi¹(✉) and Sándor M. Szilágyi²

¹ Computational Intelligence Research Group,
Sapientia - Hungarian Science University of Transylvania, Tîrgu Mureş, Romania
lalo@ms.sapientia.ro

² Department of Informatics, Petru Maior University, Tîrgu Mureş, Romania

Abstract. Most c -means clustering models have serious difficulties when facing clusters of different sizes and severely outlier data. The possibilistic c -means (PCM) algorithm can handle both problems to some extent. However, its recommended initialization using a terminal partition produced by the probabilistic fuzzy c -means does not work when severe outliers are present. This paper proposes a possibilistic c -means clustering model that uses only two parameters independently of the number of clusters, which is able to correctly handle the above mentioned obstacles. Numerical evaluation involving synthetic and standard test data sets prove the advantages of the proposed clustering model.

Keywords: Fuzzy c -means clustering
Possibilistic c -means clustering · Cluster size sensitivity
Outlier sensitivity

1 Introduction

The introduction of the fuzzy logic into c -means clustering opened new horizons toward fine data partitioning, but also rose several obstacles that proved difficult to handle simultaneously. The early probabilistic clustering models introduced by Dunn [6] and generalized by Bezdek [4] are strongly influenced by outlier data and uneven sized clusters. A solution to the problem of outliers was given by Dave [5], who relaxed the probabilistic constraint by introducing an extra cluster that attracted noisy data, but this method still creates clusters of equal diameter. The possibilistic c -means clustering algorithm [8] addresses the uneven sized clusters as well, but frequently creates coincident clusters [3]. Several solutions have been proposed to enable the probabilistic clustering models to correctly handle clusters of different weight and/or diameter (e.g. [10, 11]), but these are still sensitive to outlier data due to their probabilistic constraints. Leski [9] recently proposed the so-called fuzzy c -ordered means algorithm, which achieved high

The work of L. Szilágyi was supported by the Institute for Research Programs of the Sapientia University. The work of S. M. Szilágyi was supported by UEFISCDI Romania through grant no. PN-III-P2-2.1-BG-2016-0343, contract no. 114BG.

robustness, but dropped the classical alternative optimization scheme of the fuzzy c -means algorithm.

In this paper we propose a possibilistic fuzzy c -means clustering approach, which employs an extra noise cluster whose prototype is situated at constant distance from every input vector, and an adaptation mechanism that enables the algorithm to handle different cluster diameters.

2 Background

The fuzzy c -means algorithm. The conventional fuzzy c -means (FCM) algorithm partitions a set of object data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ into a number of c clusters based on the minimization of a quadratic objective function, defined as:

$$J_{\text{FCM}} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|_{\mathbf{A}}^2 = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2, \tag{1}$$

where \mathbf{v}_i represents the prototype or centroid of cluster i ($i = 1 \dots c$), $u_{ik} \in [0, 1]$ is the fuzzy membership function showing the degree to which vector \mathbf{x}_k belongs to cluster i , $m > 1$ is the fuzzyfication parameter, and d_{ik} represents the distance (any inner product norm defined by a symmetrical positive definite matrix \mathbf{A}) between \mathbf{x}_k and \mathbf{v}_i . FCM uses a probabilistic partition, meaning that the fuzzy memberships assigned to any input vector \mathbf{x}_k with respect to clusters satisfy the probability constraint $\sum_{i=1}^c u_{ik} = 1$. The minimization of the objective function J_{FCM} is achieved by alternately applying the optimization of J_{FCM} over $\{u_{ik}\}$ with \mathbf{v}_i fixed, $i = 1 \dots c$, and the optimization of J_{FCM} over $\{\mathbf{v}_i\}$ with u_{ik} fixed, $i = 1 \dots c, k = 1 \dots n$ [4]. Obtaining the optimization formulas involves zero gradient conditions of J_{FCM} and Langrange multipliers. Iterative optimization is applied until cluster prototypes \mathbf{v}_i ($i = 1 \dots c$) converge.

Relaxing the probabilistic constraint. The relaxation of the probabilistic constraint was a necessity provoked by the outlier sensitivity of the FCM algorithm. Here we need to mention two different ways the constraint was eliminated. Krishnapuram and Keller [8] introduced the possibilistic c -means (PCM) algorithm, which optimizes

$$J_{\text{PCM}} = \sum_{i=1}^c \sum_{k=1}^n [t_{ik}^p d_{ik}^2 + (1 - t_{ik})^p \eta_i], \tag{2}$$

constrained by $0 \leq t_{ik} \leq 1 \forall i = 1 \dots c, \forall k = 1 \dots n$, and $0 < \sum_{i=1}^c t_{ik} < c \forall k = 1 \dots n$, where $p > 1$ represents the possibilistic exponent, and parameters η_i are the penalty terms that control the diameter of the clusters. The iterative optimization algorithm of PCM objective function is derived from the zero gradient conditions of J_{PCM} . In the probabilistic fuzzy partition, the degrees of membership assigned to an input vector \mathbf{x}_k with respect to cluster i depends on the distances of the given vector to all cluster prototypes: $d_{1k}, d_{2k}, \dots, d_{ck}$. On the

other hand, in the possibilistic partition, the typicality value u_{ik} assigned to input vector \mathbf{x}_k with respect to any cluster i depends on only one distance: d_{ik} . PCM efficiently suppresses the effects of outlier data, at the price of frequently producing coincident cluster prototypes. The latter is the result of the highly independent cluster prototypes [3].

On the other hand, Dave [5] introduced a noise cluster into the FCM algorithm, which by definition is situated at a constant distance d_0 from any input vector \mathbf{x}_k ($k = 1 \dots n$). Thus, the objective function becomes

$$J_{\text{Dave}} = \sum_{i=0}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 = J_{\text{FCM}} + \sum_{k=1}^n u_{0k}^m d_0^2, \tag{3}$$

where the noise cluster is the one with index 0. The probabilistic constraint becomes $\sum_{i=0}^c u_{ik} = 1 \ \forall k = 1 \dots n$, thus the degrees of membership of any input vector with respect to the real clusters does not sum up to 1 anymore. Outliers will be attributed with high degrees of membership towards the noise class, making the algorithm insensitive to outlier data, without producing coincident clusters. However, this approach cannot handle clusters of different size or diameter.

Several fuzzy-possibilistic mixture partition models have been proposed to deal with the coincident clusters of PCM (e.g. [12, 13]). The most recent additive mixture model proposed by Pal et al. [13] called possibilistic-fuzzy c -means (PFCM) clustering minimizes

$$J_{\text{PFCM}} = \sum_{i=1}^c \sum_{k=1}^n [au_{ik}^m + bt_{ik}^p]d_{ik}^2 + \sum_{i=1}^c \eta_i \sum_{k=1}^n (1 - t_{ik})^p, \tag{4}$$

constrained by the conventional probabilistic and possibilistic conditions of FCM and PCM, respectively. Here a and b are two tradeoff parameters that control the strength of the possibilistic and probabilistic term in the mixed partition. All other parameters are the same as in FCM and PCM. This clustering model was found accurate and robust, but still sensitive to outlier data [14].

Fuzzy c -means with various cluster diameters. Komazaki and Miyamoto [7] presents a collection of solutions how the FCM algorithm can adapt to different cluster sizes and diameters. From the point of view of this paper, it is relevant to mention the FCMA algorithm by Miyamoto and Kurosawa [11], which minimizes

$$J_{\text{FCMA}} = \sum_{i=1}^c \sum_{k=1}^n \alpha_i^{1-m} u_{ik}^m d_{ik}^2, \tag{5}$$

subject to the probabilistic constraint of the fuzzy memberships u_{ik} ($i = 1 \dots c, k = 1 \dots n$), and of the extra terms α_i ($i = 1 \dots c$): $\sum_{i=1}^c \alpha_i = 1$. The optimization algorithm of J_{FCMA} can be derived from zero gradient conditions using Lagrange multipliers. Each iteration updates the probabilistic memberships u_{ik} ($i = 1 \dots c, k = 1 \dots n$), the cluster prototypes \mathbf{v}_i ($i = 1 \dots c$), and the cluster diameter terms α_i ($i = 1 \dots c$) as well. The algorithm stops when cluster prototypes stabilize.

3 Methods

In the following, we propose a fuzzy c -means clustering model with relaxed probabilistic constraint, which employs an extra noise cluster whose prototype is situated at constant distance from every input vector, and an adaptation mechanism to different cluster diameters. The proposed objective function is:

$$J = \sum_{i=0}^c \sum_{k=1}^n \alpha_i^{1-m} u_{ik}^m d_{ik}^2, \tag{6}$$

where $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|$ ($\forall i = 1 \dots c, \forall k = 1 \dots n$), $m > 1$ is the fuzzy exponent, and $d_{0k} = d_0 \ \forall k = 1 \dots n$ with d_0 predefined constant. Variables α_i ($i = 1 \dots c$) satisfy the probabilistic constraint $\sum_{i=0}^c \alpha_i = 1$, while fuzzy memberships are also constrained to the probabilistic rule, similarly to the FCM algorithm: $\sum_{i=0}^c u_{ik} = 1$ for any $k = 1 \dots n$. The minimization formulas of the objective function given in Eq. (6) are obtained using zero gradient conditions and Lagrange multipliers. Let us consider the functional

$$\mathcal{L} = J + \sum_{k=1}^n \lambda_k \left(1 - \sum_{i=0}^c u_{ik} \right) + \lambda_\alpha \left(1 - \sum_{i=0}^c \alpha_i \right), \tag{7}$$

where $\lambda_1 \dots \lambda_n$ and λ_α are the Lagrange multipliers. The zero gradient conditions with respect to the fuzzy memberships u_{ik} ($\forall i = 0 \dots c, \forall k = 1 \dots n$) imply

$$\frac{\partial \mathcal{L}}{\partial u_{ik}} = 0 \Rightarrow \alpha_i^{1-m} m u_{ik}^{m-1} d_{ik}^2 = \lambda_k, \tag{8}$$

and so

$$u_{ik} = \left(\frac{\lambda_k}{m} \right)^{1/(m-1)} \alpha_i d_{ik}^{-2/(m-1)}. \tag{9}$$

According to the probabilistic constraint of fuzzy memberships, we have:

$$\sum_{j=0}^c u_{jk} = 1 \Rightarrow 1 = \left(\frac{\lambda_k}{m} \right)^{1/(m-1)} \sum_{j=0}^c \alpha_j d_{jk}^{-2/(m-1)}. \tag{10}$$

Equations (9) and (10) allows us to eliminate the Lagrange multiplier λ_k from the formula of u_{ik} :

$$u_{ik} = \frac{u_{ik}}{1} = \frac{\left(\frac{\lambda_k}{m} \right)^{1/(m-1)} \alpha_i d_{ik}^{-2/(m-1)}}{\left(\frac{\lambda_k}{m} \right)^{1/(m-1)} \sum_{j=0}^c \alpha_j d_{jk}^{-2/(m-1)}} = \frac{\alpha_i d_{ik}^{-2/(m-1)}}{\sum_{j=0}^c \alpha_j d_{jk}^{-2/(m-1)}}. \tag{11}$$

The optimization formula for α_i ($i = 0 \dots c$) is obtained similarly. We start from the zero crossing of the partial derivative

$$\frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \Rightarrow \sum_{k=1}^n [\alpha_i^{-m} (1 - m) u_{ik}^m d_{ik}^2] = \lambda_\alpha, \tag{12}$$

Algorithm 1. The proposed algorithm

Data: Input data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$

Data: Number of clusters c , fuzzy exponent m , distance d_0 , threshold ε

Result: Possibilistic partition u_{ik} ($\forall i = 0 \dots c, \forall k = 1 \dots n$)

Initialize $\mathbf{v}_i^{(\text{new})}$ ($i = 1 \dots c$) as random input vectors that have several input vectors in their close neighborhood, to avoid outliers

$\alpha_i \leftarrow 1/(c + 1), \forall i = 0 \dots c$

repeat

$\mathbf{v}_i^{(\text{old})} \leftarrow \mathbf{v}_i^{(\text{new})}, \forall i = 1 \dots c$

Update partition u_{ik} , ($i = 0 \dots c, k = 1 \dots n$), according to Eq. (11)

Obtain new cluster prototypes $\mathbf{v}_i^{(\text{new})}$, ($i = 1 \dots c$), according to Eq. (17)

Update α_i values for any $i = 0 \dots c$, according to Eq. (16)

until $\sum_{i=1}^c \|\mathbf{v}_i^{(\text{new})} - \mathbf{v}_i^{(\text{old})}\| < \varepsilon$;

which implies

$$(1 - m)\alpha_i^{-m} \sum_{k=1}^n u_{ik}^m d_{ik}^2 = \lambda_\alpha \Rightarrow \alpha_i^m = \left(\frac{1 - m}{\lambda_\alpha}\right) \left[\sum_{k=1}^n u_{ik}^m d_{ik}^2\right], \quad (13)$$

and so we get

$$\alpha_i = \left(\frac{1 - m}{\lambda_\alpha}\right)^{1/m} \left[\sum_{k=1}^n u_{ik}^m d_{ik}^2\right]^{1/m}. \quad (14)$$

On the other hand, the probabilistic constraint $\sum_{j=0}^c \alpha_j = 1$ implies:

$$\sum_{j=0}^c \alpha_j = 1 \Rightarrow 1 = \left(\frac{1 - m}{\lambda_\alpha}\right)^{1/m} \sum_{j=0}^c \left[\sum_{k=1}^n u_{jk}^m d_{jk}^2\right]^{1/m}. \quad (15)$$

Equations (14) and (15) allows us to eliminate the Lagrange multiplier λ_α from the formula of α_i :

$$\alpha_i = \frac{\alpha_i}{1} = \frac{\left(\frac{1-m}{\lambda_\alpha}\right)^{1/m} \left[\sum_{k=1}^n u_{ik}^m d_{ik}^2\right]^{1/m}}{\left(\frac{1-m}{\lambda_\alpha}\right)^{1/m} \sum_{j=0}^c \left[\sum_{k=1}^n u_{jk}^m d_{jk}^2\right]^{1/m}} = \frac{\left[\sum_{k=1}^n u_{ik}^m d_{ik}^2\right]^{1/m}}{\sum_{j=0}^c \left[\sum_{k=1}^n u_{jk}^m d_{jk}^2\right]^{1/m}}. \quad (16)$$

The update formula of cluster prototypes \mathbf{v}_i is obtained as:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{v}_i} = 0 &\Rightarrow \sum_{k=1}^n (-2)\alpha_i^{1-m} u_{ik}^m (\mathbf{x}_k - \mathbf{v}_i) = 0 \\ &\Rightarrow \sum_{k=1}^n u_{ik}^m \mathbf{x}_k = \mathbf{v}_i \sum_{k=1}^n u_{ik}^m \\ &\Rightarrow \mathbf{v}_i = \frac{\sum_{k=1}^n u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^n u_{ik}^m}. \end{aligned} \quad (17)$$

If a defuzzified partition is desired, any input vector \mathbf{x}_k can be assigned to cluster number $\arg \max_i \{u_{ik}, i = 0 \dots c\}$. Vectors belonging to cluster number 0 are detected outliers. The proposed algorithm is summarized in Algorithm 1.

4 Results and Discussion

The proposed method was evaluated on three different data sets, and its behavior compared to FCM [4] and PFCM [13]. The first data set consisted of two groups of randomly generated two-dimensional input vectors, situated inside the circle with center at $(0, 1)$ and radius 1.2, and the circle with center at $(0, -1)$ and radius 0.6, respectively. Each group contained 100 vectors. The input vectors are exhibited in Fig. 1, together with the partitions created by the proposed algorithms and its counter candidates. Adding an extra input vector situated at $(\delta, 0)$ and treating δ as a parameter allowed us to evaluate the sensitivity of the clustering to outliers. FCM and PFCM were able to provide two valid clusters up $\delta = 132$ and $\delta = 158$, respectively. The proposed method was not influenced by high values of δ , it assigned the extra vector to the third (outlier) class.

The second data set employed by the numerical evaluation of the algorithms was the IRIS data set [1], which consist of 150 labeled feature vectors of four dimensions, organized in three groups that contain fifty vectors each. It is a reported fact, that conventional clustering models like FCM produce 133–134 correct decisions when classifying IRIS data. PFCM produced the best reported accuracy with 140 correct decisions using $a = b = 1, m = p = 3$, and initializing v_i with terminal FCM prototypes [13]. Under less advantageous circumstances, PFCM reportedly produced 136–137 correct decisions. Initially we normalized the IRIS data set, and included an outlier situated at $(\delta, \delta, \delta, \delta)$, where δ was a variable parameter. We tested the clustering models for various values of the algorithm parameters and outlier positions δ . Figure 2 shows the outcome of tests. FCM produces three valid clusters in case of $\delta < 9$, PFCM can correctly handle the outlier for values of δ up to 12, while the proposed method can deal with the outlier situated at any distance.

The third numerical test employed the WINE data set [2], which consist of 178 labeled feature vectors of 13 dimensions, organized in three groups of uneven cardinality. The WINE data set was initially normalized, and an outlier was included at the position $(\delta, \delta, \dots, \delta)$, where δ represents a variable parameter. We tested the clustering models for various values of the algorithm parameters and outlier positions δ . Figure 3 exhibits the test results. FCM and PFCM can produce three valid clusters in case of $\delta < 6$, and $\delta < 7$, respectively, while the proposed method has no difficulty with handling correctly an outlier situated at any distance.

Based on all numerical tests, we can assert that the proposed clustering model is able to perform better than its counter candidates, as: (1) it creates valid clusters according to cluster validity indexes proposed for the characterisation of such partitions; (2) it produces more correct decisions than previous algorithms; (3) it can correctly handle severe outlier vectors; (4) it is able to adapt the

diameter of clusters to the input data to some considerable extent; (5) it does not produce coincident clusters; (6) it uses a reduced number of parameters (2 instead of the $c + 1$ parameters of PCM). Results reported in Figs. 1, 2 and 3 were obtained using the following parameter settings: $m = 2$ for FCM and the

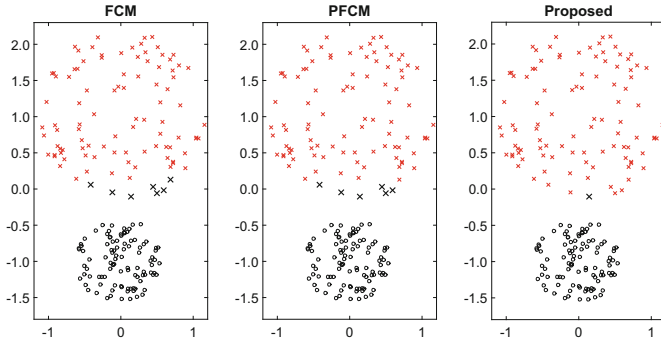


Fig. 1. The case of two groups of same cardinality but different diameter

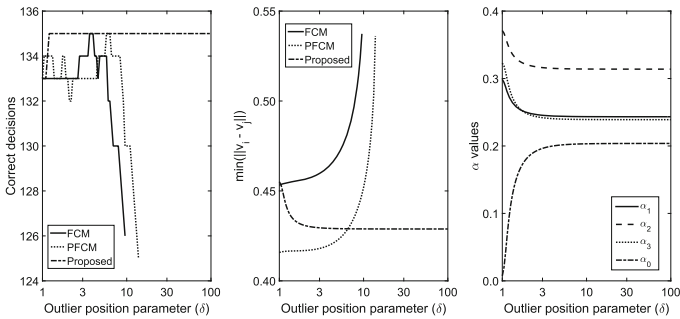


Fig. 2. Evaluation using the IRIS data set: (left) correct decisions out of 150 vs. δ , (middle) CVI vs. δ , (right) final α_i values of the proposed algorithm vs. δ

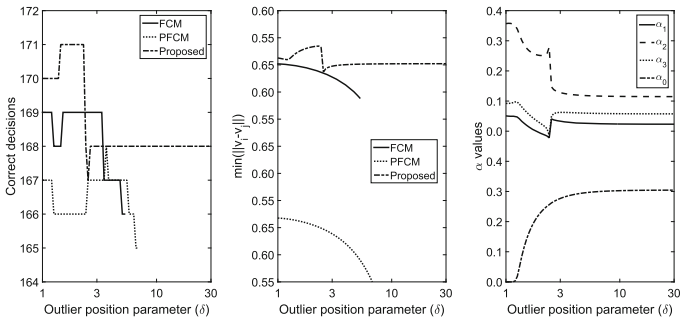


Fig. 3. Evaluation using the WINE data set: (left) correct decisions out of 150 vs. δ , (middle) CVI vs. δ , (right) final α_i values of the proposed algorithm vs. δ

proposed method, $m = p = 2$ and $a = b = 1$. Possibilistic penalty terms η_i were not established based on a terminal FCM partition, as recommended by their authors, because in many cases the terminal FCM partition was severely damaged by the presence of the outlier. Parameters η_i and d_0 were used as constants whose values were set empirically to favor correct and fine partitions.

5 Conclusions

This paper proposed a possibilistic c -means clustering model with a reduced number of parameters (two instead of $c + 1$) that can robustly handle distant outlier data. The advantageous properties of the algorithms were numerically validated using synthetic and standard test data sets. A formula for an optimal d_0 distance may further enhance the robustness of the algorithm.

References

1. Anderson, E.: The irises of the Gaspé Peninsula. *Bull. Am. Iris Soc.* **59**, 2–5 (1935)
2. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets.html>
3. Barni, M., Capellini, V., Mecocci, A.: Comments on a possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **4**, 393–396 (1996)
4. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York (1981)
5. Dave, R.N.: Characterization and detection of noise in clustering. *Pattern Recogn. Lett.* **12**, 657–664 (1991)
6. Dunn, J.C.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Cybern. Syst.* **3**(3), 32–57 (1973)
7. Komazaki, Y., Miyamoto, S.: Variables for controlling cluster sizes on fuzzy c -means. In: Torra, V., Narukawa, Y., Navarro-Arribas, G., Megias, D. (eds.) *MDAI 2013. LNCS (LNAI)*, vol. 8234, pp. 192–203. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41550-0_17
8. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1**, 98–110 (1993)
9. Leski, J.M.: Fuzzy c -ordered-means clustering. *Fuzzy Sets Syst.* **286**, 114–133 (2016)
10. Lin, P.L., Huang, P.W., Kuo, C.H., Lai, Y.H.: A size-insensitive integrity-based fuzzy c -means method for data clustering. *Pattern Recogn.* **47**(5), 2024–2056 (2014)
11. Miyamoto, S., Kurosawa, N.: Controlling cluster volume sizes in fuzzy c -means clustering. In: *SCIS and ISIS, Yokohama, Japan*, pp. 1–4 (2004)
12. Pal, N.R., Pal, K., Bezdek, J.C.: A mixed c -means clustering model. In: *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 11–21 (1997)
13. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C.: A possibilistic fuzzy c -means clustering algorithm. *IEEE Trans. Fuzzy Syst.* **13**, 517–530 (2005)
14. Szilágyi, L.: Fuzzy-possibilistic product partition: a novel robust approach to c -means clustering. In: Torra, V., Narakawa, Y., Yin, J., Long, J. (eds.) *MDAI 2011. LNCS (LNAI)*, vol. 6820, pp. 150–161. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22589-5_15