Sequential clustering: tracking down the most natural clusters

Thomas Ott[1], Albert Kern[1], Willi-Hans Steeb[1,2] and Ruedi Stoop[1]

[1] Institute of Neuroinformatics ETHZ/UNIZH, Winterthurerstrasse 190, 8057 Zürich, Switzerland
[2] School of Scientific Computing, University of Johannesburg, Auckland, RSA
E-mail: tott@ini.phys.ethz.ch, albert@ini.phys.ethz.ch, whs@na.rau.ac.za and ruedi@ini.phys.ethz.ch

Abstract. Sequential superparamagnetic clustering (SSC) is a substantial extension of the superparamagnetic clustering approach (SC). We demonstrate that the novel method is able to master the important problem of inhomogeneous classes in the feature space. By fully exploiting the non-parametric properties of SC, the method is able to find the natural clusters even if they are highly different in shape and density. In such situations, concurrent methods normally fail. We present the results from a fully automated implementation of SSC (applications to chemical data and visual scene analysis) and provide analytical evidence of why the method works.

J. Stat. Mech. (2005) P11014

Contents

1. Introduction

Distinction, grouping and naming of observations are among the most basic cultural activities [1]. In ancient times, the segmentation of the world provided a base for communication and culture. More recently, notably in science, the focus has shifted from the identification of single objects to the fusion of particular properties or objects towards bigger units (objects and classes of objects, respectively). This task of grouping together items that—in a situation-dependent sense—'belong together' is referred to as clustering. Clustering has gained increased importance in all domains of data analysis and processing, the most prominent examples of which are bioinformatics [2], chemoinformatics [3], neuroscience [4], visual scene analysis [5, 6] and the design of autonomous systems [7].

The notion of a cluster is inherently diffuse. A set of geometric figures can be clustered according to the shape or to the colour, to the combined features or to any other features. Whether a clustering result will be useful thus depends on whether the clustering is compatible with or suitable for the particular focus taken. In some cases, a data item is therefore most naturally described by a likelihood-characterized membership to different clusters (fuzzy clustering). Furthermore, clusters inherently emerge nested in hierarchies, of which only a substructure may be of interest. Many decisions relying on clustering, however, require a clear assignment of each item to one cluster or category. This is the case that we shall be interested in in the following.

In the description of clustering, we shall thus distinguish between two processing stages. In the first stage, the relevant features or attributes that characterize the items to be clustered are identified, and the similarity measure that quantifies the pairwise affinities of items, given the chosen features, is determined. This already introduces *a priori* information into the clustering so that at this point the clustering solution is already predetermined. Consequently, acquired or learnt prior information (e.g., via feedback) is implemented in the form of modified attributes or an optimized similarity measure.

In the example given above, the weight may thus shift from shape to colour, or to a new attribute, such as size. The clustering algorithm itself, which is the second stage of processing, is supposed to work in a completely unbiased way, as in most situations no *a priori* information about the number of clusters, their shapes and their internal structure (as the most relevant aspects) is available. If a reliable clustering algorithm fails to detect satisfactory solutions, our conclusion will be that the first stage of the processing needs to be improved.

A clustering algorithm's task is to bring out 'the best in possibly ambiguous data'. As the quality of a clustering result is an ill-defined concept and vague, most clustering algorithms will yield optimized results when assessed by means of an appropriate algorithm-specific measure, usually formulated in the form of an objective function. Nonetheless, some basic requirements for a good (successful and as unbiased as possible) clustering method can be listed as follows.

(a) The method should identify one unique clustering (hierarchy) containing the most 'natural' clusters. Preferably, some measure should specify how natural an obtained cluster effectively is. According to our unbiased approach, the most practical notion of a *natural cluster* is a group without any significant substructure (as quantified by the measure).

(b) The method should not assume any *a priori* information about number, shape or internal distribution of clusters.

(c) Preferably, the method should be based on the set of pairwise affinities. This allows for optimizing the results by optimizing the similarity measure.

(d) The method should be able to deal with clusters of different shapes, densities and largely unequal distances between clusters.

There exist a plethora of different clustering algorithms (for an overview see [8, 2, 9, 10]) each of which has its own advantages for certain problems. Interestingly, little attention has been paid to requirement (d). It thus comes as no surprise that many, mainly standard, algorithms have to struggle to find natural clustering solutions for inhomogeneous data distributions (see figure 1) [11]. However, inhomogeneous distributions leading to highly varying intra- and inter-cluster distances are characteristic for many real-world applications. Such inhomogeneities are often introduced when insufficient care is applied during the choosing or scaling of the attributes. By choosing even from the most sophisticated similarity measures, this neglect can often not be compensated for.

Sequential superparamagnetic clustering [3] is our solution to these problems. In addition to taking into account points (a), (b) and (c), it also offers a solution to (d). The core of the algorithm is the superparamagnetic clustering approach (SC) [12], which provides 'natural' clustering solutions on different resolution levels. The levels are controlled by a 'temperature' parameter T that controls the number of clusters. The approach, however, leaves open which resolution level provides the best clustering, in the sense of a unique choice of the 'most natural' clusters. As already pointed out in [12], the best clustering sometimes chooses clusters across different levels. The central question thus remains: According to which criterion should levels and clusters be chosen? The sequential procedure answers this question. Strictly speaking, it renders the question
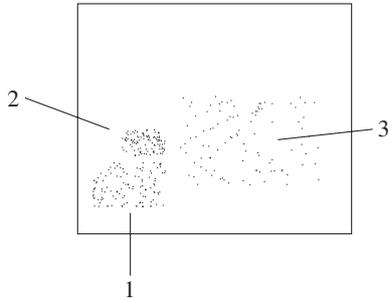
Figure 1. Toy system with three clusters of different densities.

meaningless, as the most natural clusters turn out not to be characterized by global parameters T.

The paper is organized as follows. In section 2 the concept of SSC is outlined. The advantages of SSC are demonstrated in section 3. A simple mean-field model is used to elucidate the origin of the surprising performance of this method. As an additional application, a visual scene analysis is discussed in section 4. Section 5 concludes the paper.

2. Sequential superparamagnetic clustering

2.1. Superparamagnetic clustering

Superparamagnetic clustering has been described from different viewpoints inherent to different fields, such as graph theory [6] and stochastic neural networks [14]. Here, we shall use the description via Potts spin models, as put forward in the original work by E Domany's group (see e.g. [12, 15]). We only briefly outline the algorithm; for more details we refer the reader to the aforementioned references and those that will be given below. For N items to be clustered with pairwise affinities d_{ij}, an inhomogeneous grid of Potts spins is constructed in the following way: each item i is represented by one site of the grid with Potts spin variable s_i, where $s_i \in \{1, \dots, q\}$. q is typically set to 10 or 20 [12, 3]. It is important to note that the choice of q is largely arbitrary and is not related to the number of emerging clusters. Each spin is symmetrically coupled to its k (not necessarily mutual) nearest neighbours. The choice of k will be discussed in section 3. The coupling strength J_{ij} is a decreasing function of d_{ij}, e.g.,

$$J_{ij} = J_{ji} = \frac{1}{\widehat{K}} \exp\left(\frac{-d_{ij}^2}{2a^2}\right). \tag{1}$$

$\widehat{K}$ is the average number of coupled neighbours per site (not necessarily equal to k). a is a local length scale which is set by default to the average distance between coupled spins. However, sometimes it may be advantageous to use differing length scales (see section 4). Each spin configuration is characterized by an energy expressed by the *Potts spin Hamiltonian*

$$H(s) = \sum_{(ij)} J_{ij}(1 - \delta_{s_i s_j}), \tag{2}$$

where the sum runs over all connections (ij) and s denotes a spin configuration. The system is considered in the canonical ensemble. The probability for a certain spin configuration is thus given by the Boltzmann/Gibbs distribution

$$p(s) = \frac{1}{Z} \mathrm{e}^{-H(s)/T}, \tag{3}$$

where the partition function $Z = Z(T)$ serves as a normalization factor. As the temperature T is increased, inhomogeneous Potts systems typically undergo a number of phase transitions. This is the source of inspiration for the SC clustering algorithm.

(I) For small T, the system is in the ferromagnetic phase, where spins like to be aligned.

(II) In an intermediary T-range, a superparamagnetic phase occurs: strongly coupled spins tend to be aligned, whereas weakly coupled spins behave independently. Thus, clusters of aligned spins occur, reflecting groups of similar data items. A further increase of T generally leads to a continued breaking up of these clusters into smaller clusters, so that a hierarchy of classes and subclasses is obtained.

(III) For high T, the system enters the paramagnetic phase where any order disappears and only singleton clusters remain.

At a given temperature T, clusters are identified with the help of the pair correlation: two points i and j belong to the same cluster if the pair correlation

$$G_{ij} = \sum_s p(s) \delta_{s_i s_j} \tag{4}$$

exceeds a given threshold Θ, i.e.,

$$G_{ij} > \Theta. \tag{5}$$

Whole clusters are defined by areas of the grid whose sites are connected through (5). Interestingly, the choice of Θ is not critical as long as $1/q < \Theta < (1 - 2/q)$ for larger q (e.g. $q = 10$) [15]. This is due to the fact that either $G_{ij} \approx 1$ or $G_{ij} \approx 1/q \approx 0$ for most pairs (ij), which reflects the robust self-organization at work in these systems.

For large data sets, the computation of (4) is not feasible. Therefore, the pair correlation is usually approximated by an appropriate Monte Carlo simulation, such as the Swendsen–Wang [16] or the Wolf algorithm [4]. For data sets with $N \lesssim 10\,000$, Swendsen–Wang turned out to be quite efficient, where we generally obtained reliable results for less than 230 Monte Carlo steps [3]. As an alternative to Monte Carlo methods, loopy and generalized belief propagation (BP) [13] have been proposed [6]. In [17], loopy BP has been embedded into the sequential procedure.

2.2. The sequential procedure

Although SC has been successfully applied in different fields, it does not provide a direct evidence about which clusters, among those emerging on different resolution levels T, should be selected as the natural ones. Clear clusters express themselves as regions of order that are stable over an entire phase, i.e. over a substantial range of T. The idea is thus to choose the clusters that have the largest T-range extensions (denoted by T_{cl}). Consequently, we define the *T-stability* s_T of a cluster as

$$s_T = T_{\mathrm{cl}}/T_{\mathrm{max}}, \tag{6}$$

where $T_{\max}$ is the temperature of the paramagnetic transition. In this way, s_T expresses the stability of the cluster in relation to the stability of the whole set. If the clusters are selected according to the stability s_T, the procedure will still be suboptimal, as the most natural clusters need not always be the most stable ones. Differences in shape, density and size of clusters lead to different temperature ranges of occurrence. Sparse clusters only exist at small temperatures, whereas dense clusters are more stable and decay later. As a consequence, often natural clusters emerge for short T-ranges only, after the break-up of dense superclusters at higher temperatures. Inhomogeneities in shape, density and size of clusters can thus render the recognition of the best clusters a difficult task. To overcome these difficulties, we introduced a sequential procedure [3]. In this approach, the most stable cluster in terms of s_T is extracted and it, as well as the residual set, is reclustered with readjusted weights. That is, the most stable cluster and the residual set are clustered in two new separate sessions, and the connectivity and weights are independently redetermined for each set according to the criterion of k nearest neighbours and (1). The procedure continues, resulting in a binary tree structure. The branches of the tree form sequences of sets of increasing homogeneities. Consequently, the detection of clusters becomes easier. The procedure stops in a branch if no more stable substructures can be found, i.e. if the most stable cluster detected is less stable than a threshold value s_Θ. Typically, natural clusters themselves do not have any substructures. They show a direct transition from the ferromagnetic phase to the paramagnetic phase. Therefore the temperature that marks the end of the ferromagnetic phase T_{ferro} is a good indicator for how natural a cluster intrinsically is (in contrast, s_T gives the information of how natural the cluster is in relation to the clustered mother set). T_{ferro} is also useful for judging whether the last residual set should be considered an intrinsic (i.e., natural) cluster or a set of unclustered (i.e., background) points. In the first case, T_{ferro} is of the same magnitude as for the other detected clusters, whereas in the second case, T_{ferro} is vanishingly small. In summary, the following program is executed:

(1) For the input set S, determine the weights according to the criterion of k nearest neighbours and (1) and apply SC.

(2) Determine the most stable cluster C in terms of s_T.
If $s_T > s_\Theta$: Go back to (1) and start two new sessions with $S_1 = C$ and $S_2 = S \backslash C$.
Else: Determine T_{ferro}.

(3a) If $S = S1$ in the last loop: Identify S as a natural cluster.

(3b) If $S = S2$ in the last loop or $S =$ the initial set: Use T_{ferro} to judge whether S is a natural cluster or a set of unclustered points.

We emphasize that in this way, s_Θ is the main control parameter that is set from outside. The procedure does not require one to maximize the superparamagnetic phase [18] and it also works on Potts spin graphs composed of several components. For sets with clear cluster structures, the choice of s_Θ is not critical, as we can expect that $s_T > 0$ (and $T_{\mathrm{ferro}} > 0$) only holds for the actual clusters. For sets with less clear structures, the user needs to choose an appropriate minimal stability s_Θ. In this way, the user specifies her/his notion of a 'marginally natural' cluster. Additionally, the user can specify the required minimal size of clusters which gives the possibility to optimize the procedure if additional information is available. How this is properly done is illustrated in more detail
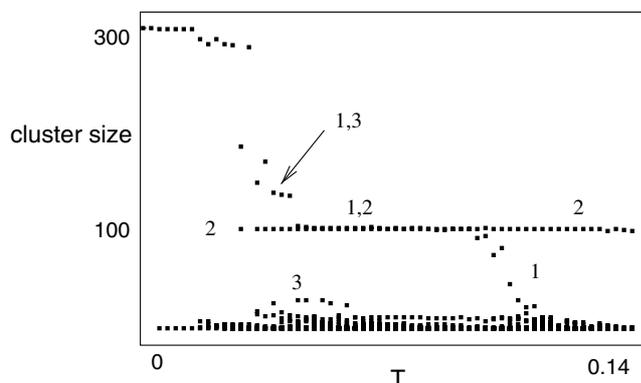
Figure 2. Results for SC using the Euclidean distance. The size of clusters occurring is drawn against the temperature T. The problems discussed in this section are best highlighted using $k \approx 10$.

in the examples discussed below. As a further benefit of the sequential procedure, the choice of the number of nearest neighbours k becomes an uncritical issue. Changing k in the range between 5 and 15 is usually without any effect on the final result. Thus, for computational reasons, $k = 5$ has been chosen as the default value in this paper.

3. Examples

3.1. A toy system

In this section, a simple toy system is used to illustrate a basic advantage of SSC. In figure 1, a 2D distribution consisting of three clusters is displayed. Whereas the clusters have the same size and shape, they differ considerably in the density.

Figure 2 shows the results obtained by SC. Remarkably, no interval $[T_{\min}, T_{\max}]$ can be found for the sparse cluster 3. After going through a short phase in a joint cluster with cluster 1, cluster 3 immediately decays into smaller units. Thus a detection of cluster 3 is intrinsically impossible. However, our automated SSC detects the three clusters without problems (see figure 3). In this diagram, extracted clusters form left branches and residual sets form right branches. In each box, the size of the set N, the length of the ferromagnetic phase T_{ferro}, the length of the T-range of the most stable cluster T_{cl} (if any) and (optionally) the paramagnetic transition temperature $T_{\max}$ are reported. For the extracted clusters, the stability values s_T are given as well. Since T_{ferro} is of the same magnitude for all three clusters, i.e. between 0.05 and 0.083, the residual set (the rightmost one) should be considered a natural cluster. When clustering the mother set, the last two clusters emerge as subclusters (with $s_T > 0$), which is consistent with the above-made cluster identification. The findings are very stable against variations of s_Θ: any value in $s_\Theta \in [0.05, 0.58]$ reflects in the same stable and clear-cut cluster structure. As a typical default value, $s_\Theta = 0.2$ can be taken.

3.2. Classifying chemical compounds

The rationale of clustering in chemistry is given by the assumption that structurally similar compounds are likely to exhibit similar properties. The search for new drugs among
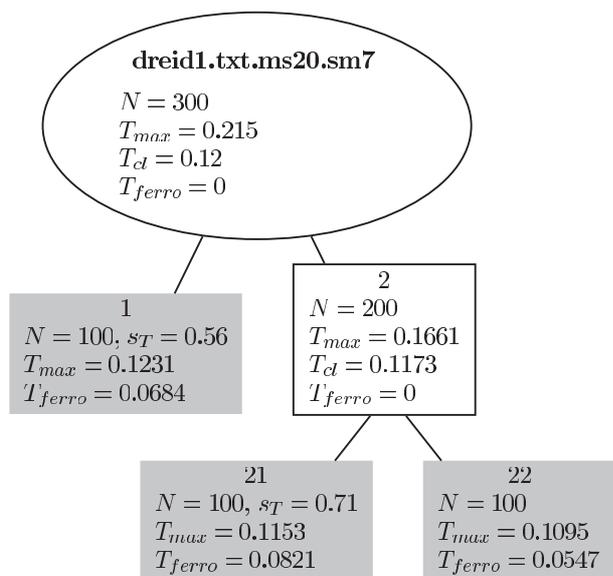
Figure 3. The binary tree structure resulting from SSC. The grey boxes indicate the detected clusters. In the heading, details of the experiment performed are given (e.g. connectivity (if different from 5), minimal size of clusters to be detected: $ms = 20$, $sm = s_\Theta = 7 * (T_{\max}/50)$ etc).

thousands of different compounds in a chemical library can thus be greatly facilitated if the library can be sectioned into natural classes of structurally similar chemicals. For optimally chosen classes, efficient testing can be achieved. In [3] the potential of SSC for chemoinformatics was investigated in detail; here we report on the results obtained from our automated implementation of SSC. The test set comprised 153 compounds from seven different chemical classes (figure 4). While the appropriate description of the compound's structure using chemical fingerprints, i.e. high-dimensional vectors, and the choice of a good similarity measure are complex issues in themselves (for details see [3]), we concentrate here on the very aspect of clustering. The results achieved by SSC were compared to results by other methods, notably Ward clustering, which in the field of chemoinformatics [19] is regarded as yielding the most reliable results. Yet, among the tested methods, SSC was the only one able to reliably reconstruct the seven chemical classes involved (up to about three misclassified compounds, depending on the chemical fingerprint descriptors and the similarity measure used [3]).

From applying SC without the sequential procedure, the origins for the failure of other algorithms become apparent. In figure 5, the clusters detected by SC are displayed. While some of the classes are easily detectable by manifesting themselves as stable clusters from the very beginning (e.g., the class of protone pump inhibitors), other classes hardly ever appear as stable clusters. Most prominently, the classes of *statins* and *corticosteroids* form a joint stable cluster (with size 45) that for large T-values quickly decays into smaller units, not revealing the actual chemical class structures. The reason for this behaviour can be found in the intra- and inter-cluster distances. For all similarity measures tested, the classes *statins* and *corticosteroids* are much more compact than all other clusters. Moreover, the distance between these two classes is found to be much smaller than the
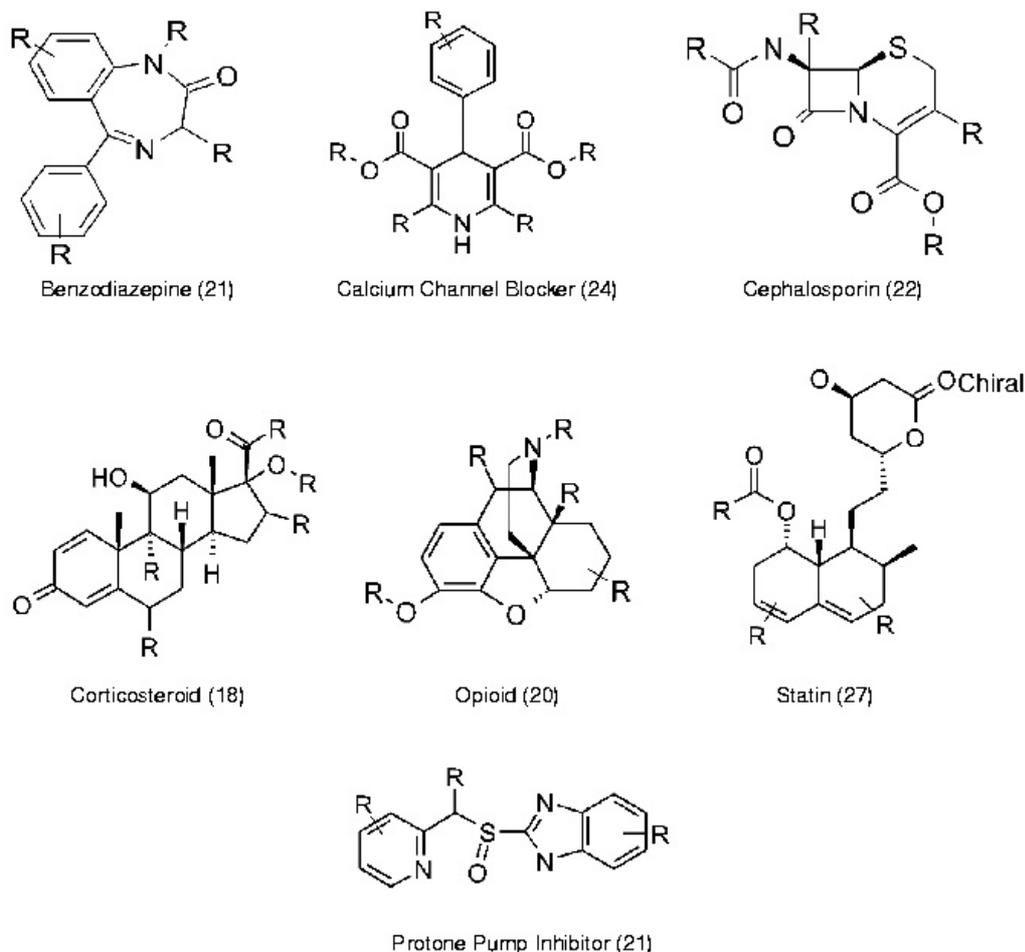
Figure 4. Seven classes of chemical compounds that provide the clustering set. For each class, the scaffold of the chemical structure is depicted and the class size is given in brackets. More details about each class are given in [3].

distances between other pairs of classes. This distance is even smaller than the average compound distance within some of the other classes. Thus, in a loose sense, from afar the two classes look like one single compact cluster. Only after zooming in (which is what we do with SSC) does the cluster reveal its actual substructure. In fact, figure 6 shows that SSC successfully identifies the seven original chemical classes.

3.3. Understanding why it works

So far, SSC has been introduced as a heuristic concept whose justification is to a large extent in its success in applications. In the following, we will provide some analytical considerations applied to a simple mean-field-like model in order to reveal the origin of this success.

In figure 7, three large clusters C_1, C_2 and C_3 have $N \gg 0$ spins each. The internal connectivity of each cluster C_l, $l = 1, 2, 3$, is as follows: Each spin is connected to k_l neighbours. The corresponding distances (or similarity values) are fixed values d_l. There is exactly one connection between the clusters. That is, no more than one of the
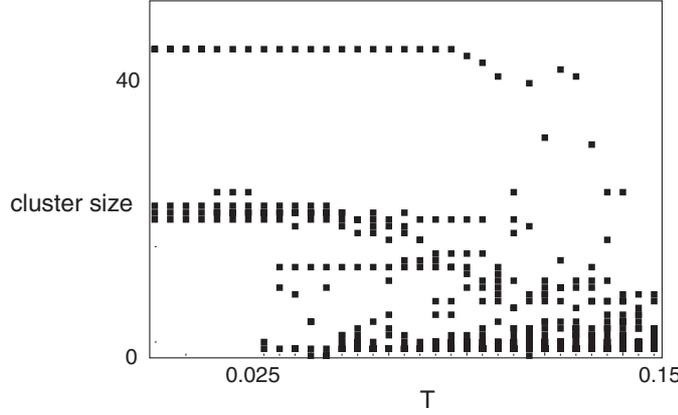
Figure 5. Clustering result achieved by SC.

spins from cluster l, $\widehat{s_{lm}}$, is connected to one of the spins of cluster m, $\widehat{s_{ml}}$, where the corresponding distances are denoted by $b_* = b_{12}$ and $b := b_{13} = b_{23}$. We generally assume that $b \geq b_* \geq d_1, d_2$ and $b \geq d_3 \geq d_1, d_2$. This situation sketches the situation encountered in the preceding examples. For simplicity, the system is treated in the familiar Ising spin formalism, i.e., the Hamiltonian is given by

$$H(s) = -\left(J^{b*}\widehat{s_{12}}\widehat{s_{21}} + J^b(\widehat{s_{13}}\widehat{s_{31}} + \widehat{s_{23}}\widehat{s_{32}}) + \sum_{l=1,2,3} \sum_{(ij)\in C_l^2} J^l s_i s_j \right), \quad (7)$$

where s_i and s_j can take on the values -1 and 1. J^l is the coupling strength within the cluster C_l corresponding to the distance d_l. J^b and J^{b*} correspond to b and b_*. The connection weights are determined according to (1). The sum (ij) is taken over all connected pairs in each cluster. Each $\widehat{s_{lm}}$ reappears in this sum as one of the $s_i \in C_l$.

If we neglect the influences from the other clusters by means of two single connections $(\widehat{s_{lm_1}}, \widehat{s_{lm_2}})$, each cluster C_l is independent. For each cluster, we can thus discriminate a ferromagnetic and a paramagnetic phase whose transition can be estimated for not too small k_l from the mean-field self-consistency equation for Ising spin models

$$\langle s_l \rangle = \tanh(T_c^l \langle s_l \rangle / T) \qquad \text{with } T_c^l = J^l k_l. \quad (8)$$

For a first estimate, we can use the critical temperature $T_c^l = J^l k_l$ as the indicator for the decay of C_l. (We assume a short decay phase of G_{ij} from 1 to 1/2. The effective T_c^l is slightly lower, depending on Θ. $\Theta = 0.7$ is a typical threshold value for $q = 2$ [17].) Furthermore, we can estimate the length of the global ferromagnetic phase from the temperature

$$T_c^b = \frac{-2J^b}{\ln(1/\Theta - 1)} = k_\Theta J^b. \quad (9)$$

At this temperature the configuration $\widehat{s_{13}} = \widehat{s_{31}}$ (and $\widehat{s_{23}} = \widehat{s_{32}}$ respectively) loses stability, i.e.,

$$p(\widehat{s_{13}} = \widehat{s_{31}}) = \frac{\exp(J^b/T)}{(\exp(J^b/T) + \exp(-J^b/T))} \leq \Theta \text{ for } T \geq T_c^b. \quad (10)$$
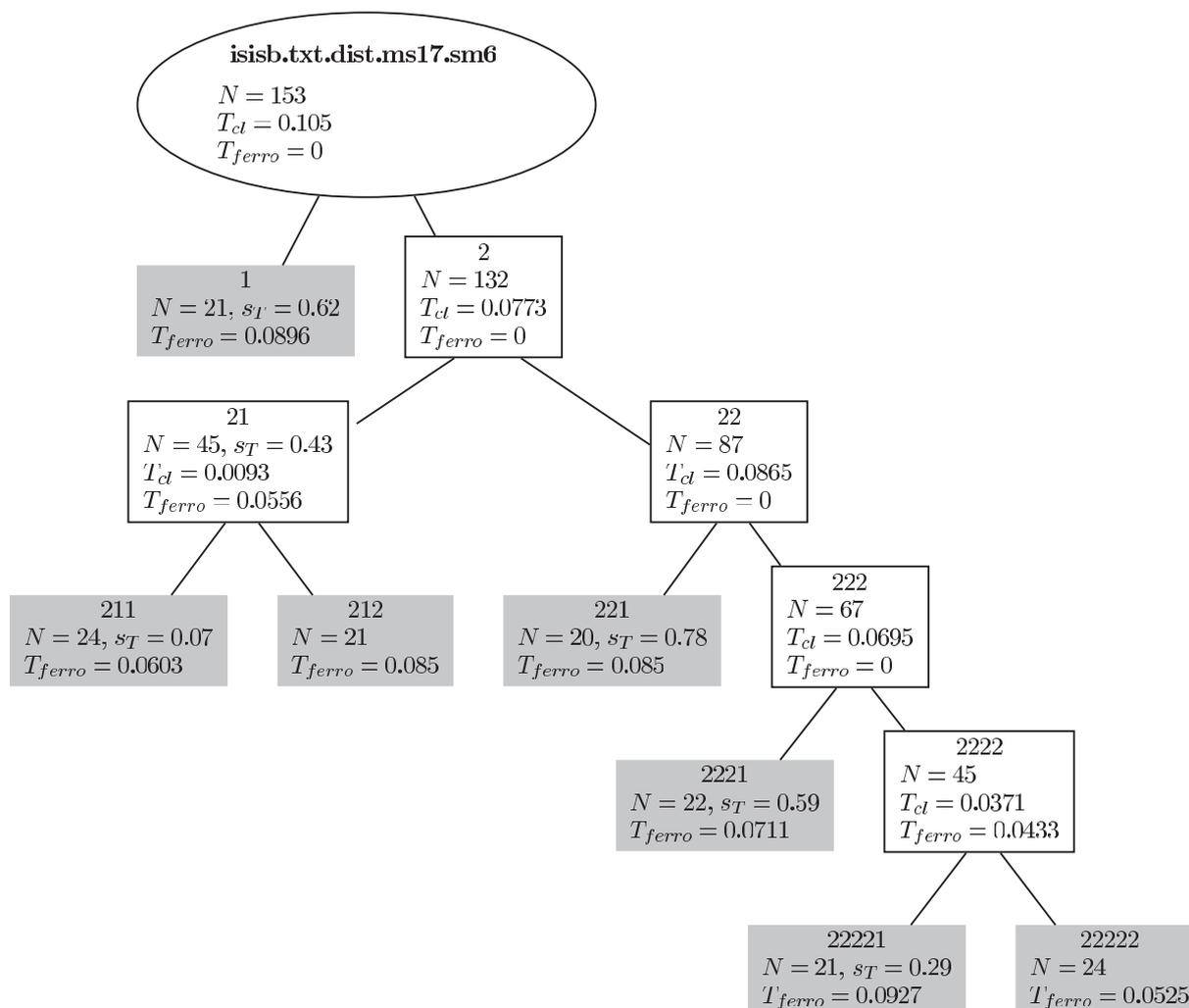
Figure 6. Our automated standard SSC successfully reveals the seven chemical classes from which the test set is composed. The only error is the assigning of three compounds to the class of *statins* instead of *corticosteroids*, yielding class sizes 24 and 21 instead of 27 and 18. A relatively small value of $s_T = 0.07$ indicates that this subdivision is less evident than other divisions in the set. For the set, ISIS binary keys with Euclidean distance were used. The results can be improved by using ISIS count keys with our extended Tanimoto similarity measure [3]. The present situation has been chosen for didactical reasons.

This estimation is justified since for these low temperatures the clusters are still fully ordered ($s_i = \widehat{s_{lm}} \ \forall s_i \in C_l$). Similarly, the conglomerate of the clusters C_1 and C_2 decays into its components at $T_c^{b_*} = k_\Theta J^{b_*}$. In figure 8, the typical phase diagram is displayed.

We now have a simple model to hand that allows us to understand some of the advantages of sequential clustering.

(a) It is clear that if $b_* \ll d_3$ then $J^3 \ll J^{b_*}$, and thus $T_c^3 < T_c^{b_*}$. In this case, C_3 decays long before C_1 and C_2 emerge as single independent clusters. In other words, a single
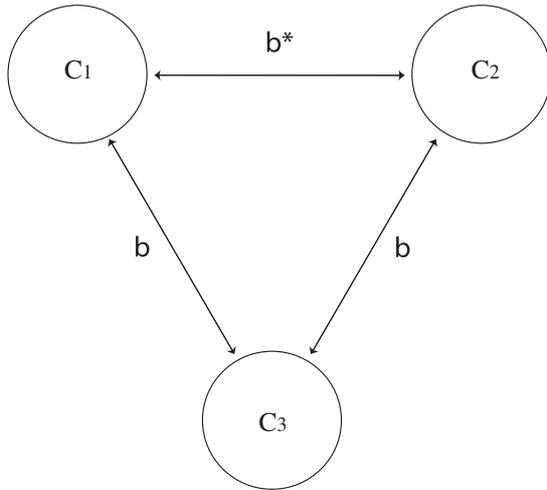
Figure 7. Model system with three clusters that are connected pairwise through one bond.
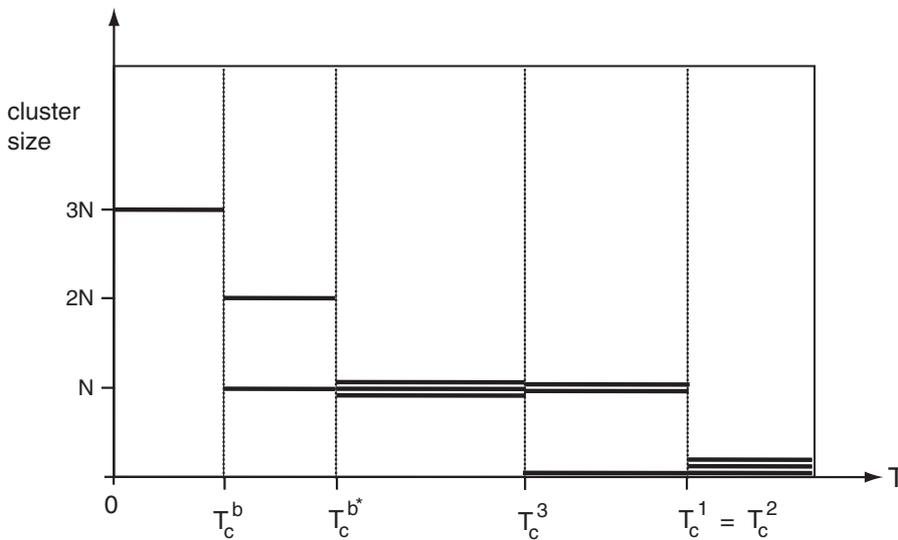
Figure 8. Typical situation for $J^b < J^{b*} < J^3 < J^1 = J^2$. At T_c^b the system enters the superparamagnetic phase and T_c^1 marks the transition to the paramagnetic phase. For convenience, Θ can be chosen such that $k_l = k_\Theta$.

level T that would contain all three clusters does not exist. Sequential clustering, however, does not rely on the specification of one (or several) resolution level(s) T.

(b) If $b \approx d_3 > b_* > d_2 \gg d_1$, then $0 \approx T_c^b \lesssim T_c^3 < T_c^{b*} < T_c^2 \ll T_c^1$. In this situation, C_1 dominates the picture and the stability of C_3, $s_T^3 = (T_c^3 - T_c^b)/T_c^1$, is marginal compared to the stability of C_1, $s_T^1 = (T_c^1 - T_c^{b*})/T_c^1$. However, the stability s_T^3 is greatly enhanced to $s_T^3 = (T_c^3 - T_c^b)/T_c^2$ (where $T_c^2 \ll T_c^1$) if C_1 is removed, allowing a clear detection of C_3. This explains why SSC could successfully detect the sparse cluster in the toy system of the first example above.

(c) The last argument in (b) holds as long as a change of the weights J, which is mainly a consequence of a changed local parameter a in (1), can be neglected. This change, however, can be crucial in other cases. Consider a situation with $d_1 = d_2 \approx b_* \ll d_3$. Then $a \gg d_1$ is to a large extent dominated by d_3. If, however, C_3 is removed and only the set containing C_1 and C_2 is clustered, we observe $a \approx d_1$. Due to (1), J^1 and thus T_c^1 is larger in the first case. However, in the second case the difference $\Delta J = J^1 - J^{b*}$ is larger as $\Delta J(a)$ is a decreasing function in $a > d_1$. The collective effects cause the stability

$$s_T^{1,2} = \frac{T_c^1 - T_c^{b*}}{T_c^1} = \frac{k_1 J^1 - k_\Theta J^{b*}}{k_1 J^1} = 1 - k\frac{J^{b*}}{J^l} = 1 - k\exp\left(\frac{d_1^2 - b_*^2}{2a^2}\right) \tag{11}$$

to decrease, as a function of a. In other words, removing C_3 enhances the stability of the clusters C_1 and C_2. These structures can therefore be found by clustering the subset containing C_1 and C_2 which has a large stability and is thus easy to detect. In contrast, we might not be able to extract C_1 and C_2 when clustering a superset including also some less dense clusters. In this case, due to a small degree of stability, the C_1 and C_2 structures may remain unrevealed. This explains why SSC could successfully uncover all the chemical classes in the example given above, whereas SC failed. In conclusion, inhomogeneities decrease the stability of single clusters. SSC counteracts this problem by successively producing sets of increased homogeneity.

4. Visual scene analysis application

Here we touch on another field of application of clustering techniques of current interest: visual scene analysis. The overall goal of visual scene analysis is to identify and classify the (relevant) objects present in the visual field (of a visual sensor, such as a biological or artificial retina or a camera). A reliable scene analysis system also plays an important role in the realization of autonomous robots. Obviously, autonomously interacting robots require some degree of visual scene understanding. Usually, the visual information is provided in the form of digital images, i.e., a two-dimensional matrix $f(x, y)$, where (x, y) denotes the spatial coordinates and $f(x, y)$ is the according feature value(s). In our simple studies, $f(x, y)$ codes for the pixel intensities of a greyscaled picture.

Clustering has been proposed as a unifying principle for all levels of image understanding [5]. Low-level analysis is associated with scene segmentation by means of pixel clustering, whereas clustering of edge elements for perceptual grouping is referred to as mid-level analysis. High-level analysis is associated with clustering of whole objects of the visual scene, in order to find object categories.

In the following, we concentrate on a simple low-level task, the separation of objects (i.e., homogeneous regions) from a noisy background. Our working example is shown in figure 9, where two black objects of sizes 65 and 25 pixels and three small white objects of sizes 12, 8 and 8 pixels are embedded. For clustering, to each pixel i a Potts spin s_i is assigned. The connectivity graph is obtained by connecting each pixel to its four adjoined neighbours. The weight of a connection is again given by (1). The distance between two connected pixels i and j is determined by the pixel values difference, i.e.,

$$d_{ij} = |f(i) - f(j)|, \tag{12}$$
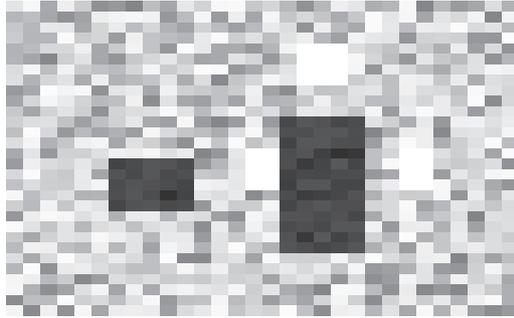
Figure 9. Different objects on a noisy background.

where $f(i) = f(x_i, y_i)$ and the pixels are numbered in some way. The task posed by figure 9 might appear factitious, but it has a pedagogic intention. We now deal with a crucial difference from the examples discussed above: the connectivity of the Potts spin system does not just reflect an actual neighbourhood relation in terms of similarity of the quantities that are expressed by the spin (i.e., the pixel values). Instead, it is forced from outside by the pixel arrangement. As a consequence, the original choice of the local parameter a (equation (1)) turns out to be problematic. However, convincing results (figure 10) are achieved over a large range of constant, but relatively small values of a. For the displayed results, $a = 0.01$ was chosen (alternatively, the spatial distance could be included in d_{ij} and a can be chosen as usual). The residual set (782 pixels) has a clearly shortened ferromagnetic phase T_{ferro} compared to the ferromagnetic phase of the extracted clusters. It is thus clearly classified as background.

5. Conclusions

Sequential superparamagnetic clustering provides a unique natural clustering solution, where the algorithm can be implemented in an automated way. Unlike other algorithms, the approach is introduced as a heuristic concept, and is not directly based on the optimization of a chosen objective function. The approach takes advantage of the robustness of self-organization processes in Potts spin systems in two respects. On the one hand, superparamagnetic clustering uses the statistical robustness properties of the Boltzmann distribution to evaluate an averaged, balanced and thus natural clustering for different resolution levels T. On the other hand, the sequential procedure reaches across the clusterings obtained for different Ts, in order to compare the stability (s_T) of all clusters found within the clustered set.

However, the sequential clustering concept does more than just search for the most stable clusters in the original set. By extracting the most stable sets and reclustering, the algorithm overcomes the problem that natural entities or classes that are inhomogeneously represented in the feature space have to be selected against their respective background. Some classes therefore are not found in the original set, but only in a subset of adequate homogeneity. This also expresses itself in the fact that, for some clusters, a global resolution level T can no longer be specified. Instead the adequate resolutions for a best unique choice of clusters become dependent on a
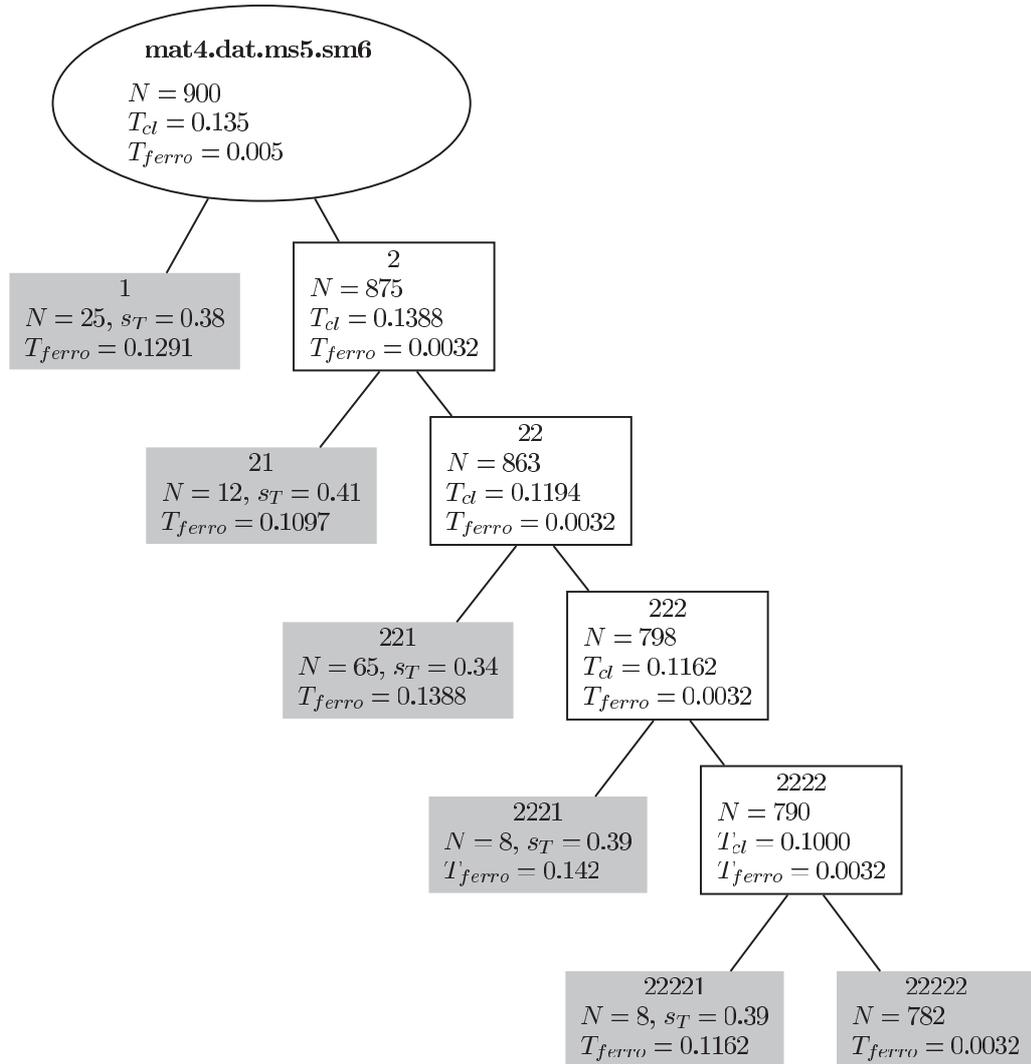
Figure 10. Clusters recognized in figure 9. The residual set (782 pixels) must be interpreted as background since $T_{\mathrm{ferro}} \approx 0$.

'local' clustering context. As a consequence of the sequential procedure, the clusters identified by the algorithm tend to be homogeneous themselves. This is consistent with the strictly non-parametric approach of our two-stage clustering paradigm that prohibits the assumption of any particular consistency or internal distribution of clusters.

The ultimate rationale behind this paradigm is the notion that clusters are not actual entities by themselves, but a fragmentary representation of some kind of real class entities out there. Clustering therefore should not just be considered a simple data or information compression. Rather, clustering is a means for accessing hidden information about object relations. Eventually, this allows for unexpected predictions based on the similar-property principle, i.e., the experience that (structurally) similar objects tend to behave similarly in many situations.

References

[1] Hampel F, *Some thoughts about classification*, 2002 *Classification, Clustering and Data Analysis* (Berlin: Springer) pp 5–26

[2] Zhao Y and Karypis G, *Clustering in life sciences*, 2003 *Methods Mol. Biol.* **224** 183–218

[3] Ott T, Kern A, Schuffenhauer A, Popov M, Acklin P, Jacoby E and Stoop R, *Sequential superparamagnetic clustering for unbiased classification of high-dimensional chemical data*, 2004 *J. Chem. Inf. Comput. Sci.* **44** 1358–64

[4] Quian Quiroga R, Nadasdy Z and Ben-Shaul Y, *Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering*, 2004 *Neural Comput.* **16** 1661–87

[5] Gdalyahu Y, Weinshall D and Werman M, *Self-organization in vision: stochastic clustering for image segmentation, perceptual grouping, and image database organization*, 2001 *IEEE Trans. Pattern Anal. Mach. Intell.* **23** 1053–74

[6] Shental N, Zomet A, Hertz T and Weiss Y, *Pairwise clustering and graphical models*, 2003 *Proc. NIPS*

[7] van der Vyver J-J, Christen M, Stoop N, Ott T, Steeb W-H and Stoop R, *Towards genuine machine autonomy*, 2004 *Robot. Auton. Syst.* **64** 151–7

[8] Downs G M and Barnard J M, *Clustering methods and their uses in computational chemistry*, 2002 *Rev. Comput. Chem.* **18** 1–40

[9] Jain A K, Murty M N and Flynn P J, *Data clustering: a review*, 1999 *ACM Comput. Surv.* **31** 264–323

[10] Buhmann J, *Data clustering and learning*, 2002 *The Handbook of Brain Theory and Neural Networks* 2nd edn, ed M A Arbib (Cambridge, MA: MIT Press)

[11] Ott T and Stoop R, *Human-like perception using sequential superparamagnetic clustering*, 2004 *Proc. NOLTA (Fukuoka)* pp 391–4

[12] Blatt M, Wiseman S and Domany E, *Superparamagnetic clustering of data*, 1996 *Phys. Rev. Lett.* **76** 3251–4

[13] Yedidia J S, Freeman W T and Weiss Y, *Understanding belief propagation and its generalizations*, 2003 *Exploring Artificial Intelligence in the New Millennium* pp 239–6, chapter 8

[14] Ott T, Steeb W-H and Stoop R, *The stochastic network approach to clustering*, 2004 *Proc. NOLTA (Fukuoka)* pp 383–6

[15] Wiseman S, Blatt M and Domany E, *Superparamagnetic clustering of data*, 1998 *Phys. Rev. E* **57** 3767–83

[16] Wiseman S, Blatt M and Domany E, *Data clustering using a model granular magnet*, 1997 *Neural Comput.* **9** 1805–42

[17] Ott T, Dauwels J and Stoop R, *Sequential clustering by loopy belief propagation*, 2005 *Proc. ECCTD (Cork)* Paper No 328

[18] Agrawal H and Domany E, *Potts ferromagnets on coexpressed gene networks: identify maximally stable partitions*, 1996 *Phys. Rev. Lett.* **90** 3251–4

[19] Willett P, 1987 *Similarity and Clustering in Chemical Information Systems* (Letchworth: Research Studies Press)