CrossMark

FOCUS

A hybrid proposal for cross-sectoral analysis of knowledge management

Álvaro Herrero[1] · Lourdes Sáiz-Bárcena[1] · Miguel A. Manzanedo[1] ·
Emilio Corchado[2]

Abstract At present time, although many theoretical formulations have been successfully proposed, there is a lack of ICT-based tools to support practical deployment of knowledge management (KM) in real settings. To bridge this gap, a hybrid artificial intelligence system is proposed in present study, aimed at gaining deeper knowledge about KM practices in four different economic sectors. By means of soft computing, companies are diagnosed according to their status regarding KM and subsequent explanations about crucial KM practices and perspectives are generated. Interesting conclusions are then derived from these explanations, allowing KM managers to optimise their decisions and obtain better results. Experimental results of real-life data from Spanish companies associated with different economic sectors validate the proposed combination of techniques.

Keywords Knowledge management · Hybrid artificial intelligence system · Unsupervised neural networks · Exploratory projection pursuit · Classification trees

✉ Lourdes Sáiz-Bárcena
lsaiz@ubu.es

Álvaro Herrero
ahcosio@ubu.es

Miguel A. Manzanedo
mmanz@ubu.es

Emilio Corchado
escorchado@usal.es

[1] Department of Civil Engineering, University of Burgos, Avenida de Cantabria, s/n, 09006 Burgos, Spain

[2] Departamento de Informática y Automática, Universidad de Salamanca, Plaza de la Merced, s/n, 37008 Salamanca, Spain

1 Introduction and previous work

Knowledge management (KM) (Herrero et al. 2010) means that organisations can capture and share the collective experience and the know-how (knowledge) of their employees and apply their knowledge in intelligent ways (Durst and Edvardsson 2012). In an environment such as today's, where everything changes at great speed and almost nothing remains static, it could be said that knowledge emerges as the key factor in any economy (Levy 2011). A firm requires both "general" knowledge and "specific/singular" knowledge, which will permit the firm to pursue excellence alongside others. This class of first-level knowledge is held by a small number of people and more select and sophisticated knowledge is required to meet the needs, in general terms.

Companies from different industrial sectors in Burgos (Spain) are analysed in present study. These companies are in a dynamic environment characterised by: high levels of competitiveness; clients with increasing demands that know their own needs and how to satisfy them; a need for personalised products and services; the existence of novel techniques that require professionals to have a knowledge and a good command; highly-qualified providers; disconcerting changes and new problems that must be addressed.

Under such circumstances, traditional sources for competitive advantages (such as physical, financial and technological assets, access to raw material or special markets, and list of clients) are not enough because these sources are available for the majority of the companies, subject to the same conditions.

To effectively compete, companies must focus on those resources and capabilities that are truly valuable: difficult to get by other companies, with a positive effect on the business, being irreplaceable for the company, highly complemen-

🖄 Springer

tary to other resources/capabilities, and low accessible to competitors. At the same time, it is crucial to promote and strengthen the development of the own knowledge, leading to a distinctive identity and personality in the processes and activities the companies carry out. KM studies these changes, and the forecasting of them, trying to respond to the above-mentioned challenges by designing and developing concepts, tools and management models.

As well as being the cause and the result of successful business activity, we may consider knowledge as one of its positive outcomes. Furthermore, its wealth lies in the variety of knowledge that the firm can amass, apply, and develop in its day-to-day activities, along with its social and collective nature. It is specially important, given that a firm's knowledge is not merely a juxtaposition of the individual knowledge of its employees, but a construct that emerges from the events experienced by the firm and its working methods, organisational routines, and shared values, developed over time (Sáiz et al. 2010).

According to that, it is vital to have tools supporting KM and facilitating its application in the reality of a firm. At a theoretical level, research has advanced quite a lot over the last 15 years. Nevertheless, to the best of the authors' knowledge, proposals of application tools and even empirical studies that allow KM to be studied and systematised are very scarce. Up to now, a heterogeneous set of KM technologies (Maier and Remus 2002; Maurer and Tochtermann 2002; Rollett 2003; Nielsen and Michailova 2007; Ruggles 2009) has been proposed from industrial vendors (Hyperwave 2008), (OPENTEXT 2008), (IBM 2008) as well as from academia (Woitsch and Karagiannis 2002; Dustdar 2002; Wang 2004; Chen et al. 2005; Du Plessis 2007; Sun 2010; Xu et al. 2010; Lerro et al. 2012; Singh and Gupta 2014; Yao et al. 2015).

In recent years, the deployment of information technology has become a crucial tool for enterprises to achieve a competitive advantage and organisational innovation (Shu-Mei 2008; Chang et al. 2012; Kang and Hau 2014; Khedhaouria and Jamal 2015; Tan and Wong 2015). However, few authors have approached the application to KM of ICT in general terms and artificial intelligence more specifically. Among them, some have addressed the topic KM performance evaluation from the fuzzy logic perspective: Wang et al. (2016) proposes a tool based on triangular fuzzy numbers to measure indexes and facilitate the performance evaluation with a group support system, while Kan et al. (2016) apply an induced interval-valued intuitionistic uncertain linguistic ordered weighted geometric operator. In Trappey et al. (2013) an ontology-based adaptation of back-propagation neural networks is applied to support knowledge sharing in the research and development field (patents for collaborative product development).

Authors of present paper have previously combined intelligent techniques to address KM in the automotive industry versus wall-painting companies (Herrero et al. 2010) and power sector (Sáiz-Bárcena et al. 2015). Apart from analysing KM in companies from different sectors and from a cross-sectorial perspective, present paper differentiates from previous work in the target tasks and applied methods, as described in Sect. 2. It extends the seminal work (Herrero et al. 2015) by considering a whole set of KM data and more sectors. Additionally, companies are not only diagnosed by means of a neural visualisation, but an interesting characterisation work is also developed, by means of classification trees.

In this work, authors wish to make a partial contribution to overcome the lack of KM tools, by a proposal that can digest the large amount of information that empirical research brings together and process that information in an automated or semi-automated way. Present study aims at analysing the effect (at company, economic sector, and cross-sectoral levels) of some KM practices on competitiveness, success and survival. The study at the sector level is important because both opportunities and menaces are the same for the companies in a certain sector. Among the analysed practices, it is worth mentioning the competitive advantages, critical capabilities and pieces of knowledge, used resources, formalise and document the available knowledge, needed ways of learning, which one of those are more profitable, how do they have to design efficient strategies for knowledge transfer, share, and capitalisation. As a tool supporting KM practice, a hybrid artificial intelligent system (HAIS) (Corchado et al. 2009) is proposed in present work. The proposed HAIS supports KM managers in understanding the KM status and the best KM practices under the frame of a certain company and sector, mainly by combining artificial neural networks and classification trees, as described in Sect. 2. To validate the proposed tool, it has been tested against real-life datasets that come from different economic sectors.

To sum up, the study intends to generate interesting guidelines about some questions that have not been addressed up to now in the KM field and lead company activity to take advantage of KM. The rest of this paper is organised as follows: The proposed hybrid approach is presented in Sect. 2, while details about the data and the experiments are provided in Sect. 3. Section 4 presents the derived conclusions and points out for future work.

2 A HAIS for knowledge management

To address the problem previously explained, present research proposes a HAIS combining the following techniques:
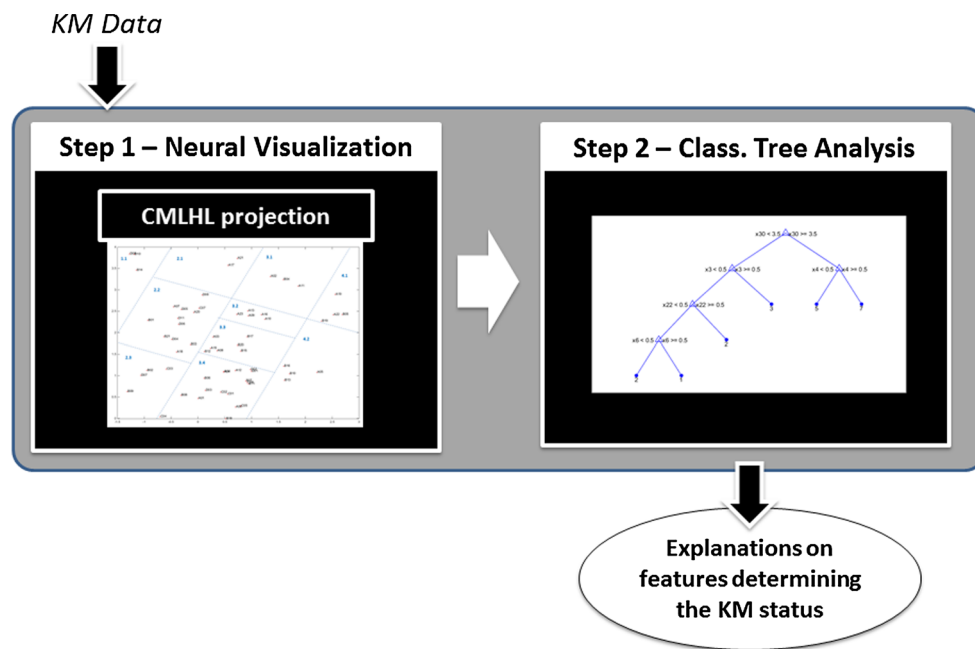
Fig. 1 Process of the proposed HAIS

- Artificial neural networks (described in Sect. 2.1): are applied to generate intuitive visualisations of KM data by reducing their dimensionality. Thanks to such a visualisation, the inner structure of the dataset is revealed and the different companies are diagnosed according to their position in the obtained projections by KM experts. Based on the diagnosis of the KM status, a company is assigned a label that is subsequently employed to build the classification trees (see Sect. 3).

- Classification trees (described in Sect. 2.2): are applied to identify the data features and their values that determine the KM status, once diagnosed by means of the neural networks. This is very interesting as the decision expressions associated with tree nodes give explanations about the classification of companies according to their KM status. As a result, interesting conclusions about the features and values that let us distinguish between the different status a company might be (see Sect. 3 for details in a real-life situation).

These techniques are applied under the frame of the process depicted in Fig. 1 and described below.

Unsupervised projection is the kind of neural model selected to perform step 1 due to its ability to provide us with deep knowledge about the ordering of a dataset whose structure is not known in advance. It means that, from the original multidimensional dataset it can be obtained a 2D projection of the data, to intuitively see the structure of the dataset. Cooperative maximum likelihood Hebbian learning (CMLHL) (Corchado and Fyfe 2003) has been chosen

because it reduces the data dimensionality while preserving the topology in the original data set (see Sect. 2.1). From the wide range of available projection models, CMLHL has been applied as it proved to obtain more informative projections than the other ones (Herrero et al. 2009; Herrero et al. 2010; Herrero et al. 2015). On the other hand, for step 2, a model giving reasons about how the original features determine the structure identified in step 1 was required. This was the main reason to select classification trees, that are able to provide explanations about the classification carried out (see Sect. 2.2).

2.1 Neural visualisation: cooperative maximum likelihood Hebbian learning

Projection models (Friedman and Tukey 1974) operate on the spatial coordinates of high-dimensional data, in order to project them onto lower dimensional spaces. The main goal is to identify the patterns that exist across dimensional boundaries by identifying "interesting" directions, in terms of any specific index or projection. Such indexes or projections are, for example, based on the identification of directions that account for the largest variance of a data set—i.e. principal component analysis (Hotelling 1933; Pearson 1901)—or the identification of higher-order statistics such as the skew or kurtosis index—i.e. exploratory projection pursuit (Friedman and Tukey 1974). Having identified the most interesting projections, the data are then projected onto a lower dimensional subspace, plotted onto two or three dimensions, which makes it possible to examine its structure with the naked eye.

The combination of projection techniques together with the use of scatter plot matrices is a very useful visualisation tool to investigate the intrinsic structure of multidimensional data sets, allowing experts to study the relations between different components, factors or projections, depending on the technique that is applied.

CMLHL is based on maximum likelihood Hebbian learning (MLHL) (Corchado et al. 2004), and introduces the application of lateral connections derived from the rectified Gaussian distribution (Seung et al. 1998). Considering an N-dimensional input vector (x), and an M-dimensional output vector (y), with W_{ij} being the weight (linking input j to output i), then CMLHL can be expressed as:

1. Feed-forward step:

$$y_i = \sum_{j=1}^{N} W_{ij} x_j, \quad \forall i \quad (1)$$

2. Lateral activation passing:

$$y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ \quad (2)$$

3. Feedback step:

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \quad \forall j \quad (3)$$

4. Weight change:

$$\Delta W_{ij} = \eta \cdot y_i \cdot \text{sign}(e_j) |e_j|^{p-1} \quad (4)$$

where η is the learning rate, τ is the "strength" of the lateral connections, b the bias parameter, p a parameter related to the energy function and A is a symmetric matrix used to modify the response to the data. The effect of this matrix is based on the relation between the distances separating the output neurons. This neural projection model has been applied in present work as it proved to be the one obtaining the most informative projections for this kind of data (Herrero et al. 2015). From the KM data to be analysed (see Sect. 3 for further details), obtained projections are studied to identify the KM status of each company.

2.2 Classification tree analysis

Classification and regression tree (CART) analysis (Breiman et al. 1984) is well known and has developed since its introduction in the 1980s to predict and to explain the responses of a categorical dependent variable. Its purpose is to estimate data membership or cases in a categorical dependent variable class, from the measurements of one or more predictor variables.

The hierarchical nature of a classification tree means that the relationship of a leaf to the tree on which it grows can be described by the hierarchy of splits in the branches (starting from the root) leading to the last branch from which the leaf hangs.

The main targets for CART analysis, some of which are similar to those of other classification techniques, are (Safavian and Landgrebe 1991):

- Properly classify as much of the training dataset as possible.
- Generalise beyond the training dataset so that previously unseen data may be classified with the minimum possible error.
- Easily update the tree on the availability of further training data.
- Simplify the tree structure as much as possible.

Classification trees are usually constructed in two phases (Safavian and Landgrebe 1991): in the first (the growth phase), an overly large classification tree is constructed from the training data. In the second (the pruning phase), the final size of the tree is determined with the goal of minimising the error of the tree. Different pruning methods have developed over time, as obtaining a tree of the proper size (as small as possible) is an important issue (Sreerama 1998).

CART applies recursive binary splitting to uncover high-dimensional dataset structures. Thus, it will partition the input space into many disjoint sets, where the response measure values of a set share greater similarity than the values of different sets. The recursive partitioning of data-sample learning may be performed, as the class label and the value of the predictor variables are known in each case. Each partition is represented by a node in the binary tree that is generated. Many different measures can be applied to determine the best way of splitting the input space, most of which are based on the degree of impurity or error of the tree nodes. The applied measures are:

- Gini's diversity index: that is defined for a certain node (t) as follows.

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2 \quad (5)$$

where $p(i|t)$ is the fraction of data belonging to class i at a given node t, and c is the total amount of classes. A pure node (node with just one class) has a Gini index of 0; otherwise the Gini index is positive.

- Deviance: that is defined for a certain node (t) as follows.

$$\text{Deviance}(t) = -\sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t) \qquad (6)$$

where $p(i|t)$ defined the same as for the Gini index, and $0 \log_2 0 = 0$. As in the case of Gini index, a deviance of 0 means that the node is pure; otherwise, the deviance is positive.

- Twoing rule: differentiating from two previous ones, Twoing is not a purity measure. It is a measure for deciding how to split a node, trying to maximise:

$$P(L) P(R) \left(\sum_{i=0}^{c-1} |L(i) - R(i)| \right)^2 \qquad (7)$$

where $P(L)$ and $P(R)$ are the fractions of observations that split to the left and right branches of a node, respectively, while $L(i)$ and $R(i)$ are the fractions of members of class i in the left and right child nodes after a split, respectively.

One of the simplest measures to assess the quality of CART analysis is its resubstitution error, defined as the difference between the response to training data and the predicted response made by the tree based on the input training data. Low resubstitution error rates do not necessarily mean good predictions for new data.

Classification trees automatically constructed from data have been successfully used in many real-world situations (Budnik and Krawczyk 2013; Chang and Chien 2013; Hossain and Piantanakulchai 2013; Jung et al. 2016; Mather and Johnson 2015; Ramachandiram and Pazhanivelan 2015; Sadeghi et al. 2013). Their effectiveness has been widely compared with other automated data exploration methods and human experts (Sreerama 1998). CART analysis performed in MATLAB (The MathWorks 2016) has been chosen for the present work as binary trees ease the interpretation and explanation of classification results, which is a key issue in the problem under consideration. To the best of the authors knowledge, classification trees have not been previously used in tools supporting KM in companies.

3 Experiments and results

As previously mentioned, the proposed HAIS has been applied to analyse data coming from companies in different sectors. The analysed data, as well as the obtained results are described in this section.

3.1 Dataset

For the proposed analysis, the dataset previously applied in Herrero et al. (2015) has been extended, gathering data from 62 companies in four different sectors (Electrical and Telecommunications, Timber, Food and Private Academies). The headquarters of all the studied companies are located in Burgos (Spain). Many different features have been collected for each company through a survey, asking questions about three main areas:

1. Learning and Knowledge.
2. KM Practices
3. Contribution and Competitiveness

Data from the first area (learning and knowledge) are organised in some different subareas: competitive advantages, capabilities for competitive advantages, KM elements, ways of learning, knowledge documentation and given situations. Information about these features and their values is shown in Table 1.

Data from the second area (KM Practices) are organised in the following subareas: present situation regarding KM, perceived benefits of KM, number of people in the KM team, difficulties of KM implementation, level of KM implementation. Information about these features and their values is shown in Table 2.

Data from the third area (Contribution and Competitiveness) are organised in the following subareas: Influence of KM, Clients/supplier opinion about KM, Staff opinion about KM, Results time, and Amount of employees. Information about these features and their values is shown in Table 3.

3.2 Neural projections

The obtained projections, together with the conclusions derived from them are described in this section. Firstly, CMLHL projection of the whole dataset (comprising data from the three areas) is shown in Fig. 2. Scatterplot matrix of first three components is depicted in Fig. 2a while component pair 1–2 is in Fig. 2b. Each company in the dataset is depicted as a single point (red dot in Fig. 2a and black asterisk in Fig. 2b).

To ease the analysis, each company in the dataset was assigned and ID and these codes are depicted in the figures below (3 and 5–7), according to the projection of its values for the different features in the reduced-dimensionality space. Figure 3 shows the component pair 1–2 of the CMLHL projection, where the dataset structure can be identified with the naked eye. The different groups in this structure is depicted in Fig. 3 and numbers (1.1, 2.1, 2.2 …4.2) have been assigned to each one of them to ease referencing.

Table 1 Data features related to area 1 (Learning and Knowledge)

#	Feature name	Value range
Competitive advantages		
1	Product/service	0 (absent) and 1 (present)
2	Customer service	0 (absent) and 1 (present)
3	Innovation	0 (absent) and 1 (present)
4	Company's image/brand	0 (absent) and 1 (present)
5	Employees	0 (absent) and 1 (present)
6	Agility/adaptation	0 (absent) and 1 (present)
7	Technology	0 (absent) and 1 (present)
8	Company's management/organisation	0 (absent) and 1 (present)
Capabilities for competitive advantages		
9	Human resources	0 (absent) and 1 (present)
10	Knowledge/experience/know-how	0 (absent) and 1 (present)
11	Innovation/Competitiveness/Design	0 (absent) and 1 (present)
12	Organisation	0 (absent) and 1 (present)
13	Company's nature	0 (absent) and 1 (present)
KM elements		
14	Patents and brands	0 (absent) and 1 (present)
15	Processes	0 (absent) and 1 (present)
16	Experience	0 (absent) and 1 (present)
17	Technology	0 (absent) and 1 (present)
18	Innovation	0 (absent) and 1 (present)
Ways of learning		
19	Several employees working together	0 (absent) and 1 (present)
20	Experienced employees working with unexperienced ones	0 (absent) and 1 (present)
21	Some employees get deep knowledge in specialised areas	0 (absent) and 1 (present)
22	Internal and external training courses	0 (absent) and 1 (present)
23	New duties are assumed by employees on a regular basis	0 (absent) and 1 (present)
24	New employees, specialist in a certain issue, join the company	0 (absent) and 1 (present)

Table 1 continued

#	Feature name	Value Range
25	Collaboration with competitor companies	0 (absent) and 1 (present)
26	Opinion of clients and suppliers is taken into account	0 (absent) and 1 (present)
27	External staff is subcontracted	0 (absent) and 1 (present)
Knowledge documentation		
28	Level of knowledge documentation	1 (low) to 4 (high)
Given situations		
29	When a problem arises, the company knows who should be contacted	1 (no) to 4 (totally)
30	New ideas and proposals are welcomed	1 (no) to 4 (totally)
31	There are collaboration and knowledge exchange among employees, clients and suppliers	1 (no) to 4 (totally)
32	Experienced employees teach the other ones	1 (no) to 4 (totally)

Thanks to the neural projection, companies are grouped according to similar characteristics in the original space defining a map associated with the three analysed areas. Taking into account the structure of the companies data in Fig. 3 and once a KM expert analyses the data associated with the first area (Learning and Knowledge) it can be said that companies in best situations are clearly placed in the upper side of the projection, while companies in the worst situations are located at the bottom of the projection. More precisely, companies in best situations are those in groups 1.1 and 1.2 (in Fig. 3). Companies in groups 2.2 and 3.1 are in a good situation and those in group 3.2 are in a good/medium situation. Groups 2.3 and 3.3 gather companies in a medium situation while companies in groups 3.4 and 4.2 are in the worst situation.

When taking into account the second area (KM Practices), structure in Fig. 3 is not as clearly related to the companies values as in the case of area 1. However, companies with similar characteristics are visualised in similar ways and hence, some conclusions can be drawn; companies in best situations are located in the upper right side of Fig. 3 (groups 2.1, 3.1, and 4.1), where all the relevant features are present and high scores are obtained except for "Present level of KM implementation". Companies in the worst situation are located at the bottom of Fig. 3 (groups 2.3 and 3.4). Ranging between these two opposite situations, there are some companies in an intermediate situation (groups 1.1, 2.2, 3.2, 3.3, and 4.2).

Finally, the third area (contribution and competitiveness) has also been studied, with positive results. Although there is not a clear tendency for some of the features

Table 2 Data features related to area 2 (KM Practices)

#	Feature name	Value range
Present situation regarding KM		
1	Present situation of the company regarding KM implementation	1 (not initiated) to 4 (completed)
Perceived benefits of KM		
2	Better management	0 (absent) and 1 (present)
3	Knowledge sharing	0 (absent) and 1 (present)
4	Costs reduction	0 (absent) and 1 (present)
5	Access to knowledge	0 (absent) and 1 (present)
6	Better communication	0 (absent) and 1 (present)
7	Better results	0 (absent) and 1 (present)
8	Increased training for employees	0 (absent) and 1 (present)
9	Increased innovation	0 (absent) and 1 (present)
10	Others	
Number of people in the KM team		
11	Number of people that are in the KM team now	1 (none) to 3 (more than 4)
Difficulties of KM implementation		
12	Lack of knowledge about KM	0 (absent) and 1 (present)
13	Hierarchical organisation of the company	0 (absent) and 1 (present)
14	Lack of compromise	0 (absent) and 1 (present)
15	Too much work	0 (absent) and 1 (present)
16	Opposition to changes	0 (absent) and 1 (present)
17	Absence of KM culture	0 (absent) and 1 (present)
18	More staff is needed	0 (absent) and 1 (present)
19	Lack of technology	0 (absent) and 1 (present)
Level of KM implementation		
20	Present level of KM implementation	1 (inexistent) to 4 (high)

Table 3 Data features related to area 3 (Contribution and Competitiveness)

#	Feature name	Value range
Influence of KM on		
1	Product/service quality	0 (absent) and 1 (present)
2	Client satisfaction	0 (absent) and 1 (present)
3	Response capability	0 (absent) and 1 (present)
4	Innovation capability	0 (absent) and 1 (present)
5	Staff satisfaction	0 (absent) and 1 (present)
6	Quick response	0 (absent) and 1 (present)
7	Costs reduction	0 (absent) and 1 (present)
8	Higher sales	0 (absent) and 1 (present)
Clients/supplier opinion about KM		
9	Clients/supplier opinion about KM	1 (indifferent) to 5 (positive)
Staff opinion about KM		
10	Staff opinion about KM	1 (indifferent) to 5 (positive)
Results time		
11	Required time to obtain results from KM	0 (I do not know) to 3 (more than 12 months)
Amount of employees		
12	Total amount of employees	1 (less than 5) to 6 (more than 100)

("Clients/supplier opinion about KM" and "Staff opinion about KM"), that is not the case of the most representative subarea ("Influence of KM"), whose impact on the data ordering is strong and precisely shown in Fig. 3. Thus, companies in best situations according to this subarea are located in the upper left side of the projection (groups 1.1, 2.1, 3.1, and

4.1). Moving to the right and down, companies in an intermediate situation are found (groups 2.2, 2.3, 3.2, and 3.3), finally identifying companies in the worst situation (groups 3.4 and 4.2).

From a general analysis considering the three areas, it can be said that similar companies are grouped. That is the case of companies in best situations (groups 1.1, 2.1, 3.1, and 4.2 from Fig. 3) and those in worst situations (groups 3.4 and 4.2).

As general conclusions considering the sector level are also pursued, companies have also been visualised according to their sector in the component pair 1–2 of the CMLHL projection, as can be seen in Fig. 4.

Two interesting conclusions are obtained from Fig. 4: whether there exists any kind or ordering according to the sector of the company and what is the activity of companies in the best, intermediate and worst situations. Companies from the Electrical and Telecommunications and Private Academies
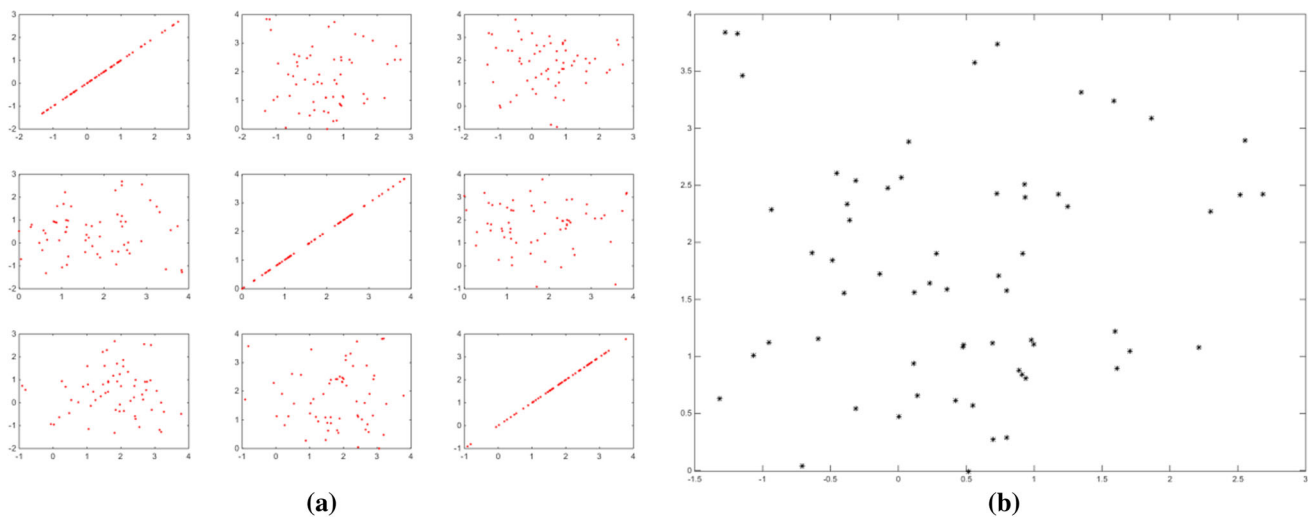
Fig. 2 CMLHL projection of the whole dataset. **a** Scatterplot matrix of three main components. **b** Component 1–2
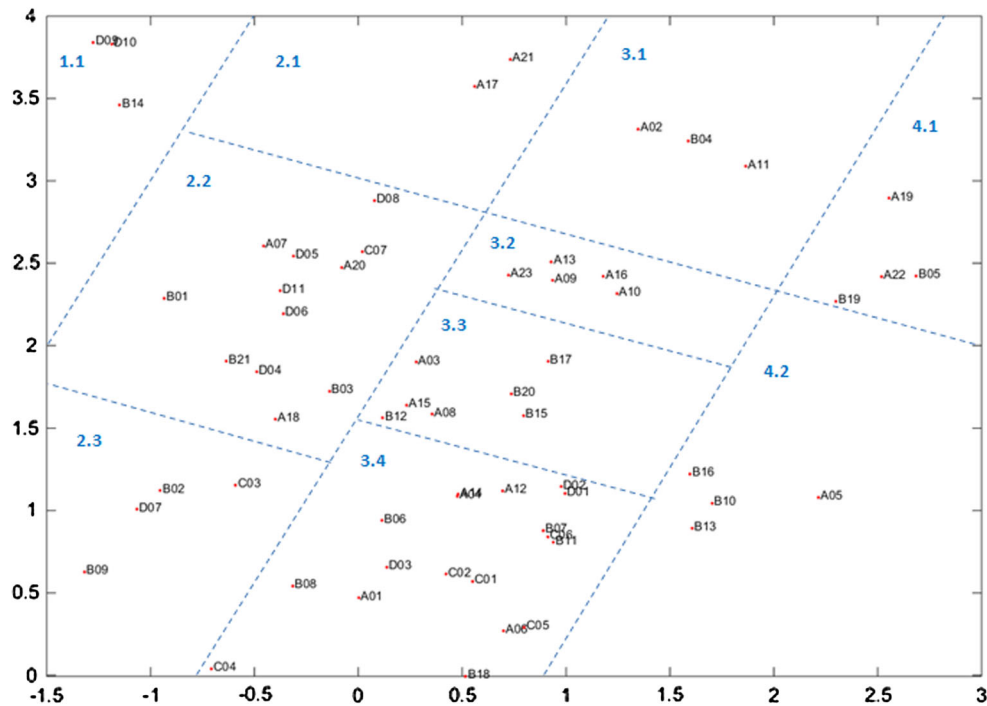
Fig. 3 CMLHL projection (component 1–2) of the whole dataset—labelled groups

sectors are mainly located in central and upper left positions of Fig. 4. Companies from the Timber sector are places in central-left positions, while companies from the Food sector are located at the central-bottom position in Fig. 4. In the best groups (excellent or good situations), that are 1.1, 2.1, 3.1 and 4.1, there are six companies from the Electrical and Telecommunications sector (26 % of the analysed companies from this sector) as well as four companies from the Timber sector (19 %) and two companies from the Private Academies (18 %) sector. From remaining situations, it can be said that most of the companies from all sectors are located in inter-

mediate positions. This means that such companies do still require improvement of their KM situation.

For a comprehensive analysis of the dataset and more interesting conclusions to be drawn, it has also been analysed area by area. According to that, CMLHL projections of data associated with each one of the three areas were obtained and are shown below. Group identification and analysis has been performed for each one of the areas.

From the analysis of groups identified in the CMLHL visualisation of the features associated with area 1 (Fig. 5), it can be said that the KM status of the companies located in the
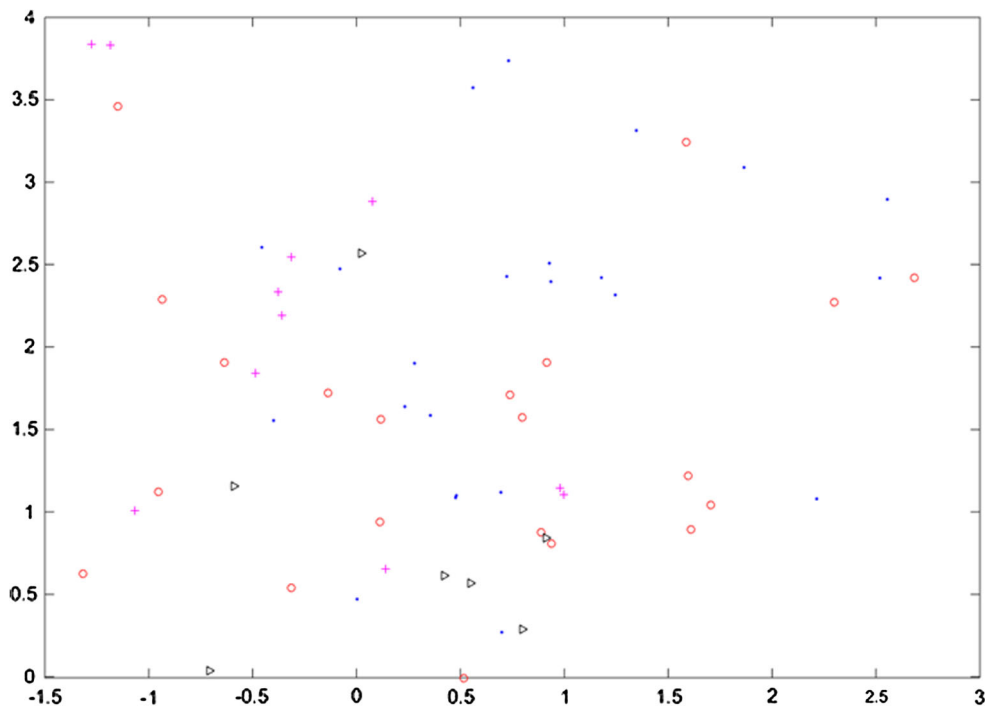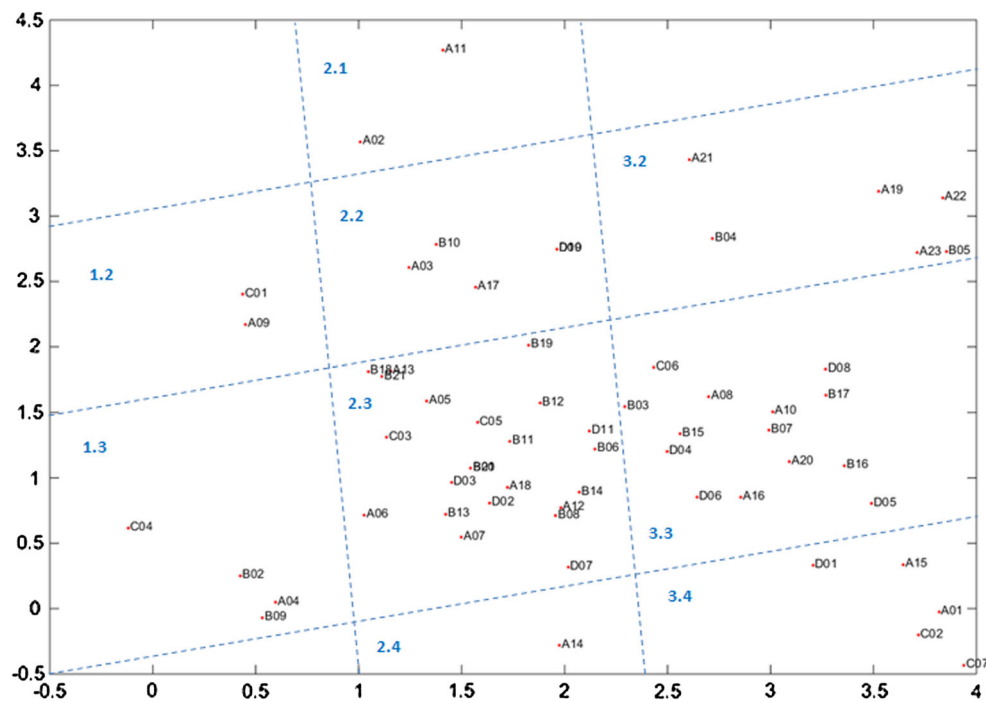
Fig. 4 CMLHL projection (component 1–2) of the whole dataset—sector visualisation. Sector legend: *blue dot* Electrical and Telecommunications, *red circle* Timber, *black triangle* Food, *pink cross* Private Academies (color figure online)
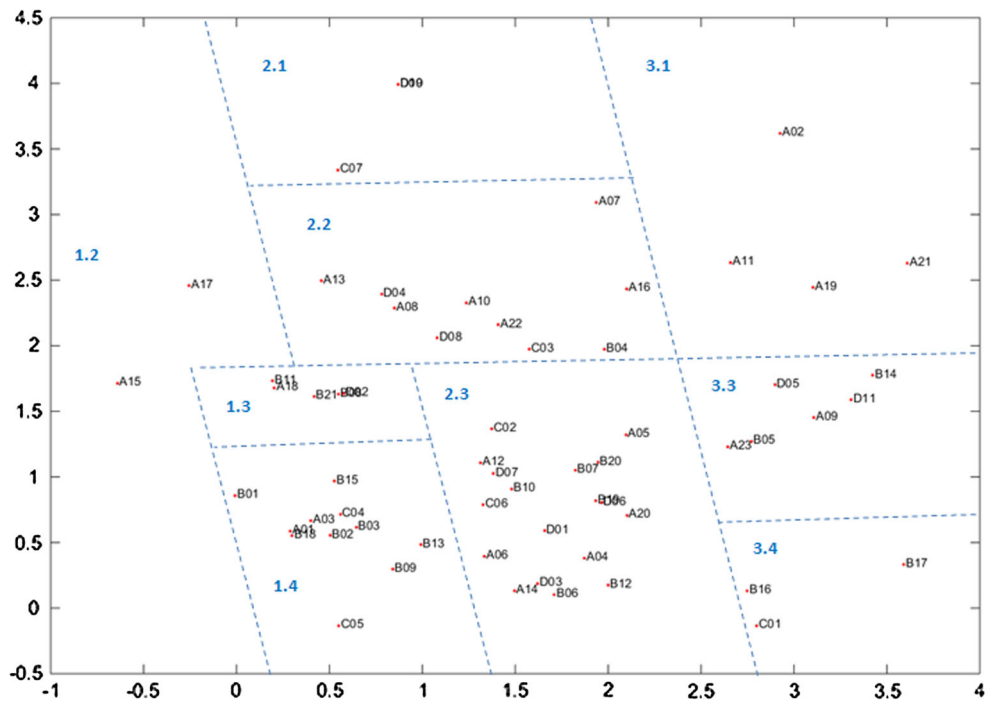
Fig. 5 CMLHL projection (component 1–2) of the features associated with area 1—labelled groups

upper right hand of the projection is excellent for this area. Groups 2.1 and 3.1 are the ones containing such companies (4 companies from the Electrical and Telecommunications, 1 from the Food and 2 from the Private Academies sector). The

status of intermediate (moving to the left and down) companies is worse, reaching the worst situations those companies located in the bottom-left side of the projection. The companies in an avoidable situation are 29 (8 are from the Electrical

Fig. 6 CMLHL projection (component 1–2) of the features associated with area 2—labelled groups

and Telecommunications, 13 from the Timber, 4 from the Food and 4 from the Private Academies sector).

From the analysis of groups identified in the CMLHL visualisation of the features associated with area 2 (Fig. 6), it can be said that, as in the case of Fig. 5, companies located in the upper right hand of the projection are those in an excellent situation for area 2. Groups 2.1, 2.2 and 3.2 are the ones containing the 13 companies in the best situation (8 from the Electrical and Telecommunications, 3 from the Timber, and 2 from the Private Academies sector). Coherently, companies located in the bottom-left side of the projection are the 5 ones with worst KM status (2 are from the Electrical and Telecommunications, 2 from the Timber, and 1 from the Food sector).

Differentiating from two previous area-based projections, in the CMLHL visualisation of the features associated with area 3 (Fig. 7), companies with an excellent KM status are located at the bottom (slightly right) of the projection. This region comprises groups 2.3 and 3.2, and the 5 associated companies (1 is from the Electrical and Telecommunications, 2 from the Timber, and 2 from the Private Academies sector). Companies with worst KM status are those located in the upper left side of the projection (groups 1.1 and 1.2), including 5 from the Electrical and Telecommunications, 6 from the Timber, 2 from the Food and 1 from the Private Academies sector.

For subsequent analysis based on classification trees, the different situations associated with the KM status of com-

panies (as identified through CMLHL projections) were numbered, according to:

1. Bad.
2. Bad/Medium.
3. Medium.
4. Medium/Good.
5. Good.
6. Excellent/Good.
7. Excellent.

These numbers are the class labels used in order to build the classification trees shown in the following section.

3.3 Classification tree analysis

Once every single company is diagnosed according to its KM status, classification trees were built. For comparison purposes, three different criteria for node splitting were applied for each one of the analysed datasets. Shown trees (Fig. 8, 9, 10) have been selected as being the ones with lowest resubstitution error (see Table 4).

From the comparison of the resubstitution errors obtained by the different split criteria under analysis in present study, it can be concluded that none of them obtained the lowest rates in all cases. Regarding the question about which one is the best criterion, it can be said that it depends on the analysed datasets (area).
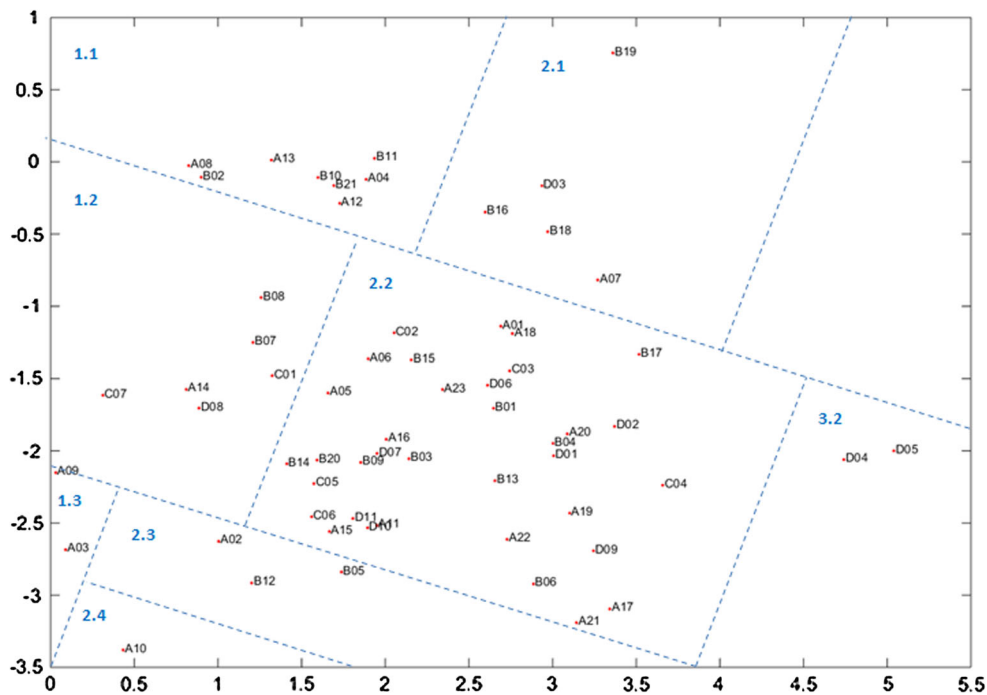
Fig. 7 CMLHL projection (component 1–2) of the features associated with area 3—labelled groups

Table 4 Resubstitution error of classification trees

Area	Gini diversity index	Twoing	Deviance
1	**0.24194**	0.24194	0.25806
2	0.25806	**0.24194**	0.27419
3	0.27419	0.27419	**0.25806**

The less significant values of each area are in bold

The tree built on the features associated with area 1 according to the Gini diversity index was pruned and its final state is shown in Fig. 8. Features used to build the tree (labelled according to Table 1 and Fig. 8) are:

- x3: Competitive advantages—Innovation
- x4: Competitive advantages—Company's image/brand
- x6: Competitive advantages—Agility/adaptation
- x22: Ways of learning—Internal and external training courses
- x30: Given situations—New ideas and proposals are welcomed.

From the decision explanations of the nodes in the tree depicted in Fig. 8, interesting conclusions can be drawn about the features associated with area 1. The first one is that the most discriminating feature is "New ideas and proposals are welcomed" from the subarea of "Given situations". According to the values assigned to this feature (see Table 1), if a company scores 4 in this feature, it would be in the
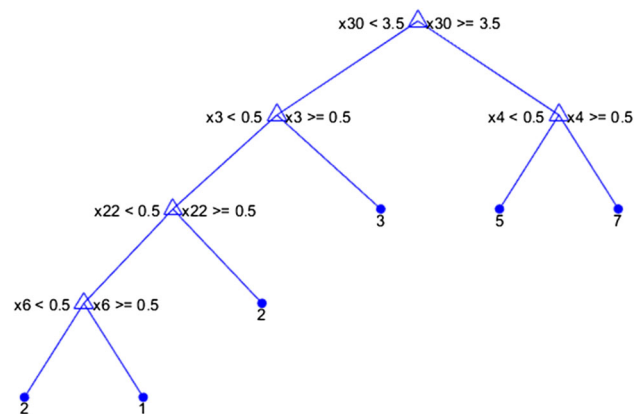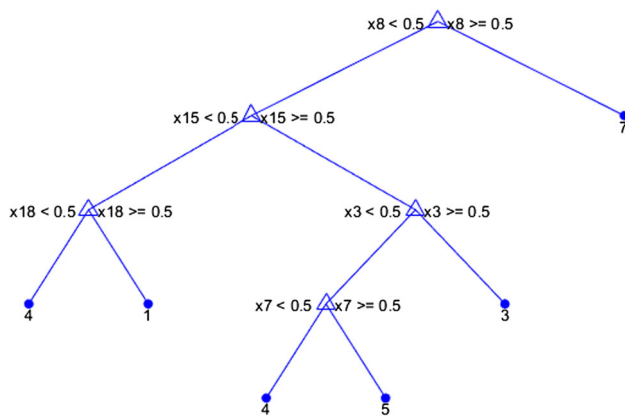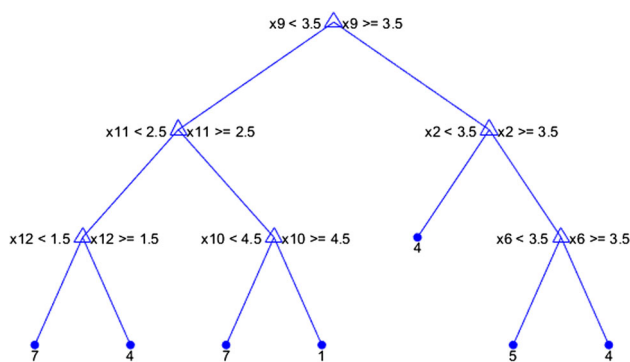

Fig. 8 Pruned tree for the features associated with area 1 (Gini diversity index)

best situations (good and excellent) while remaining values (1, 2, and 3) are associated with those companies in medium/bad situations (status lower than 4). Additionally, if the feature "Company's image/brand" is present in the "Competitive advantages" subarea for the company, it is in the optimum situation (groups 2.1 and 3.1 in projection shown in Fig. 5). If the company scores 4 in the first feature (New ideas and proposals are welcomed) but the feature "Company's image/brand" is not present, it is in a good (5 out of 7) situation (groups 2.2 and 3.3 in Fig. 5). Company's image/brand stands for the perception that customers have about the identity of the company. A high perception is associated, in general terms, with positive feelings and emotions

Fig. 9 Pruned tree for the features associated with area 2 (Twoing)

Fig. 10 Pruned tree for the features associated with area 3 (Deviance)

from potential consumers about the company and, consequently, high sales.

On the other hand, when the value associated with the feature "New ideas and proposals are welcomed" is lower than 4, and the Innovation competitive advantage is present (feature from the subarea "Competitive advantages"), the company is in an intermediate status (groups 1.2, 1.3, and 3.4 in Fig. 5). At the same time, if the value associated with the feature "New ideas and proposals are welcomed" is lower than 4, but the "New ideas and proposals are welcomed" feature is not present as a competitive advantage, the company is in a bad situation (lower than 3). From the different situations associated with these cases, the feature "Internal and external training courses" (from the "Ways of learning" subarea) is present, the company is in a Bad/Medium situation (group 2.3 in Fig. 5). In the same status are companies for which features "Internal and external training courses" (from the "Ways of learning" subarea) and "Agility/adaptation" (from the "Competitive advantages" subarea) are not present. This in an alarm situation where there are many risks linked to learning and knowledge. Lastly, in the worst situation (group 1.4 in Fig. 5) are companies for which the feature "Internal

and external training courses" (from the "Ways of learning" subarea) is not present but "Agility/adaptation" (from the "Competitive advantages" subarea) is.

These results from Fig. 8 are very interesting for companies committed to KM. They let us clearly identify which competitive advantages are more relevant: Innovation, Company's image/brand, and Agility/adaptation. This means that the other competitive advantages under analysis (Product/service, Customer service, Employees, Technology, and Company's management/organisation) have a lower importance when distinguishing between good and bad situations. Additionally, from the nine different ways of learning under analysis (listed in Table 1), the most explanatory one is the "Internal and external training courses". Finally, the given situation that greatly affects the KM status is "New ideas and proposals are welcomed".

The tree built on the features associated with area 2 according to the Twoing criterion was pruned and its final state is shown in Fig. 9. Features used to build the tree (labelled according to Table 2 and Fig. 9) are:

- x3: Perceived benefits of KM—Knowledge sharing
- x7: Perceived benefits of KM—Better results
- x8: Perceived benefits of KM—Increased training for employees
- x15: Difficulties of KM implementation—Too much work
- x18: Difficulties of KM implementation—More staff is needed.

From the decision explanations of the nodes in the tree depicted in Fig. 9, interesting conclusions can be drawn about the features associated with area 2 (KM Practices). The most discriminating feature in this case is "Increased training for employees" (from the "Perceived benefits of KM" subarea). When this feature is present (groups 2.1, 2.2, and 3.2 in Fig. 6) the companies are always in the best (Excellent) KM status. When this features is not present, that is, KM does not get a better training for employees, and the features "Too much work" (from the "Difficulties of KM implementation" subarea) and "Knowledge sharing" (from the "Perceived benefits of KM" subarea) are present, the company is in an intermediate status (group 3.4 in Fig. 6).

When the previously described situation is given, replacing "Knowledge sharing" by "Better results", the status of the company improves up to a good situation (group 3.3 in Fig. 6). On the contrary, if the "Better results" feature is not present, company is in a slightly worse (but still good) situation (Medium/Good). In the same situation are companies when the following features are not present: "Increased training for employees", "Too much work", and "More staff is needed" (from the "Difficulties of KM implementation" subarea).

The remaining situation is the worst one ("bad"), given when the "Increased training for employees", and "Too much work" features are not present, but "More staff is needed" is present. It is associated with extreme situations, where immediate corrective actions are required (groups 1.3, and 2.4 in Fig. 6).

In the case of KM practices, the classification tree results let us identify the "Perceived benefits of KM" and "Difficulties of KM implementation" subareas as the most relevant ones. On the one hand, in the "Perceived benefits of KM" subarea, a total of 8 features have been considered (see Table 2), being "Knowledge sharing", "Better results", and "Increased training for employees" the most influencing ones. On the other hand, in the "Difficulties of KM implementation" subarea, eight different features have been also considered (see Table 2). Among them, the most important ones according to the classification tree are "Too much work", and "More staff is needed".

The tree built on the features associated with area 3 according to the Deviance criterion was pruned and its final state is shown in Fig. 10. Features used to build the tree (labelled according to Table 3 and Fig. 10) are:

- x2: Influence of KM on—Client satisfaction
- x6: Influence of KM on—Quick response
- x9: Clients/supplier opinion about KM—Clients/supplier opinion about KM
- x10: Staff opinion about KM—Staff opinion about KM
- x11: Results time—Required time to obtain results from KM
- x12: Amount of employees—Total amount of employees.

The following conclusions are drawn from the analysis of the decision explanations of the nodes in the tree depicted in Fig. 10, associated with area 3 (Contribution and Competitiveness). In this case, the node in the root of the three is associated with the feature "Clients/supplier opinion about KM" from the "Clients/supplier opinion about KM" subarea. If this feature takes a value of 4, the company is in "Medium/Good" or "Good" situation. Under such circumstances, when the influence of KM on client satisfaction is lower than 4 (from 1 to 3), the company is in a "Medium/Good" situation (group 2.2 in Fig. 7). In this same status are companies that score the highest values (4) for the features "Clients/supplier opinion about KM", "Influence of KM on Client satisfaction", and "Influence of KM on quick response". If lower values are assigned to this last feature ("Influence of KM on quick response"), the company is in a good status (group 2.3 in Fig. 7).

About the left side of the three, it can be said that when the feature "Clients/supplier opinion about KM" takes a value lower than 4 and the "Required time to obtain results from KM" and "Staff opinion about KM" features take the highest values (3 and 5 respectively), the company is in the worst situation (groups 1.1 and 1.2 in Fig. 7). If the last feature ("Staff opinion about KM") takes lowest values (1–4), the company is in the best status (group 3.2 in Fig. 7).

Finally, when at the same time the feature "Clients/supplier opinion about KM" takes a value lower than 4, the feature "Required time to obtain results from KM" takes a value lower than 3, and the "Total amount of employees" takes a value higher than 4 (up to 6), the company is in a good/medium status (group 2.2 in Fig. 7). If the last feature ("Total amount of employees") takes a value lower than 5, the company is in the best situation (group 3.2 in Fig. 7).

As in the case of previous areas, the analysis of classification tree lets us identify the most discriminating features from the "Contribution and Competitiveness" area. From the "Influence of KM on" subarea, the most important features are "Client satisfaction" and "Quick response". The subareas (and hence the only feature they contain) "Clients/supplier opinion about KM" and "Staff opinion about KM" are identified as important ones as well.

As it can be seen from the examples shown in this section, based on the classification trees it is easy to clearly identify the requisites (associated with each one of the areas) for a company to be in the excellent status from the KM standpoint, differentiating from those companies in a worse situation. Company managers could use the proposed HAIS to obtain important information to know the KM status of their company, as well as information stating the elements of KM management to be modified or incorporated to the company.

4 Conclusions

The main objective of present work has always been to propose and validate a HAIS applied to the field of KM. According to that, Neural Networks and Classification Trees are combined on real-life data whose structure was not known in advance.

To investigate about the KM situation of companies, three main areas have been analysed: Learning and Knowledge, KM Practice, and Contribution and Competitiveness. Each one of them comprises many features that may take different values. The validation of proposed HAIS has been performed on a real-life setting, with Spanish companies from Electrical and Telecommunications, Timber, Food and Private Academies sectors.

From the KM perspective, present work aims at going one step further on KM research, bridging the gap between theoretical formulations and reliable and appropriate tools to support not only diagnosis and knowledge about the present situation of both companies and sectors, but also identifying the most important elements for the success of KM. This

means, key information to take suitable decisions aimed at defining the company strategy in terms of KM, designing the most appropriate system and deploying it, identifying the required elements and main obstacles, knowing the needed resources and valuing the benefits for company survival and competitiveness.

After the experimental setup, it can be said that initial targets have been achieved, as the applied techniques let us identify the features that are most discriminative for each one of the three analysed areas, apart from grouping and ordering the companies according to the different characteristics determining their KM status. A proper interpretation of results is very valuable in present research, as required for obtaining valid and interesting results not only for advancing research on critical aspects of KM, but mainly for company managers. The main reason for that is that present research supports decision taking linked to competitive advantages generated by KM, different ways of learning, given situations when dealing with knowledge, KM benefits, difficulties for KM deployment, its influence on the improvement of company elements and opinion of clients, providers and staff.

Main results of present work can be identified as:

- The status of companies in the areas of "Learning and Knowledge", "KM Practice", and "Contribution and Competitiveness" has been identified. This is useful for recognising the situation for every company and taking subsequent decisions.
- Analysed companies have been assigned to different groups clearly identified and associated with situations of a different nature, ranging from excellent to bad. Groups of companies in a certain region of CMLHL projections are described by their characteristics and corrective actions can be adapted to each one of them to improve their situation and hence, move to groups associated to better status.
- The most discriminative features have been identified and, hence, they are considered as the most representative ones for each one of the three analysed areas. It is not only features but also their values that have been identified to diagnose and improve the situation of a company. This information may be also used for decision taking regarding KM and for contributing to the survival and competitiveness of a company.

Although data only from Burgos region (Spain) have been studied, the applied methodology and techniques have proved to be valid, appropriate and novel in present study. As a result, they can also be applied to data coming from different locations, companies and sectors.

As a conclusion, the applied combination of artificial intelligence techniques allows identifying, grouping, discovering and valuing KM features considered in present research. This provides companies with required information for being conscious about KM and progress on successful deployment. The research carried out revealed that promoting new ideas and proposals is one of the most important actions to success in the KM field. Present research has also identified as important actions those related to increasing training for employees and taking into account the opinion of clients. As a result of a successful KM system, companies obtain competitive advantages such as: innovation, branding, and flexibility. Additionally, benefits are: high level of knowledge sharing among staff, improved results for the company, improved customer satisfaction, and quick response to new challenges. Thanks to obtained results, a wide range of possibilities are opened for the analysis of this kind of data and, what is more important, to get valuable knowledge from these data to support successful KM. It is a major contribution to KM where there is a lack of empirical studies and tools to support such tasks.

References

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont 358

Budnik M, Krawczyk B (2013) On optimal settings of classification tree ensembles for medical decision support. Health Inform J 19:3–15. doi:10.1177/1460458212446096

Corchado E, Fyfe C (2003) Connectionist techniques for the identification and suppression of interfering underlying factors. Int J Pattern Recognit Artif Intell 17:1447–1466

Corchado E, MacDonald D, Fyfe C (2004) Maximum and minimum likelihood hebbian learning for exploratory projection pursuit. Data Min Knowl Discov 8:203–225

Corchado E, Wu X, Oja E, Herrero Á, Baruque B (eds) (2009) Hybrid artificial intelligence systems, vol 5572. Lecture Notes in Artificial Intelligence. Springer, Berlin/Heidelberg

Chang CM, Hsu MH, Yen CH (2012) Factors affecting knowledge management success: the fit perspective. J Knowl Manag 16:847–861. doi:10.1108/13673271211276155

Chang LY, Chien JT (2013) Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Saf Sci 51:17–22. doi:10.1016/j.ssci.2012.06.017

Chen L, Shadbolt NR, Tao F, Goble C, Puleston C, Cox SJ (2005) Semantics-assisted problem solving on the semantic grid. Comput Intell 21:157–176. doi:10.1111/j.0824-7935.2005.00269.x

Du Plessis M (2007) The role of knowledge management in innovation. J Knowl Manag 11:20–29

Durst S, Edvardsson IR (2012) Knowledge management in SMEs: a literature review. J Knowl Manag 16:879–903

Dustdar S (NO) Collaborative Knowledge Flow—Improving process-awareness and traceability of work activities. In: 4th International conference on practical aspects of knowledge management (2002) LNCS. Springer, NO

Friedman JH, Tukey JW (1974) A projection pursuit algorithm for exploratory data-analysis. IEEE Trans Comput 23:881–890

Herrero Á, Corchado E, Gastaldo P, Zunino R (2009) Neural projection techniques for the visual inspection of network traffic. Neurocomputing 72:3649–3658

Herrero Á, Corchado E, Sáiz-Bárcena L, Manzanedo MÁ (2015) Analysis of knowledge management in industrial sectors by means of neural models. In: Herrero Á, Sedano J, Baruque B, Quintián H, Corchado E (eds) 10th International conference on soft computing models in industrial and environmental applications, vol 368. Advances in intelligent systems and computing. Springer International Publishing, pp 65–75. doi:10.1007/978-3-319-19719-7_6

Herrero Á, Corchado E, Sáiz L, Abraham A (2010) DIPKIP: a connectionist knowledge management system to identify knowledge deficits in practical cases. Comput Intell 26:26–56

Hossain MM, Piantanakulchai M (2013) Groundwater arsenic contamination risk prediction using GIS and classification tree method. Eng Geol 156:37–45. doi:10.1016/j.enggeo.2013.01.007

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–444

Hyperwave Knowledge Management Software. http://www.hyperwave.com. Accessed 16 Oct 2008

IBM Lotus, Knowledge Management Software. http://www-01.ibm.com/software/lotus/. Accessed 16 Oct 2008

Jung S, Qin X, Oh C (2016) Improving strategic policies for pedestrian safety enhancement using classification tree modeling. Transp Res Part A Policy Pract 85:53–64. doi:10.1016/j.tra.2016.01.002

Kan S, Guo F, Li S (2016) An approach to evaluating the knowledge management performance with interval-valued intuitionistic uncertain linguistic information. J Intell Fuzzy Syst 30:1557–1565

Kang M, Hau YS (2014) Multi-level analysis of knowledge transfer: a knowledge recipient's perspective. J Knowl Manag 18:758–776. doi:10.1108/JKM-12-2013-0511

Khedhaouria A, Jamal A (2015) Sourcing knowledge for innovation: knowledge reuse and creation in project teams. J Knowl Manag 19:932–948. doi:10.1108/JKM-01-2015-0039

Lerro A, Iacobone FA, Schiuma G (2012) Knowledge assets assessment strategies: organizational value, processes, approaches and evaluation architectures. J Knowl Manag 16:563–575. doi:10.1108/13673271211246149

Levy M (2011) Knowledge retention: minimizing organizational business loss. J Knowl Manag 15:582–600

Maier R, Remus U (2002) Defining process-oriented knowledge management strategies. Knowl Process Manag 9:103–118

Mather AL, Johnson RL (2015) Event-based prediction of stream turbidity using a combined cluster analysis and classification tree approach. J Hydrol 530:751–761. doi:10.1016/j.jhydrol.2015.10.032

Maurer H, Tochtermann K (2002) On a new powerful model for knowledge management and its applications. J Univ Comput Sci 8:85–96

Nielsen BB, Michailova S (2007) Knowledge Management systems in multinational corporations: typology and transitional dynamics. Long Range Plan 40:314–340

OPENTEXT Livelink, Knowledge Management Software. http://www.opentext.com. Accessed 16 Oct 2008

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572

Ramachandiram K, Pazhanivelan S (2015) Determination of nitrogen and water stress with hyper spectral reflectance on maize using classification tree analysis. J Agrometeorol 17:213–218

Rollett H (2003) Knowledge management: processes and technologies. Kluwer, Dordrecht

Ruggles RL (2009) Knowledge management tools. Butterworth-Heinemann, Oxford

Sadeghi R, Zarkami R, Sabetraftar K, Van Damme P (2013) Application of genetic algorithm and greedy stepwise to select input variables in classification tree models for the prediction of habitat requirements of *Azolla filiculoides* (Lam.) in Anzali Wetland. Iran Ecol Model 251:44–53. doi:10.1016/j.ecolmodel.2012.12.010

Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 21:660–674

Sáiz-Bárcena L, Herrero Á, Manzanedo del Campo MÁ, Martínez RDO (2015) Easing knowledge management in the power sector by means of a neuro-genetic system. Int J Bio-Inspired Comput 7:170–175

Sáiz L, Manzanedo MÁ, del Olmo R, Alcalde R (2010) Proposal of a knowledge management model for power industry. An empirical application. Dir Organ Rev Ing Organ 42:31–37

Seung HS, Socci ND, Lee D (1998) The rectified Gaussian distribution advances in neural information. Process Syst 10:350–356

Shu-Mei T (2008) The effects of information technology on knowledge management systems. Expert Syst Appl 35:150–160. doi:10.1016/j.eswa.2007.06.011

Singh RM, Gupta M (2014) Knowledge management in teams: empirical integration and development of a scale. J Knowl Manag 18:777–794. doi:10.1108/JKM-11-2013-0450

Sreerama KM (1998) Automatic construction of decision trees from data: a multi-disciplinary survey. Data Min Knowl Discov 2:345–389. doi:10.1023/a:1009744630224

Sun P (2010) Five critical knowledge management organizational themes. J Knowl Manag 14:507–523. doi:10.1108/13673271011059491

Tan LP, Wong KY (2015) Linkage between knowledge management and manufacturing performance: a structural equation modeling approach. J Knowl Manag 19:814–835. doi:10.1108/JKM-11-2014-0487

The MathWorks I, Natick, Massachusetts, United States. MATLAB (2016)

Trappey AJC, Trappey CV, Chiang T-A, Huang Y-H (2013) Ontology-based neural network for patent knowledge management in design collaboration. Int J Prod Res 51:1992–2005. doi:10.1080/00207543.2012.701775

Wang J, Ding D, Liu O, Li M (2016) A synthetic method for knowledge management performance evaluation based on triangular fuzzy number and group support systems. Appl Soft Comput 39:11–20. doi:10.1016/j.asoc.2015.09.041

Wang Z (2004) Knowledge systems engineering: a new discipline of knowledge management and enabling. Int J Knowl Syst Sci 1:9–16

Woitsch R, Karagiannis D (2002) Process-oriented knowledge management systems based on KM-services: the PROMOTE approach. Int J Intell Syst Acc Finance Manag 11:253–267

Xu J, Houssin R, Caillaud E, Gardoni M (2010) Macro process of knowledge management for continuous innovation. J Knowl Manag 14:573–591. doi:10.1108/13673271011059536

Yao Y, Wang Y, Xing L, Xu H (2015) An optimization method of technological processes to complex products using knowledge-based genetic algorithm. J Knowl Manag 19:82–94. doi:10.1108/JKM-11-2014-0454