Kano Glückstad, Fumiko ; Herlau, Tue; Schmidt, Mikkel N.; Mørup, Morten

Post print of *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on Signal Image Technology & Internet Based Systems*, Page(s): 233-240
Stable URL to publisher: http://dx.doi.org/10.1109/SITIS.2013.48

Unsupervised knowledge structuring Application of Infinite Relational Models to the FCA visualization

Fumiko Kano Glückstad
Dept. of International Business Communication
Copenhagen Business School, CBS
Frederiksberg, Denmark
Email: fkg.ibc@cbs.dk

Tue Herlau, Mikkel N. Schmidt, Morten Mørup
Section for Cognitive Systems, DTU Compute
Technical University of Denmark, DTU
Kgs Lyngby, Denmark
Email: {tuhe, mnsc, mmor}@dtu.dk

***Abstract*—This work presents a conceptual framework for learning an ontological structure of domain knowledge, which combines Jaccard similarity coefficient with the Infinite Relational Model (IRM) by (Kemp et al. 2006) and its extended model, i.e. the normal-Infinite Relational Model (n-IRM) by (Herlau et al. 2012). The proposed approach is applied to a dataset where legal concepts related to the Japanese educational system are defined by the Japanese authorities according to the International Standard Classification of Education (ISCED). Results indicate that the proposed approach effectively structures features for defining groups of concepts in several levels (i.e., concept, category, abstract category levels) from which an ontological structure is systematically visualized as a lattice graph based on the Formal Concept Analysis (FCA) by (Ganter and Wille 1997).**

***Keywords*-ontology learning; knowledge structuring; semantic representation; unsupervised machine learning; Infinite Relational Model; Formal Concept Analysis;**

I. Introduction

One approach to clarify and distinguish the meaning of different concepts within a specific domain is to characterize concepts in the form of their categories and relations to other concepts, i.e., using a domain specific ontology (concept system) [1]. In recent years, several automated clustering methods for learning a concept hierarchy (ontology) have been introduced, e.g. [2], [3], [4]. In particular, [3] employs a restructuring lattice theory, the so-called Formal Concept Analysis (FCA) [5] that is widely used as a visual data analysis tool in diverse research domains, e.g. [6] in ontology merging, [7] in multilingual concept analysis, and [8], [9] in social network structure visualization. These works employing the FCA point out that the complexity of lattice representing a hierarchical structure might cause difficulties in interpreting contextual relations in large amount of input data.

In this paper, we propose a statistical approach combining Bayesian relational models, i.e. Infinite Relational Model (IRM) [10] and its extension, the normal-Infinite Relational Model (n-IRM) [11] for structuring domain knowledge before visualizing it with the FCA. The main contribution of this work is to demonstrate the principle of this approach. Thus, we apply this approach to a simple small-sized dataset. The IRM has previously been applied in several different areas, including learning ontologies [10], analyzing text corpora [12], [13], collaborative filtering for movie recommendation [12], and analyzing neuro imaging data [14].

The approach proposed in this paper is an extension of a work presented in [15], a cross-categorization approach that maps legal concepts between two different legal systems. In this paper, we test applicabilities of the cross-categorization approach to the automatic knowledge structuring of a single dataset representing the Japanese educational system. We further investigate how this approach combined with the FCA contributes to identifying and visualizing contextual relations.

In the following we describe the data we use in our experiments and briefly review the statistical methods used in the proposed approach. Next, we evaluate our clustering method and discuss the FCA visualization of the results. We conclude by discussing some strenghts and weaknesses of the proposed approach.

II. Methods

A. Data sources

To evaluate our proposed method, we used a data set[1] from the UNESCO Institute for Statistics. In particular, we used the part of the data relating to the Japanese educational system. The data comprises 54 Japanese educational programmes. Each programme is associated with number of features such as *[ISCED level]*, *[Theoretical starting age]*, *[Theoretical duration of the programme]*, and *[Programme specifically designed for adults]*, etc. Each of these features were treated as categorical values and mapped to a 83-dimensional binary feature matrix. For example, the educational programme with the English name *Elementary school* had the feature *[Theoretical starting age]* equal to “6”, which was mapped to a binary feature *[Theoretical starting age = 6]* equal to “1”. Other educational programmes with

[1]The data is available from http://www.uis.unesco.org/education/ISCEDmappings

different starting ages would have this feature equal to "0". Based on this, we formed a 54×83 dimensional binary concept-by-feature matrix denoted $\mathbf{R}$.

B. Similarity computation

A central component in our approach is to measure the similarity between two concepts, each of which is represented by a set of binary features. To measure similarity we use the Jaccard index [16] which can be formulated as follows:

$$\mathrm{sim}(x, y) = \frac{|X \bigcap Y|}{|X \bigcup Y|}. \tag{1}$$

In the equation, x and y represent two concepts, and X and Y denote their respective binary feature sets. In words, the Jaccard index measures similarity as the ratio of the number of shared features and the total number of distinct features possessed by either concept. The Jaccard index ranges between 0 and 1, where 0 indicates two concepts with no shared features and 1 indicates two concepts with identical features. In the present paper the concepts x and y are Japanese educational programmes, and the binary feature sets represent properties of these programmes. Computing the Jaccard index between each pair of the 54 concepts represented in the concepts-by-feature matrix, we formed a 54×54 concept-by-concept similarity matrix denoted $\mathbf{W}$. Due to the symmetry of the Jaccard index, this matrix was symmetric by construction.

C. Relational Models

To analyze the two constructed data matrices, we employed two different relational models. The IRM was used to analyze the bipartite concept-by-feature matrix, and the n-IRM was used to analyze the unipartite concept-by-concept similarity matrix. In the following we review these two statistical methods.

1) Infinite Relational Model: The infinite relational model (IRM) introduced by Kemp et al. [10] is a general modeling framework that can be used to perform cluster analysis on multiple types of relational data. Here we use the IRM to compute a bipartite clustering of the binary concept-by-feature matrix. The output of the IRM is a two-mode-clustering of the concepts and the features: We let $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ denote the cluster assignments of the two modes respectively. Conditioned on the clustering, the binary data is modelled by a Bernoulli distribution with a separate success-rate parameter for each pair of clusters in mode 1 and 2: We let $\eta_{\ell m}$ denote the probability that any concept belonging to concept-cluster ℓ possesses any feature belonging to feature-cluster m. The parameters in the model are thus the two sets of cluster assignments as well as the feature probabilities. The IRM is fully specifies by assigning prior distributions over the clusters: The priors for the two cluster assignments is a so-called Chinese Restaurant Process (CRP) [10], [17]. The CRP can be thought of a defining a probability distribution over partitionings of the data, and thus the CRP provides a mechanism to automatically infer the number of clusters. As a prior over the feature probabilities, a Beta distribution is used. Thus, the generative model for the bi-partite binary IRM can be compactly written as:

$$\begin{aligned}
\mathbf{z}^{(1)} &\sim \mathrm{CRP}(\gamma^{(1)}) && \textit{clustering first mode,}\\
\mathbf{z}^{(2)} &\sim \mathrm{CRP}(\gamma^{(2)}) && \textit{clustering second mode,}\\
\eta_{\ell m} &\sim \mathrm{Beta}(\beta_0^+, \beta_0^-) && \textit{feature probabilities,}\\
R_{ij} &\sim \mathrm{Bernoulli}\left(\eta_{z_i^{(1)} z_j^{(2)}}\right) && \textit{feature assignments.}
\end{aligned}$$

To summarize the generative model we i) first partition the concepts and features into clusters, ii) next define the probability of a feature in feature-cluster m appearing in concepts in concept-cluster ℓ, and iii) finally generate the concept-feature matrix by generating each binary feature independently conditioned on the above quantities.

Both the CRP and the Beta priors have further hyper-parameters which must be chosen in order to fully specify the model. The hyper-parameter of the CRP is called the concentration parameter and influences the expected number of clusters: In our experiments using a rule of thumb, we set the concentration parameter to the logarithm of the number of entities, i.e. $\gamma^{(1)} = \log(54)$ and $\gamma^{(2)} = \log(83)$. The Beta prior influences the expected number of active features: Here we set $\beta_0^+ = \beta_0^- = 1$, specifying a uniform prior over the feature probabilities.

2) Normal-Infinite Relational Model: The Normal-Infinite Relational Model (n-IRM) [11] is a recent extension of the IRM in which the data matrix is modelled as following a Normal rather than a Bernoulli distribution. We used this to perform a cluster analysis of the concept-by-concept similarity matrix, in which each entry is a real number between zero and one describing the similarity between two concepts. The output of the cluster analysis is a cluster assignment $\mathbf{z}$ that groups the concepts. Conditioned on the clustering, the n-IRM models the observed data matrix using a Normal distribution with separate mean and precision (inverse variance) parameters for each pair of concepts. We let $m_{\ell m}$ and $\lambda_{\ell m}$ denote the mean and precision of the similarities between concepts in cluster ℓ and m. As prior distribution over the cluster assignments we use a CRP and as prior over the mean and precision parameters we use a Normal-Gamma distribution. The generative model can thus be summarized as:

$$\begin{aligned}
\mathbf{z} &\sim \mathrm{CRP}(\gamma) && \textit{clustering,}\\
\lambda_{\ell m} &\sim \mathrm{Gamma}(\alpha_0, \beta_0) && \textit{precision,}\\
m_{\ell m} &\sim \mathrm{Normal}(m_0, (\kappa_0 \lambda_{\ell m})^{-1}) && \textit{mean,}\\
W_{ij} &\sim \mathrm{Normal}(m_{z_i z_j}, \lambda_{z_i z_j}^{-1}) && \textit{similarity.}
\end{aligned}$$

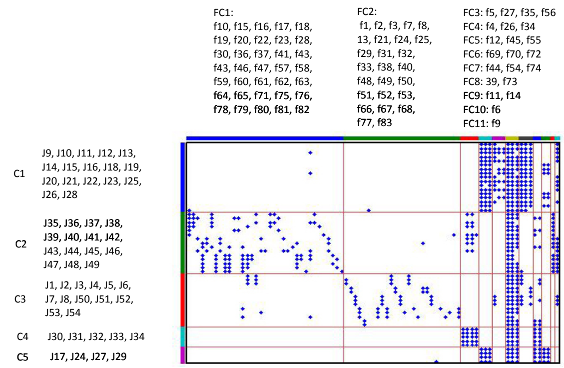
As in our IRM analysis we set the concentration parameter of the CRP to the logarithm of the number of concepts, $\gamma = \log(53)$. The hyper-parameters for the mean and precision were selected as $m_0 = 0$, $\kappa_0 = 1$, $\alpha_0 = 10$ and $\beta_0^{-1} = 10$ to yield a reasonably uninformative prior.

3) Inference by Markov chain Monte Carlo: Inference in the IRM as well as the n-IRM models entail computing (or more correctly approximating) the posterior distribution of the cluster assignment parameters. For the IRM we use the Markov chain Monte Carlo (MCMC) method described in [14] which is based on Gibbs sampling combined with a split-merge Metropolis-Hastings algorithm. In the n-IRM, we used the MCMC inference procedure described by Herlau et al. [11]. In all experiments we ran the MCMC samplers for 1 000 iterations, and discarded the first 500 realizations for burnin. The outcome of the MCMC inference is a set of posterior samples that represent the posterior distribution of the cluster assignments: In the further analysis and graphical display of the results we used the single realization that attained the highest likelihood.

III. Evaluation

Fig. 1 shows the clustering results obtained from the proposed approach combining the n-IRM and the IRM. The square plot called *"54 concepts: unsorted"* in the middle of Fig. 1 overviews similarity relations computed by the Jaccard index, i.e., corresponds to the similarity matrix $\mathbf{W}$. The plot shows similarity scores of all combination of 54 educational concepts in the gray scale where similarity scores close to 1 gets darker gray color. The other square plot *"Sorted concept clusters: C1-C7"* below the *"54 concepts: unsorted"* plot shows sorted concept clusters obtained by the n-IRM. The members of each concept clusters are listed in the table above the plots. The stabilities of the obtained concept clusters are, based on Normal Mutual Information (NMI) measure [18], quantified as 0.93 in average and 0.02 in standard deviation for 10 times run. The NMI result implies that the obtained clusters are fairly stable. By fixing these *C1-C7* as $\mathbf{z}^{(1)}$ mode, the bipartite IRM is applied to the original binary matrix $\mathbf{R}$ named as *"83 features: unsorted"*. The result of the bipartite IRM application is shown as *"Sorted feature clusters: FC1-FC12"*. The members of each feature clusters (FCs) are listed in the table at the upper-right side of Fig. 1.The NMI scores for the obtained FCs are, 0.89 in average and 0.01 in standard deviation. The three tables at the bottom of Fig. 1 show mean values and standard deviations of the Jaccard indices within each intersection between *C1-C7*, as well as density values (i.e. $\boldsymbol{\eta}$) of each intersection between *C1-C7* and *FC1-FC12*. Thresholds set for the respective type of values are highlighted in these tables.

For the comparison purpose, we applied the bipartite IRM directly to the binary matrix $\mathbf{R}$, *"83 features: unsorted"*, consisting of 54 Japanese terms and 83 features. As shown



Density	FC1	FC2	FC3	FC4	FC5	FC6	FC7	FC8	FC9	FC10	FC11
C1	0.007	0.005	0.014	0.962	0.500	0.940	0.961	0.028	0.361	0.053	0.579
C2	0.197	0.008	0.258	0.022	0.021	0.894	0.043	0.250	0.031	0.875	0.588
C3	0.022	0.145	0.019	0.024	0.098	0.951	0.268	0.286	0.037	0.067	0.267
C4	0.006	0.008	0.950	0.059	0.059	0.941	0.059	0.909	0.083	0.143	0.143
C5	0.007	0.019	0.056	0.923	0.071	0.923	0.071	0.900	0.900	0.167	0.167

Figure 2. Clustering results directly obtained by the IRM

in Fig. 2, the direct application of the bipartite IRM identifies only five concept clusters *C1-C5*. The stabilities of the obtained concept clusters computed by the NMI measure resulted in 0.87 in average and 0.02 in standard deviation for the concept clusters, and 0.81 in average and 0.01 in standard deviation for the feature clusters. This implies that the clustering result is slightly unstable compared to the clusters obtained by the n-IRM computation.

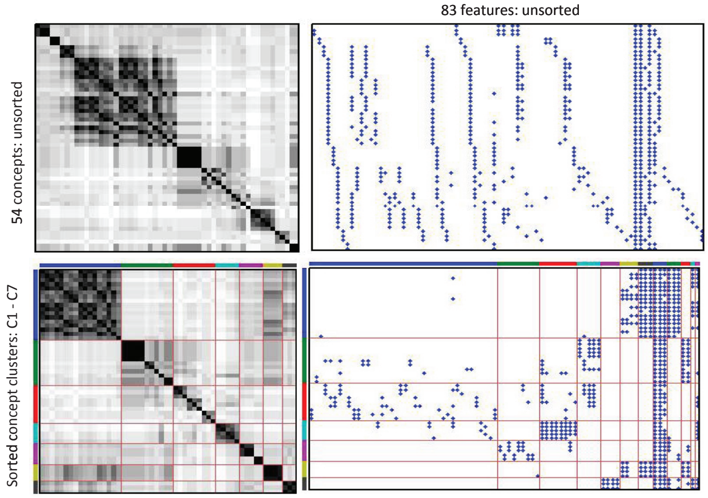
When studying details of the diagonal line in the *"Sorted concept clusters: C1-C7"* plot in Fig. 1, *C1 (upper secondary education concepts)*, *C4 (Master level university education concepts)*, *C6 (short-term upper secondary education concepts)* and *C7 (lower secondary education concepts)* clusters are colored in the darker gray, while *C2*, *C3*, and *C5* respectively consist of several smaller clusters within the concept clusters. This phenomenon can be seen in the tables where mean values of *C1*, *C4*, *C6* and *C7* are higher (intuitively, all of them are above 0.65) and their standard deviations are lower (all of them are below 0.2), while mean values of *C2*, *C3*, and *C5* are lower (below 0.55) and their standard deviatoins are higher (above 0.2). This phenomenon can be identified in the *"Sorted concept clusters: C1-C7"* plot that the concept clusters *C2*, *C3*, and *C5* respectively contain several smaller sub-clusters. The observation of the *"Sorted feature clusters: FC1-FC12"* plot further support this phenomenon that features belonging to each feature cluster related to *C2*, *C3*, and *C5* are more scattered compared to the feature clusters related to *C1*, *C4*, *C6* and *C7*. The mean table further indicate that *C1* has slightly related with *C6*, since the mean values of the intersections between *C1* and *C6* is slightly higher compared to the other intersections. This is also observable from the *"Sorted concept clusters: C1-C7"* plot.

The clustering results indicate that the proposed approach uncovers relations in several levels, e.g., cluster-cluster re-

Concept cluster	Concepts
C1	'J9:Upper secondary school, full day general course ' 'J10:Upper Secondary school, day/evening general course' 'J11:Upper secondary school, correspondence general course' 'J12:Upper secondary school, full day integrated course (general) ' 'J13:Upper secondary school, day/evening integrated course (general) ' 'J14:Upper secondary school, full day specialized course ' 'J15:Upper secondary school, day/evening specialized course' 'J16:Upper secondary school, correspondence specialised course' 'J18:Secondary education school (upper division), full day general course' 'J19:Secondary education school (upper division), day/evening general course' 'J20:Secondary education school (upper division), full day integrated course (general)' 'J21:Secondary education school (upper division), day/evening integrated course (general)' 'J22:Secondary education school (upper division), full day specialized course' 'J23:Secondary education school (Upper division), day/evening specialized course' 'J25:School for Special Needs Education, upper secondary department, general course' 'J26:School for Special Needs Education, upper secondary department, Specialized course' 'J28:College of technology, regular course '
C2	J30:Upper secondary school, (full day, day/evening), advanced (general, integrated, specialized)' J31:Secondary education (Upper division), (full day, day/evening), advanced (general, integrated, specialized)' J32:School for Special Needs Education, upper secondary department, advanced (general, specialized)' 'J33:Junior college, short-term course' 'J34:University, short-term course' 'J36:Junior college, advanced course' 'J39:College of technology, advanced course' 'J40:Specialised training college, post-secondary course' 'J43:University, advanced course' 'J53:Specialised training college, general course' 'J54:Miscellaneous schools'
C3	'J35:Junior college, regular course' 'J37:Junior college, correspondence course' 'J38:College of technology, regular course ' 'J41:University, undergraduate ' J42:University, undergraduate of pharmacy, medicine, dentistry and veterinary medicine' 'J44:University, undergraduate, correspondence course ' 'J50:University, graduate school, Doctor''s course' J51:University, graduate school, Doctor''s course of pharmacy, medicine, dentistry and veterinary medicine' 'J52:University, graduate school, Doctor''s course correspondence course'
C4	'J45:University, graduate school, Master''s course correspondence course' 'J46:University, professional graduate school, Professional course correspondence course ' 'J47:University, graduate school, Master''s course' 'J48:University, professional graduate school, Professional course' 'J49:University, professional graduate school, Graduate law school'
C5	'J1:Day nursery' 'J2:Kindergarten' 'J3:School for Special Needs Education, kindergarten department' 'J4:Elementary school' 'J5:School for Special Needs Education, elementary department'
C6	'J17:Upper secondary school, (full day/evening school), short-term course (general, integrated, specialized)' J24:Secondary education (Upper division), full day, day/evening short-term (general, integrated, specialized)' J27:School for Special Needs Education, upper secondary department short-term (general, specialized)' 'J29:Specialized training college, upper secondary course'
C7	'J6:Lower secondary school' 'J7:Secondary education school(lower division)' 'J8:School for Special Needs Education, lower secondary department'

Feature cluster	Features
FC1	'f7:ISCED 6' 'f8:ISCED NC' 'f13:programme orientation Pre vocational' 'f15:theoretical cumulative at ISCED 5: short' 'f16:theoretical cumulative at ISCED 5: medium' 'f18:qualification structure at ISCED 5: 1st degree' 'f19:qualification structure at ISCED 5: intermediate' 'f21:qualification structure at ISCED 5: 3rd degree ' 'f22:qualification structure at ISCED 5: BA' 'f24:qualification structure at ISCED 5: PhD' 'f29:Minimum entrance requirement (ISCED level or other) 5A 2nd' 'f30:Minimum entrance requirement (ISCED level or other) 5B' 'f36:starting age: 20' 'f38:starting age: 24' 'f42:duration: 2+' 'f46:duration: 4' 'f47:duration: 4+' 'f48:duration: 5' 'f49:duration: 5+' 'f57:cumulative ducatiion: 14' 'f58:cumulative ducation: 14-15' 'f59:cumulative ducation: 15+' 'f60:cumulative ducatiion: 16' 'f61:cumulative ducation: 16+' 'f62:cumulative ducation: 17+' 'f64:cumulative ducation: 18+' 'f65:cumulative ducation: 19' 'f66:cumulative ducation: 21' 'f67:cumulative ducation: 21+' 'f68:cumulative ducation: 22' 'f71:specifically designed for part-time attendance' 'f75:certification: associate degree ' 'f76:certification: associate' 'f77:certification: diploma (senmonshi)' 'f78:certification: advanced diploma (Kodo senmonshi)' 'f79:cerification_ Bachelor''s degree' 'f80:certification: Master''s degree' 'f81:certification: Professional degree' 'f82:certification: Juris Doctor' 'f83:certification: Doctor''s degree'
FC2	'f1:ISCED 0' 'f2:ISCED 1' 'f31:starting age: 3-5' 'f32:starting age: 6' 'f40:duration: 1-3' 'f43:duration: 2-3' 'f50:duration: 6' 'f51:cumulative ducatiion: 6' 'f53:cumulative ducation: 10+'
FC3	'f6:ISCED 5' 'f17:theoretical cumulative at ISCED 5: long' 'f20:qualification structure at ISCED 5: 2nd degree' 'f23:qualification structure at ISCED 5: MA' 'f28:Minimum entrance requirement (ISCED level or other) 5A 1st' 'f37:starting age: 22' 'f41:duration: 2' 'f63:cumulative duration: 18'
FC4	'f5:ISCED 4' 'f10:programme destination B' 'f27:Minimum entrance requirement (ISCED level or other) 3' 'f35:starting age: 18' 'f56:cumulative ducation: 13+'
FC5	'f3:ISCED 2' 'f25:Minimum entrance requirement (ISCED level or other) 1' 'f33:starting age: 12' 'f52:cumulative ducation: 9'
FC6	'f11:programme destination C' 'f14:programme orientation vocational' 'f45:duration 3 +' 'f55:cumulative ducation: 12+'
FC7	'f4:ISCED 3' 'f26:Minimum entrance requirement (ISCED level or other) 2' 'f34:starting age: 15'
FC8	'f69:no work based element ' 'f70: not specifically designed for adults ' 'f72:not specifically designed for part-time attendance'
FC9	'f44:duration: 3' 'f54:cumulative ducatiion: 12' 'f74:certification: certificate of graduation'
FC10	'f39:duration: 1+' 'f73:certification: certificate of completion'
FC11	'f12:programme orientation General'
FC12	'f9:programme destination A'

83 features: unsorted



Sorted feature clusters: FC1 - FC12

Mean	C1	C2	C3	C4	C5	C6	C7
C1	**0.725**	0.165	0.137	0.155	0.233	0.429	0.340
C2	0.165	**0.528**	0.222	0.172	0.232	0.343	0.203
C3	0.137	0.222	**0.278**	0.183	0.165	0.138	0.151
C4	0.155	0.172	0.183	**0.672**	0.168	0.123	0.185
C5	0.233	0.232	0.165	0.168	**0.486**	0.217	0.254
C6	0.429	0.343	0.138	0.123	0.217	**0.955**	0.187
C7	0.340	0.203	0.151	0.185	0.254	0.187	**0.889**

Density	FC1	FC2	FC3	FC4	FC5	FC6	FC7	FC8	FC9	FC10	FC11	FC12
C1	0.006	0.006	0.007	0.023	0.014	0.414	0.962	0.943	0.962	0.028	0.579	0.579
C2	0.036	0.010	0.089	0.474	0.022	0.022	0.029	0.971	0.029	0.917	0.077	0.154
C3	0.149	0.048	0.149	0.340	0.026	0.026	0.034	0.862	0.034	0.050	0.091	0.364
C4	0.064	0.021	0.929	0.037	0.045	0.045	0.059	0.824	0.118	0.083	0.143	0.857
C5	0.005	0.426	0.095	0.037	0.045	0.045	0.059	0.941	0.353	0.250	0.143	0.143
C6	0.006	0.053	0.029	0.045	0.056	0.500	0.929	0.929	0.071	0.900	0.167	0.167
C7	0.008	0.034	0.038	0.059	0.929	0.071	0.091	0.909	0.545	0.250	0.800	0.800

SD	C1	C2	C3	C4	C5	C6	C7
C1	**0.154**	0.033	0.035	0.052	0.034	0.103	0.095
C2	0.033	**0.254**	0.075	0.089	0.045	0.062	0.044
C3	0.035	0.075	**0.202**	0.063	0.043	0.030	0.035
C4	0.052	0.089	0.063	**0.131**	0.045	0.031	0.050
C5	0.034	0.045	0.043	0.045	**0.344**	0.025	0.035
C6	0.103	0.062	0.030	0.031	0.025	**0.045**	0.032
C7	0.095	0.044	0.035	0.050	0.035	0.032	**0.079**

Mean value equal or over: 0.400

Standard Deviation equal or over: 0.200

Density value equal or over: 0.400

Figure 1. Clustering results obtained by the n-IRM combined with the IRM

lations, sub-cluster relations within a concept cluster etc. as well as feature structures influencing each cluster formation. This implies that the information obtained from the proposed approach is useful for visually constructing a concept hierarchy, which is demonstrated in the next section.

IV. Visualization

As mentioned in the introduction, the FCA [5] is a convenient tool for visual data analysis. The FCA considers a relation that connects objects and features possessed by the objects. A *formal concept* of the context is defined as $C = (G, M, I)$ where G and M respectively refer to a set of objects and a set of features, and I denote relations between G and M. For example, in case of the FCA lattice *"FCA: C5 cluster internal structure"* in Fig. 4, the context *C5* is represented as G: *(J1, J2, J3, J4, J5)* and M: *(f1, f2, f31,.... f73, f74)* with sets of their individual relations I. When an object, e.g., *"J4: Elementary school"* belonging to G (expressed as $g \in G$) has the feature *"f2: ISCED1"* $(m \in M)$, this specific relation is expressed as gIm. The set of features $B \subseteq M$, e.g., *("f2: ISCED1", "f32: starting age 6", "f50: duration 6", "f51: cumulative duration 6", "f69: no work based element", "f70:not design for adults", "f72: not designed for part-time attendance", "f74: certification of graduation")* shared by a set of objects, e.g., *("J4: Elementary school", "J5: School for Special Needs, elementary")*, is defined as $\acute{A} = \{m \in M \mid gIm \ for \ all \ g \in A\}$. In the same way, the set of objects $A \subseteq G$ *("J2: Kindergarten", "J3: School for Special Needs Education, kindergarten")* shared by all features, e.g. *("f1: ISCED0", "f31: starting age:3-5", "f40: duration:1-3", "f41: duration:2-3", "f41: duration:2", "f44: duration:3", "f69: no work based element", "f70:not design for adults", "f72: not designed for part-time attendance", "f73: certification: certificate of completion")*, is expressed as $\acute{B} = \{g \in G \mid gIm \ for \ all \ m \in B\}$. [5] calls A and B as the extent and the intent of the concept (A, B) in their literature. A *formal concept* of the context (G, M, I) is represented by (A, B) defined as $A \subseteq G, B \subseteq M, \acute{A} = B, \acute{B} = A$. The set of all formal concepts in context (G, M, I) is depicted as a complete lattice called *Gallois lattice*. The algorithms for depicting a Gallois lattice are described in details in literatures in [5] as well as in [19].

Fig. 3 shows a lattice graph directly produced from the matrix $\mathbf{R}$ consisting of 54 Japanese educational terms and their 83 features, drawn by a tool called Concept Explorer (ConExp) [20] [2]. It means the context C is represented as G: *(J1, J2, ... J54)* and M: *(f1, f2, f3, ... f83)* with their relations I. Thus in the drawn lattice graph in Fig. 3, the white labels and the gray labels respectively refer to objects and features. Each node in the lattice graph is considered as *formal concept*. When a node is blue and black color, both objects and features are attached to the concept. If a node is white and black color, only objects are attached to the concept. A concept is supposed to inherit features from the upper connected edges. Thus it is possible to analyze what features are possessed by a specific concept by tracing the ascending paths in the lattice graph. As pointed out by [6], [8], [9], the drawn lattice graph in Fig. 3 is rather complex and it is difficult to overview the expressed relations.

On the other hand, the diagrams in Fig. 4 are produced based on our proposed approach, i.e., information extracted in Fig. 1 explained in the previous section. In Fig. 4, the lattice graph called *"FCA: Overall cluster structure"* is created based on the values in the density table in Fig. 1. In order to draw the lattice graph, the density values equal or above 0.4 are considered as *true*. The *"FCA: Overall cluster structure"* graph in Fig. 4 illustrates the results obtained from the n-IRM in an appropriate way that e.g. *C1 (upper secondary education concepts)*, *C7 (lower secondary education concepts)* and *C4 (Master level university education concepts)* are indirectly connected by sharing *FC12 ("f9: programme destination A")*, while *C7* and *C1* is distinguished from *C4* by commonly sharing *FC9 ("f44: duration 3", "f54: cumulative duration: 12", "f74: certification of graduation")* and *FC11 ("f12: programme orientation general")*. The difference between *C1* and *C7* are that *C1* commonly shares *FC7 ("f4: ISCED 3", "f26: minimum entrance ISCED 2", "f34: starting age 15")* and *FC6 ("f11: programme destination C", "f14: programme orientation vocational", "f45: duration 3+", "f55: cumulative duration 12+")* with *C6*, whereas *C7* possesses *FC5 ("f3: ISCED 2", "f25: minimum entrance ISCED 1", "f33: starting age 12", "f52: cumulative duration 9")*.

As indicated in the SD table in Fig. 1, *C2*, *C3*, and *C5* are relatively uneven clusters consisting of smaller sub-groups of concepts. Thus the sub-structures within each concept cluster are analyzed by applying the FCA. More specifically, the FCA is applied to sub concept-feature matrices consisting of the members of each concept cluster (*C2*, *C3*, and *C5*) and their features, where the dots scattered in the *"Sorted feature clusters: FC1-FC12"* in Fig. 1 are considered as *true* values. The three lattices, *"FCA: C2 cluster internal structure"*, *"FCA: C3 cluster internal structure"*, and *"FCA: C5 cluster internal structure"* in Fig. 4 illustrate the internal structure of each concept cluster.

For example, the *"FCA: C2 cluster internal structure"* graph in Fig. 4 shows that the objects *"J30: Upper secondary school, advanced"*, *"J31: Secondary education school, upper division, advance"*, *"J32: School for Special Needs, upper secondary, advance"* and *"J34: University, short-term"* are grouped together by sharing a feature *"f5: ISCED4"* and other features attached to its ascending edges. In the same way, *"J53: Specialized training college, general"* and *"J54: Miscellaneous school"* are grouped by sharing *"f8: ISCED-NC"* and features attached to its ascending edges. Whereas *"J36: Junior college, advanced"* is directly

[2] http://conexp.sourceforge.net/

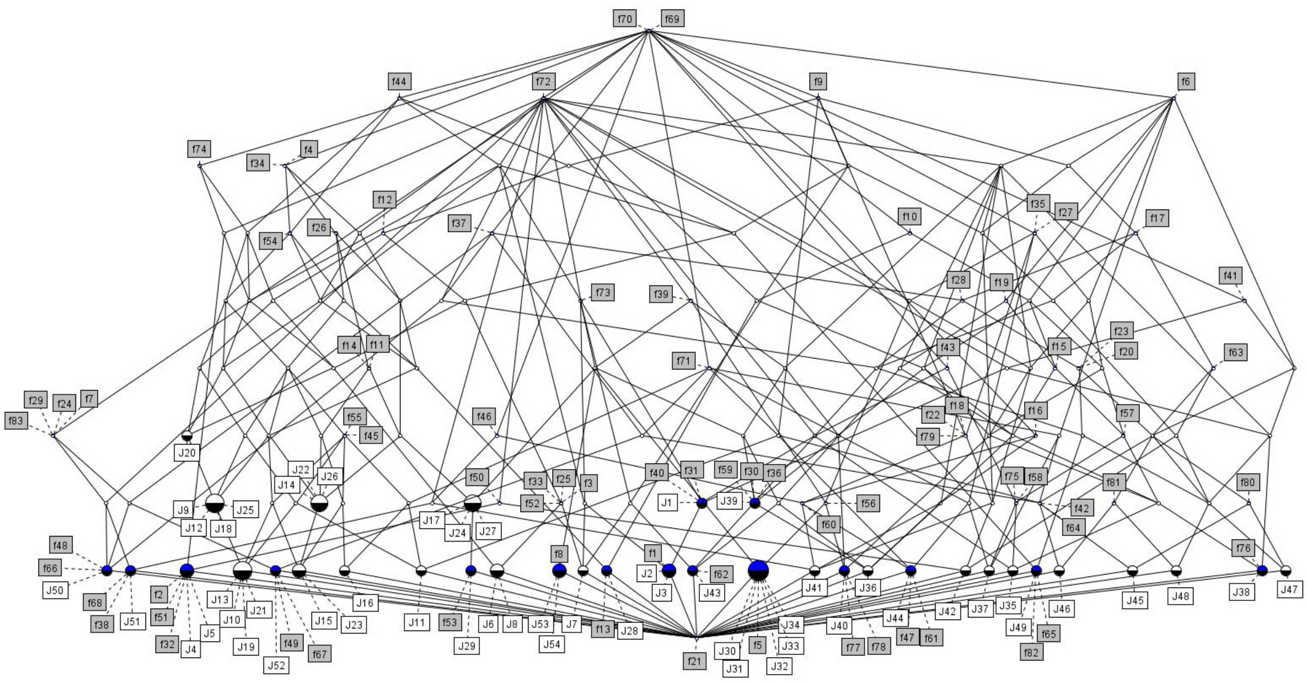


Figure 3. Lattice representing the original object-feature matrix

connected with *"J39: College of technology, advanced"* by sharing features *"f59: cumulative duration 15+"*, *"f36: starting age 20"*, *"f30: minimum entrance requirement-5B"* and their ascending features, these two concepts *J39* is distinguished from *J36* by possessing *"f10: programme destination B"*. In this way, each uneven concept cluster can be scrutinized by visually inspecting the lattice graphs. One notable point is that the sub-clusters identified in the *"Sorted concept clusters: C1-C7"* plot in Fig. 1 are sub-groups or a group of concepts that are closely connected in the lattices shown in Fig. 4.

V. Discussion

The results presented in the previous sections indicate that the concept clusters obtained by the n-IRM applied to the Jaccard index scores are fine-grained and stable compared to the results obtained from the direct application of the IRM to the concept by feature matrix $\mathbf{R}$. The *"Sorted concept clusters: C1-C7"* plot combined with the *"Sorted feature clusters: FC1-FC12"* plot in Fig. 1 overviews a knowledge structure at several levels, at the levels of individual concepts, sub-groups of concepts existing in a specific concept cluster, feature structures forming each concept cluster and more abstract concept cluster levels. These relations and structures are effectively visualized by ConExp, a publicly available automatic lattice creation tool. The lattice graphs created from the results of our proposed approach in Fig. 4 enable us to visually inspect relations between concepts and features in a more systematic manner, compared to the lattice graph directly created from the original concept-feature matrix shown in Fig. 3.

One notable point in this work is that thresholds of the density values for generating the *"FCA: Overall cluster structure"* graph are arbitrarily selected. By changing the thresholds, the shape of the lattice graph is substantially influenced. One of the challenges is how to identify appropriate thresholds. In this respect, an extended FCA such as Fuzzy Formal Concept Analysis [21] could be one of the possible alternative solutions to be considered in the future.

In this paper, we combined the n-IRM with the Jaccard index that equally consider all features possessed by an object. Another notable point is that our proposed approach accommodates any feature-based similarity measures, e.g., a similarity measure that differentiates degrees of importance of a feature possessed by an object. For example, in our previous work [15], we employed a Bayesian generalization model by [22] which considers a feature possessed by many concepts as less important, and vice versa. The clustering results obtained by the n-IRM was substantially influenced by the employed similarity measures. Thus further investigation is needed to identify what type of similarity measure is appropriate for what type of applications. Another future challenge would be to integrate the steps combining the unipartite n-IRM and the bipartite IRM. By integrating these steps, the unsupervised knowledge structuring performance is expected to be optimized. Finally, the dataset employed in this work has been simple and relatively small. Hence it is necessary to test the proposed approach with different sizes

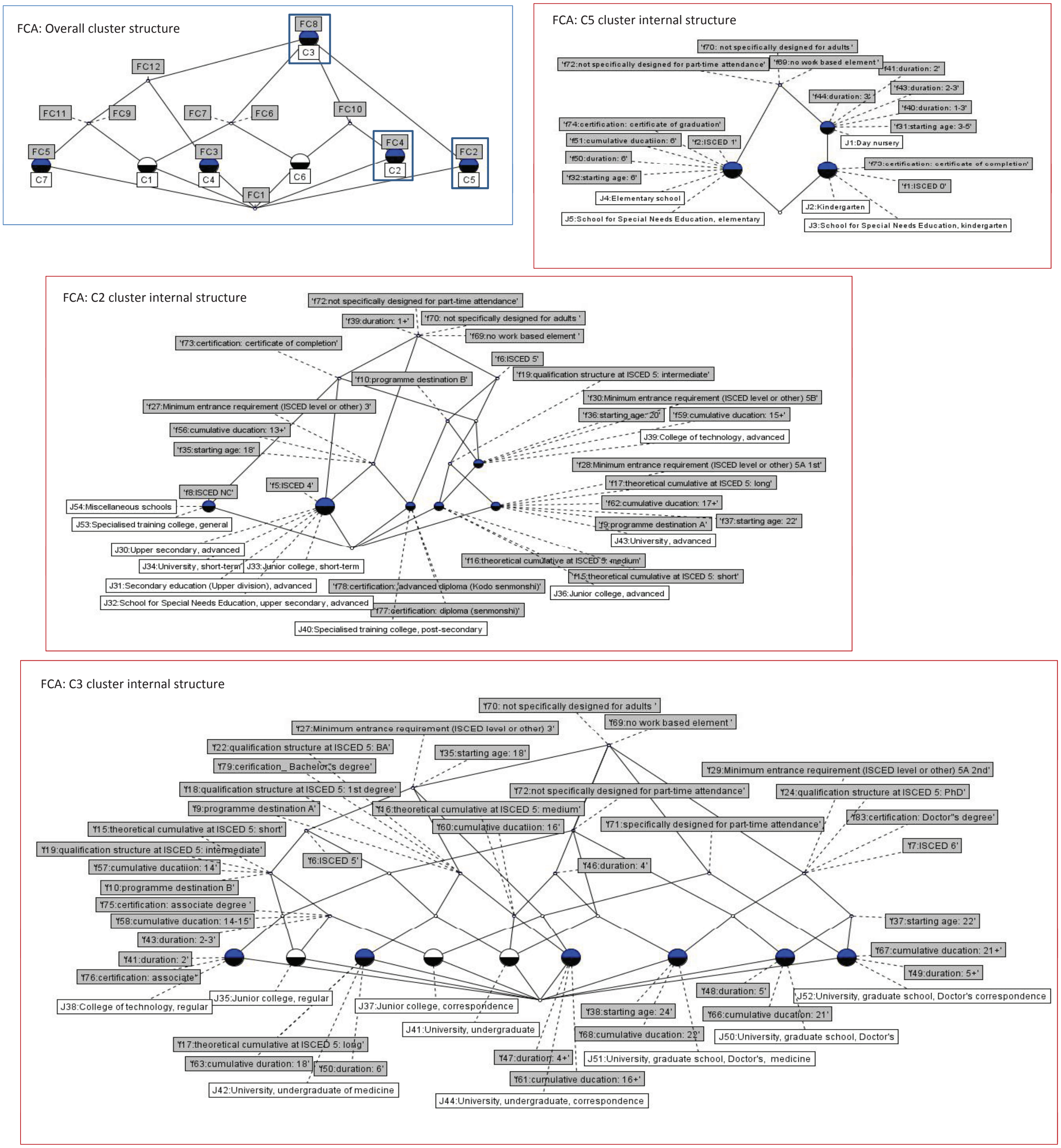


Figure 4. Lattices representing an overall structure of concept- and feature clusters and internal structures within each concept cluster

and types of datasets for validating the applicability of our approach in different scenarios in the future.

VI. Conclusions

In this paper, we presented a conceptual framework for learning a hierarchical structure of domain knowledge, which combines the Jaccard index with the IRM and n-IRM models. The results presented in this paper indicate that the proposed approach effectively clusters relations between concepts and features, and structures domain knowledge at several levels, i.e., at the levels of individual concepts, subgroups of concepts existing in a specific concept cluster, feature structures forming each concept cluster, and more abstract concept cluster levels. These structures and relations were visualized as lattice graphs created by the Formal Concept Analysis (FCA) [5]. In contrast to the direct application of the FCA to the original dataset, the proposed approach combined with the FCA contributes to effectively identify and visualize contextual relations hidden in the dataset consisting of Japanese educational concepts and their characteristic features.

References

[1] B. N. Madsen, H. E. Thomsen, J. Halskov, and T. Lassen, "Automatic ontology construction for a national term bank," in *Proceedings of Terminology and Knowledge Engineering Conference Dublin, Ireland*, 2010.

[2] L. Karoui, M. Aufaure, and N. Bennacer, "Ontology discovery from web pages: Application to tourism," in *The Workshop of Knowledge Discovery and Ontologies*, 2004.

[3] P. Cimiano, H. A., and S. S., "Learning concept hierarchies from text corpora using formal concept analysis," *CoRR*, vol. abs/1109.2140, 2011.

[4] M. Touati and M. Chavent, "Div: A divisive and symbolic clustering method," in *IXemes journees de la societe Francophone de Classification*, 2001.

[5] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, 1st ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1997.

[6] R. Ichise and H. Takeda, "Discovery of relationships among concept hierarchies using formal concept analysis," in *Proceedings of the 20th Annual Conference of the Japanese Society for Artificial Intelligence*, 2010.

[7] M. Janssen, *SIMuLLDA: A Multilingual Lexical Database Application Using a Structured Interlingua*, 2002.

[8] V. Snásel, Z. Horák, and A. Abraham, "Understanding social networks using formal concept analysis," in *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03*, 2008, pp. 390–393.

[9] S. Andrews and C. Orphanides, "Analysis of large data sets using formal concept lattices," in *Proceedings of the 7th International Conference on Concept Lattices and Their Applications (CLA) 2010*. University of Seville, 2010, pp. 104–115, iSBN 978-84614-4027-6.

[10] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinite relational model," in *The Twenty-First National Conference on Artificial Intelligence*, 2006.

[11] T. Herlau, M. Mørup, M. N. Schmidt, and L. K. Hansen, "Modelling dense relational data," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Santander, Spain*, 2012.

[12] Z. Xu, V. Tresp, K. Yu, and H. P. Kriegel, "Infinite hidden relational models," in *Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence, Cambridge, MA*, 2006.

[13] T. J. Hansen, M. Mørup, and L. K. Hansen, "Non-parametric co-clustering of large scale sparse bipartite networks on the gpu," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), IEEE*, 2011.

[14] M. Mørup, K. H. Madsen, A. M. Dogonowski, H. Siebner, and L. K. Hansen, "Infinite relational modeling of functional connectivity in resting state fmri," in *Proceedings of Neural Information Processing Systems*, 2010.

[15] F. K. Glückstad, T. Herlau, M. N. Schmidt, and M. Mørup, "Cross-categorization of legal concepts across boundaries of legal systems: in consideration of inferential links," in *Artificial Intelligence and Law*, under review.

[16] P. Jaccard, "Distribution de la flore alpine dans le bassin des dranses et dans quelques regions voisines," in *Bulletin de la societe vaudoise des sciences naturelles,vol.37*, 1901, pp. 241–272.

[17] J. Pitman, "Combinatorial stochastic processes," in *Booktitles for Saint Flour Summer School*, 2002.

[18] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification." Journal of Statistical Mechanics: Theory and Experiment, 2005.

[19] M. Liquiere, *Some Links Between Formal Concept Analysis and Graph Mining*. Wiley-Interscience, 2006.

[20] S. A. Yevtushenko, "System of data analysis "concept explorer" (in russian)," in *Proceedings of the 7th national conference on Artificial Intelligence KII-2000, Russia*, 2000, pp. 127–134.

[21] T. T. Quan, S. C. Hui, T. Hoang, C. Thanh, T. Quan, S. C. Hui, and T. H. Cao, "A fuzzy fca-based approach to conceptual clustering for automatic generation of concept hierarchy on uncertainty data," *Cybernetics and Intelligent Systems, 2004 IEEE*, pp. 578–583, 2004.

[22] J. B. Tenenbaum and T. L. Griffiths, "Generalization, similarity, and bayesian inference," in *Behavioral and Brain Sciences*, 2001, no. 4, pp. 629–640.