Agreement Contexts in Formal Concept Analysis

Richard Cole[1] and Peter Becker[1]

School of Information Technology and Electrical Engineering (ITEE)
The University of Queensland
QLD 4072, Australia
rcole@itee.uq.edu.au, pbecker@itee.uq.edu.au

Abstract. This paper describes a technique in which some columns of an *n*-ary relation are interpreted as defining a *situation*. For example when considering movies, critics and reviews we talk about the situation when the critic is a particular critic. We then construct a formal context called an *agreement context* designed to show that which is in common between the situations. We elaborate this idea in two ways: (i) by combining different types of situation; and (ii) using conceptual scales to introduce intervals of agreement.

1 Introduction

Formal concept analysis as a technique of data analysis has a long and successful history of application in various domains. Through these applications, and in concert with theoretical developments a work-flow of conceptual data analysis has been established that may be summarised as: (i) collect high quality, meaningful input data and organise it as a many-valued context, (ii) create a collection of conceptual scales with which to interpret the values in the many-valued context, and (iii) explore the landscape created by the application of these conceptual scales to the data in order to form and test hypotheses about the data and retrieve specific instances.

Classic conceptual data analysis requires that the input data can be organised as a many valued context. This assumes a functional dependency: a value has to be unique for any given object-attribute pair. Such a context can be represented in a form similar to a spreadsheet — each object is assigned a row, each attribute a column, and each cell may contain at most a single value.

Much of the data that we would wish to analyse though exists within relational database in a form that cannot be readily converted into the form of a many-valued context since either the functional dependency is lacking or the relation to be analysed is of higher order than three. A number of approaches have been defined that extend FCA to handle higher order relational data in various ways. Notable examples include: a triadic approach to concept analysis [WL95], multi-contexts [Wil96] and power context families[Wil97], relational graphs [Wil02], and reverse pivoting [HS01].

This paper provides a quite different approach to the analysis of high order relations by considering some columns of the relation a defining a *situation*. By situation we mean that part of the relation defines a context in which something can be true, while not necessarily being true in general. For example by considering a movie critic as defining a situation, we can ask within that situation if a particular movie is considered good

Movie	**Critic**	**Rating**
The League Of Extraordinary Gentlemen	Margaret	2.5
The League Of Extraordinary Gentlemen	David	1
Matchstick Men	Margaret	4.5
Matchstick Men	David	4.5
The Wannabes	Margaret	2
The Wannabes	David	3.5
Swimming Pool	Margaret	4
Swimming Pool	David	3.5
American Splendor	Margaret	4.5
American Splendor	David	4
28 Days Later	Margaret	4
28 Days Later	David	3
Bad Boys II	Margaret	2
Bad Boys II	David	3
Japanese Story	Margaret	4.5
Japanese Story	David	4.5

Table 1. Movie ratings taken from the “SBS Movie Show” by two reviewers, Margaret Pomeranz and David Stratton, organised as a ternary relation R_{ms}

or not. This notion of situation is similar in some ways to a situation within situation theory [Bar89], or a local logic within the logic of distributed systems [BS97].

Throughout the paper we will use ratings of movies by two reviewers, Margaret Pomeranz and David Stratton, from a TV show called “The Movie Show”[1]. Table 1 shows the eight movies we will use in our first example and refer to as R_{ms}, other examples use an extended set of 35 movies whose reviews were conducted around the same time which we refer to as R_{ml}.

The rest of the paper is structured as follows: Section 2 introduces the notion of an *agreement context* and demonstrates its use via an example. using the data from Table 1. Section 3 explains how the notion of a agreement contexts may be combined with conceptual scaling, first as a mechanism to scale the output of the agreement context and then secondly as a way to scale relation before looking for agreement. Section 4 provides a conclusion and outlook.

2 Agreement Contexts

In order to understand the basic intuition behind agreement contexts, consider the columns *movie*, *critic* and *rating* of the example introduced in Section 1. The data contained in these columns can be seen as a relation, $R \subseteq D_1 \times D_2 \times D_3$, in which the dimensions, $D_1, \ldots, D_3$ cover movies, critics, and ratings respectively.

Now consider what is in common between the ratings assigned by a group of critics, $S \subseteq D_2$. A movie rating pair $(m, r) \in D_1 \times D_3$ is in common between critics $s_1, s_2 \in$

[1] `http://www.sbs.com.au/movieshow/`

S if $(m, s_1, r) \in R$ and $(m, s_2, r) \in R$. The following approach considers each critic as defining a situation and then generates a concept lattice containing the agreement between each subset of the defined situations.

First we introduce two relational projections:

Definition 1. *Given an n-ary relation R, let π_i^n and $\overline{\pi}_i^n$ be operations on tuples of R defined as:*

$$\pi_i^n(g_1, \ldots, g_n) = (g_i)$$
$$\overline{\pi}_i^n(g_1, \ldots, g_n) = (g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n)$$

Likewise define π_i^n and $\overline{\pi}_i^n$ as operations on n-ary relation given by:

$$\pi_i^n R = \{\pi_i^n g \mid g \in R\} \quad \textit{and} \quad \overline{\pi}_i^n R = \{\overline{\pi}_i^n g \mid g \in R\}$$

Given these definitions we can generalize the idea above to create a formal context to express agreement.

Definition 2. *For any relation R, we define $\mathbb{K}_{R,i}$ as:*

$$\mathbb{K}_{R,i} := (\overline{\pi}_i^n R, \pi_i^n R, I_i)$$

where

$$(g, m) \in I_i \;:\Leftrightarrow\; \exists t \in R : g = \overline{\pi}_i^n t \;\textit{ and }\; m = \pi_i^n t$$

This structure is not suited for apposition, an operation on contexts needed to apply nesting and zooming as used in programs such as ToscanaJ [BH04] and CEM [CS00]. Given $\mathbb{K}_i = (G, M_i, I_i)$, for $i \in 1, 2$ the apposition, $\mathbb{K}_1|\mathbb{K}_2$, is defined [GW99] as:

$$\mathbb{K}_1 \mid \mathbb{K}_2 := (G, \dot{M}_1 \cup \dot{M}_2, \dot{I}_1 \cup \dot{I}_2)$$

where $\dot{M}_i = \{i\} \times M_i$ and $\dot{I}_i = \{(g, (i, m)) \in G \times M_i \mid (g, m) \in I_i\}$ for $i \in 1, 2$.

In order to construct an apposition, both contexts must share a common object set, and this does not hold for the family of contexts $(\mathbb{K}_{R,i})_{i=1,\ldots,n}$. To obtain a common object set we introduce the following definition which uses the tuples of R as objects:

Definition 3. *The i*-th agreement context *of the relation R is the formal context $\mathcal{M}(R, i) := (R, \pi_i^n R, \nabla_i R)$ where*

$$\nabla_i R := \{(g, m) \in R \times \pi_i^n R \mid \exists h \in R : \overline{\pi}_i^n h = \overline{\pi}_i^n g \;\textit{ and }\; \pi_i^n h = m\}$$

The concept lattice of the agreement context $\mathcal{M}(R, i)$ is isomorphic to the concept lattice of the context $\mathbb{K}_{R,i}$ since $\overline{\pi}_i^n$ is an intent preserving surjective mapping from the objects of $\mathcal{M}(R, i)$ to the objects of $\mathbb{K}_{R,i}$. Since the object set for all agreement contexts of one relation is the same, apposition can be applied and thus nesting and zooming is possible.

Example: A agreement context was generated from the data in Table 1 by considering the critic field as a situation and the concept lattice of this context is shown in Fig. 1.

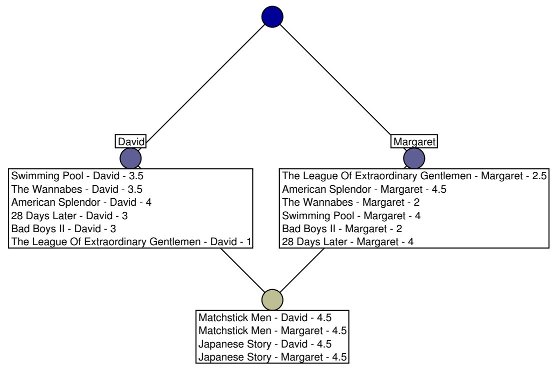

Fig. 1. Lattice for the agreement context $\mathcal{M}(R_{ms}, 2)$

The concept lattice contains four concepts, the bottom concept has as extent tuples in which the ratings are common between Margaret and David.

The extent for the concept labeled Margaret contains (i) those reviews particular to Margaret, (ii) the reviews of Margaret that match a review of David, and (iii) the reviews of David that match a review of Margaret. The concept labeled David similarly contains those reviews that match Margaret's, those of Margaret that match one of David's, and those particular to David.

Two entries occurs for each movie, one for each critic in Fig. 1. As we explained earlier it is possible to recover $\mathbb{K}_{R,i}$ by applying the projection $\overline{\pi}_i^n$ to each extent. In our example this corresponds to removing the critics from the object labels and then removing duplicate labels. If this were done the bottom concept would contain just "Matchstick Men - 4.5" and "Japanese Story - 4.5".

3 Agreement Contexts and Conceptual Scaling

Conceptual scaling means defining a formal context by which values from some domain may be interpreted. As an example consider the commonly used interordinal scale, I_n. This scale is used to interpret numerical values in a certain range by constructing intervals between n selected boundaries.

Many-valued contexts are defined in [GW89] as a tuple (G, M, W, I) where G, M, and W are sets, and $I \subseteq G \times M \times W$ is a relation such that w is functionally dependent on g and m, i.e. $(g, m, w_1), (g, m, w_2) \in I$ implies $w_1 = w_2$. A conceptual scale, for an attribute m is a formal context (G_s, M_s, I_s) with $G_s \supseteq \{w \in W \mid (g, m) \in I\}$. The conceptual scale can be applied to the many valued context to derive a formal context (G, M_s, J), where $(g, m) \in J \;:\Leftrightarrow\; \exists w \in W : (g, m, w) \in I, (w, m) \in I_s$.

In the following we will describe two approaches of combining conceptual scaling with agreement contexts: first using conceptual scaling as means of visualising the agreement extents and secondly applying conceptual scaling to a relation before the agreement operations to find agreement induced by the scale.

3.1 Applying conceptual scales to agreement extents

Definition 4. *Given a relation* $R \subseteq D_1 \times \ldots \times D_n$*, and a conceptual scale* $\mathbb{S} := (G_s, M_s, I_s)$*, with* $G_s \supseteq D_i$ *the* context derived with respect to agreement scaling *is* (R, M_s, J) *where*

$$(g, m) \in J :\Leftrightarrow \exists w \in D_i : (g, w) \in \nabla_i R, \text{ and } (w, m) \in I_s$$

Standard conceptual scaling relies on the functional dependency of the many-valued context in order to achieve the result that the context derived with respect to plain scaling can be ∨-embedded into the scale lattice. Definition 4 drops this requirement and thus we need to make use of a construction introduced by Stumme and Hereth [HS01], namely the power scale.

Definition 5. *A* power scale *of a conceptual scale* (G_s, M_s, I_s)*, is the formal context* $(\mathfrak{P}(G_s), M_s, I_s^{\mathfrak{P}})$*, where*

$$(A, m) \in I_s^{\mathfrak{P}} :\Leftrightarrow \exists g \in A : (g, m) \in I_s$$

for $A \subseteq G_s$*, and* $m \in M$.

The concept lattice of a context derived with respect to agreement scaling from a relation R and a conceptual scale $\mathbb{S}$ can be ∨-embedded into the concept lattice of the power scale of $\mathbb{S}$, and if $\nabla_i R$ defines a functional relationship then the derived context can be ∨-embedded into $\underline{\mathfrak{B}}(\mathbb{S})$, thus agreement scaling can be consistently combined with conceptual scaling in order to provide an interpretation for the domain to which it is applied.

Example: The example in Fig. 2 is derived from our "SBS Movie Show" example using agreement scaling with two conceptual scales. The outer scale is a boolean scale for Margaret and David, while the inner scale an interordinal scale I_6. Since the rating is functionally dependent on the movie and the reviewer we can use I_6 rather than having to resort to the power scale of I_6.

The bottom concept of the outer scale contains the movies whose ratings were in common between the two reviewers. For instance both reviewers agreed that "Legally Blonde 2" should have a rating of 2 stars. Similarly both critics agreed that "Matchstick Men" should deserved a review of 4.5 stars. A limitation of this approach is that the bottom outer concept contains only movie ratings that are in exact agreement between the two critics. This can be seen by considering the concepts for Margaret and David. Both critics gave low scores to Take Away — David gave it 1 while Margaret gave it 0.5. Although these ratings are close they do not result in an agreement in this structure. The approach outlined in the next section attempts to remedy this.

3.2 Finding the agreement in a conceptual scale

The approach presented so far shows exact agreement and does not account for how similar two ratings may be. In the example used one might ask a question like: "What is the range of ratings given for a particular movie?". This can be achieved by first applying the conceptual scale to a relation and then creating the agreement context.

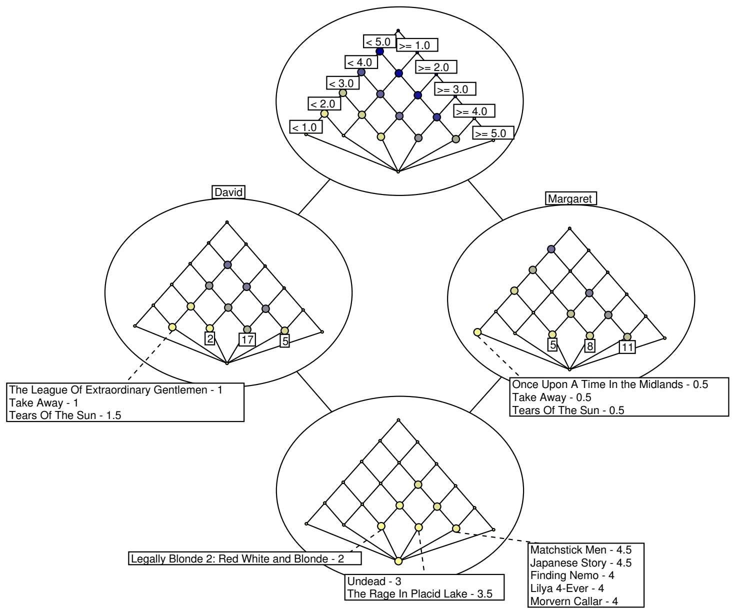

Fig. 2. A nesting of two conceptual scales of the B_2 and I_6 type, applied via agreement scaling to the second and third column of R_{ml}. The object labels contain contingents projected through $\overline{\pi}^3_2$.

Definition 6. *Given an n-ary relation* $R \subseteq D_1 \times \ldots \times D_n$ *and a conceptual scale* $\mathbb{S} := (G_s, M_s, I_s)$*, with* $G_s \supseteq D_i$ *the* relation derived with respect to plain scaling *is:*

$$R \otimes_i \mathbb{S} := \left\{ g \in \times_{k=1}^{i-1} D_k \times M_s \times_{k=i+1}^{n} D_k \middle| \exists h \in R : \overline{\pi}^i_n h = \overline{\pi}^i_n g \text{ and } (\pi^i_n h, \pi^i_n g) \in I_s \right\}$$

In terms of relational algebra this is a join of R with I_s: the i-th column of the relation gets replaced with the attributes of the conceptual scale. Creating agreement contexts for this relation finds agreement based on the attributes in the conceptual scale, rather than exact matches within the column begin scaled. By using different conceptual scales the commonality can be modeled with different granularities and dimensions.

Example: Fig. 3 shows the commonality between Margaret and David using the interordinal scale from the example above, but this time scaling is performed prior to constructing the agreement context.

The first step to create this diagram is to conceptually scale R_{ml} by using the same I_6 we used before. From this scaled relation we construct the agreement context and select the bottom concept whose extent corresponds to the reviews in common between Margaret and David. We generate the 3-rd agreement context which corresponds to considering the scale attributes as defining an agreement context. The resulting concept lattice is shown in Figure 3 with the following modification: We project out both the rating and the critic from the concept extents. This modification is valid since the critics are already determined to be both Margaret and David for each movie rating pair, and the ratings for each movie are specified by the concept intent.

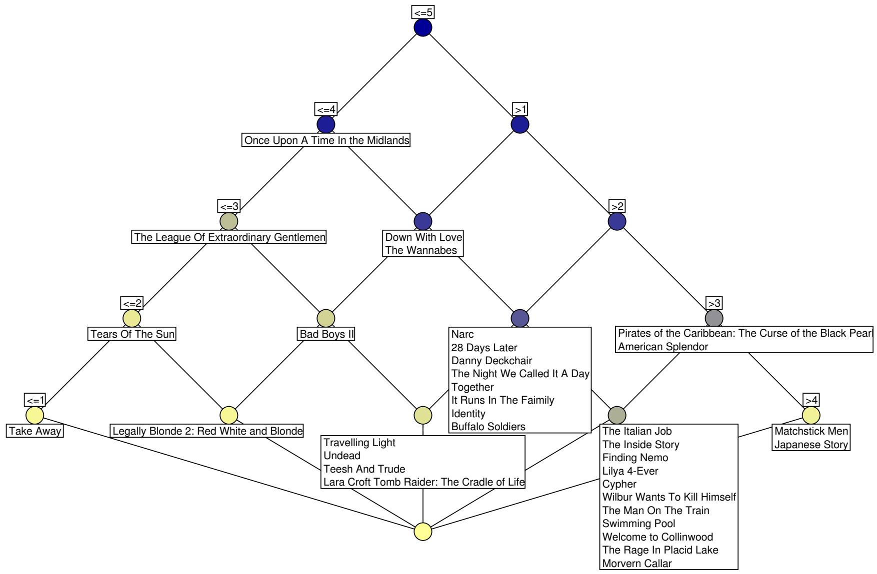

Fig. 3. Ratings in common between Margaret's and David's generated from the extent of the bottom concept of $\underline{\mathfrak{B}}(\mathcal{M}(R_{ml} \otimes_3 I_6, 2))$. From the extent of this bottom concept an agreement context with respect to third column — the scale attributes ≤ 1, ≤ 2 etc. — has been constructed. It's concept lattice is shown with extents projected to contain only movie titles.

Note that the result is structurally an I_5 even though the scaling was done with an I_6. This occurred because the lattice was generated using a data driven process where unrealised concepts are not drawn.

The diagram in Fig. 3 gives a good overview of the interval of agreement between the two reviewers for each movie. The movies at the widest layer of the lattice correspond to total agreement between the reviewers within the given scale. For movies in this layer the reviews can vary by at most in half a point since the narrowest interval in the scale is one point wide. The higher a movie is in this diagram, the less agreement there is. "Once Upon A Time In The Midlands" has the least agreement, the two critics agree that its rating is less than or equal to 4 points, but nothing more — Margaret gave half a point, David three and a half.

4 Conclusion and Outlook

For the purposes of experimenting with the approach we wrote a prototype called Tupleware[2]. This program allows to import relational data from different sources, such as text files, relational databases, RDF files or command line calls to other storage systems. The results can be exported to ToscanaJ and thus nesting and zooming are supported in a limited fashion.

[2] `http://tockit.sourceforge.net/tupleware`

We have given an approach of interpreting dimensions of a relation as a situation and then exploring the agreement between these situations using Formal Concept Analysis. With the prototype we implemented we showed that the approach is suited to the analysis of agreement between two movie critics and that the resulting diagrams visualize different aspects of the agreement in an intuitive fashion.

The work outlined in this paper can be extended in different ways. In terms of problem domains more problems can be approached, such as comparing results from different sensors or analysing larger data sets with more or wider dimensions. In terms of the mathematisation the question of how functional dependencies between columns of the relation affect the suitability of columns to be selected as defining situations is yet to be fully explored. In terms of tool support the current prototype is still quite limited in its support for the features presented in this paper and in terms of user guidance.

References

[Bar89] J. Barwise. *The situation in logic*. Center for the Study of Language and Information, Stanford, CA, 1989.

[BH04] P. Becker and J. Hereth Correia. The ToscanaJ suite for implementing Conceptual Information Systems. In *Formal Concept Analysis – State of the Art*, Berlin – Heidelberg – New York, 2004. Springer. To appear.

[BS97] J. Barwise and J. Seligman. *Information flow : the logic of distributed systems*. Cambridge University Press, Cambridge ; New York, 1997.

[CS00] R. Cole and G. Stumme. CEM: A Conceptual Email Manager. In B. Ganter and G. Mineau, editors, *Conceptual Structures: Logical, Linguistic, and Computational Issues*, number 1867 in LNAI, pages 438–452. Springer Verlag, Berlin – Heidelberg – New York, 2000.

[GW89] B. Ganter and R. Wille. Conceptual scaling. In F. Roberts, editor, *Applications of combinatorics and graph theory to the biological and social sciences*, pages 139–167. Springer-Verlag, New York, 1989.

[GW99] B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, 1999.

[HS01] J. Hereth and G. Stumme. Reverse pivoting in conceptual information systems. In Harry S. Delugach and Gerd Stumme, editors, *Conceptual Structures: Broadening the Base*, volume 2120 of *LNAI*, pages 202–215. Springer, Berlin – Heidelberg – New York, 2001.

[Wil96] R. Wille. Conceptual structures of multi-contexts. In P. Eklund, G. Ellis, and G. Mann, editors, *Conceptual Structures: Knowledge Representation as Interlingua*, number 1114 in LNAI, pages 23–39. Springer Verlag, Berlin, 1996.

[Wil97] R. Wille. Conceptual graphs and formal concept analysis. In D. Lukose, H. Delugach, M. Keeler, L. Searle, and J. F. Sowa, editors, *Conceptual structures: Fulfilling Peirce's dream.*, volume 1257 of *LNAI*, pages 317–331. Springer, Berlin – Heidelberg – New York, 1997.

[Wil02] R. Wille. Relational graphs: A structure for representing relations. In U. Priss, D. Corbett, and G. Anagelova, editors, *Conceptual Structures: Integration and Interfaces*, volume 2393 of *LNAI*, pages 34–47. Springer, Berlin – Heidelberg – New York, 2002.

[WL95] R. Wille and F. Lehmann. A triadic approach to formal concept analysis. In Gerard Ellis, R. Levinson, W. Rich, and J. Sowa, editors, *Conceptual structures: applications, implementation and theory*, volume 954 of *LNAI*, pages 32–43. Springer, Berlin – Heidelberg – New York, 1995.