A Mass Assignment Approach to Granular Association Rules for Multiple Taxonomies

Trevor Martin[1,2], Yun Shen[1] and Ben Azvine[2]

[1] AI Group, University of Bristol, BS8 1TR UK
[2] Intelligent Systems Lab, BT, Adastral Park, Ipswich IP5 3RE, UK
{Trevor.Martin, Yun.Shen}@bristol.ac.uk, Ben.Azvine@bt.com

Abstract. The use of hierarchical taxonomies to organise information (or sets of objects) is a common approach for the semantic web and elsewhere, and is based on progressively finer granulations of objects. In many cases, seemingly crisp granulation disguises the fact that categories are based on loosely defined concepts which are better modelled by allowing graded membership. A related problem arises when different taxonomies are used, with different structures, as the integration process may also lead to fuzzy categories. Care is needed when information systems use fuzzy sets to model graded membership in categories - the fuzzy sets are not disjunctive possibility distributions, but must be interpreted conjunctively. We clarify this distinction and show how an extended mass assignment framework can be used to extract relations between fuzzy categories. These relations are association rules and are useful when integrating multiple information sources categorised according to different hierarchies. Our association rules do not suffer from problems associated with use of fuzzy cardinalities. An example of discovering associated film genres is given.

Keywords: fuzzy, granules, association rules, hierarchies, mass assignments, semantic web, iPHI

1 1 Introduction

The use of taxonomic hierarchies to organise information and sets of objects into manageable chunks (granules) is widespread. Granules were informally defined by Zadeh [1] as a way of decomposing a whole into parts, generally in a hierarchical way. We can regard a hierarchical categorisation as a series of progressively finer granulations, allowing us to represent problems at the appropriate level of granularity.

The idea of a taxonomy serves as an organisational principle for libraries, for document repositories, for corporate structure, for the grouping of species and very many other applications. It is therefore no surprise to note that the semantic web adopts hierarchical taxonomies as a fundamental structure, using the *subClassOf* construct. Although in principle the idea of a taxonomic hierarchy is crisply defined, in practice there is often a degree of arbitrariness in its definition. For example, we might divide the countries of the world by continent at the top level of a taxonomic hierarchy. However, continents do not have crisp definitions - Europe contains some definite members (e.g. France, Germany) but at the Eastern and South-Eastern border, the question of which countries belong / do not belong is less clear. Iceland is generally included in Europe despite being physically closer to Greenland (part of North America). Thus although the word "Europe" denotes a set of countries (i.e. it is a granule) and can be used as the basis for communication between humans, it does not have an unambiguous definition in terms of the elements that belong to the set. Different "authorities" adopt different definitions - the set of countries eligible to enter European football competitions differs from the set of countries eligible to enter the Eurovision song contest, for example.

Of course, mathematical and some legal taxonomic structures are generally very precisely defined - the class of polyhedra further subdivides into triangles, quadrilaterals, etc and triangles may be subdivided into equilateral, isosceles etc. Such definitions admit no uncertainty. Most information systems model the world in some way, and need to represent categories which correspond to the loosely defined classes used by humans in natural language. For example, a company may wish to divide adults into customers and non-customers, and then sub-divide these into high-value customers, dissatisfied customers, potential customers, etc. Such categories are not necessarily distinct (i.e. they may be a covering rather than a partition) but more importantly, membership in these categories is graded - customer *X* may be highly dissatisfied and about to find a new supplier whilst customer *Y* is only mildly dissatisfied. We argue that most hierarchical taxonomies involve graded or loosely defined categories, but the nature of computerised information systems means that a more-or-less arbitrary decision has to be made on borderline cases, giving the taxonomy the appearance of a crisp, well-defined hierarchy. This may not be a problem as long as a rigorous and consistent criterion for membership is used (e.g. a dissatisfied customer is defined as one who has made at least two calls complaining about service), but the lack of subjectivity in a definition is rare. The use of graded membership (fuzziness) in categories enhances their expressive power and usefulness.

A related problem arises when trying to combine multiple sources of information that have been categorised in some way (often hierarchically). For example, the category of "vintage wine" has a different (but objective) definition, depending on the country of origin. To a purist, vintage wines are made from grapes harvested in a single year – however, the European Union allows up to 5% of the grapes to be harvested in a different year, the USA allows 15% in some cases and 5% in others, while other countries such as Chile and South Africa may allow up to 25%. Thus even taking a simple (crisp) granulation of wines into vintage and non-vintage categories can lead to problems if we try to integrate different sources.

In this paper we describe a new method for calculating association rules to find correspondences between fuzzy granules in different hierarchies (with the same underlying universe). We discuss the semantics of fuzzy sets when used to describe granules, and introduce a mass assignment-based method to rank association rules and show that the new method gives more satisfactory results than approaches based on fuzzy cardinalities. Ongoing work is focused on comparison of this approach to others (e.g. on ontology merging benchmarks), and with application to merging classified directory content.

2 Background

This work take place in the context of the iPHI system (intelligent Personal Hierarchies for Information) [2] which aims to combine and integrate multiple sources of information and to configure access to the information based on an individual's personal categories. We assume here that the underlying entities (instances) that are being categorised are known unambiguously - when integrating multiple sources, this is often not the case. We have outlined SOFT (the Structured Object Fusion Toolkit) elsewhere [3] as one solution to this problem.

2.1 Fuzzy Sets in Information Systems

Many authors (e.g. [4]) have proposed the use of fuzzy sets to model uncertain values in databases and other knowledge based applications . The standard interpretation of a fuzzy set in this context is as a *possibility distribution* - that is to say it represents a single valued attribute which is not known exactly. For example we might use the fuzzy set *tall* to represent the height of a specific person or *low* to represent the value shown on a dice. The fuzzy sets *tall* and *low* admit a range of values, to a greater or lesser degree; the actual value is taken from the range. Knowing that a dice value *val* is *even* restricts the possible values to *val=2 XOR val=4 XOR val=6* (where *XOR* is an exclusive or). If a fuzzy set on the same universe is defined as *low = {1/1, 2/1, 3/0.4}* then knowing the value *val* is *low* restricts the possible values to *val=1 XOR val=2 XOR val=3* with corresponding memberships.

The conjunctive interpretation of a fuzzy set occurs when the attribute can have multiple values. For example, a person may be able to speak several languages; we could model this as a fuzzy set of languages, where membership would depend on the degree of fluency. This is formally a relation rather than a function on the underlying sets. Our position is to make a distinction between the conjunctive interpretation - modelled by a fuzzy relation – and the disjunctive interpretation – modelled by a possibility distribution. To emphasise the distinction, we use the notation

$$F(a) = \{x/\mu(x) \mid x \in U\}$$

to denote a single valued attribute F of some object a (i.e. a possibility distribution over a universe U) and

$$R(a) = [x/\chi(x) \mid x \in U]$$

to denote a multi-valued attribute (relation). Granules represent the latter case, since we have multiple values that satisfy the predicate to a greater or lesser degree.

2.2 Association Rules

In creating association rules within transaction databases (e.g. [5], see also [7] for a clear overview), the standard approach is to consider a table in which columns correspond to items and each row is a transaction. A column contains 1 if the item was bought, and 0 otherwise. The aim of association rule mining is to determine whether or not there are links between two disjoint subsets of items – for example, do customers generally buy biscuits and cheese when beer, lager and wine are bought?

Let X denote the set of items, so that any transaction can be represented as $tr \subseteq X$ and we have a multiset Tr of transactions. We must also specify two non-overlapping subsets of X, s and t. An association rule is of the form S => T where S (resp T) is the set of transactions containing the items s (resp t). The rule is interpreted as stating that when the items in s appear in a transaction, it is likely that the items in t will also appear i.e. it is not an implication in the formal logical sense.

Most authors use two measures to assess the significance of association rules, although these measures can be misleading in some circumstances. The support of a rule is the fraction of transactions in which both S and T appear, and the confidence of a rule is an estimate (based on the samples) of the conditional probability of T given S

$$Support(S, T) = |S \cap T|$$

and

$$Conf(S, T) = \frac{|S \cap T|}{|S|}$$

where we operate on multisets rather than sets. Typically a threshold is chosen for the support, so that only frequently occurring sets of items s and t are considered; a second threshold filters out rules of low confidence.

Various approaches to fuzzifying association rules have been proposed e.g. [6-8]. The standard extension to the fuzzy case is to treat the (multi-) sets S, T as fuzzy and find the intersection and cardinality using a t-norm and sigma-count respectively.

$$Conf(S,T) = \frac{\sum_{x \in X} \mu_{S \cap T}(x)}{\sum_{x \in X} \mu_S(x)}$$

Note that many authors just refer to fuzzy sets, rather than multisets.

As pointed out by [7], using min and the sigma count for cardinality can be unsatisfactory because it does not distinguish between several tuples with low memberships and few tuples with high memberships - for example,

$$S = [x_1/1, x_2/0.01,\ x_3/0.01, \ldots, x_{1000}/0.01\ ]$$
$$T = [x_1/0.01, x_2/1,\ x_3/0.01, \ldots, x_{1000}/0.01\ ]$$

leads to

$$Conf(S,T) = \frac{1000 \times 0.01}{1 + 999 \times 0.01} \approx 0.91$$

which is extremely high for two almost disjoint sets (this example originally appeared in [9]). Using a fuzzy cardinality (i.e. a fuzzy set over the possible cardinality values) is also potentially problematic.

For these reasons, we propose the use of mass assignment theory in calculating the support and confidence of association rules between fuzzy categories.

The fuzziness in our approach arises because we allow partial membership in categories – for example, instead of looking for an association between biscuits and beer, we might look for an association between *alcoholic drinks* and *snack foods*. It is important to note that we are dealing with conjunctive fuzzy sets (monadic fuzzy relations) here. Mass assignment theory is normally applied to fuzzy sets representing possibility distributions and the operation of finding the conditional probability of one fuzzy sets given another is known as semantic unification [10]. This rests on the underlying assumption of a single valued attribute – a different approach is required to find the conditional probability when we are dealing with set-valued attributes.

2.3 Mass Assignments

A mass assignment [11] (see also [12]) is a distribution over a power set, representing disjunctive uncertainty about a value. For a universe U

$$m : P(U) \to [0,1]$$
$$\sum_{X \subseteq U} m(X) = 1 \qquad (1)$$

The mass assignment is related to a fuzzy set (possibility distribution) A as follows:
Let μ_A be the membership function of A with range

$R(\mu_A) = \{\mu_A^1, \mu_A^2, \ldots, \mu_A^m\}$

such that $\mu_A^1 > \mu_A^2 > \ldots > \mu_A^m$

and Ai be the alpha-cuts at these values i.e.

$A_i = \{x | \mu_A(x) \geq \mu_A^i\}$

(also known as the focal elements)
Then

$$m_A(A_i) = \mu_A^i - \mu_A^{i+1} \qquad (2)$$

Given a fuzzy set A, the corresponding mass assignment can be written as

$M(A) = \{A_i : m_A(A_i) \mid A_i \subseteq A\}$

where conventionally only the focal elements (non-zero masses) are listed in the mass assignment. The mass assignment represents a family of probability distributions on U, with the restrictions

$$p : U \to [0,1]$$
$$\sum_{x \in U} p(x) = 1 \qquad (3)$$
$$m(\{x\}) \leq p(x) \leq \sum_{x \in X} m(X)$$

For example, if $X = \{a, b, c, d\}$ and A is the fuzzy set

{a/1, b/0.8, c/0.3, d/0.2}

then

$M(A) = \{\{a\} : 0.2, \{a,b\} : 0.5, \{a,b,c\} : 0.1, \{a,b,c,d\} : 0.2\}$

In the example above, *p(a) = 0.4, p(b) = 0.3, p(c) =0.1, p(d) = 0.2* is a possible distribution, obtained by allocating the mass of 0.5 on the set $\{a, b\}$ to a (0.2) and b (0.3), and so on. We can also give a mass assignment definition of the cardinality of a fuzzy set as a distribution over integers

$$p(|A| = n) = \sum_{\substack{A_i \subseteq A \\ |A_i| = n}} m_A(A_i)$$

for $0 \leq n \leq |U|$

In the example above, $p(|A| = 1) = 0.2$, $p(|A| = 2) = 0.5$, etc. Clearly in this framework, the cardinality of a fuzzy set can be left as a distribution over integer values, or an expected value can be produced from this distribution in the usual way. A similar definition of fuzzy cardinality was proposed by [13], also motivated by the problem of fuzzy association rules.

Baldwin introduced the least prejudiced distribution (lpd) which is a specific distribution satisfying (3) above but also obeying

$$lpd_A(x) = \sum_{x \in A_i} \frac{m(A_i)}{|A_i|} \tag{4}$$

where $|A|$ indicates the cardinality of the set A and the summation is over all focal elements containing x.

Informally, wherever mass is associated with a non-singleton focal element, it is shared equally between the members of the set. Clearly a least prejudiced distribution is a restriction of the original assignment.

The steps from lpd to mass assignment and then to fuzzy set can be reversed, so that we can derive a unique fuzzy set for any frequency distribution on a finite universe, by assuming the relative frequencies are the least prejudiced distribution (proof in [14]).

If the relative frequencies are written

$$L_A = \{L_A(x_1), L_A(x_2), \ldots, L_A(x_n)\}$$

such that

$$L_A(x_1) > L_A(x_2) > \ldots > L_A(x_n)$$

then we can define

$$A_i = \{x | x \in U \wedge L_A(x) \geq L_A(x_i)\}$$

and the fuzzy set memberships are given by

$$\mu_A(x_i) = |A_i| \times L_A(x_i) + \sum_{j=i+1}^{n} \left(|A_j| - |A_{j-1}|\right) \times L_A(x_j)$$

2.4 Fuzzy relations and mass assignments

A relation is a conjunctive set of ordered n-tuples i.e. it represents a conjunction of n ground clauses. For example, if U is the set of dice scores then we could define a predicate *differBy4or5* on $U \times U$ as the set of pairs

[(1, 6), (1, 5), (2, 6), (5, 1), (6, 1), (6, 2)]

This is a conjunctive set in that each pair satisfies the predicate. In a similar way, a fuzzy relation represents a set of n-tuples that satisfy a predicate to a specified degree. Thus *differByLargeAmount* could be represented by

[(1, 6)/1, (1, 5)/0.6, (2, 6)/0.6, (5, 1)/0.6, (6, 1)/1, (6, 2)/0.6]

2.5 Mass-based association rules

We consider two granules, represented as monadic fuzzy relations *S* and *T* on the same domain, and wish to calculate the degree of association between them. For example, consider a database of sales employees, salaries and sales figures. We can categorise employees according to whether their salaries are *high*, *medium* or *low* and also according to whether their sales figures are *good*, *moderate* or *poor*. A mining task might be to find out whether the *good* sales figures are achieved by the *highly paid* employees. For example, given the table

name	sales	salary
a	100	1000
b	80	400
c	50	800
d	20	700

we might define the monadic fuzzy relations

S = goodSales = [a/1, b/0.8, c/0.5, d/0.2]

and

T = highSalary = [a/1, b/0.4, c/0.8, d/0.7]

These represent sets of values (1-tuples) that all satisfy the related predicate to a degree. The confidence in an association rule can be calculated as follows:

For a source granule

$$S = \left[x_1 / \chi_S(x_1),\ x_2 / \chi_S(x_2), \ldots,\ x_{|S|} / \chi_S\left(x_{|S|}\right) \right]$$

and a target granule

$$T = \left[x_1 / \chi_T(x_1),\ x_2 / \chi_T(x_2), \ldots,\ x_{|T|} / \chi_T\left(x_{|T|}\right) \right]$$

we can define the corresponding mass assignments as follows. Let the set of distinct memberships in *S* be

$$\left\{ \chi_S^{(1)}, \chi_S^{(2)}, \ldots, \chi_S^{(n_S)} \right\}$$

where

$$\chi_S^{(1)} > \chi_S^{(2)} > \ldots > \chi_S^{(n_S)}$$

and $n_S \leq |S|$

Let

$$S_1 = \left\{ \left[x \middle| \chi_S(x) = \chi_S^{(1)} \right] \right\}$$

$$S_i = \left\{ \left[x \middle| \chi_S(x) \geq \chi_S^{(i)} \right] \right\} \cup S_{i-1} \qquad 1 < i \leq n_S$$

Then the mass assignment corresponding to S is

$\left\{ S_i : m_S(S_i) \right\}, \quad 1 \leq i \leq n_S$

where $m_S(S_k) = \chi_S^{(k)} - \chi_S^{(k+1)}$

and we define

$$\chi_S^{(i)} = 0 \quad \mathit{if} \quad i > n_S$$

For example, the fuzzy relation

S = [a/1, b/0.8, c/0.5, d/0.2]

has the corresponding mass assignment

$$M_S = \{\{[a]\}:0.2, \{[a],[a,b]\}:0.3, \{[a],[a,b],[a,b,c]\}:0.3, \{[a],[a,b],[a,b,c],[a,b,c,d]\}:0.2\}$$

The mass assignment corresponds to a distribution on the power set of relations, and we can define the least prejudiced distribution in the same way as for the standard mass assignment. In the example above

$$L_S = \{[a]:0.5, [a,b]:0.3, [a,b,c]:0.15, [a,b,c,d]:0.05\}$$

We can now calculate the confidence in the association between the granules *S* and *T* using mass assignment theory. In general, this will be an interval as we are free to move mass (consistently) between elements of each S_i .and T_j

For two mass assignments

$$M_S = \{\{S_{p_i}\}:m_S(S_i)\}, \quad 1 \le p_i \le i \le n_S$$

$$M_T = \{\{T_{q_j}\}:m_T(S_j)\}, \quad 1 \le q_j \le j \le n_T$$

the composite mass assignment is

$$\begin{aligned} M_C &= M_S \oplus M_T \\ &= \{X : m_C(X)\} \end{aligned}$$

where m_C is specified by the composite mass allocation function $C(i, j, S_{p_i}, T_{q_j})$ subject to

$$\sum_{j=1}^{n_T} \sum_{\substack{1 \le q_j \le j \\ 1 \le p_i \le i}} C(i, j, S_{p_i}, T_{q_j}) = m_S(S_i)$$

$$\sum_{i=1}^{n_S} \sum_{\substack{1 \le p_i \le i \\ 1 \le q_j \le j}} C(i, j, S_{p_i}, T_{q_j}) = m_T(T_j)$$

This can be visualised using a mass tableau (see [11]) Each row (column) represents a focal element of the mass assignment, and is split into sub-rows (sub-columns). The mass associated with a row (column) is shown at the far left (top) and can be distributed amongst the sub-rows (sub-columns). For example consider the granules

S = [*a/1, b/0.8, c/0.5, d/0.2*] and

T = [*a/1, b/0.4, c/0.8, d/0.7*]

The rule confidence is given by equation (5)

$$Conf(S \rightarrow T) = \left(\frac{\sum_{i=1}^{n_S} \sum_{j=1}^{n_T} \sum_{\substack{1 \le q_j \le j \\ 1 \le p_i \le i}} C(i, j, S_{p_i}, T_{q_j}) \times |S_{p_i} \cap T_{q_j}|}{\sum_{i=1}^{n_S} \sum_{j=1}^{n_T} \sum_{\substack{1 \le q_j \le j \\ 1 \le p_i \le i}} C(i, j, S_{p_i}, T_{q_j}) \times |S_{p_i}|} \right) \qquad (5)$$

		0.2	0.1		0.3			0.4			
		a	*a*	*ac*	*a*	*ac*	*acd*	*a*	*ac*	*acd*	*abcd*
0.2	*a*	0.2									
0.3	*a*		0.1								
	ab										0.2
0.3	*a*				0.3						
	ab										
	abc										
0.2	*a*							0.2			
	ab										
	abc										
	abcd										

(a) $$Conf(S \to T) = \frac{0.2\times1+0.1\times1+0.2\times2+0.3\times1+0.2\times1}{0.2\times1+0.1\times1+0.2\times2+0.3\times1+0.2\times1}$$
$$=1$$

		0.2	0.1		0.3			0.4			
		a	*a*	*ac*	*a*	*ac*	*acd*	*a*	*ac*	*acd*	*abcd*
0.2	*a*	0.2									
0.3	*a*										
	ab		0.1					0.2			
0.3	*a*										
	ab										
	abc				0.3						
0.2	*a*										
	ab										
	abc										
	abcd							0.2			

(b) $$Conf(S \to T) = \frac{0.2\times1+0.1\times1+0.2\times1+0.3\times1+0.2\times1}{0.2\times1+0.1\times2+0.2\times2+0.3\times3+0.2\times4}$$
$$=0.4$$

Fig 1 - Composite mass allocation (a) maximising and (b) minimising association rule confidence

Clearly the mass can be allocated in many ways, subject to the column constraints and it is not always straightforward to find the minimum and maximum confidences arising from different composite mass allocations. Two extreme examples are shown in Fig 1, so that the confidence in the association rule between the two granules lies in the interval [0.4, 1]. In general there can be considerable computation involved in finding the maximum and minimum confidences for a rule. When ranking association rules it is preferable to have a single figure for confidence, rather than an interval which can lead to ambiguity in the ordering.

We can redistribute the mass according to the least prejudiced distribution i.e. split the mass in each row (column) equally between its sub-rows (sub-columns) and taking the product as the mass in each cell. In this case, the calculation is simplified by (a) combining rows (columns) with the same label and (b) re-ordering the summations. This enables us to calculate association confidences with roughly *O(n)* complexity, rather than *O(n^4)* where *n* is the number of focal elements in the source granule S. The confidence is then given by

$$Conf_{LPD}(S,T) = \frac{\sum_{i=1}^{n_S}\sum_{j=1}^{n_T} LPD_S(S_i) \times LPD_T(T_j) \times |S_i \cap T_j|}{\sum_{i=1}^{n_S} LPD_S(S_i) \times |S_i|} \qquad (6)$$

(due to the nested structure of the sets, the numerator does not require a double summation but can be calculated by stepping through the cells on the leading diagonal). If we choose the least prejudiced distribution and re-arrange sub-rows into single rows with the same label (also columns) we obtain the following intersections

	0.45 a	0.25 ac	0.2 acd	0.1 abcd
0.5 a	a	a	a	a
0.3 ab	a	a	a	ab
0.15 abc	a	ac	ac	abc
0.05 abcd	a	ac	acd	abcd

and the numerator for the rule confidence is
0.5 × (0.45+0.25+0.2+0.1) × 1
+ 0.3 × (0.45+0.25+0.2) × 1 + 0.3 × 0.1 × 2
+ 0.15 × 0.45 × 1 + 0.15 × (0.25+0.2) × 2+ 0.15 × 0.1 × 3
+ 0.05 × 0.45 × 1 + 0.05 × 0.25 × 2+ 0.05 × 0.2 × 3 + 0.05 × 0.1 × 4

giving a confidence of 0.67 - lying in the interval shown in Fig 1 (obviously). Using the LPD allows us to replace the calculation in eq 5 with straightforward calculations of the expected values of the cardinality of the source set and the intersection.

The example above gives a similar result to the cardinality-based method, but this is not always the case. For example if

$$S = [x_1/1, x_2/0.01,\ x_3/0.01, \ldots, x_{1000}/0.01\ ]$$

$$T = [x_1/0.01, x_2/1,\ x_3/0.01, \ldots, x_{1000}/0.01\ ]$$

then a fuzzy cardinality based approach gives a confidence of 10/10.99 ≈ 0.91 whereas our approach gives approximately 10^{-5}. Clearly this is a far more reasonable answer, as there are no elements with strong membership in both granules.

3 Experiment

We have carried out preliminary tests on the approach by finding associations between movie genres from different online sources. Ongoing work is focusing on finding associations between music genres, categories in different classified business directories and also on comparative studies using the ontology matching benchmarks, where suitable instance data is available.

The two online movie databases IMDB and Rotten Tomatoes have been used in previous work [15] to test instance matching methods. We have used the SOFT method to establish correspondence between the (roughly) 95000 movies in the databases. Within these two sources, movies are assigned to one or more genres and our task is to find strong associations between genres. The genres form a fairly flat hierarchy, although in principle one would expect genres to form a deeper hierarchical structure (e.g. comedy could be sub-divided into slapstick, satire, situation comedy, etc). At this stage, there is no benchmark for comparison but the results are intuitively reasonable as shown in Fig 2.

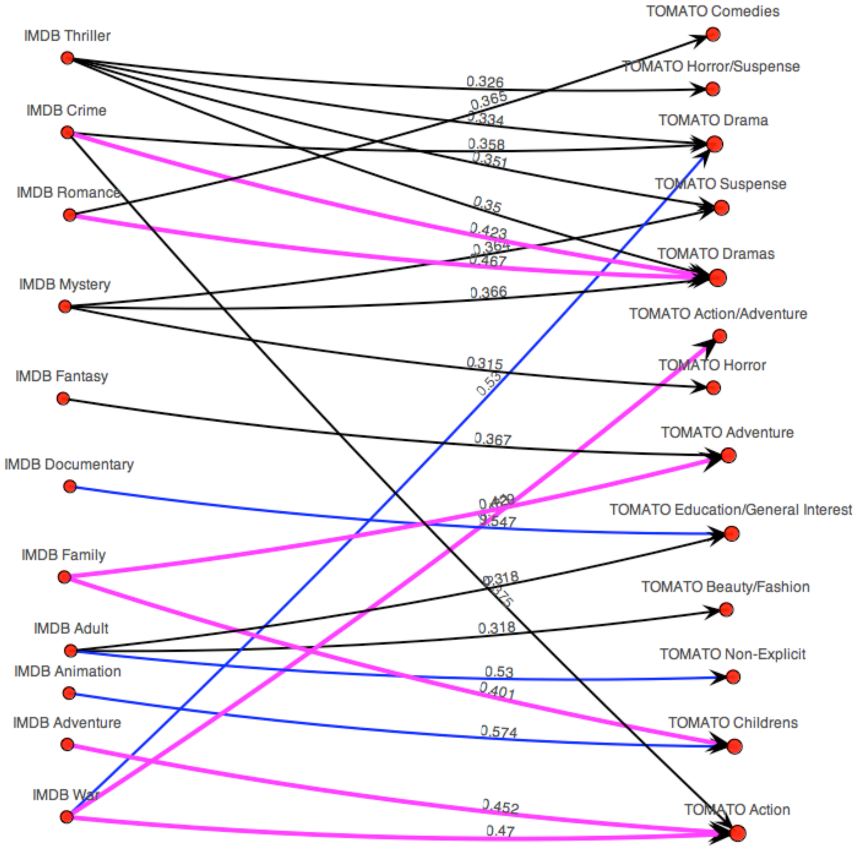

Fig 2 - strong associations from source IMDB genres (left) to target Rotten Tomato genres (right). Edge labels denotes the association strength.

4 Summary

We have described a new method for generating association rules between granules in different information hierarchies. These rules enable us to find related categories without leading to spurious relations suggested by association rules based on fuzzy cardinalities. Results were presented for discovery of links between film genres in different classification hierarchies, giving intuitively reasonable associations. The new method is currently undergoing further tests, looking at benchmark instance-matching problems, finding associations between music genres and finding links between categories in different classified business directories.

Acknowledgement : this work was partly funded by BT and the Defence Technology Centre Data and Information Fusion. We would like to thank the referees for their careful reviews and helpful suggestions.

5 References

[1] Zadeh, L. A., "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 90, pp. 111-127, 1997.

[2] Martin, T. P. and B. Azvine, "Acquisition of Soft Taxonomies for Intelligent Personal Hierarchies and the Soft Semantic Web," *BT Technology Journal*, vol. 21, pp. 113-122, 2003.

[3] Martin, T. P. and B. Azvine, "Soft Integration of Information with Semantic Gaps," in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed.: Elsevier, 2005.

[4] Bosc, P. and B. Bouchon-Meunier, "Databases and Fuzziness - Introduction," *International Journal of Intelligent Systems*, vol. 9, pp. 419, 1994.

[5] Agrawal, R. and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," presented at Very large data bases, Santiago, 1994.

[6] Bosc, P. and O. Pivert, "On Some Fuzzy Extensions of Association Rules," presented at IFSA world congress, Vancouver, Canada, 2001.

[7] Dubois, D., E. Hullermeier, and H. Prade, "A systematic approach to the assessment of fuzzy association rules," *Data Mining and Knowledge Discovery*, vol. 13, pp. 167-192, 2006.

[8] Kacprzyk, J. and S. Zadrozny, "Linguistic Summarization of Data Sets Using Association Rules," presented at Fuzzy systems; Exploring new frontiers, St Louis, MO, 2003.

[9] Martin-Bautista, M. J., M. A. Vila, H. L. Larsen, and D. Sanchez, "Measuring Effectiveness in Fuzzy Information Retrieval," presented at Flexible Query Answering Systems (FQAS), 2000.

[10] Baldwin, J. F., J. Lawry, and T. P. Martin, "Efficient Algorithms for Semantic Unification," presented at Information Processing and the Management of Uncertainty, Spain, 1996.

[11] Baldwin, J. F., "The Management of Fuzzy and Probabilistic Uncertainties for Knowledge Based Systems," in *Encyclopedia of AI*, S. A. Shapiro, Ed., 2nd ed: John Wiley, 1992, pp. 528-537.

[12] Dubois, D. and H. Prade, "On Several Representations of an Uncertain Body of Evidence," in *Fuzzy Information and Decision Processes*, M. M. Gupta and E. Sanchez, Eds.: North Holland, 1982.

[13] Delgado, M., D. Sanchez, M. J. Martin-Bautista, and M. A. Vila, "A probabilistic definition of a nonconvex fuzzy cardinality," *Fuzzy Sets and Systems*, vol. 126, pp. 177-190, 2002.

[14] Baldwin, J. F., J. Lawry, and T. P. Martin, "A Mass Assignment Theory of the Probability of Fuzzy Events," *Fuzzy Sets and Systems*, vol. 83, pp. 353-367, 1996.

[15] Martin, T. P. and Y. Shen, "Improving access to multimedia using multi-source hierarchical meta-data," in *Adaptive Multimedia Retrieval: User, Context, and Feedback*, vol. LNCS vol 3877, *LNCS*: Springer, 2006, pp. 266 - 278.