Granular computing is an emerging computing paradigm of information processing that concerns the processing of complex information entities called "information granules ", which arise in the process of data abstraction and derivation of knowledge from information or data. Generally speaking, information granules are collections of entities that usually originate at the numeric level and are arranged together due to their similarity , functional or physical adjacency, indistinguishability, coherency, or the like.

At present, granular computing is more a theoretical perspective than a coherent set of methods or principles. As a theoretical perspective, it encourages an approach to data that recognizes and exploits the knowledge present in data at various levels of resolution or scales. In this sense, it encompasses all methods which provide flexibility and adaptability in the resolution at which knowledge or information is extracted and represented.

Types of granulation

As mentioned above, granular computing is not an algorithm or process; there is no particular method that is called "granular computing". It is rather an approach to looking at data that recognizes how different and interesting regularities in the data can appear at different levels of granularity, much as different features become salient in satellite images of greater or lesser resolution. On a low-resolution satellite image, for example, one might notice interesting cloud patterns representing cyclones or other large-scale weather phenomena, while in a higher-resolution image, one misses these large-scale atmospheric phenomena but instead notices smaller-scale phenomena, such as the interesting pattern that is the streets of Manhattan . The same is generally true of all data: At different resolutions or granularities, different features and relationships emerge. The aim of granular computing is to try to take advantage of this fact in designing more effective machine-learning and reasoning systems.

There are several types of granularity that are often encountered in data mining and machine learning , and we review them below:

Value granulation (discretization/quantization)

One type of granulation is the quantization of variables. It is very common that in data mining or machine-learning applications the resolution of variables needs to be decreased in order to extract meaningful regularities. An example of this would be a variable such as "outside temperature" (temp), which in a given application might be recorded to several decimal places of precision (depending on the sensing apparatus). However, for purposes of extracting relationships between "outside temperature" and, say, "number of health-club applications" (club), it will generally be advantageous to quantize "outside temperature" into a smaller number of intervals.

Motivations

There are several interrelated reasons for granulating variables in this fashion:

Based on prior domain knowledge , there is no expectation that minute variations in temperature (e.g., the difference between 80–80.7 °F (26.7–27.1 °C)) could have an influence on behaviors driving the number of health-club applications. For this reason, any "regularity" which our learning algorithms might detect at this level of resolution would have to be spurious , as an artifact of overfitting. By coarsening the temperature variable into intervals the difference between which we do anticipate (based on prior domain knowledge) might influence number of health-club applications, we eliminate the possibility of detecting these spurious patterns. Thus, in this case,

reducing resolution is a method of controlling overfitting .

By reducing the number of intervals in the temperature variable (i.e., increasing its grain size), we increase the amount of sample data indexed by each interval designation. Thus, by coarsening the variable, we increase sample sizes and achieve better statistical estimation. In this sense, increasing granularity provides an antidote to the so-called curse of dimensionality , which relates to the exponential decrease in statistical power with increase in number of dimensions or variable cardinality.

Independent of prior domain knowledge, it is often the case that meaningful regularities (i.e., which can be detected by a given learning methodology, representational language, etc.) may exist at one level of resolution and not at another.

For example, a simple learner or pattern recognition system may seek to extract regularities satisfying a conditional probability threshold such as $p(Y=y_j|X=x_i) \geq \alpha$. In the special case where $\alpha = 1$, this recognition system is essentially detecting logical implication of the form $X=x_i \rightarrow Y=y_j$ or, in words, "if $X=x_i$, then $Y=y_j$ ". The system's ability to recognize such implications (or, in general, conditional probabilities exceeding threshold) is partially contingent on the resolution with which the system analyzes the variables.

As an example of this last point, consider the feature space shown to the right. The variables may each be regarded at two different resolutions. Variable X may be regarded at a high (quaternary) resolution wherein it takes on the four values $\{x_1,x_2,x_3,x_4\}$ or at a lower (binary) resolution wherein it takes on the two values $\{X_1,X_2\}$. Similarly, variable Y may be regarded at a high (quaternary) resolution or at a lower (binary) resolution, where it takes on the values $\{y_1,y_2,y_3,y_4\}$ or $\{Y_1,Y_2\}$, respectively. At the high resolution, there are no detectable implications of the form $X=x_i \rightarrow Y=y_j$, since every x_i is associated with more than one y_j , and thus, for all x_i , $p(Y=y_j|X=x_i)<1.$ However, at the low (binary) variable resolution, two bilateral implications become detectable: $X=X_1 \leftrightarrow Y=Y_1$ and $X=X_2 \leftrightarrow Y=Y_2$, since every X_1 occurs iff Y_1 and X_2 occurs iff Y_2 . Thus, a pattern recognition system scanning for implications of this kind would find them at the binary variable resolution, but would fail to find them at the higher quaternary variable resolution.

Issues and methods

It is not feasible to exhaustively test all possible discretization resolutions on all variables in order to see which combination of resolutions yields interesting or significant results. Instead, the feature space must be preprocessed (often by an entropy analysis of some kind) so that some guidance can be given as to how the discretization process should proceed. Moreover, one cannot generally achieve good results by naively analyzing and discretizing each variable independently, since this may obliterate the very interactions that we had hoped to discover.

A sample of papers that address the problem of variable discretization in general, and multiple-variable discretization in particular, is as follows: Chiu, Wong & Cheung (1991) , Bay (2001) , Liu et al. (2002) , Wang & Liu (1998) , Zighed, Rabaséda & Rakotomalala (1998) , Catlett (1991) , Dougherty, Kohavi & Sahami (1995) , Monti & Cooper (1999) , Fayyad & Irani (1993) , Chiu, Cheung & Wong (1990) , Nguyen & Nguyen (1998) , Grzymala-Busse & Stefanowski (2001) , Ting (1994) , Ludl & Widmer (2000) , Pfahringer (1995) , An & Cercone (1999) , Chiu & Cheung (1989) , Chmielewski & Grzymala-Busse (1996) , Lee & Shin (1994) , Liu & Wellman (2002) , Liu & Wellman (2004) .

Variable granulation (clustering/aggregation/transformation)

Variable granulation is a term that could describe a variety of techniques, most of which are aimed at reducing dimensionality, redundancy, and storage requirements. We briefly describe some of the ideas here, and present pointers to the literature.

Variable transformation

A number of classical methods, such as principal component analysis , multidimensional scaling , factor analysis , and structural equation modeling , and their relatives, fall under the genus of "variable transformation." Also in this category are more modern areas of study such as dimensionality reduction , projection pursuit , and independent component analysis . The common goal of these methods in general is to find a representation of the data in terms of new variables, which are a linear or nonlinear transformation of the original variables, and in which important statistical relationships emerge. The resulting variable sets are almost always smaller than the original variable set, and hence these methods can be loosely said to impose a granulation on the feature space. These dimensionality reduction methods are all reviewed in the standard texts, such as Duda, Hart & Stork (2001) , Witten & Frank (2005) , and Hastie, Tibshirani & Friedman (2001) .

Variable aggregation

A different class of variable granulation methods derive more from data clustering methodologies than from the linear systems theory informing the above methods. It was noted fairly early that one may consider "clustering" related variables in just the same way that one considers clustering related data. In data clustering, one identifies a group of similar entities (using a " measure of similarity " suitable to the domain — Martino, Giuliani & Rizzi (2018)), and then in some sense replaces those entities with a prototype of some kind. The prototype may be the simple average of the data in the identified cluster, or some other representative measure. But the key idea is that in subsequent operations, we may be able to use the single prototype for the data cluster (along with perhaps a statistical model describing how exemplars are derived from the prototype) to stand in for the much larger set of exemplars. These prototypes are generally such as to capture most of the information of interest concerning the entities.

Similarly, it is reasonable to ask whether a large set of variables might be aggregated into a smaller set of prototype variables that capture the most salient relationships between the variables. Although variable clustering methods based on linear correlation have been proposed (Duda, Hart & Stork 2001 ; Rencher 2002), more powerful methods of variable clustering are based on the mutual information between variables. Watanabe has shown (Watanabe 1960 ; Watanabe 1969) that for any set of variables one can construct a polytomic (i.e., n-ary) tree representing a series of variable agglomerations in which the ultimate "total" correlation among the complete variable set is the sum of the "partial" correlations exhibited by each agglomerating subset (see figure). Watanabe suggests that an observer might seek to thus partition a system in such a way as to minimize the interdependence between the parts "... as if they were looking for a natural division or a hidden crack."

One practical approach to building such a tree is to successively choose for agglomeration the two variables (either atomic variables or previously agglomerated variables) which have the highest pairwise mutual information (Kraskov et al. 2003). The product of each agglomeration is a new (constructed) variable that reflects the local joint distribution of the two agglomerating variables, and thus possesses an entropy equal to their joint entropy .

(From a procedural standpoint, this agglomeration step involves replacing two columns in the attribute-value table—representing the two agglomerating variables—with a single column that has a unique value for every unique combination of values in the replaced columns (Kraskov et al. 2003). No information is lost by such an operation; however, if one is exploring the data for inter-variable relationships, it would generally not be desirable to merge redundant variables in this way, since in such a context it is likely to be precisely the redundancy or dependency between variables that is of interest; and once redundant variables are merged, their relationship to one another can no longer be studied.

System granulation (aggregation)

In database systems , aggregations (see e.g. OLAP aggregation and Business intelligence systems) result in transforming original data tables (often called information systems) into the tables with different semantics of rows and columns, wherein the rows correspond to the groups (granules) of original tuples and the columns express aggregated information about original values within each of the groups. Such aggregations are usually based on SQL and its extensions. The resulting granules usually correspond to the groups of original tuples with the same values (or ranges) over some pre-selected original columns.

There are also other approaches wherein the groups are defined basing on, e.g., physical adjacency of rows. For example, Infobright implemented a database engine wherein data was partitioned onto rough rows , each consisting of 64K of physically consecutive (or almost consecutive) rows. Rough rows were automatically labeled with compact information about their values on data columns, often involving multi-column and multi-table relationships. It resulted in a higher layer of granulated information where objects corresponded to rough rows and attributes - to various aspects of rough information. Database operations could be efficiently supported within such a new framework, with an access to the original data pieces still available (Slezak et al. 2013).

Concept granulation (component analysis)

The origins of the granular computing ideology are to be found in the rough sets and fuzzy sets literatures. One of the key insights of rough set research—although by no means unique to it—is that, in general, the selection of different sets of features or variables will yield different concept granulations. Here, as in elementary rough set theory, by "concept" we mean a set of entities that are indistinguishable or indiscernible to the observer (i.e., a simple concept), or a set of entities that is composed from such simple concepts (i.e., a complex concept). To put it in other words, by projecting a data set (value-attribute system) onto different sets of variables, we recognize alternative sets of equivalence-class "concepts" in the data, and these different sets of concepts will in general be conducive to the extraction of different relationships and regularities.

Equivalence class granulation

We illustrate with an example. Consider the attribute-value system below:

When the full set of attributes $P = \{ P_1 , P_2 , P_3 , P_4 , P_5 \}$ is considered, we see that we have the following seven equivalence classes or primitive (simple) concepts:

Thus, the two objects within the first equivalence class, $\{ O_1 , O_2 \}$, cannot be distinguished from one another based on the available attributes, and the three objects within the second equivalence class, $\{ O_3 , O_7 , O_{10} \}$, cannot be distinguished from one another based on the available attributes. The remaining five objects are each discernible from all other objects. Now, let us imagine a projection of the attribute value system onto attribute P_1 alone, which would represent, for example, the view from an observer which is only capable of detecting this single attribute. Then we obtain the following much coarser equivalence class structure.

This is in a certain regard the same structure as before, but at a lower degree of resolution (larger grain size). Just as in the case of value granulation (discretization/quantization) , it is possible that relationships (dependencies) may emerge at one level of granularity that are not present at another. As an example of this, we can consider the effect of concept granulation on the measure known as attribute dependency (a simpler relative of the mutual information).

To establish this notion of dependency (see also rough sets), let $[x]_Q = \{ Q_1 , Q_2 , Q_3 , \dots , Q_N \}$ represent a particular concept granulation, where each Q_i is an equivalence class from the concept structure induced by attribute set Q . For example, if the attribute set Q consists of attribute P_1 alone, as above, then the concept structure $[x]_Q$ will be composed of

The dependency of attribute set Q on another attribute set P , $\gamma_P (Q)$, is given by

That is, for each equivalence class Q_i in $[x]_Q,$ we add up the size of its "lower approximation" (see rough sets) by the attributes in P, i.e., $\underline{P}Q_i.$ More simply, this approximation is the number of objects which on attribute set P can be positively identified as belonging to target set $Q_i.$ Added across all equivalence classes in $[x]_Q,$ the numerator above represents the total number of objects which—based on attribute set P —can be positively categorized according to the classification induced by attributes Q. The dependency ratio therefore expresses the proportion (within the entire universe) of such classifiable objects, in a sense capturing the "synchronization" of the two concept structures $[x]_Q$ and $[x]_P.$ The dependency $\gamma_P(Q)$ "can be interpreted as a proportion of such objects in the information system for which it suffices to know the values of attributes in P to determine the values of attributes in Q " (Ziarko & Shan 1995).

Having gotten definitions now out of the way, we can make the simple observation that the choice of concept granularity (i.e., choice of attributes) will influence the detected dependencies among attributes. Consider again the attribute value table from above:

Consider the dependency of attribute set $Q=\{P_4,P_5\}$ on attribute set $P=\{P_2,P_3\}.$ That is, we wish to know what proportion of objects can be correctly classified into classes of $[x]_Q$ based on knowledge of $[x]_P.$ The equivalence classes of $[x]_Q$ and of $[x]_P$ are shown below.

The objects that can be definitively categorized according to concept structure $[x]_Q$ based on $[x]_P$ are those in the set $\{O_1,O_2,O_3,O_7,O_8,O_{10}\},$ and since there are six of these, the dependency of Q on P, $\gamma_P(Q)=6/10.$ This might be considered an interesting dependency in its own right, but perhaps in a particular data mining application only stronger dependencies are desired.

We might then consider the dependency of the smaller attribute set $Q=\{P_4\}$ on the attribute set $P=\{P_2,P_3\}.$ The move from $Q=\{P_4,P_5\}$ to $Q=\{P_4\}$ induces a coarsening of the class structure $[x]_Q,$ as will be seen shortly. We wish again to know what proportion of objects can be correctly classified into the (now larger) classes of $[x]_Q$ based on knowledge of $[x]_P.$ The equivalence classes of the new $[x]_Q$ and of $[x]_P$ are shown below.

Clearly, $[x]_Q$ has a coarser granularity than it did earlier. The objects that can now be definitively categorized according to the concept structure $[x]_Q$ based on $[x]_P$ constitute the complete universe $\{O_1,O_2,\ldots,O_{10}\},$ and thus the dependency of Q on P, $\gamma_P(Q)=1.$ That is, knowledge of membership according to category set $[x]_P$ is adequate to determine category membership in $[x]_Q$ with complete certainty; In this case we might say that $P\rightarrow Q.$ Thus, by coarsening the concept structure, we were able to find a stronger (deterministic) dependency. However, we also note that the classes induced in $[x]_Q$ from the reduction in resolution necessary to obtain this deterministic dependency are now themselves large and few in number; as a result, the dependency we found, while strong, may be less valuable to us than the weaker dependency found earlier under the higher resolution view of $[x]_Q.$

In general it is not possible to test all sets of attributes to see which induced concept structures yield the strongest dependencies, and this search must be therefore be guided with some intelligence. Papers which discuss this issue, and others relating to intelligent use of granulation, are those by Y.Y. Yao and Lotfi Zadeh listed in the #References below.

Component granulation

Another perspective on concept granulation may be obtained from work on parametric models of categories. In mixture model learning, for example, a set of data is explained as a mixture of distinct Gaussian (or other) distributions. Thus, a large amount of data is "replaced" by a small number of distributions. The choice of the number of these distributions, and their size, can again be viewed as a problem of concept granulation . In general, a better fit to the data is obtained by a larger number of distributions or parameters, but in order to extract meaningful patterns, it is necessary to constrain the number of distributions, thus deliberately coarsening the concept resolution. Finding the "right" concept resolution is a tricky problem for which many methods have been proposed (e.g., AIC , BIC , MDL , etc.), and these are frequently considered under the rubric of " model regularization ".

Different interpretations of granular computing

Granular computing can be conceived as a framework of theories, methodologies, techniques, and tools that make use of information granules in the process of problem solving. In this sense, granular computing is used as an umbrella term to cover topics that have been studied in various fields in isolation. By examining all of these existing studies in light of the unified framework of granular computing and extracting their commonalities, it may be possible to develop a general theory for problem solving.

In a more philosophical sense, granular computing can describe a way of thinking that relies on the human ability to perceive the real world under various levels of granularity (i.e., abstraction) in order to abstract and consider only those things that serve a specific interest and to switch among different granularities. By focusing on different levels of granularity, one can obtain different levels of knowledge, as well as a greater understanding of the inherent knowledge structure. Granular computing is thus essential in human problem solving and hence has a very significant impact on the design and implementation of intelligent systems.

See also

Rough Sets , Discretization

Type-2 Fuzzy Sets and Systems

References

An, Aijun; Cercone, Nick (1999), "Discretization of continuous attributes for learning classification rules", in Ning Zhong; Lizhu Zhou (eds.), Methodologies for Knowledge Discovery and Data Mining: Proceedings of the Third Pacific-Asia Conference, PAKDD-99 , Lecture Notes in Computer Science, vol. 1574, Beijing, China , pp. 509– 514, doi : 10.1007/3-540-48912-6_69 , ISBN 978-3-540-65866-5 {{ citation }} : CS1 maint: location missing publisher (link) .

Bargiela, A. and Pedrycz, W. (2003) Granular Computing. An introduction , Kluwer Academic Publishers

Bay, Stephen D. (2001), "Multivariate discretization for set mining", Knowledge and Information Systems , 3 (4): 491– 512, CiteSeerX 10.1.1.217.921 , doi : 10.1007/PL00011680 , S2CID 10945544 .

Catlett, J. (1991), "On changing continuous attributes into ordered discrete attributes" , in Y. Kodratoff (ed.), Machine Learning—EWSL-91: European Working Session on Learning , Porto, Portugal , pp. 164– 178, ISBN 9780387538167 {{ citation }} : CS1 maint: location missing publisher (link) .

Chiu, David K. Y.; Cheung, Benny (1989), "Hierarchical maximum entropy discretization", in Ryszard Janicki; Waldemar W. Koczkodaj (eds.), Computing and Information: Proceedings of the International Conference on Computing and Information (ICCI '89) , Toronto, Ontario , Canada: North-Holland, pp. 237– 242 .

Chiu, David K. Y.; Cheung, Benny; Wong, Andrew K. C. (1990), "Information synthesis based on hierarchical maximum entropy discretization", Journal of Experimental and Theoretical Artificial Intelligence , 2 (2): 117– 129, doi : 10.1080/09528139008953718 .

Chiu, David K. Y.; Wong, Andrew K. C.; Cheung, Benny (1991), "Information discovery through hierarchical maximum entropy discretization and synthesis", in Gregory Piatetsky-Shapiro; William J. Frawley (eds.), Knowledge Discovery in Databases , Cambridge, MA : MIT Press, pp. 126– 140 .

Chmielewski, Michal R.; Grzymala-Busse, Jerzy W. (1996), "Global discretization of continuous attributes as preprocessing for machine learning" (PDF) , International Journal of Approximate Reasoning , 15 (4): 319– 331, doi : 10.1016/s0888-613x(96)00074-6 .

Dougherty, James; Kohavi, Ron; Sahami, Mehran (1995), "Supervised and unsupervised discretization of continuous features" , in Armand Prieditis; Stuart Russell (eds.), Machine Learning: Proceedings of the Twelfth International Conference (ICML 1995) , Tahoe City, CA : Morgan Kaufmann, pp. 194– 202 .

Duda, Richard O.; Hart, Peter E.; Stork, David G. (2001), Pattern Classification (2nd ed.), New York City : John Wiley & Sons, ISBN 978-0-471-05669-0

Fayyad, Usama M.; Irani, Keki B. (1993), "Multi-interval discretization of continuous-valued attributes for classification learning", Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93) , Chambéry, France , pp. 1022– 1027 {{ citation }} : CS1 maint: location missing publisher (link) .

Grzymala-Busse, Jerzy W.; Stefanowski, Jerzy (2001), "Three discretization methods for rule induction", International Journal of Intelligent Systems , 16 (1): 29– 38, CiteSeerX 10.1.1.330.2975 , doi : 10.1002/1098-111X(200101)16:1<29::AID-INT4>3.0.CO;2-0 .

Hastie, Trevor ; Tibshirani, Robert ; Friedman, Jerome (2001), The Elements of Statistical Learning: Data Mining, Inference, and Prediction , New York City : Springer, ISBN 978-0-387-84857-0

Kraskov, Alexander; Stögbauer, Harald; Andrzejak, Ralph G.; Grassberger, Peter (2003), Hierarchical clustering based on mutual information , arXiv : q-bio/0311039 , Bibcode : 2003q.bio....11039K .

Lee, Changhwan; Shin, Dong-Guk (1994), "A context-sensitive discretization of numeric attributes for classification learning", in A. G. Cohn (ed.), Proceedings of the 11th European Conference on Artificial Intelligence (ECAI 94) , NL , pp. 428– 432 {{ citation }} : CS1 maint: location missing publisher (link) .

Liu, Chao-Lin; Wellman, Michael (2002), "Evaluation of Bayesian networks with flexible state-space abstraction methods", International Journal of Approximate Reasoning , 30 (1): 1– 39, CiteSeerX 10.1.1.127.7040 , doi : 10.1016/S0888-613X(01)00067-6 , S2CID 17529419 .

Liu, Chao-Lin; Wellman, Michael (2004), "Bounding probabilistic relationships in Bayesian networks using qualitative influences: Methods and applications", International Journal of Approximate Reasoning , 36 (1): 31– 73, doi : 10.1016/j.ijar.2003.06.002 .

Liu, Huan; Hussain, Farhad; Tan, Chew Lim; Dasii, Manoranjan (2002), "Discretization: An enabling technique", Data Mining and Knowledge Discovery , 6 (4): 393– 423, doi : 10.1023/A:1016304305535 , S2CID 207609303 .

Ludl, Marcus-Christopher; Widmer, Gerhard (2000), "Relative unsupervised discretization for association rule mining", in Djamel A. Zighed; Jan Komorowski; Jan Zytkow (eds.), Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2000) , Lecture Notes in Computer Science, vol. 1910, Lyon, France , pp. 148– 158, doi : 10.1007/3-540-45372-5_15 , ISBN 978-3-540-41066-9 {{ citation }} : CS1 maint: location missing publisher (link) .

Monti, Stefano; Cooper, Gregory F. (1999), "A latent variable model for multivariate discretization" , Uncertainty 99: The 7th International Workshop on Artificial Intelligence and Statistics , Fort Lauderdale, FL {{ citation }} : CS1 maint: location missing publisher (link) .

Martino, Alessio; Giuliani, Alessandro; Rizzi, Antonello (2018), "Granular Computing Techniques for Bioinformatics Pattern Recognition Problems in Non-metric Spaces", in Pedrycz W.; Chen SM. (eds.), Computational Intelligence for Pattern Recognition , Studies in Computational Intelligence, vol. 777, Springer International Publishing, pp. 53– 81, doi : 10.1007/978-3-319-89629-8_3 , ISBN

978-3-319-89628-1 .

Nguyen, Hung Son; Nguyen, Sinh Hoa (1998), "Discretization methods in data mining", in Lech Polkowski; Andrzej Skowron (eds.), Rough Sets in Knowledge Discovery 1: Methodology and Applications , Heidelberg : Physica-Verlag, pp. 451– 482 .

Pfahringer, Bernhard (1995), "Compression-based discretization of continuous attributes" , in Armand Prieditis; Stuart Russell (eds.), Machine Learning: Proceedings of the Twelfth International Conference (ICML 1995) , Tahoe City, CA : Morgan Kaufmann, pp. 456– 463 .

Rencher, Alvin C. (2002), Methods of Multivariate Analysis , New York City : Wiley .

Simon, Herbert A.; Ando, Albert (1963), "Aggregation of variables in dynamic systems", in Albert Ando; Franklin M. Fisher; Herbert A. Simon (eds.), Essays on the Structure of Social Science Models , Cambridge, MA: MIT Press, pp. 64– 91

Simon, Herbert A. (1996), "The architecture of complexity: Hierarchic systems", in Herbert A. Simon (ed.), The Sciences of the Artificial (2nd ed.), Cambridge, MA: MIT Press, pp. 183– 216

Slezak, Dominik; Synak, Piotr; Wojna, Arkadiusz; Wroblewski, Jakub (2013), "Two Database Related Interpretations of Rough Approximations: Data Organization and Query Execution", Fundamenta Informaticae , 127 (1– 4): 445– 459, doi : 10.3233/FI-2013-920 .

Ting, Kai Ming (1994), Discretization of continuous-valued attributes and instance-based learning (Technical Report No.491) , Sydney : Basser Department of Computer Science .

Wang, Ke; Liu, Bing (1998), "Concurrent discretization of multiple attributes" , in Springer (ed.), Proceedings of the 5th Pacific Rim International Conference on Artificial Intelligence , London : Springer-Verlag, pp. 250– 259 .

Watanabe, Satosi (1960), "Information theoretical analysis of multivariate correlation", IBM Journal of Research and Development , 4 (1): 66– 82, doi : 10.1147/rd.41.0066 .

Watanabe, Satosi (1969), Knowing and Guessing: A Quantitative Study of Inference and Information , New York City : Wiley .

Witten, Ian H.; Frank, Eibe (2005), Data Mining: Practical Machine Learning Tools and Techniques (2 ed.), Amsterdam : Morgan Kaufmann, archived from the original on 2020-11-27 , retrieved 2007-02-11

Yao, Y.Y. (2004) "A Partition Model of Granular Computing", Lecture Notes in Computer Science (to appear)

Yao, Y. Y. (2001). "On modeling data mining with granular computing" . Proceedings of the 25th Annual International Computer Software and Applications Conference (COMPSAC 2001) . pp. 638– 643.

Yao, Yiyu (2006). "Granular computing for data mining" (PDF) . In Dasarathy, Belur V. (ed.). Proceedings of the SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security . Archived from the original (PDF) on 2007-04-18.

Yao, J. T.; Yao, Y. Y. (2002). "Induction of classification rules by granular computing" (PDF) . Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing (TSCTC'02) . London, UK: Springer-Verlag. pp. 331– 338.

Zadeh, L.A. (1997) "Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic" , Fuzzy Sets and Systems , 90:111-127

Zighed, D. A.; Rabaséda, S.; Rakotomalala, R. (1998), "FUSINTER: A method for discretization of continuous attributes" , International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems , 6 (3): 307– 326, doi : 10.1142/s0218488598000264 .