RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **98 (**1), 2004, pp. 85–94
Ciencias de la Computación / Computational Sciences

Numerical Taxonomy: A Missing Link for Case-Based Reasoning and Autonomous Agents

John A. Campbell

Abstract. Numerical taxonomy, which uses numerical methods to classify and relate items whose properties are non-numerical, is suggested as both an advantageous tool to support case-based reasoning and a means for agents to exploit knowledge that is best expressed in cases. The basic features of numerical taxonomy are explained, and discussed in application to a problem where human agents with differing views obtain solutions by negotiation and by reference to knowledge that is essentially case-like: allocation of frequencies for shortwave radio broadcasting.

Taxonomía numérica: un eslabón perdido para el razonamiento basado en casos y agentes autónomos

Resumen. Se propone la taxonomía numérica, que emplea métodos numéricos para clasificar elementos cuyas propiedades no son numéricas y para establecer relaciones entre ellos, no sólo como una herramienta ventajosa a fin de proporcionar soporte al razonamiento basado en casos, sino también como un medio para que los agentes exploten el conocimiento que se expresa de la forma más adecuada como casos. Se explican las características básicas de la taxonomía numérica, y se expone su aplicación a un problema en el que unos agentes humanos con diferentes puntos de vista obtienen soluciones por negociación y por referencia al conocimiento que, en su esencia, se parece a los casos: la asignación de frecuencias para las emisiones por radio de onda corta.

1. Introduction and Motivation

Case-based reasoning (CBR) uses records of extended particular episodes as knowledge that can be exploited to derive suggested solutions for new problems. It is most relevant in artificial intelligence (AI) when these records exist in subjects for which the available generalised knowledge, expressed in traditional representations of AI (rules, logics, semantic nets etc.) and computer science (algorithms), is insufficient to solve practical problems in those subjects.

CBR is a variety of reasoning by analogy. Children meet analogical reasoning in elementary schools in many countries in the form of questions, usually involving matching of colours and geometrical shapes, phrased as A is to B as C is to ?. Answering the question requires first establising a correspondence or mapping M between A and C, and then substituting for ? the result of applying the same mapping M to B. CBR does essentially the same thing, though its A, B ... are sets of components (e.g. logical assertions) and M is a set of mappings between them. In CBR, a case has at least two parts. The parts of type A and C

Presentado por Luis M. Laita.
Recibido: February 18, 2004. *Aceptado*: October 13, 2004.
Palabras clave / Keywords: numerical taxonomy, case-based reasoning, autonomous agents, radio frequency allocation.
Mathematics Subject Classifications: 62H30, 68T99, 92B10.

are statements of problems or situations to be resolved, and B and ? are statements of actual or suggested solutions for the problems.

This is a very basic simplification of CBR. For example, most cases have at least a third part which contains an *a posteriori* critique, explanation or justification for the success or lack of success of the solution. The first general CBR textbook, by Kolodner [9], still gives an accurate picture of the field and its main issues.

Autonomous agents and multi-agent systems (MAS) are among the most popular and active topics in artificial intelligence (AI) at present. This may not be surprising: they originated from the research area that was originally called distributed AI. Agent is also a popular name for programs with no AI ancestry, e.g. programs acting on behalf of a user to negotiate purchases or other financial transactions on the Web.

Although the *representation* and use of knowledge is a central part of classical AI, typical papers and textbooks on MAS (e.g. Weiss [17]) give it very little special attention; for example, they tend to take it for granted that first-order predicate logic will be the medium of representation. Instead, other issues such as inter-agent communication, coordination, distribution of tasks, and negotiation. are in the foreground. This profile of emphasis is particularly well exemplified in [15]. The implication is that the design choices associated with these and related topics are primarily responsible for how an MAS behaves. (Indeed, interesting MAS behaviours can be obtained from purely reactive agents - simple software objects - and subsumption architectures [2] making use of them). When knowledge as such is accommodated in agent designs, it is usually algorithmic or may occasionally be explicitly rule-like: that is, generalised rather than episodic or particular knowledge.

This raises the question of what agents can do in situations where the main knowledge is episodic and particular, i.e. case-like. Although Acquire and use cases is an obvious answer, it has not yet been followed up thoroughly - possibly because the overheads in building case bases and applying CBR have been seen as too great for efficient computation in MAS of any significant size. Where MAS and CBR are mentioned together, it is in connection with democratic decision-making among agents as to which agent should be assigned a particular task [6]. Agents case bases then consist only of simple records noting which agents have done well on instances of past tasks. But there are many situations where wider case knowledge can benefit agents and MAS. This paper considers an application - frequency assignment in shortwave broadcasting - where wider knowledge exists and is relevant, and summarises some results.

There is one distinctive feature of the work, which contributes something new and useful to both CBR and MAS. This is the use of numerical taxonomy. It restates in a clear and intuitive way some of the less formal approaches to retrieving cases from a case base according to their similarity with a given new problem. It is computationally cheap (and therefore allows agents to use CBR without heavy overheads). Finally, it offers a medium for interaction between agents that can communicate both negotiating information and knowledge.

2. Numerical Taxonomy

Taxonomists study the classification of items on the basis of similarities among those of their properties that experts consider to be significant. Historically, their field has been most important where an abundance of heterogeneous data has become available before the existence of theoretical frameworks to provide - or at least suggest - tight forms of organisation and generalisation which can make clear the relative positions of the different items. Taxonomic treatment of data in this spirit occurred first in botany and zoology, some centuries ago. Most recently, and with the same justification, it has started to occur in bioinformatics, e.g. in construction and maintenance of repositories of knowledge about animal viruses (as in the AVIS system [8]) and about configurations and related properties of proteins [4,5].

In the language of computer science, taxonomy has only a weak type discipline (or none): an item can have properties covering an arbitrary range of types. Numerical taxonomy is more strongly typed. Each property is of type real or integer, so that a metric function giving a distance between any two items can be computed from the values of their properties in the same way as a Pythagorean distance between two locations is computed from their coordinates.

The two fundamental operations of numerical taxonomy are:

* converting properties that are not already of type integer or real into integer-valued or real-valued properties;

* establishing metrics on those properties, so as to reproduce experts opinions about the relative distances of items in their conceptual space (i.e. with small distances indicating similarity of items, and larger distances indicating less similarity).

Several good textbooks, e.g. Sneath and Sokal [16], explain these operations and the scope of the subject in detail.

Case bases in AI have much in common with sets of botanical data in the early days of botany: plenty of heterogeneous (including non-numerical) data about items, no overall theoretical framework expressing generalisations (otherwise the knowledge in the cases would be expressed instead in one of the alternative representations used in AI to fomulate generalisations). Cases remain examples of particular rather than general knowledge. They are therefore somewhat analogous to botanical specimens collected on field trips before the time of Linnaeus (the father of modern botanical generalisation).

This paper argues that the analogy can be taken further, and that a useful and computationally cheap way of handling case knowledge, for agents in particular, is to apply the techniques of numerical taxonomy to it.

2.1. Converting Arbitrary Properties to Numerical Form

Many properties that are not explicitly numerical have an implicit numerical flavour because they are expressed essentially through quantitative adjectives or adverbs (no X, very weakly X, fairly X, extremely X etc.). When these occur as arguments of logical predicates, fuzzy logic [7] is a popular and effective way of dealing with them. For numerical taxonomy, it is possible to use just the basic apparatus of fuzzy logic, e.g. in place of the distributions that represent these modifiers and lead to logical deduction through the evaluation of convolution integrals, it is enough (at least, in typical CBR applications) to regard the means of the distributions as the desired numerical values. In fuzzy logic, these values all fall in the range [0.0, 1.0].

That example stands for any situation where the values refer to one conceptual dimension (X-ness, say) in the multidimensional classification of items. If one-dimensionality cannot be taken for granted on a set of qualitative values, but a rough notion of less and greater can be stated, draw a directed graph with each value as a point, and a directed edge from p to q if value p is less than value q according to that notion. If the result is a line graph including all the points, a single dimension is established, and numerical values (subject to the ordering) can then be assigned to all of the points, ideally by obtaining the opinion of an expert in the subject of the application. If any other kind of graph is produced, this is an indication that more than one distinct dimension is hidden in the values. With the graph as a guide or stimulus to thought, each such dimension can then be identified (again, expert advice is helpful here), its values redefined and reassigned if necessary, and the general procedures for dealing with non-numerical values repeated on it.

The set bristly, spiky, lumpy, fuzzy, furry, bushy, bald, stubbly, hairy illustrates this situation. Treating it as above shows that two distinct materials (hair, and some other surface feature) and therefore two dimensions are involved, that some terms should be changed (e.g. it is advisable to reserve bald to refer to an absence of hair, and to replace the term by a new one such as smooth where an absence of the other surface feature is meant), and that the terms in each dimension have a linear order.

Boolean-valued properties are not unusual in numerical taxonomy, but unless they are the dominant majority of the properties of items in an application (unlikely in CBR), they are treated as integers and combined in purely numerical operations with the numerical values of other properties when distances between items are computed.

According to textbooks in numerical taxonomy, a more traditional way to treat multiple (m ¿ 2) possible non-numerical values is to rewrite the m-valued property as m-1 Boolean properties. As an example, if m = 4, the values of the 3 derived properties for each of the possible original values can be 0 0 0, 1 0 0, 1 1 0 and 1 1 1. However, this approach should be treated with caution for CBR, where getting the significant variables right and avoiding double counting and overweighting or underweighting are all important. The other methods above are preferable.

Avoiding double counting means not using A and B as separate properties or dimensions if current knowledge about the items says that there is some clear dependence f between them. That is, if, given A, f(A) determines B. Then, only one of them should be chosen as a distinguishing property in the evaluation of distances between items. Of course, it may be that A and B seem independent now but will prove to be interdependent in the light of future scientific developments. But those developments are likely to be revealed first by inconsistencies between present and future data, e.g. unsatisfactory results from CBR on all the cases available at some future time. While results continue to be of good quality, this is circumstantial evidence that no significant errors of double counting have been made.

When all the properties of an item have been reduced to numerical values, one further step is needed before the computation of distances between items can begin. This is the normalisation of the values: scaling and translation so that they can contribute on an equal footing to that computation. The basic normalisation is to ensure that, for example, a qualitative property whose numerical reduction according to the fuzzy-logic formulation lies in [0.0, 1.0] is treated equally with another property, explicitly quantitative (e.g. power of a radio transmitter), whose range in the given units (kW) and the given data may be [0.1, 500]. In effect, there may be secondary changes in normalisation when weights for various dimensions in a distance (metric) function are changed to ensure that CBR used on a training set of problems gives correct results. However, this is a part of the determination of an appropriate metric, which belongs in section 2.2.

2.2. Determination of a Metric

If an item A is specified by n numerical properties $a_1, a_2 ...$, it can be regarded as located at a point in an n-dimensional space. If there is no reason to believe that the space has any irregularities, As distance d from any other item B in the space is given by some simple metric function of the differences between corresponding coordinates, where the functional treatment of all the coordinates is the same, i.e.

$$d(A, B) = F[k_1 G(a_1 - b_1) + k_2 G(a_2 - b_2) + ... + k_n G(a_n - b_)],$$

and where the quantities k are constants, typically 1 unless changed by the training activity just mentioned. It is rare to find any reason in a taxonomic application to use any expression other than a Manhattan (F(x) = 1, G(x) = —x—) or Pythagorean (F = square root, G = square) metric. The latter is used in this application.

Evidently, impressions of relative distances of complex items (like cases) A, B ... from each other in some single notional space are subjective, even though there should be broad agreement among experts as to which items are close (similar) to any A and which items are far from A. In effect, this is because experts knowledge, in subjects not reducible mainly to general principles, is incomplete or partly implicit or both. In such subjects, apprentices learn by trial and error, and by adjusting their estimates of similarity. Even experts fine-tune their knowledge in the same way, in the face of new experience.

In numerical taxonomy, this process is equivalent to adjusting the values of the constants k in the metric function d above. Given a set of items A, B ... , one chooses a subset that is reasonably representative, and adjusts the constants until the values of d reproduce the knowledge about relative degrees of similarity of the items acceptably. In situations where new items arrive often, this subset may be the existing set itself. Because the point of the exercise is to find a metric that has predictive power and that can be tested, the subset should be small (analogous to a training set in machine learning) if new items cannot be expected to arrive often. Prediction then becomes computation of distances for the items not in the training set, and testing is examination of these distances by experts.

This rather subjective procedure has one constraint. Even if the determination of a metric involves only a subset of the items, it is advisable to compute and examine the metric for all pairs of items. That is because a necessary condition for any metric to make sense is that the so-called triangle inequality must always be satisfied: for any A, B and C, d(A,B) + d(B,C) cannot be less than d(A,C). (Consider the impossibility of making a map containing towns A, B ... , given a table of distances between towns where the triangle inequality is violated for even one triple A B C). Computationally this is not a burden for typical CBR, where the number of cases is likely to be of the order of 100 (rather than, say, 1000), because recomputation simply involves elementary operations with changed values of constants k. The values of properties of cases do not change.

3. The Frequency-Allocation Problem in Shortwave Radio Broadcasting

Shortwave broadcasting occurs in limited ranges or bands of frequencies. In the most popular bands, international demand from broadcasters exceeds the supply of frequencies, which means that listeners can experience problems of interference between unwanted signals and stations they wish to hear. Even so, the situation would be much worse if an administrative procedure to coordinate allocations of frequencies did not exist.

The basic procedure is that broadcasters in a country present their requests for frequencies at least annually to their national PTT (or equivalent more modern name) authority, which then deals with a part of the International Telecommunications Union on their behalf. All the eventual frequency assignments are lodged with the IFRB (International Frequency Registration Board), which makes them quasi-official by publishing them.

The present way of resolving different national requests that are contradictory, e.g. proposals from Russia and the USA to broadcast with transmitters of high power to the same geographical area on the same frequency at the same time, is not systematic, and sometimes obscure. In practice there is some informal consultation and negotiation (outside the national and IFRB administrative procedures) between broadcasters, checking of current and past experience as a means of avoiding obvious sources or risks of interference, etc. Negotiation is helped by the fact that the specialists and consultants in frequency management are not numerous and generally know where to find each other internationally. (It is not helped by the fact that some stations and national PTT authorities pay no attention to the basic procedure, fail to submit up-to-date information, or are even not fully aware of what shortwave broadcasting - licensed or unlicensed - exists in their countries).

Despite the best intentions of the participants, assignments of frequencies can still lead to problems for listeners: unreliable reception because the received signal is not strong enough or regular enough, and interference. Some instances of these problems are outside the broadcasters control, e.g. when listeners receiving equipment is of low quality, which is an economic fact of life in much of Asia and Africa. Frequency allocation involves many factors, including factors of this kind, which resist objective modelling. It is therefore as much an art as a science.

A good source of evidence for both the art and the science, and at the same time a running record of the changes that have occurred annually, is the Passport to World Band Radio [11] handbook (PWBR). Its listings, ordered by frequency, are in an easy-to-read graphical format that can show convincingly why the risk of interference is always present. They have been presented in the same format for 20 years. Also, they rely on actual monitoring of the shortwave bands and not only on information available to or from the IFRB.

Explicitly the PWBR listings are data, but they also signify much implicit knowledge. For example, there are assignments of one frequency to two or more stations at the same time, where the justifications can be expressed as expert-system-like rules (e.g. a station with a purely national audience and with no more than 1 kW of power can coexist on a frequency with any stations located in other continents - though even this has exceptions, which can be stated in modified rules that are less easy to understand intuitively) or through numerical computations based on the physics of electromagnetic radiation in the ionosphere [1,3]. But the most important implicit knowledge refers probably more to the art than the science of the field. Because of this, it is case-like.

In principle, use of methods derived from electromagnetic physics eliminates what a listener would regard as mistaken allocations. Yet there are changes from year to year in PWBR that cannot be traced to objective causes such as a need to move from one band to another to optimise the signal received in the target area for the listeners during variations in the 11-year sunspot cycle [1], or political decisions to cease broadcasting (e.g. the BBC abandoned its broadcasts to North America in 2001, and the Danish national radio organisation withdrew completely from shortwave broadcasting at the end of 2003). That is, some assignments turn out to be mistakes, or at least poor choices.

With the help of practical experience of shortwave reception, it is possible to examine the PWBR listings for any year and identify allocations on any frequency that are potentially poor or mistaken. (This does not necessarily mean that the experts have faltered: e.g. sometimes they are unaware of a source of interference because the PTT authority in its home country has kept no record of its existence and has therefore not told the IFRB about it). A simple test of this identification is to compare the data with the PWBR listing for the same frequency in the next year. If the questionable allocation has changed and there is no apparent extraneous reason for the change, the two situations plus the reasons for the identification cover exactly the kinds of information represented in a case as used in case-based reasoning in artificial intelligence.

The process can be iterated over three or more years if there is still an apparent weakness after an observed change. When this happens, it leads to an enrichment of the knowledge held in the case base.

4. Some Considerations of Case-Based Reasoning

4.1. Indexing versus Metrics

Indexing is a basic topic in CBR, for the same reasons as in the administration of libraries and archives. Kolodner [9] presents the most detailed coverage of case indexing, which amounts to recommendations on how to choose good sets of index terms, plus (in chapter 6.4) a description of one useful method [14]. The same kind of approach also occurs in the teaching of cataloguing procedures to librarians (where experts often resolve problems of ambiguity in classification by referring to past *cases* of cataloguing similarly troublesome items!).

Choice of the properties that figure in metric functions in numerical taxonomy should respect the same recommendations. In effect, a request for retrieval of cases is a set of desired values of the relevant properties, i.e. a specification of an ideal item A in the metric above. Cases B in a case base are then retrieved if their d(A,B) is minimal or below some threshold of distance.

A metric function does the same job as an index. In principle, and in practice for the shortwave radio application, it can therefore operate on an unindexed case base. Having an index in addition to a metric function gives the advantage of faster computation when a case base is large. Otherwise (e.g. when multiple agents with different criteria of significance and therefore different metrics need to access the same case base), relying simply on numerical taxonomy and diverting the effort formerly devoted to indexing into defining good metric functions is an attractive alternative for this part of CBR.

4.2. Case Structure

Syntactically a case is simple, as various textbooks (e.g. [9], with the cooking example on page 172) imply. The notation is familiar to anyone with experience of predicate calculus or the Prolog programming language.

The part that states a problem expresses it as a set of assertions of the form ¡f x y ...¿, where f is a predicate or an instruction to do or achieve something in the solution, and x, y ... are constants needed for the evaluation of f. The part that contains the solution reads like a plan (e.g. see [12], section IV), i.e. an ordered set, in the same notation, ideally with its components labelled to allow easy reference from a third part. Where this third part with a critique or comments on the solution is included, the notation is again the same, and some of its x,y ... constants are labels for the components in the solution that are being assessed.

As mentioned in section 3, the first part of a case in the present application is a record of the data on radio stations on a particular frequency that are listed in two successive years of PWBR plus a statement of the problem and what items in the data are responsible for it. Also, indications of whether the problem in the first year has been changed (e.g. removed) in the second year, and which agents are concerned with it, are given. The solution part of the case states what actions are most likely to have caused the changes, and a skeleton scenario for the negotiation (described in detail in section 4.3) that could have led to the choice of the actions. A scenario states which of the agents mentioned in the first part could have been active at each iteration of the negotiation. This is necessarily an imaginative exercise (though informed by significant past experience with shortwave radio reception), because there is no practicable way to discover what negotiations, if any, took place during the annual activities described in section 3.

4.3. Where do Agents come in?

There are multiple players in the frequency-allocation game whose annual endpoints are published by the IFRB and included (except where they are contradicted too sharply by reality) in the PWBR listings. Usually a player represents (the interests of) one broadcasting organisation. Here, an agent represents one specialised view of shortwave reception, which sometimes corresponds with just one of those interests.

There are differences in the players interests and expertises, but they have two things in common: their knowledge rests ultimately on particular experience, i.e. case-like knowledge, and the basic cases and their contents are available to all, e.g. in IFRB publications. Where they differ is in their interpretations of cases - a case that contains a good outcome for one interest may be bad from anothers point of view, or irrelevant to anothers expertise - and in their views of the publicly-recorded rationale for the choice of some part of the solution expressed in a case. (When the official minutes of a contentious meeting are published, some of those attending the meeting, while not disagreeing that what was actually decided has been recorded correctly, often have their own strong opinions on what the rest of a correct set of minutes should have said).

With the help of numerical taxonomy, the former of these differences can be represented by allowing each player (agent) to have its own metric over properties of the cases. Each metric is refined independently of all the others, by the method described in section 2.2. In the shortwave example, for simplicity, the case base is common to all agents, but if in some applications it is natural to have additional cases that are private to individuals or groups of agents, the same CBR and taxonomic procedures can still be used on those cases in the same way.

Here and in many other multiple-agent applications where agents with different goals and perspectives may want different things, negotiation leading to agreement on what should actually be done is the central part of typical MAS computations. Most MAS research on negotiation (e.g. [10,13]) relies on market and economic analogies, which require fairly direct expression of economic criteria like payoff functions. But in applications where experts do not reason naturally in such terms, and state their preferences merely by referring to similarities between a current situation and their past experiences, metrics and cases are much better suited to capturing their knowledge for use in agents.

The agents that participate in the negotiation represent different expertises, leading to different views about how much of a problem (including none) the situation creates. In the shortwave application, 18 agents are used, representing knowledge on: low-power local/regional broadcasting, risks of ionospheric disturbances, risks of auroral disturbances for trans-Polar broadcasts, high-power international broadcasting organisations, listeners with insensitive receivers, listeners with receivers that cannot separate adjacent frequencies cleanly, and conditions for reception in 12 parts of the world.

An agents initial contribution to the negotiation is I recommend assignment A, and in support of it I bid case C whose distance from the present situation according to my metric is D. In effect, this contains knowledge as well as a numerical bid, and other agents can make use of it, e.g. in changing their own metrics eventually to take revised account of C (although this kind of learning has not been attempted in the present work), or keeping records of agents performance which can show correlations between their bids and those of their recommendations that were worth accepting because they did not cause problems to the other agents.

With respect to a given state of the problem, after one round of bids there will be a set of case and distance data C and D supplied by the bidding agents. Each of those agents then computes the distances, according to its own metric, to each case C bid by any other agent. Although simple to understand, this information admits many ways of continuing the negotiation - potentially a rewarding field for more research.

In the present work, if the metric of an agent x whose distance for its own bid is D gives a value less than tD (t = 1.1 at present) for any case bid by another agent, that case is also regarded as acceptable to x. Further, a weighted mean distance W associated with each case among the bids is computed, with the bidders distance to the case being 2 if there are up to 4 bidding agents or n-2 if n ¿ 4 agents are bidding. The case(s) mentioned the smallest number of times in the process above, and the case with the largest weighted mean distance, are eliminated. Any agent whose associated cases have all disappeared as a result is regarded as bidding the member of the remaining mentioned cases with the smallest distance according to its own metric, but with that cases associated W as the distance included in its bid. If no agent has lost all its cases, there is a further step or steps of elimination, as above. The entire process is repeated until one case survives. The solution adopted is then the solution proposed by the agent that justified it by quoting the surviving case in its initial bid.

5. The Frequency-Assignment Application in Practice

The most crowded shortwave broadcasting bands, thus the best sources of frequency-assignment problems, are around wavelengths of 49 and 31 metres. These include frequency ranges of 6015-6160 and 9565-9710 kHz respectively. Most stations operate on frequencies that are multiples of 5 kHz.

Each such frequency in these ranges has one case associated with it, drawn from the data published in PWBR in 2001 and 2002 (and, when further non-trivial changes have occurred there in 2003, an additional case relates the 2002 and 2003 data). A case states the assignments on its frequency, the changes from the first year to the second, possible defects in the assignments, a possible skeleton scenario (see section 4.3) for some assignments, and the most likely reason(s) for changes. The reasons are supplied by the author, who is not a professional in frequency assignment, but has about 40 years of experience in monitoring of the shortwave bands and 20 years of association with PWBR as a consulting editor.

The 60 frequencies selected have given rise to a case base containing 74 cases.

5.1. Some Examples of Frequency Allocation

The negotiation process described in section 4.3 has been tested, for given frequencies, by the input of some or all of the assignments on each of those frequencies in the 2002 PWBR plus invented new requests that would have caused problems if the new and old items had been assigned the same frequency in 2003. Each problem is resolved by the assignment of some of the broadcasts to different frequencies. These resolutions were then checked against the actual PWBR entries for the relevant frequencies for 2002 and 2003. While the subject has no clearly right answers to serve as benchmarks, all the results made sense and none contained anything that was evidently dubious or wrong.

Identification and handling of problems is illustrated by three examples.

After metric-based retrieval of cases of where interference between stations broadcasting to and those broadcasting in the same general geographical area was significant in the determination of similarity, the fact that Deutsche Welle used an Irkutsk relay transmitter in 2002 in Chinese from 2300 to 2345 UTC on 9645 kHz while a Chinese domestic station in Beijing was also in operation there was identified as a possible problem for the listeners. 9605 kHz was suggested as a problem-free alternative. Deutsche Welle seemed to agree about the problem: in the 2003 PWBR listings, this programme had moved - but to 9690 kHz, which is occupied by yet another Chinese station. If the relay is intended for the coastal provinces, the more westerly location of that station, in Xian, may have removed most of the original problem, e.g. as Deutsche Welle is still scheduled to use 9690 kHz in 2004. Nevertheless, the recommendation of 9605 kHz, through CBR, appears to be an even better solution, with no problem of interference implied by either the 2003 or 2004 listings.

Anoher problem of interference from a listeners point of view occurs at 6150 kHz in PWBR for 2002. In this instance, the actual similarity is indicated mainly by a different dimension in the metric, which measures potential interference for listeners in a region R qualitatively by a combination of transmitter powers and distances of stations not broadcasting to R. The problem is that a 10 kW programme from Colombo in Tamil for northern Sri Lanka from 0200 to 1000 UTC has to compete with regional broadcasts, not intended for Sri Lanka, from Singapore with 250 kW and Iran with 500 kW. The alternative found, free from interference in both 2002 and 2003, was 6090 kHz. The Sri Lankan broadcast appears to have been abandoned after 2002: there is no obvious alternative assignment for it in 2003 or 2004. It may be that the broadcast was ended after a truce in the fighting in the north. But if it happened because shortwave seemed to be ineffective in reaching the Tamil audience, the 6090 kHz alternative suggests that that was probably a premature conclusion.

Radio Veritas, a 10 kW church-sponsored station in Liberia, was able to resume broadcasting from 0600 to 1700 UTC after the level of unrest there had fallen low enough to give it some degree of safety. An allocation for it is published in the 2002 PWBR. For the present work, this was treated as a new assignment exercise: for operation in the 49-metre band, what frequency would make a case as distant as possible from the cases that represented problems? The answer, still good in 2004, was 6115 kHz. An obvious reason why 6090 kHz was not competitive was that it was occupied at the same hours by a station in Kaduna with an intended coverage of (at least) all of the north of Nigeria and therefore the high power allocation of 250 kW. But in fact the PWBR entry for 2002 puts Radio Veritas on 6090 kHz. It is strange, but the same entry continues in 2003 and 2004. Either CBR is not as good as human expert knowledge here, or the Liberian audience must live very close to the transmitter or be remarkably patient. (It may also be relevant that some organisations with limited budgets use old crystal-controlled transmitters, with each quartz crystal providing just one frequency, and own very few satisfactory crystals).

These examples illustrate the more general point that CBR where cases are accessed with the help of metrics rather than indexes is capable of making suggestions that are reasonable (or, at worst, require expert-level refutation) and that deploy its case knowledge well.

6. Conclusion

This paper reports work amounting to a proof of concept for two ideas:

* that numerical taxonomy is a useful tool, whose content is easy to understand and whose relevance is easy to justify, for case-based reasoning,

  and

* that when particular or episodic knowledge is significant in an area where autonomous-agent software designs seem good for problem-solving, agents can exploit this knowledge effectively when it is expressed in cases and handled with the help of numerical taxonomy.

In addition to showing how to implement the ideas, the paper identifies related topics that deserve further research: development of metrics on (case) knowledge bases, especially to represent different views of the same knowledge; use of cases by agents as parts of their bids or proposals in multiple-agent negotiation; case-based learning among agents, including acquisition of knowledge from case information that individual agents quote when they are negotiating; finding good general frameworks for negotiations that involve cases.

References

[1] G. Braun. 1986. Planning and engineering of shortwave links. John Wiley & Sons Ltd.

[2] R.A. Brooks. 1986. A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation 2, pp. 14-23.

[3] J.R.G. Corbett. 1997. Where there is no telephone. Baptist Missionary Society, Didcot, England.

[4] S. Dietmann and L. Holm. 2001. Automated identification of structural homology. Nature Structural Biology 8, pp. 953-957.

[5] L. Holm and C. Sander. 1996. Mapping the protein universe. Science 273, pp. 595-602.

[6] M.H. Huhns and L.M. Stephens. 1999. Multiagent systems and societies of agents. Chapter 2 of Weiss (reference [17]), pp. 79-120.

[7] G.J. Klir and B. Yuan. 1995. Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall.

[8] A.S. Kolaskar and P.S. Naik. 2000. Online identification of viruses. Journal of Microbial Immunology 33, pp. 69-78.

[9] J. Kolodner. 1993. Case-based reasoning. Morgan Kaufmann.

[10] S. Kraus, J. Wilkenfeld and G. Zlotkin. 1995. Multiagent negotiation under time constraints. Artificial Intelligence 75, pp. 297-345.

[11] L. Magne (ed). 2004. Passport to world band radio. IBS Ltd., Penns Park, Pennsylvania. Earlier annual editions have been published since 1984.

[12] S. Russell and P. Norvig. 1995. Artificial intelligence: a modern approach. Prentice Hall.

[13] T.W. Sandholm and V.R. Lesser. 1997. Coalitions among computationally bounded agents. Artificial Intelligence 94, pp. 99-137.

[14] R.C. Schank and R. Osgood. 1990. A content theory of memory indexing. Technical Report no. 2, Institute for the Learning Sciences, Northwestern University.

[15] M.P. Singh, A.S. Rao and M.P. Georgeff. 1999. Formal methods in DAI: logic-based representation and reasoning. Chapter 8 of Weiss (reference [17]), pp. 331-376.

[16] P.A. Sneath and R.R. Sokal. 1973. Numerical taxonomy: the theory and practice of numerical classification. Freeman.

[17] G. Weiss (ed.). 1999. Multiagent systems: a modern approach to distributed artificial intelligence. MIT Press.

John A. Campbell
Department of Computer Science
University College London
Gower Street
London WC1E 6BT, England
jac@cs.ucl.ac.uk