Semantic Pattern Matching in NetBases

M A Heather, Sutherland Building, Newcastle Polytechnic
Newcastle upon Tyne, England NE1 8ST
B N Rossiter, Computing Laboratory, Newcastle University
Newcastle upon Tyne, England NE1 7RU
B.N.Rossiter@newcastle.ac.uk

1991

Abstract

NetBases are the counterpart of databases but store information dynamically rather than passively. Meaning is embedded in the syntactical ordering of the components of the information system. In general the components are abstract entities which are representable in forms appropriate for human signification. In a business textbase, the components are ideas and concepts represented in symbols arranged to make words. In CAD/CAM information systems, they are designs, shapes, manufacturing processes, etc, stored in graphical form or algorithmic procedures. In banking, they are transactions represented in traditional database methods. In NetBase systems like those to be found in commerce and industry, a semantic query recovers information by a process of pattern matching on meaning.

The theory of NetBases may be represented as a directed net system in a topological information space. An information net is a pair $(\Gamma, \geq)$ consisting of a function Γ whose domain is directed by the ordering $\geq$. The family of all neighbourhoods of any point in the space is ordered by inclusion $\subset$. A finite ordered sequence constitutes both the address and the information content itself in a generalised form – the concept or idea. Recovering the meaning by pattern matching at the semantic level is achieved by following paths of association in the directed net system to reach convergence– the neural process of recollection.

1 The NetBase Memory

The NetBase is the neural-model counterpart to databases on conventional computers. Like databases the NetBase stores information by providing an organized view of data but a more flexible one. The organization however is conceptually quite different. A traditional database is a reductive tool while the NetBase is integrative.

There is a whole class of problems which amount to pattern matching at the semantic level in information systems with a wide range of applications in challenging areas where little progress has been made with traditional techniques. NetBases belong to a quite different methodology than databases and may be able to take over in application areas where databases fail [2].

2 Neural NetBases and Databases Compared

2.1 Inadequacies of Databases

The prime objective of a database management system is to make application programs independent of the physical structure of the data. A conceptual schema or model with a global logical definition of the data structure relates to the internal (physical) definition by a mapping from the logical level to the physical level. This provides a very clean environment for pure data with advantages of centralized control, data independence, the avoidance of redundancy, a common access language, standard access for all users, security controls, integrity checks, customized user views, etc. However it has turned out, rather as with analogous work with expert systems, that this clean environment is too clinical and too sterile for real world applications dealing with heterogeneous real-world data of any size and complexity.

The main problems are that:

- the assumptions of atomistic and independent data are invalid;
- the rigid schemata have to be completely designed initially which distort and tightly constrain all future processes;

2.2 Neural NetBases as an Alternative to Databases

Any machine emulation requires a model of the processes to be carried out and usually some model of the structure of the data. It is sometimes said that the neural model is a model without a model. It is true that it does not require any prior model of the data structures or of the processes which are to take place that are usually needed in any machine emulation. Except perhaps that all cells can be connected together at least indirectly and influence one another. This greatly advances the traditional database network model in a different direction without having to rely on artificial constructs needed in hierarchical and relational models. Neural-NetBases can now be thought of as an alternative rather than a successor to Databases.

The neural-model characteristics important for the functions of memory and recall in NetBases include the following:

- the merging and combining of data and procedure – the prescriptive with the descriptive;
- contentious redundancy and contradictions both contained and utilized;
- the resilience and fault tolerance;
- the differentiation of permutable orderings;
- the responding to context-sensitivity;
- inscrutable rationality based on intuitionistic logics;
- full recognition of non-commutation and cross-coupling effects;
- no requirement for an aprioristic data model;
- the capability for self-organizing and unsupervised learning;
- the particularism within a holistic approach;
- the faculties of an associative content-addressable memory; and
- computational asynchronous concurrency and mutual communication.

2.3 Examples of Neural NetBases

It seems that there is a class of medium that exhibits the above characteristics. The most notable seems to be the cerebral medium itself. While neural behaviour and processes may be the most celebrated there are many others which are more accessible to investigation. One is natural language, which shows many, if not all, of these properties. Textbases where the components are ideas and concepts represented in symbols arranged to make words are becoming more and more important as a storehouse and workshop for business information. Textbases stretch databases to their limits even with added object-oriented features [5]. Natural language is an external manifestation of mental processes and thought that can exist in the human brain. It has been studied intensively but is still far from being understood.

Other examples analogous to the cerebral medium are only just emerging in an external form. In CAD/CAM information systems, there are designs, shapes, manufacturing processes, stored in graphical form or algorithmic procedures. These consist of reusable generic features, algebraic pen positions, scale factors, rotation operations about arbitrary axes, etc. The abstracted forms all have meaning and purpose at what is termed in natural language the semantic level. This is the semasiological use in linguistics of the term 'semantics' rather than the way it is employed with traditional computer languages.

In banking, there are transactions represented in more usual database terms but within them lies a semantic layer of information embedded in the relationships between the transactions. So for instance, spotting fraud on an EFT/POS system requires a comparison of individual behaviour with habits of the general public. This is an operation at the semantic level. In NetBase systems like those to be found in commerce and industry, a semantic query recovers information by a process of pattern matching on meaning.

3 Theory for Semantic Queries

3.1 Non-Classical Semantics

Classifications such as syntax, semantics and pragmatics are reductionist aids [1] which merge together in the neural net methods. Pattern matching at the semantic level is in essence the same process as that already successfully demonstrated at the syntactical level with speech, handwriting and so on, where physical form is represented in the structure and behaviour of the components of a connectionist system.

Lying at the root of many of the characteristics outlined above is their non-linearity arising from the fundamental point that data is not independent. Most formal models involve assumptions that lose information because they cannot recognise this dependence, the overlap, the redundancy that provide the non-deterministic features of information. Classical logic, statistics and any model relying on real number axioms are deficient in this respect. However, some recent branches of mathematics like topology can be employed which do not need these classical assumptions.

3.2 Directed Neural Net Information Systems

We can be general by defining a directed neural net information system as a topological space $(X, \Im)$ where X is a set and $\Im$ is the family of open sets of X. X does not need to be finite or discrete. X may be ordered locally with a binary inclusion operator $\subset$. That is a more general ordering than the fully or partially ordered set, taking the usual definition for a neighbourhood, of an element $x \in X$, as any set (which need not be open) that contains an open set of which x is a member. The $\subset$ directs the family of all neighbourhoods of any point in the space, because the intersection of two neighbourhoods is a neighbourhood which follows both in the ordinary $\subset$. An information net is a pair $(\Gamma, \geq)$ consisting of a function Γ whose domain is directed by the ordering $\geq$ [3].

The requirements for an ordering on the set involving a binary relation $\geq$ that directs the set are:

1. If x, y and z are members of X such that for all $x \geq y$ and $y \geq z$ then $x \geq z$ (transitivity);

2. If for $x \in X$ then $x \geq x$ (reflexivity); and

3. If x and y are members of X, then there is z in X such that $z \geq x$ and $z \geq y$ (special property).

This provides the syntactical ordering of the elements which constitutes the associative semantic pattern whether the elements are characters of the alphabet, banking transactions, CAD/CAM representations, etc. An ordered sequence of such elements constitutes an address (in database terms) and also the data itself shown in a generalised form – the concept or idea. This relationship between the address and its contents as an identity mapping is an essential point made by Kohonen [4]. The actual process of recollection is to follow the paths in the directed net system to reach convergence. This is a Moore-Smith convergence, details of which can be found in [3].

References

[1] Heather, M A, Legal Structures for Law Machines: in C. Ciampi **1** 117-127, Artificial Intelligence and Legal Information Systems. North Holland, 1982.

[2] Heather, M A, Normative Neural Nets in a Connectionist Model of the Common Law. 3rd Int Conf Logica Informatica Diritto, Expert Systems in Law. Florence, 1989.

[3] Kelley, John L, General Topology, Van Nostrand, Princeton 1955.

[4] Kohonen, Teuvo, Self-Organization and Associative Memory, **8**, Springer Series in Information Sciences, Springer-Verlag, Berlin, 1984.

[5] Rossiter, B N, & Heather, M A, Towards the Object-Oriented Textbase, Newcastle University Technical Report Series no 297, Newcastle upon Tyne, 1990.