Case Retrieval Nets: Basic Ideas and Extensions

Mario Lenz and Hans-Dieter Burkhard

Dept. of Computer Science, Humboldt University Berlin, Axel-Springer-Str. 54a, 10117 Berlin, Germany, Email: {lenz,hdb}@informatik.hu-berlin.de

Abstract. An efficient retrieval of a relatively small number of relevant cases from a huge case base is a crucial subtask of Case-Based Reasoning. In this article, we present Case Retrieval Nets (CRNs), a memory model that has recently been developed for this task. The main idea is to apply a spreading activation process to a net-like case memory in order to retrieve cases being similar to a posed query case. We summarize the basic ideas of CRNs, suggest some useful extensions, and present some initial experimental results which suggest that CRNs can successfully handle case bases larger than considered usually in the CBR community.

Keywords: Case-based Reasoning, Case Retrieval, Spreading Activation

1 Introduction

Within the field of case-based reasoning, a major area of research in recent years has been the development of techniques allowing for an efficient and yet flexible retrieval of relevant cases. This has led to a number of sophisticated techniques for this subtask, as for example indexing techniques ([13]); kd-trees ([17, 18]); the heuristic "Fish-and-Sink" approach ([14]); the CRASH memory model ([3]); and Knowledge-directed Spreading Activation (KDSA, [19]).

As an alternative technique especially suitable for the design of decision support systems ([4, 8]), we developed the concept of Case Retrieval Nets (CRNs) as a memory structure supporting efficient but nevertheless flexible case retrieval. In contrast to ABS (an earlier model using spreading activation, cf. [9]), CRNs are more generally applicable and do rest on a sound formal framework which allows for the investigation of properties. Besides the theoretical investigations, we already performed a number of empirical tests with CRNs. The complete documentation of the results can be found in [7].

Before we go into the details about CRNs, we would like to outline what we consider as requirements for a case retrieval technique. Basically, there are three conditions to be met: efficiency, completeness, and flexibility (cf. also [3, 17]):

**Efficiency** concerns the effort required to access relevant cases. Access of these cases should avoid exhaustive search in memory.

**Completeness** assures that every sufficiently similar case in memory will be found during retrieval. (The corresponding problem of correctness is often solved as a secondary selection step over the retrieved cases.)

**Flexibility** expresses that there are no inherent restrictions concerning the circumstances under which a particular piece of knowledge can be recalled.

In general, these conditions express contradictory goals: Most techniques for pre-structuring the case memory increase efficiency but limit the flexibility.
2 Basic ideas of Case Retrieval Nets: An Example

The most fundamental item in the context of CRNs are so-called Information Entities (IEs). These may represent any basic knowledge item, such as a particular attribute-value-pair. A case then consists of a set of such IEs, and the case base is a net with nodes for the IEs observed in the domain and additional nodes denoting the particular cases. IE nodes may be connected by similarity arcs, and a case node is reachable from its constituting IE nodes via relevance arcs. Different degrees of similarity and relevancy may be expressed by varying arcs weights. Given this structure, case retrieval is performed by

- activating the IEs given in the query case,
- propagating this activation according to similarity through the net of IEs,
- and collecting the achieved activation in the associated case nodes.

Fig. 1. Example of a CRN in the TRAVEL AGENCY domain.

The idea is illustrated for the TRAVEL AGENCY domain ([6]) in Figure 1: A case is a special travel offer, denoted by a case descriptor, e.g. <Offer 20219>. It consists of a set of corresponding IEs giving the specification of that offer, in case of <Offer 20219> the IE nodes <Type:Swimming>, <Price:980>, <Place:Matala>, <Region:Crete>, <Distance to beach:500 m> are associated to that case node. Asking for an offer in Crete for swimming and not too far from the beach, the IE nodes <Type:Swimming>, <Distance to beach:200 m> and <Region:Crete> are initially activated. By similarity, the IE nodes <Region:Malta> and <Distance to beach:500 m> will be activated in the next step, but the amount of activation depends on arc weights. Finally, the three offers <Offer 20024>, <Offer 20219>, <Offer 500122> will each get some activation depending on the incoming activations of IE nodes and their relevances. The highest activated cases may be proposed to the customer. A first list of proposals might include alternative solutions which are pruned after the customer decided for either of them.

This simple example already points out some features of the memory model: