Title

Planits: The Case-based Reasoner As A Planning Tool For Intelligent Transportation Systems

Authors

Khattak, Asad
Kanafani, Adib

Publication Date

1995

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA. BERKELEY

PLANiTS: The Case-based Reasoner as a Planning Tool for Intelligent Transportation Systems

Asad Khattak
Adib Kanafani

California PATH Research Report
UCB-ITS-PRR-95-23

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

July 1995

ISSN 10551425

Institute of Transportation Studies
University of California at Berkeley

PLANiTS: THE CASE-BASED REASONER AS A PLANNING TOOL FOR INTELLIGENT TRANSPORTATION SYSTEMS

Asad Khattak

Adib Kanafani

109 McLaughlin Hall
Berkeley, California 94720

Phone (5 10) 642-3585
Fax (510) 643-5456

Submitted to:

The California Department of Transportation

(MOU-52)

Draft Final Report

Volume 3

March 8, 1995

ABSTRACT

This paper develops a data synthesis methodology in PLANiTS (Planning and Analysis Integration for Intelligent Transportation systems) using case-based reasoning. The reasoner contains mechanisms for matching, ranking and analyzing past cases in relation to current cases. The current cases consist of transportation improvement actions, performance measures and environments defined in terms of spatial, temporal and user dimensions. PLANiTS users can apply increasing levels of stringency to match cases. We also discuss issues related to computer implementation and the limitations of case-based reasoning.

Key words: transportation planning; case-based reasoning; intelligent transportation systems.

SUMMARY

Intelligent transportation systems may support traveler decisions and improve transportation system performance. To plan for systematic testing and deployment of such systems, a framework for a new transportation planning methodology known as PLANiTS is being developed. An important component of the methodology deals with synthesizing and processing existing knowledge about transportation improvement actions, including new transportation technologies, and presenting it to individuals involved in planning. The intention is to support future planning decisions which will inevitably deal with deployment of new technologies.

To synthesize and process existing transportation knowledge, we propose the use of Case-Based Reasoning (CBR), a relatively new paradigm in Artificial Intelligence. Given a planning context where individuals are exploring the impacts of transportation improvement actions in terms of certain measures of performance, the Case-Based Reasoner determines the similarity of past cases to the present situation, retrieves relevant cases from computer memory and informs the user about the impacts of past solutions (transportation improvement actions) and their success or failure. For example, if the impact of Advanced Transportation Management and Information Systems or ATMIS is being considered on system-wide delay, then CBR will determine the similarity of ATMIS cases implemented elsewhere.

The cases are stored in "case base." All cases must contain transportation improvement action(s), performance measure(s) and the context or environment. The user must define the current case in terms of actions, performance measures and context and their spatial, temporal and user dimensions. A matcher will provide a listing of similar cases by accessing them in the case memory. PLANiTS users can match cases at different levels of stringency. A ranker will rank them in accordance with their similarity to the current case. To predict performance measures when several similar cases are retrieved, an analyzer provides estimates; and an adviser compiles evidence from the cases and makes (qualitative) recommendations.

An important limitation of CBR in the context of transportation planning is that no two cases are exactly similar. Judging similarity of cases is often difficult, and sometimes arbitrary, because the methods for comparing cases are not well developed. The paper points out this and other important issues in developing Case-Based Reasoning for transportation planning.

ACKNOWLEDGEMENT

This research was conducted at the Institute of Transportation Studies, University of California at Berkeley. It was funded by the California Department of Transportation, as a part of the PATH Program. PLANiTS team members include: Adib Kanafani, Asad Khattak, Randall Cayford, David Lovell, Thananjeyan Paramsothy, Rosella Picado, Joy Dahlgren, Melanie Crotty, Nicholas Vlahos and Marvin Manheim.

TABLE OF CONTENTS

INTRODUCTION

New transportation technologies offer solutions to the continuing problems of traffic congestion, air quality, safety and accessibility. To understand the extent of new technology impacts, we can learn from similar technology implementations of the past. Specifically, experiences with appropriate and inappropriate technology implementations provide a rich resource. It is important to learn from successes as well as mistakes and failures; an important source of innovative new solutions is probably retrospection about mistakes. In this regard, we propose to use case-based reasoning as a transportation planning support tool that determines similarity of historical cases to the current case, provides information on whether similar past cases were successful and warns against potential mistakes/failures.

Transportation planning occurs in a complex political environment, where people must analyze and judge several aspects of a problem based on available information. For example, before implementing intelligent transportation systems, participants need to know their potential benefits. Currently, there is well-founded skepticism about the role of new transportation technologies among people involved in the transportation planning process. Case-Based Reasoning (CBR) allows participants to be skeptical by examining past mistakes and successes and learning from them. They can obtain insights regarding the role of new transportation technologies and estimate their impacts/benefits. Initially, implementation decisions can be made by examining evidence from (often limited) field operational tests along with experiences from actual implementation of similar precursor technologies. Successful field operational tests and implementations can be replicated in appropriate environments, whereas past mistakes can be anticipated and avoided.

We have proposed a methodology called PLANiTS (Planning and Analysis Integration for Intelligent Transportation Systems) which is a comprehensive tool designed to meet the needs of the emerging planning processes. A detailed description of PLANiTS appears in Kanafani, Khattak, and Dahlgren (1994) and Vlahos, Khattak, Kanafani, and Manheim (1994). PLANiTS is designed to be used by agencies and citizens involved in transportation planning. To complement structured analysis and modeling, we have proposed a knowledge base that synthesizes data in meaningful ways. Specifically, a case base will contain synthesized transportation knowledge. The basic functions provided are the presentation of information regarding previous cases to participants, analysis of transportation improvement action impacts, and the provision of judgements on similarity. PLANiTS case-base reasoner supports learning from experience with transportation improvement actions.

CBR helps participants analyze the impacts of actions in terms of performance measures by displaying similar cases. Matching relevant cases allow users to draw inferences regarding values of performance measures and obtain insights in the current context. Other functions include ranking historical cases in terms of their similarity to the current case, using statistical techniques to estimate parameters for the similar cases, examining subjective information about past cases, particularly past mistakes that can be avoided, and storing new cases in the computer memory.

CBR takes advantage of the existing cases in transportation to seed a case-based system, e.g., experiences with ATMIS precursor technologies such as information collection, processing and dissemination technologies. The case-base becomes richer as the results from field operational tests on new technologies are compiled and synthesized.

The following section provides an overview of PLANiTS and discusses the application of CBR in PLANiTS. The section on conceptual structure discusses issues related to case representation, causal models and similarity. Next the implementation of CBR along with an overview of CBR architecture is provided. Finally, the limitations of the methodology are discussed.

PLANiTS OVERVIEW

PLANiTS Components

PLANiTS has various bases described in Kanafani et al. (1994) and discussed briefly below (Figure 1):

- The Policy and Goals Base contains mandates objectives and constraints communicated in terms of appropriate policy factors to be satisfied and measures of performance to be evaluated.
- The Strategy and Action Base contains a catalogue of possible actions and rules that can recommend competing and associated actions.
- The Data and Knowledge Base provides access to data bases and has knowledge in terms of theoretical and empirically established relationships between transportation objects.
- The Methods and Tools Base contains transportation models and generic methods of analysis, and utilities and tools, e.g., for network analysis.

The glue that binds PLANiTS components together and allows deliberative planning and analysis processes to occur is the Planning Vector, **PV,** which contains three sub-vectors:

- Action Vector, A, which contains the proposed set of actions that are the subject of the planning process.

- Criteria Vector, Y, which contains the measures of performance representing the goals for which the actions are proposed.
- Environment Vector, E, which contains the descriptors of the context that are relevant to the subject actions and impacts.

Thus, we have **PV =** [A, Y, E]

The planning vector allows users to select transportation actions, performance measures and environmental descriptors. Users specify each of these elements in terms of their spatial, temporal and user dimensions and analyze the planning vector with models, case-based reasoning and expert systems. During the process of planning vector specification and analysis, people at different locations can communicate by sending and receiving messages and sharing the planning vector.

The knowledge and the methods bases are used to analyze the elements of the planning vector and the results used to inform the decision making process involved in programming projects.

Case-Based Reasoning in PLANiTS

CBR is part of the knowledge base in PLANiTS. To present cases that can give meaningful solutions and insights, it relies on rules and mathematical operations. CBR determines the similarity of past cases to the current situation, and presents relevant cases stored in the computer memory and informs the user on how similar situations were addressed and whether the previous solutions (actions) were successful in achieving their objectives (Kolodner 1993; CBR 1989, 1990). However, the methods used to determine similarity are not well developed. Judging similarity of cases is often difficult, and sometimes arbitrary (research on this topic is limited, see Kolodner 1993 for a discussion of “preference heuristics”). Furthermore, there is a need to develop a general structure for representing and retrieving cases to support users’ action generation, criteria selection and evaluation activities in PLANiTS.

We use High Occupancy Vehicle (HOV) and Advanced Transportation Management and Information Systems (ATMIS) examples throughout this paper. We choose the HOV lane’s example because such lanes constituted a large portion of the recent planning projects in California and they can potentially integrate new technologies, e.g., real-time rideshare matching systems. Also the idea of priority for certain vehicles can be extended to include AHS (Automated Highway Systems). That is, automated highways may allow certain vehicle types to travel faster than others.

The ATMIS example is appropriate because it represents a set of new technologies that are closest to implementation. Moreover, there is some experience with precursor ATMIS technologies that can seed the case base.

CONCEPTUAL STRUCTURE

A Structure for Representing Cases

A historical case must contain the elements of the planning vector. Cases consist of actions implemented in an environment and their consequent impacts. The three elements of the planning vector, namely actions, environment and performance measures are defined in terms of space, time and user/traveler dimensions. For example, an HOV lane action (or an automated lane of the future) must have a location, a time of operation and a vehicle occupancy threshold. It will influence certain users traveling on a network during a certain time. The environment may consist of the geographic area where the HOV lane was implemented, the time of its implementation and the rideshare opportunities in the area.

The descriptors can be expressed more elaborately as follows:

- Spatial dimensions
  - Nodes and links
  - Technology
    - * Vehicles (type, number demanding service)
    - * Intelligent transportation systems
- Temporal dimensions
  - Yearly, monthly, daily
  - Peak, off-peak
- User/traveler dimensions
  - Traveler decisions (participation in labor force, home/office location, mode, route)
  - Traveler attributes (income, gender)
  - Travel purpose (work, shop)

Actions and performance measures each have a hierarchy. In PLANiTS, cases are indexed according to their hierarchy as illustrated below:

Define action clusters,

$$A_i = \{a_1, a_2, a_3, \ldots, a_i\}$$

In the transportation planning context,

$$A_i = \{\text{roadway, transit, bicycle/pedestrian, intermodal/freight}\}$$

The sub-cluster level is,

$$A_{j|i} = \{a_{1|1}, a_{1|2}, a_{1|3},\ldots, a_{1|i} \\ a_{2|1}, a_{2|2}, a_{2|3},\ldots, a_{2|i} \\ a_{j|1}, a_{j|2}, a_{j|3},\ldots, a_{j|i}\}$$

For example,

$$A_{j|i} = \{\text{HOV|Roadway, Bus priority|Transit,} \\ \text{ATMIS|Roadway, Transit information system|Transit,}\ldots\}$$

The reason for structuring the actions as a matrix (rather than a vector) is that certain sub-types of actions, e.g., intelligent transportation systems, may be represented as rows. Further, actions are specified in terms of three descriptors.

$$[A_{j|i}]^{S',T',U'} = [\text{HOV|Roadway}]^{S_1,\ldots,S_k\ T_1,\ldots,T_k\ U_1,\ldots,U_k}$$

Where,

S' symbolizes spatial dimensions--HOV links $S_1,\ldots,S_p, S_{p+1}$ can be whether HOV lane is separated and S_k can be the number of HOV lanes in this example

T' symbolizes temporal dimensions--times of HOV lane operation $T_1,\ldots,T_k$,

U' symbolizes user dimensions--vehicle occupancy thresholds at different times of operation $U_1,\ldots,U_k$.

For the ATMIS example, this specification is:

$$[A_{j|i}]^{S',T',U'} = \{\text{ ATMIS|Roadway}]^{S_1,\ldots,S_k\ T_1,\ldots,T_k\ U_1,\ldots,U_k}$$

Where,

$S_1,\ldots,S_k$ = network links where ATMIS is implemented,

$T_1,\ldots,T_k$ = times when real-time information is available,

$U_1,...,U_k$ = traveler decisions: equipped with advanced traveler information device or unequipped.

Notice that for each $[A_{j|i}]^{S',T',U'}$ there will be n cases denoted by $[A_{j|i}]_n^{S',T',U'}$. Now define performance clusters,

$$Y_q = \{y_1, y_2, y_3, ..., y_q\}$$

For example,

$$Y_q = \{\text{congestion, air quality, safety, accessibility,...}\}$$

The next level is,

$$\begin{aligned} Y_{r|q} = \{ & y_{1|1}, y_{1|2}, y_{1|3}, ..., y_{1|q} \\ & y_{2|1}, y_{2|2}, y_{2|3}, ..., y_{2|q} \\ & y_{r|1}, y_{r|2}, y_{r|3}, ..., y_{r|q} \} \end{aligned}$$

For example,

$Y_{r|q}$ = {Person-delay|Congestion, Carbon-Monoxide|Air Quality, . . .
travel time|Congestion, Ozone|Air Quality,...}

Further,

$$[Y_{r|q}]^{S',T',U'} = [\text{Person-delay|Congestion}]^{S_1,...,S_k\, T_1,...,T_k\, U_1,...,U_k}$$

Where,

S' symbolizes spatial dimensions--links $S_1,...,S_k$ where person-delay was measured/estimated in the historical case,

T' symbolizes temporal dimensions--times of day $T_1,...,T_k$ when delay was measured/estimated,

U' symbolizes user dimensions--delay disaggregated by HOV eligible and non-eligible $U_1,...,U_k$.

For the ATMIS example, this specification is:

$[Y_{r|q}]^{S',T',U'} =$ $[\text{Person-delay|Congestion}]^{S_1,...,S_k\ T_1,...,T_k\ U_1,...,U_k}$

Where,

$S_1,...,S_k$ = Person-delay on network links where real-time travel information is disseminated,

$T_1,...,T_k$ = Person-delay during peak period/off-peak period incidents,

$U_1,...,U_k$ = Person-delay for ATIS equipped and unequipped travelers.

The n cases are denoted by: $[Y_{r|q}]_n^{S',T',U'}$

The environment, $E_{a,y}$, is the context where actions are implemented and system performance impacts estimated and/or measured. That is, it defines more generally where and when the action was implemented and who did it impact? The environment is a filter that increases in its stringency as illustrated below:

- Acceptable location where the ***historical*** case was implemented may vary from the country (e.g., USA) to the State (e.g., California) on to the particular city where the ***current*** will be implemented.
- Acceptable time when the historical case was implemented may vary from no time constraint to the very recent (e.g., cases implemented within the past few years) to.
- Acceptable population density may have increasingly tighter bounds relative to the current case.

Other information stored with each case is as follows:

- Case description (C,)
- Modeling intensity of study (M,) (Scale: 0=low . . . 1=High)
- Data Source (S,) (measured or estimated through model)
- Data Validity (V,) (Scale: 0=invalid...1=valid)
- Quality of study (Q_n) (Scale: O=bad... l=good)
- Case knowledge (I'_n) (success/failure, mistakes, community reactions to action, lessons learnt)
- Case prescriptions (P',)

  The following prescriptions could be stored with specific cases:

  - The HOV lane implemented in location X was not effective in reducing traffic congestion due to lack of enforcement, therefore, consider allocating sufficient resources for HOV enforcement.

- The ATMIS implemented in area X was not effective in reducing traffic congestion because community opposition to increased neighborhood traffic rendered surplus capacity on alternate routes unusable, therefore, do not plan on using such surplus capacity for route diversions.

If the retrieved cases contain such prescriptions, then users can examine them. If multiple (retrieved) historical cases show similar patterns, then stronger recommendations can be made. Ultimately, PLANiTS CBR will provide a synthesis of multiple retrieved prescriptions.

Relationships between variables

We may have knowledge of important relationships that hold for cases. Relationships can be of the following type (Kanafani, Khattak and Dahlgren 1994):

- R_c, Theoretical causal relationship, e.g., higher modal travel costs reduce the possibility of its use or greater number of workers in a household cause more trips.
- R_a Theoretical a-causal relationship, e.g., technology X dominates technology Y
- R_e Observed (empirical) relationship (may be theoretical or unexplained)

The causal relationships that are present in the planning vector are:

$$[Y_{r|q}]_n^{S',T',U'} \; R_c \quad \{ [A_{j|i}]_n^{S',T',U'} \quad [E_{a,y}]_n^{S',T',U'} \}$$

In words, this implies that when an action (HOV lane) is implemented in an environment (area), it can theoretically cause certain impacts (reduction in person-delay); noting that the three elements are defined in terms of their spatial, temporal and user dimensions. Replacing R_c with R_e would mean that the relationship(s) also holds empirically. The causal relationships reflect a-priori knowledge about the important relationships. Although such knowledge may be limited, it is needed for matching at the descriptor level, as explained below.

Similarity Measures

Similarity is a continuum that can vary from a clone to completely dissimilar. Importantly, similarity depends on the frame of reference. For example, at one level all humans are similar and at another everyone is unique. Similarity comes from sharing certain attributes and behaviors while the dissimilarity comes from differences across attributes and behaviors. For case-based reasoning, the ***degree*** of similarity between current and historical cases and ***the relative*** importance of the planning vector descriptors are relevant.

At the simplest level, a historical case is similar to the current case when it exactly matches the current case at the specified cluster, sub-cluster or descriptor level (Table 1). One can examine the presence or absence of planning vector elements and/or their descriptors to determine similarity. For the current case, PLANiTS users specify the actions $[A_{j|i}]_c^{S', T', U'}$ and the environment $[E_{a,y}]_c^{S', T', U'}$ and define the desired performance measures $[Y_{r|q}]_c^{S', T', U'}$. The current case can then be compared with each historical case. If the historical case, satisfies certain conditions, then it is retrieved and presented. Table 2 presents a systematic way to examine various possibilities for matching.

At the cluster level, one can match A_i; for example, present all historical cases that relate to highway actions; or match Y_q, for example, present all cases that relate to traffic congestion. Matching both A_i and Y_q may retrieve highway cases that must provide information on traffic congestion. Given A_i and Y_q, one can also match the environment filter, for example present all cases that occurred within a region and a time period and relate to highways with traffic congestion as the performance measure. Similarly, cases can be matched at the sub-cluster level with increased stringency. Matching $A_{j|i}$ may, for example, retrieve all HOV lane cases whereas, $Y_{r|q}$ may retrieve all cases that have the person-delay as criteria. The environment filter can increase stringency as shown in the Table. Finally, at the descriptor level, only cases that match exactly on spatial, temporal and user dimensions are retrieved. At this level, the retrieved case must contain the important (causal) dimensions for the action and environment, specified in the current case, as well as the performance measure at the disaggregate level specified by the participants, e.g., person-delay at such and such location and time of day, incurred by certain types of users.

With this structure, the aggregation of performance measures can occur simply (aggregation is needed to get an overall estimate of the impacts). For example, if $[Y_{r|q}]_n^{S',T',U'}$ means delay (Y_r) given congestion cluster (q), for a historical case n, on links (s), during peak periods (t) for HOV eligible users (u), then the total person-delay can be obtained by aggregating:

$$\text{Total delay} = \sum_{S_1,\ldots,S_k} \sum_{T_1,\ldots,T_k} \sum_{U_1,\ldots,U_k} [Y_{r|q}]_n$$

When the variables being matched are continuous, e.g., length of a facility, then exact matches will be unlikely. Similarity can be operationalized in such a situation by examining the distance between the values of descriptors (and then specifying the thresholds/limits of acceptability). One measure of distance between the value of a current case descriptor and stored case descriptor can be written in a non-dimensional form as follows:

$$D_A = |[A_{jli}]_n^{S_k} - [A_{jli}]_c^{S_k}| \,/\, |[A_{jli}]_n^{S_k}| + |[A_{jli}]_c^{S_k}|$$

and

$$D_E = |[E_{a,y}]_n^{S_k} - [E_{a,y}]_c^{S_k}| \,/\, |[E_{a,y}]_n^{S_k}| + |[E_{a,y}]_c^{S_k}|$$

Where, D_A, D_E= Distance between the values of the current and stored cases on the spatial descriptor. These distances can be similarly calculated for the temporal and spatial descriptors.

Another type of similarity relates to relative dimensions. Such similarity is manifested when the proportions between two cases are similar (however, their actual dimensions are different). Cases are judged similar to the current case if the ratios of their respective descriptors are (nearly) constant. That is, whether the following condition is satisfied,

$$[A_{jli}]_n^{S_1} \,/\, [A_{jli}]_c^{S_1} \approx [A_{jli}]_n^{S_2} \,/\, [A_{jli}]_c^{S_2} \approx [A_{jli}]_n^{S_k} \,/\, [A_{jli}]_c^{S_k}$$

This implies that the historical case is proportionally similar to the current case in terms of its physical (spatial) dimensions. The historical case can be judged similar if it is a dimensionally smaller or larger replica of the current case. The same logic can be extended to temporal and user dimensions.

Table 3 shows the use of similarity measures to compute scores when descriptor importance values are known. The objective is to select the closest historical case and use its Y_{rlq} value (person-delay) appropriately aggregated to predict the impacts for the current case. Assume that there are n cases in the computer memory that are candidates for matching at the descriptor level and that the importance of action and environment descriptors in influencing person-delay is known. To estimate the "Similarity Index" scores for each of the n cases, we compute the degree of match between the current and historical case and multiply it by the importance of the descriptor. We sum the scores and normalize the answer by the sum of importance ratings. Then select the case with the highest score. It is important to keep in mind that the matching operations can vary. Other more complicated similarity measures can be formulated, however, at this point our purpose is to identify this as an area of future research. Also, there is a need to develop importance criteria based on theory and empirical evidence.

IMPLEMENTATION OF CBR IN PLANiTS

In the PLANiTS prototype, users define their current case that consists of actions, environment and performance measures in terms of spatial, temporal and user dimensions. Then participants can begin using the case-based reasoner (other options include structured models and expert systems). Users can browse through all available cases by opening an information window about the selected case. Currently, the information that appears includes, a text description of the case (C,), the planning vector elements and M_n, S_n, V_n, Q_n, I'_n, P',.

The participants can adjust the degree of matching stringency for each element of the planning vector. In the prototype, participants can choose three levels of stringency: low, medium and high corresponding to the cluster, sub-cluster and descriptor level matching respectively. The cases are reexamined each time the stringency is adjusted and only the matching cases appear in the list. Figure 2 presents a view of the main CBR window in PLANiTS. Choosing to match action will cause the reasoner to limit cases where, for low fit, the action cluster must match, for medium fit, an HOV or ATMIS case must match and for high fit, an HOV case must match vehicle occupancy level and number of HOV lanes, while an ATMIS case must match in-vehicle or roadside sign parameter.

Choosing to use the environment filter limits cases based on geographic location and temporal dimensions. Low fit matches all cases in the US (given that the current case is always in the San Francisco Bay Area), medium fit matches recent cases in the State (cases in California that are no more than ten years old), while high fit matches very recent cases in close vicinity (cases in the San Francisco Bay Area, that are no more than a couple years old).

Choosing to match performance measures causes PLANiTS to limit cases to those which, for low fit, match the primary category of the measure of performance (e.g., congestion). A medium fit limits cases to those that match the sub-cluster level (person-delay). A high fit limits cases to person-delay on specific links and at certain times.

Following is a description of how cases can be selected for further processing and manipulation:

Rank

The historical cases will be ranked in terms of their similarity with the current case. The rank function can sort historical cases based on aggregate score. That is, case summaries are ordered according to a Similarity Index, which can be based simply on the number of matched descriptors or a combination of the degree of match and importance weights attached to action and environment descriptors.

Analyze

The analyze function allows participants to perform statistical operations on similar cases, such as averaging and summation. The analyzer combines data from the historical cases. Typically, confidence in a result increases when several cases point in the same direction and vice versa. Case similarity analysis uses small sample statistics (χ^2, t- and F-distribution) to analyze results. The result of this analysis (e.g., performance measure values) can be presented using descriptive statistics such as the number of very similar cases.

The analyzer can be particularly useful when predictions are needed. The predictions may be relatively straight forward when the results from cases are equivalent; if not then, modification of the historical cases may be considered. For example, if the congestion impacts of a three-person per vehicle HOV lane are needed, while the similar historical cases relate to two and four persons per vehicle HOV lanes, then the analyzer may simply interpolate to predict performance measures for the current case (this of course assumes a causal model where the key factor influencing the performance measure is the occupancy threshold).

Resolve

It is useful to resolve differences between the historical cases and the current case, if possible. When comparing the current case to historical cases, several possibilities may arise. A descriptor data item may be absent in the historical case but present in the current case or present in the historical case but the value seems abnormal. In the first instance, users can be asked to: (i) find a proxy in the historical case as a substitute, (ii) include the data item by assumption, (iii) ignore the historical case altogether (iv) warn the users regarding the absence (particularly if important data items are missing in the historical case). In the second instance, a plausible explanation would be needed. If such an explanation cannot be found, then the confidence in the comparison would be low and either a warning to this effect may be issued or the historical case can be ignored altogether.

Advise

The prescriptive reasoning mechanism synthesizes evidence from historical cases, e.g., whether an action has potential for negative impacts such as igniting community conflicts, creating congestion elsewhere, or falling much short of expectations. The mechanism warns against potential problems and sensitizes participants to unexpected outcomes.

At the simplest level, the advisor examines the number of successful and unsuccessful similar cases. If successful similar cases (i.e., cases where the performance measures of interest responded favorably to the action) exceed the unsuccessful cases, then the mechanism can infer that the action has a higher possibility of success and recommend it.

Compare and Remove

The compare function allows comparison of the current case (i.e., participant specified planning vector values of the descriptors) with the selected historical case(s) in a Table. The cases are compared in terms of their action and environment descriptors. The performance data (present only in the historical case but unknown for the current case) can also be viewed.

The remove function will allow the removal of any selected case that the user feels is not applicable (the case is removed from the list of similar cases but not permanently from the computer memory).

Accept

After participants choose particular levels of stringency and are satisfied with the similarity of specific cases, they can accept the performance measure estimates. This means that the case performance measure outcomes for the historical cases are averaged and saved as the evaluation of the current performance measure. This value can be compared with structured modeling estimates and reviewed during future deliberation on impacts.

Assimilate

New cases are acquired as information from studies, field operational tests and implementations of technologies become available. The new cases are stored in the memory in accordance with the PLANiTS case base structure.

LIMITATIONS

An important limitation of CBR in transportation planning is that no two cases are exactly similar. As we have seen, judging similarity is often difficult. Further, if two cases are judged similar enough, it is likely that there will be wide variability in the performance of cases when they are used to obtain insights for the current case. The variability can be due to errors in measurement, differences in model types used for evaluation and their specifications, differences in contexts not captured in the analysis and differences in researchers’ and users’ judgements across historical cases. There are no easy procedures that deal with such issues. We recognize the issue of variation in apparently similar cases and hope to develop methods that distinguish between cases with increasing accuracy.

SUMMARY

To support transportation planning processes, case-based reasoning can be used along with structured models, semi-structured expert systems and unstructured electronic support for human interactions. In this paper, we develop a methodology to represent and match similar past cases. We have structured case representation in terms of action, performance measure and environment clusters, sub-clusters and descriptors. The main functions include matching, ranking, analyzing, comparing and assimilating. Given current case specifications, the matcher provides a listing of similar cases by locating them in the case-base. Participants can match with increasing levels of stringency. A ranker can rank similar cases in terms of their relative similarity to the current case. An analyzer provides statistical analysis tools for predictions of performance measures and an advisor compiles evidence from several historical cases. After an action is implemented and results on its performance become available, an assimilator acquires and stores the information for future use.

Developing the PLANiTS case-based reasoning further is challenging. The areas that need development include causal models and similarity analysis. Furthermore, besides HOV lanes and ATMIS, many more actions need to be considered. Similarly, a comprehensive set of evaluation criteria needs to be examined.

Full development is ambitious requiring many more years, However, when fully developed, CBR will likely enhance a participants' creativity in exploring innovative new technology solutions such as ITS by providing information on how other technologies/transportation improvement actions have performed.

REFERENCES

CBR, ***Proceedings, Case-based Reasoning Workshop*** held in Clearwater Beach, Florida 10- 13 May 1988. ed. J. Kolodner. Sponsored by Defense Advanced Research Projects Agency, Information Science and Technology Office, Florida, 1988.

CBR, ***Proceedings, Case-based Reasoning Workshop*** held in Pensacola Beach, Florida 3 1 May-2 June 1989. Sponsored by Defense Advanced Research Projects Agency, Information Science and

Technology Office, Florida 1989.

Kanafani A., A. Khattak, M.and J. Dahlgren, “ A planning methodology for intelligent urban transportation systems,” Presented at the 73rd Annual Meeting of the Transportation Research Board, Preprint No. 940418, Washington, D. C., ***Transportation Research, C*** Vol. 2, No. 4, 1994.

Kolodner, J. ***Case-Based Reasoning,*** Morgan Kaufmann Publishers, San Mateo, California, 1993.

Vlahos N., A. Khattak, M. Manheim, and A. Kanafani, “The role of teamwork in a planning methodology for intelligent transportation systems,” Presented at the 73rd Annual Meeting of the Transportation Research Board, Preprint No. 940491, Washington, D. C., ***Transportation Research, C*** Vol. 2, No. 4, 1994.

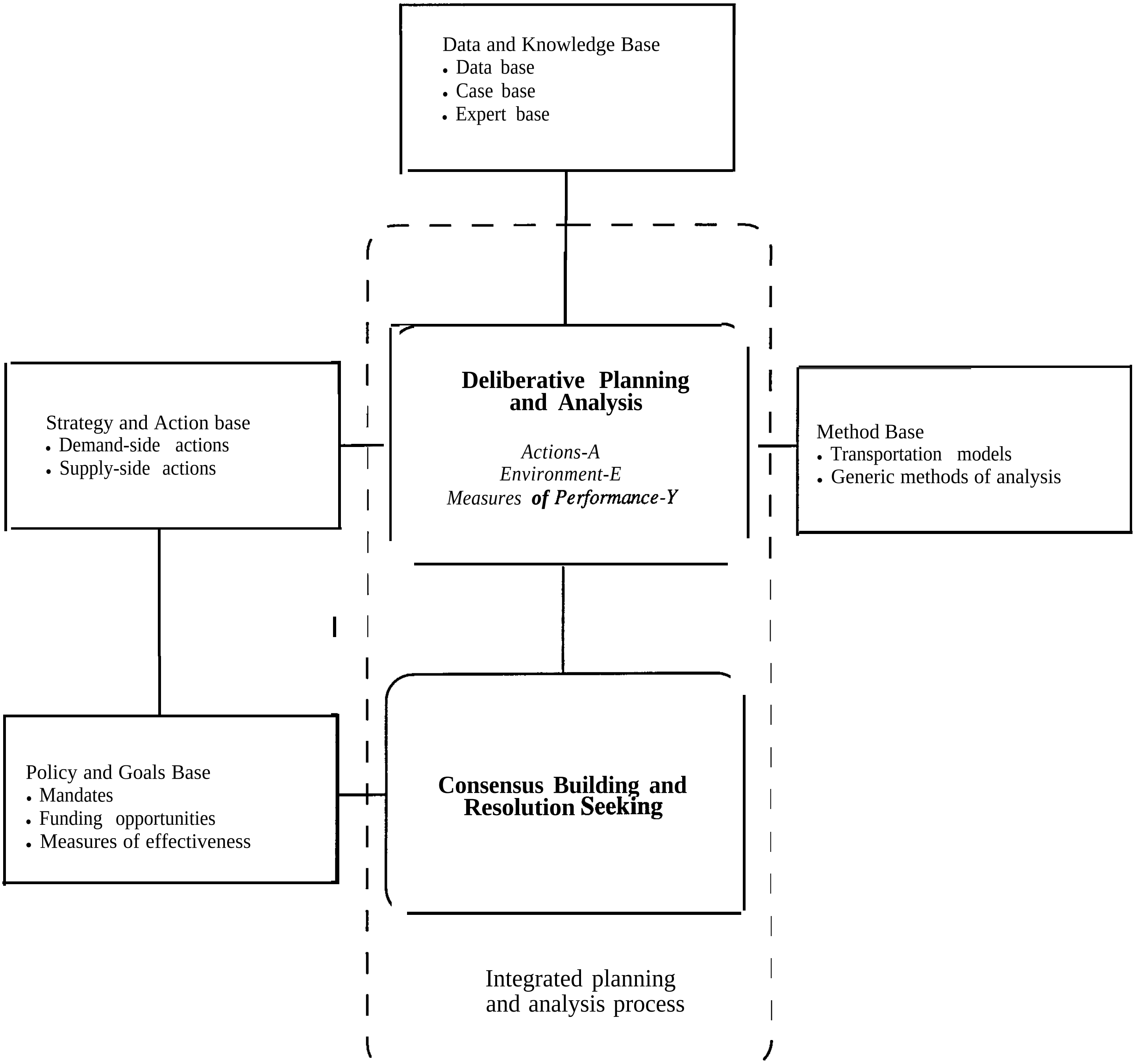

Figure 1. PLANiTS components. (Source: Kanafani, Khattak and Dahlgren 1994)

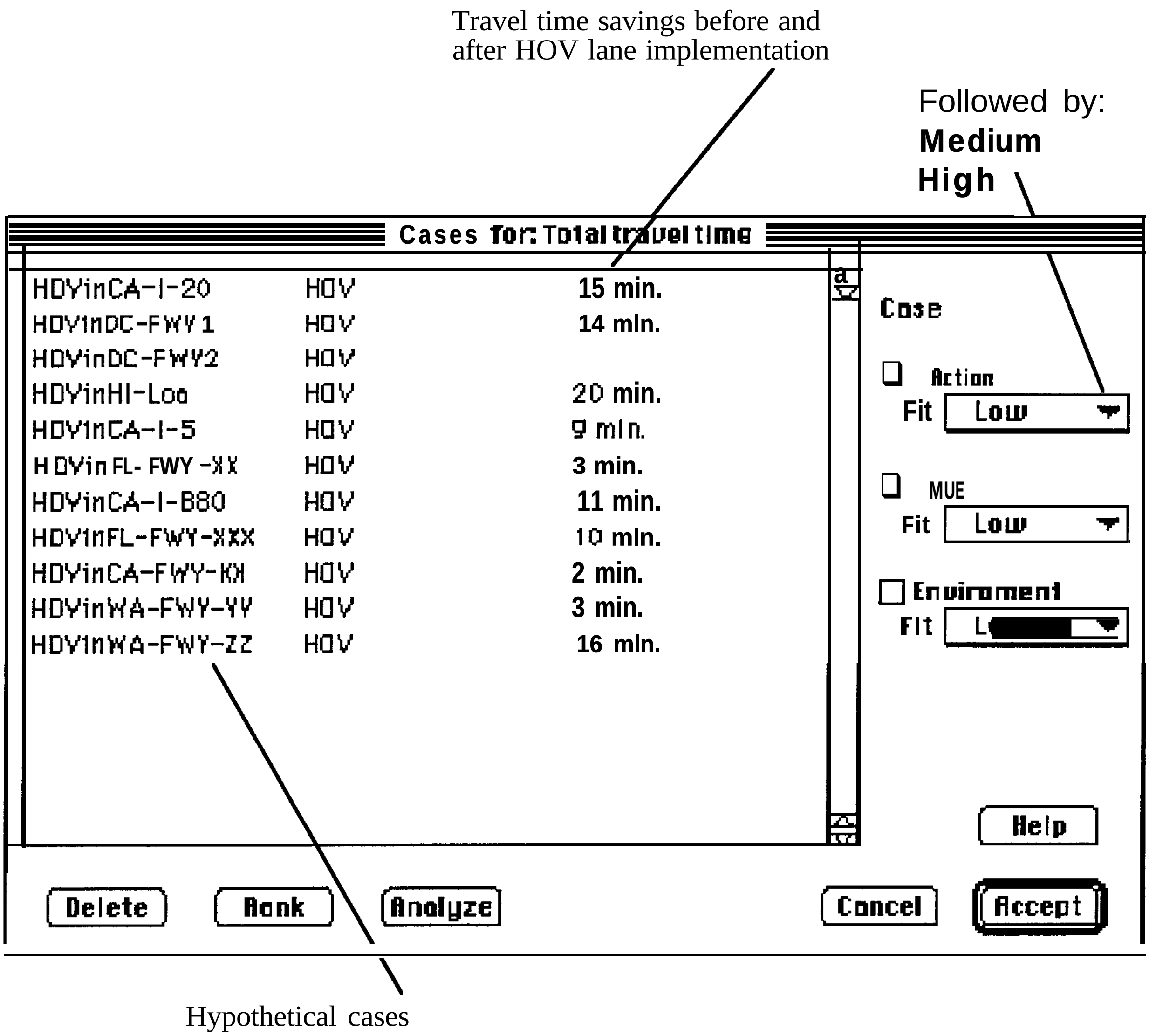

Figure 2. The Case-Based Reasoner.

Table 1. Cases matched according the planning vector elements: Descriptor level (hypothetical data).

	Actions $[A_{jli}]_n^{S, T, U}$			Environment $[E]_n^{S, T, U}$		
	Number of HOV lanes	Time of operation (peak/all day)	occupancy threshold	Geographic area (Match on same State)	Year implemented. (Match on 1990+)	Urban Area size (Match on +/- 50%)
Current Case	1	peak only	2+	I-80 Bay Area	1996 (proposed)	6 Million
Historical Case 1	1 Match	peak only Match	2+ Match	LA Match	1991 Match	8 Million Match
Historical Case 2	2 Not Match	peak only Match	4+ Not Match	Chicago Not Match	1989 Not Match	6 Million Match
Historical Case 3	1 Match	All day Not Match	Buses only Not Match	Orlando Not Match	1984 Not Match	2 Million Not Match
Historical Case n	1 Match	All day Not Match	Buses only Not Match	Seattle Not Match	1987 Not Match	1 Million Not Match

Notes:
Based on number of matches Case 1 is the closest to the current case

Table 2. Examples of case-based reasoning at the three levels

	Actions	Performance measure	Environment
Cluster Level Matching	$[A_i]$ Ex: All highway cases	$[Y_r]$ Ex: All cases related to congestion	$[E,\ _y]^{S, T, U}$ Ex:' All cases in region X_1 that occurred in Y_1 years and relate to urban area size $\geq Z_1$
Sub-Cluster Level Matching	$[A_{j\|i}]$ Ex: All HOV lane cases	$[Y_{q\|r}]$ Ex: All cases related to person delay	$[E_{a,y}]^{S, T, U}$ Ex: All cases in region X_1 -X_2 that occurred in Y_1 - Y_2 years and relate to urban area size $\geq Z_1$ and $\leq Z_2$
Descriptor Level Matching	$[A_{j\|i}]^{S, T, U}$ Ex: All HOV lane cases with s_1 (number) HOV lanes, t_1 (24) hour operation and user occupancy threshold u_1	$[Y_{q\|r}]^{S, T, U}$ Ex: Person-delay cases disaggregated by links (s), times of day (t) and occupancy levels (u)	$[E_{a,y}]^{S, T, U}$ Ex: All cases in region X_1 - (X_2 + X3) that occurred in Y_1 - (Y_2 + Y_3) years and relate to urban area size $\geq Z_1$ and $\leq Z_3$ where $Z_2 > Z_3$

Table 3. Calculation of Similarity Index scores

	Actions/ spatial descriptors	Actions/ temporal descriptors	Actions/ user descriptors	Environment/ spatial descriptors	Environment/ temporal descriptors	Environment/ user descriptors
Current case c	$[A_{j\mid i}]_c{}^{S_k}$	$[A_{j\mid i}]_c{}^{T_k}$	$[A_{j\mid i}]_c{}^{U_k}$	$[E_{a,y}]_c{}^{S_k}$	$[E_{a,y}]_c{}^{T_k}$	$[E_{a,y}]_c{}^{U_k}$
Historical case n	$[A_{j\mid i}]_n{}^{S_k}$	$[A_{j\mid i}]_n{}^{T_k}$	$[A_{j\mid i}]_n{}^{U_k}$	$[E_{a,y}]_n{}^{S_k}$	$[E_{a,y}]_n{}^{T_k}$	$[E_{a,y}]_n{}^{U_k}$
HOV example	No. of HOV lanes	Time of Operation	occupancy threshold	Geographic location	Year implemented	Urban area size
Match Criteria $M_w=$	$\lvert[A_{j\mid i}]_n{}^{S_k} - [A_{j\mid i}]_c{}^{S_k}\rvert$	$\lvert[A_{j\mid i}]_n{}^{T_k} - [A_{j\mid i}]_c{}^{T_k}\rvert$	$\lvert[A_{j\mid i}]_n{}^{U_k} - [A_{j\mid i}]_c{}^{U_k}\rvert$	$\lvert[E_{a,y}]_n{}^{S_k} - [E_{a,y}]_c{}^{S_k}\rvert$	$\lvert[E_{a,y}]_n{}^{T_k} - [E_{a,y}]_c{}^{T_k}\rvert$	$\lvert[E_{a,y}]_n{}^{U_k} - [E_{a,y}]_c{}^{U_k}\rvert$
Importance $I_w=$	$[I_{j\mid i}]^{S_k}$	$[I_{j\mid i}]^{T_k}$	$[I_{j\mid i}]^{U_k}$	$[I_{a,y}]^{S_k}$	$[I_{a,y}]^{T_k}$	$[I_{a,y}]^{U_k}$

Operations:
Similarity Index = $\sum_{w=1,\ldots x} M_w I_w / \sum_{w=1,\ldots x} I_w$

Notes:
Other match criteria are possible, e.g., $\lvert[A_{j\mid i}]_n{}^{S_k} - [A_{j\mid i}]_c{}^{S_k}\rvert / \{\lvert[A_{j\mid i}]_n{}^{S_k}\rvert + \lvert[A_{j\mid i}]_c{}^{S_k}\rvert\}$ instead Of $\lvert[A_{j\mid i}]_n{}^{S_k} - [A_{j\mid i}]_c{}^{S_k}\rvert$