Available online at www.sciencedirect.com
SCIENCE DIRECT
Advanced Engineering Informatics 20 (2006) 23–30
ADVANCED ENGINEERING INFORMATICS
www.elsevier.com/locate/aei

Computer based pedestrian landscape design using decision tree templates

Baranidharan Raman [a,*], Jody R. Naderi [b]

[a] *Department of Computer Science, Texas A&M University, 520 Harvey R.Bright Building, College Station, TX 77840, USA*
[b] *Department of Landscape Architecture and Urban Planning, Texas A&M University, College Station, TX 77840, USA*

Received 18 March 2004; revised 1 April 2005; accepted 23 August 2005

Abstract

Machine Learning algorithms can act as a valuable analytical tool in design research. In this paper, we demonstrate the application of a decision tree learning algorithm for designing pedestrian landscapes that encourage walking for health. The domain knowledge was captured using intercept surveys that queried responses to cognitive, physical and social attributes that influence pedestrian spatial analysis. Decision trees extracted from the knowledge base were used in the design of pedestrian landscapes, which were tested in a transportation simulator. The observed match between the change in the participants' response to manipulation of physical variables in the simulated world with those predicted by the decision rules indicate the appropriateness of applying decision tree rules as guidelines during the process of pedestrian landscape design and research.

Keywords: Machine learning; Decision tree templates; Pedestrian landscape design; Simulation

1. Introduction

Learning and design are cognitive processes fundamental to problem solving. Learning is the process of adaptation based on the past and current events, which improves the performance on a specific task. Design on the other hand is the process of applying the learned knowledge to seek solutions that satisfy current requirements and meet predicted future needs. This complementing aspect of learning and design raises a fundamental question. Can the complexity of the design process and the outcome be facilitated with inferences learned from related experience? In other words, how can we improve design learning and prediction using learning algorithms capable of inductively modeling human experience? This is the research question that is being investigated in this paper.

In the context of design, machine learning has been primarily used to extract symbolic knowledge from previous design data using inductive learning approaches [1], recall and reuse of specific previous design experience through case-based reasoning [2], and addressing combinatorial optimization problems in design with genetic algorithms [3]. A range of design problems from simple discrimination of feasible designs from infeasible ones [4] to complex tasks like optimizing architectural floor plans [3], completing partial room designs [5], and generating bridge designs [6,7] have already been investigated using these approaches. Readers are referred to [8 and references therein] for a thorough review of the applications of machine learning methods in design.

In this paper, we apply learning to guide generation of site-specific use-specific pedestrian landscape design. To understand how this might be done, let us consider the example of designing pedestrian landscapes based on the motive for taking the walk: commuting or health purposes. Depending on their motive, people may evaluate the environment using different set of features and decision rules. In our case study in Texas, a person walking for health may choose a route because of the presence of water features like fountain, lake, pond etc. in the vicinity of the path or by the amount of trees along the path, and may use the presence of these features to determine whether the path is suitable. A commuter on the other hand might use features like the weather condition, the traffic and the width of the walk for his evaluation of the quality of the walking route. A possible role for machine learning becomes extraction of symbolic structures (decision trees in this case) from a grass roots knowledge base of landscape evaluations made by the people walking for these two distinct walking purposes. The extracted rules can then be formulated as use-specific design principles for recovering/designing better pedestrian

* Corresponding author. Tel.: +1 9798450531; fax: +1 9798478578.
E-mail addresses: barani@tamu.edu (B. Raman), jnaderi@tamu.edu (J.R. Naderi).

1474-0346/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aei.2005.08.002

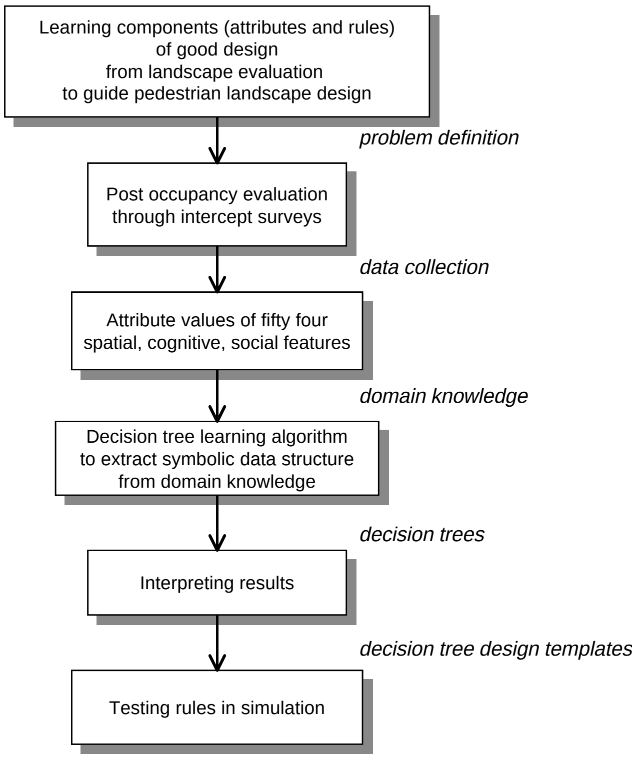

Fig. 1. Process flowchart indicating various phases involved in applying learning to guide pedestrian landscape design.

landscapes that are intended to encourage walking for both commuters as well as for people walking for health.

From the above example, we can identify the following six steps for applying machine learning (ML) techniques in design [1]: (i) formulation of the learning problem in a particular design context, (ii) preparing the inputs, (iii) developing a representation for the input information, (iv) selecting a learning algorithm, (v) selecting operational parameters, and, (vi) analyzing results. Fig. 1 shows a flowchart of this process with an additional simulation testing stage used in this study.

2. Pedestrian landscape knowledge base

The pedestrian landscape knowledge base was set up by disaggregating and evaluating the parts and synthesis of common roadside pedestrian environments. The underlying assumption is that the pedestrian environment can be considered as a combination of several attributes or features that represent some intrinsic property of the environment. In theory, manipulation of the attributes will affect human behavior and perception of the space. Researchers have demonstrated that certain landscape dimensions can be used successfully to prepare an evaluation, and measurement of the aesthetic impact [9]. The use of physical environmental attributes for predicting public perceptions has already been demonstrated [10]. People's perception and response to common elements captures the necessary information about the environment [11]. A landmark study in the field of urban design of the physical attributes and dimensions of pedestrian intensive environments and the attributes of great streets are provided in Ref. [12]. A detailed discussion of the pedestrian environmental attributes consideration is presented in Ref. [13].

We extracted a list of 50 cognitive, physical and social attributes related to walking environments from literature review of health, transportation, engineering and planning domains to generate a grass roots survey for on-site data collection. A target attribute that captured the *overall evaluation* of the landscape as either 'good for walking' or 'not good for walking' was included in the survey to identify the predisposition of the person being interviewed for the particular walking site.

Table 1 shows the complete list of the attributes that were used in this study. Fifty-four intercept surveys conducted at six different locations in College Station, Texas queried responses to the identified attributes and captured the grass roots domain knowledge. Each attribute was measured using scales of specific choices, although margin notes and general comments were welcomed. The data from the surveys were arranged in the UCI machine learning repository format to input the domain knowledge to the learning algorithm.

3. Decision tree learning

The prominent machine learning algorithms can be roughly categorized as instance-based learners like nearest neighbors, symbolic learners like ID3 decision trees, statistical learners like naive Bayes classifiers, connectionist approaches like back-propagation network, or ensemble learning that combines many of these strategies. A study on the role of machine learning for well defined design problems and their limitations when applied to real world tasks has been discussed in Ref. [1].

The learning model required for our task must fulfill the following requirements: (i) must be easy to understand as it is a tool for non-AI researchers, (ii) a symbolic model, as a quantitative assessment methods for measuring scenic impacts of landscapes is complex [9], (iii) non-parametric and deterministic model (iv) must have a feature selection scheme to identify the most important features and arrange them in the order of importance. This final requirement is important because all possible categories of features were added for evaluating the landscape, therefore, the learner must not be a passive component, i.e. must not just use all the information provided to come up with the hypothesis model but select information for developing a consistent hypothesis. The ID3 decision tree was selected for performing this task as it satisfies these primary requirements.

3.1. PED-Learner

The decision tree learning algorithm performs inductive inference and produces a set of if-then decision-making rules [14]. The ID3 decision tree algorithm [15] builds a bottom-up hierarchical model of the concept with the most relevant feature as the root and less relevant features at the upper levels (near the leaves) of the decision tree. Here, each branch of the tree corresponds to one decision-making rule. ID3 tries to develop hypotheses that consistently explain the training examples.

Table 1
Fifty attributes used for capturing the domain knowledge

Attribute	Value	Attribute	Value
Motive	Commute, health, spiritual, combination, all	Color	1-Violet blue purple, 2-yellow orange red, 3-black, 4-brown beige tan, 5-white gray, 6-green
Initial bias	Good place, ok place, not a good place	Focal point	Yes, no
Optimum time	1.Early morning, 2.late morning, 3.mid-day, 4. late afternoon, 5.dusk, 6.night	Light condition	Uniform brightness, uniform shadow, dramatic shadows, cloudy, dappled sunlight, combination, all
Weather condition	Sunny, cold, rainy, windy, hot, humid, dark, combination, all	Handicap access	Yes, no
Availability of seating	Not enough, sufficient seating, more than needed	Safe	Yes, no
Width of walk	Too narrow, narrow, normal, wide, too wide	Safe while crossing street	Yes, no
Length of walk	Too short, adequate, long, too long	Quantity of trees	Too many, not enough, adequate
Width of surrounding	Very narrow, narrow, normal, wide, very wide, varies along way, not applicable	Quantity of lights	Too many, not enough, adequate
Proximity of parking	Interfering with pedestrian movement, not interfering, not applicable	Access drinking water	Not enough, adequate
Availability of Parking	Ample, less than enough, none	Quantity of shelters	Too many, not enough, adequate
Slope of the walk	Steep, flat, just right	Quantity of restrooms	Too many, not enough, adequate
Edge of space	Well-defined, ill-defined, urban/built, natural, combination of edges	Quantity of trashcans	Too many, not enough, adequate
Path edge	Well-defined, Ill-defined, random	Views and vistas	Too many, not enough, adequate
Resiliency of path surface	Too hard, just right, too flexible	Quantity of greenery	Too many, enough, bare
Roughness of path	Too smooth, too rough, just right	Marked entries/exits	Too big, adequate, too small, none
Music	Too little, just right, too much	Quantity of focal points	Too many, adequate, less than enough
Traffic sound	Too little, ok, too much	Signage	Too much, too few, adequate
People sound	Too little, ok, too much	Quantity of building	Too many, enough, too few
Water sound	Too little, ok, too much	Presence of water features	Too much, too little, perfect
Birds sound	Too little, ok, too much	Presence of wildlife	Too much, too little, perfect, null
Construction sound	Too little, ok, too much	Traffic	No impact from traffic, too much from traffic, too close to traffic, too close to traffic sound, combination, all
Industrial sound	Too little, ok, too much	Traffic components	Trucks, cars, mixed
Silence	Too little, ok, too much	Maintenance	Good, poor
Good smell	Yes, no	Land use	Residential, school, office, park, commercial, combination, all
Bad smell	Yes, no	Overall evaluation	Good, not good

Decision Trees use only those features that are required to completely classify the training set and removes all other features [16]. This feature selection property makes it more suitable for this task, which might have many redundant and irrelevant features. The ID3 algorithm uses the statistical measure of information gain to evaluate features at each step. Information gain of feature 'A_i' in training sample S is given by the following equation,

$$\mathrm{Gain}(S,A) = \mathrm{Entropy}(S) - \sum_{v\varepsilon \text{ values } (A)} \frac{S_v}{S} \mathrm{Entropy}\ (S_v) \tag{1}$$

where, entropy gives a measure of uncertainty or impurity in the data and information gain of an attribute 'A' gives its effectiveness in classifying the data. If a feature 'A' can assume 'n' different values and P_1, P_2, $P_3 \ldots P_n$ are the probabilities of the n outcomes, then entropy is given by the following equation,

$$\mathrm{Entropy}(A) = -P_1 \log_2 P_1 - P_2 \log_2 P_2 - \cdots - P_n \log_2 P_n. \tag{2}$$

The PED-Learner is an adapted version of ID3 decision tree algorithm with one modification made to suit generation of specific design hypotheses. The general ID3 decision tree algorithm seeks hypotheses that provide solutions for a very general unconstrained design question, i.e. what constitutes the good or bad walking environment. For empirical purposes, design researchers are more interested in constrained and more specific questions e.g. which physical attributes contribute to designing transportation landscapes that encourage walking for health?

The desired mapping can be easily achieved by setting the root of the decision tree to a particular design question or variable of interest. For our design question, the root of the decision tree was set to the attribute *motive of walk*. The effect of this variable was tested on the decision to walk. If the designer determined that the relationship between two design variables was considered a given, the mapping could incorporate the condition by fixing a desired number of variables at the top of the tree and generate the decisions made under this condition. This capability allows the PED-Learner to test multiple design parameters and generate conceptual design options accordingly.

The pseudo-code of the PED-Learner algorithm is given below:

```
ID3_PED-Learner (examples, target, attributes)
if all examples belong to same class
   return the class label
else
    if the decision tree level 'k' is less than the number of
    constraint attributes-1
      make the kth constraint attribute as the root of the
      subtree
    else
       select the attribute (Ai) that has the highest
       information gain make Ai as the root of the subtree
    end if
    for each value v of Ai create a branch and distribute the
    examples based on their value of Ai
    if there are no examples corresponding to any value of Ai
        create a leaf node with most common class of
        examples associated with the root of the subtree
    else
      ID3_PED-Learner (examples(v), target, Attributes-Ai)
    end if
      end for
    end if
```

3.2. Extracted decision trees

In our pilot test of the PED-Learner, we posed the question of what physical variables in the environment encouraged people to walk for health purposes versus commuting purposes. On site intercept surveys were conducted by interviewing 50 participants walking in College Station, Texas. Fig. 2(a) shows the decision tree extracted from these post-occupancy evaluation surveys for health walking and commuting. The results indicate the emergence of a purpose-driven bias. *Motive* is a very important cognitive feature that provided participants with contextual information for evaluation of pedestrian environments. We learned that adequate water features play an important role in pedestrian preferences when selecting an environment to walk for health purposes. We state our findings as 'if-then' rules as follows: *If* a location does not have 'adequate' water features present (e.g. not 'too many' and not 'too few'), then an 'adequate' quantity of trees can compensate (e.g. not 'too many' and not 'too few'), *then* the environment is considered 'good' for the purpose of walking for health. On the other hand, commuters in our study used the *weather conditions* and the *width of the walk* as primary factors in determining whether the pedestrian environment was good for walking.

Next, we examined an architectural component of the pedestrian experience to determine how to manipulate the variable to encourage walking for health. In design theory, the *edge of space* is considered to be a very important spatial feature that impacts human comfort. The presence of edge contributes to a positive evaluation of the landscape as indicated in Ref. [17] with respect to legibility of space and way finding at the scale of the city. Adaptation of prospect-refuge theory to architectural design [18] concludes similarly that edges are important spatial experiences that influence a sense of security and identity. The importance of edge as a preferred physical construct of the environment appears in the field of ecology, where maximum species diversity and numbers concentrate at edges between landscape units [19]. We used the PED-Learner to evaluate perception of edge in relation to walking for health. The type of edge affects the relevance and role of other variables in the decision by a pedestrian as to whether a place is good for walking. Fig. 2(b) shows the decision tree extracted from the post-occupancy evaluation surveys for different types of *edge of space*. As outlined, the results from the PED-Learner analysis of the field surveys indicate if a location has a natural edge of space, then the environment is considered 'good' by people walking for health. Commuters on the other hand looked for well-defined *edge of space* with adequate *views and vistas* and good separation from traffic.

4. Decision tree templates for simulating pedestrian worlds

The hypotheses generated by the PED-Learner were tested in simulated worlds developed at Texas Transportation Institute (TTI) using the Driving Environment Simulator (DES) that operates from a Hyperion platform (Fig. 3). The driving simulator comprises of four integral components:

- A full-size 1995 Saturn SL automobiles, outfitted with computers, potentiometers and torque motors connected to accelerators, brakes and steering,
- Computers to collect data and generated images,
- Three high-resolution projectors, and
- Three high-reflectance screens.

The simulated world was developed from a series of AutoCad drawings that altered distance from traffic and edge treatment and width of sidewalk. The basic structure of the landscape was catalogued in the Hyperion platform at TTI as a series of tiles that could be connected together like a Lego set as shown in Fig. 4. From this base, static objects and kinetic objects are added to create the replica of the world we want to emulate. A simulated world that was created by piecing the tiles together is shown in Fig. 5.

A random sequence of the six worlds was experienced by 26 participants. In addition to our test question of which landscape configuration encouraged their decision to walk, the participants were asked other questions related to whether they would let their children walk in any of the simulated environments [20].

4.1. Results

In our initial work with the PED-Learner, several attributes of the physical world were influencing the decision to walk for

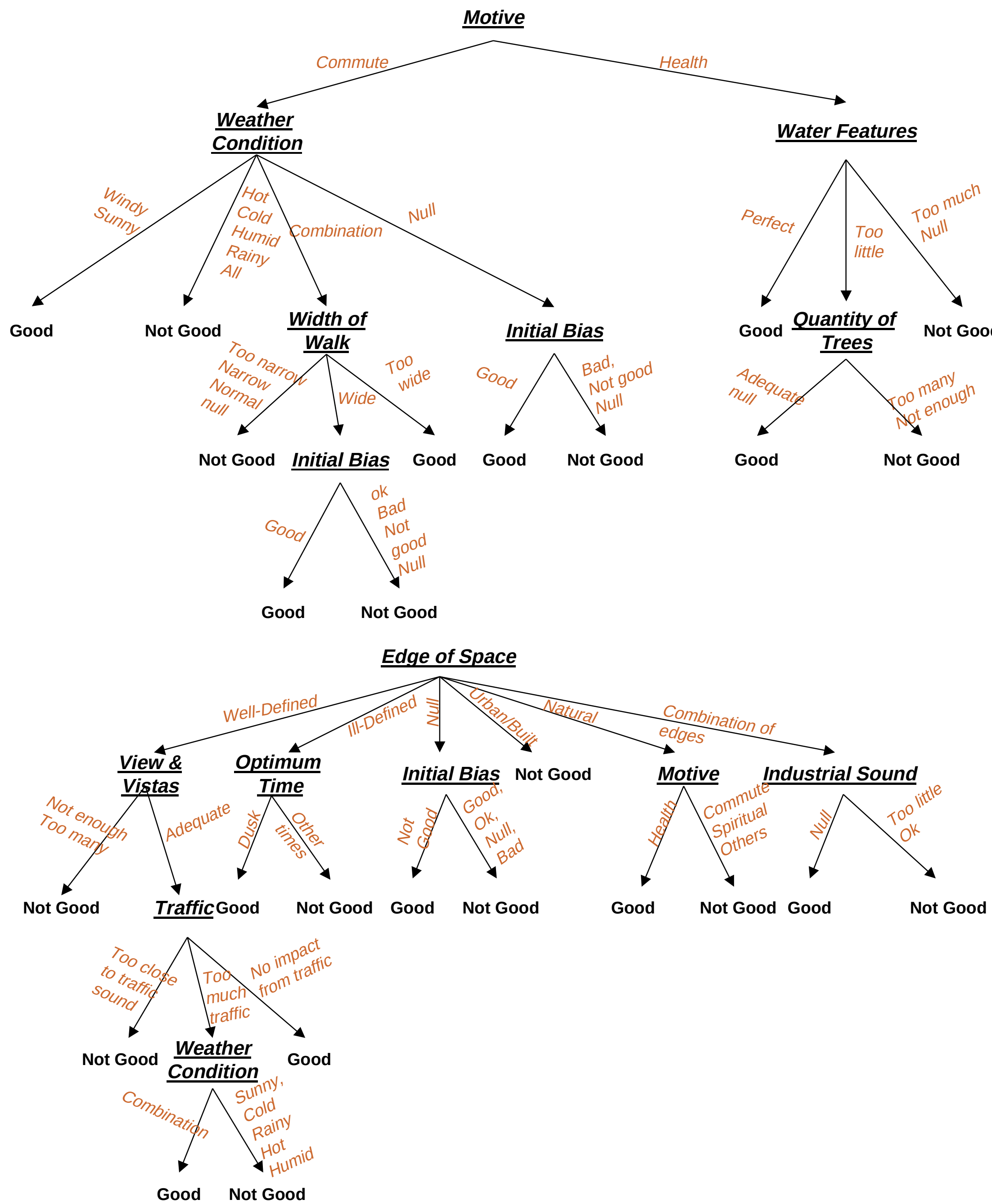

Fig. 2. Decision tree extracted from the pedestrian post-occupancy evaluation surveys: (a) use-specific decision trees for health wakers and commuters, (b) site-specific decision tree for different types of edges of spaces.

health. The presence of water and the presence of trees were features that were directly related to the motive of walking for health. Other 'if-then' rules emerged as well, related to whether a walking environment was considered good regardless of motive. Three rules were selected for the purpose of validation in the simulator:

1. Ill-defined edge of space is not good walking environment.
2. A well-defined edge of the space, with adequate view and vistas, too close to traffic is not a good walking environment.
3. A good walking environment must have a well-defined edge of the space, with adequate view and vistas, and separated from traffic.

Fig. 6 shows the three decision rules and the simulated pedestrian worlds corresponding to these three rules. The simulated worlds mimicked the Texas landscape where the original field surveys were presented. Variables were limited to those identified in the hypotheses generated by the PED-Learner. The simulation was driven through in a pretrial by volunteer participants and is now being developed as a walking simulation.

Fig. 3. Components of Driving Environment Simulator: 1995 Saturn, computer for data collection and image generation, three projectors and three screens.

In the simulated world, the volunteers drove through transportation landscapes created using the analysis by the PED-Learner. The simulated worlds were operated with the Driving Simulator at Texas Transportation Institute. The hyperdrive software was used which facilitated landscape architectural alterations of the transportation landscape and surrounding community landscape. Each volunteer participant in the pre-trial drove through four trials[1], which were randomly presented to the 27 participants. We reviewed the results of the pretrial to determine the relationship with the PED-Learner analyses relative to the field intercept surveys.

The regression analysis of the simulation trials showed significant correlation between the perception of 'edge' and willingness to walk ($p<.0001$). The perception of overall safety was correlated positively with the presence of a well-defined edge ($p<.0001$) and increased lateral distance from vehicular traffic. This result was consistent with PED-Learner rules 2 and 3, even though the simulation trial was attempting to validate the experience of the simulated world with the real world experience. While the validation was successful, the unexpected level of significance leads us to further examination of the relationship of curbside planting and urban design as it relates to the driving modality.

5. Discussions

The PED-Learner is a particularly valuable tool for designers because of the flexibility it affords in testing options and different design scenarios. The studio design tradition for both teaching and practice is to investigate alternative solutions to problems to assure that the best design solution is generated. The algorithm facilitates introducing one or multiple user-specific variables as start points (root) in the decision so that the designer can predict the impact of trade-offs necessitated by other factors such as narrow transportation corridors, construction cost and maintenance considerations. Given sufficient data, the algorithm facilitates determining what would be best for people walking who have cancer, for instance, and are therefore, sensitive to sunlight. In this case, the root would be the variable of *motivation* and the next level variable of *sunlight* would be fixed as well. The resulting decisions would be based on fixing the notion that the reason for walking was for health and that the person was sensitive to sunlight. The most important variable to this person might be the presence or absence of trees. Because of this capacity to evaluate physical attributes of walking environments from the perspective of specific user groups who are seeking particular physical, psychological and/or spiritual outcomes, the PED-Learner provides the designer with input necessary to design environments with increased user satisfaction. This makes the PED-Learner particularly useful in health design where the constraints of patient conditions may warrant different priorities in the design of the walking environment, and allows the designer to capture the unexpected relationships between physical variables. For instance, counter–intuitive relationship may emerge between health patients and physical environment. Someone with diabetes walking for health may

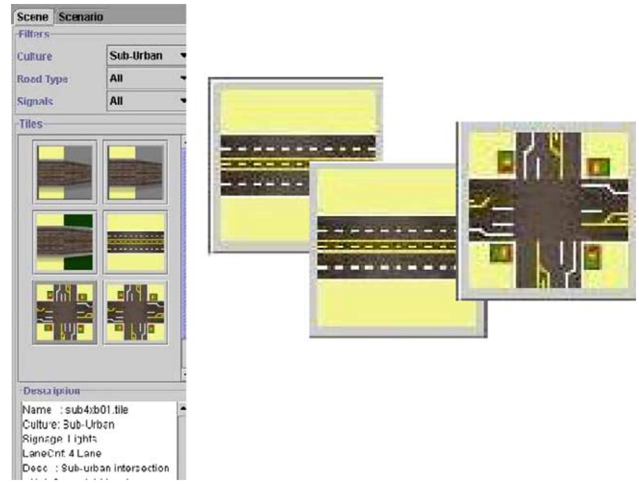

Fig. 4. A range of landscape tiles that could be put together for creating artificial worlds.

Fig. 5. A simulated pedestrian world of transportation corridor and neighboring community assembled by piecing tiles together.

[1] For simulation video streams of the pedestrian trials refer Movies 1–4.

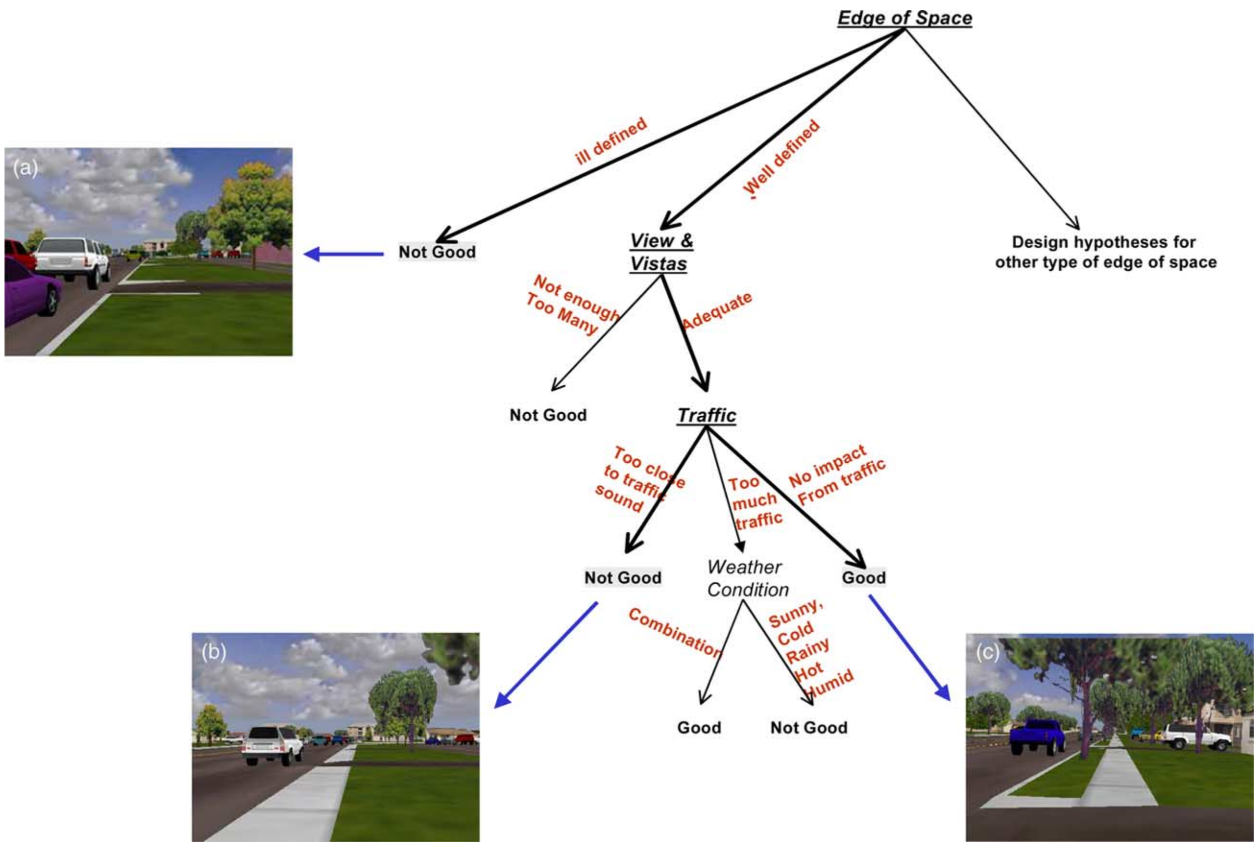

Fig. 6. Site-specific decision rules and the corresponding simulated worlds: (a) ill-defined edge of space (b) well-defined edge of space with adequate views and vistas and no separation from traffic, and (c) well-defined edge of space with adequate views and vistas and well separated from traffic.

need to have access to water, and may also consider the width of sidewalk an important factor in their decision to walk.

6. Conclusions and future directions

Machine Learning algorithms can act as a valuable analytical tool in the field of urban landscape design. The analytical activity of a designer is continuous throughout the design process whether building decision-making models or evaluation models. Clearly, the opportunity to design with evidence from a wide range of field conditions and respondents benefits all students and practitioners of landscape architectural design. Working in the trans-disciplinary domain where both Artificial Intelligence and Landscape Architecture brought together different aspects of cognition, namely learning and designing, made a deeper awareness of the relationship between human behavior and the environment. Contributions to Machine Learning comes not just from the adaptation to the decision tree learning algorithm for design purpose but from the fact that learning algorithms can be made use of in the empirical real-world purpose of designing pedestrian environments to achieve health outcomes. Depending upon the natural constraints, the algorithm can be adjusted to generate multiple hypotheses, which are site-specific. The application of decision tree learning algorithm to a complex landscape design problem is validated by virtue of making the information and knowledge in the database available to the designer or design student. This is also invaluable to the design researcher or practitioner who may not be highly experienced with the design issue under consideration. In addition, this use of technology for design was particularly beneficial in trying to understand the variables at work in a dangerous environment with no risk to human participants.

As a teaching tool, the PED-Learner allows students to understand the relationship between physical attributes of an environment and how they are synthesized and evaluated by users in every design project. The student can witness through the application how the motivations of people and their needs affect their response to the environment. The designer can use the PED-Learner to gain additional understanding of the cognitive responses that people have to the environment. Additionally, the decision making process regarding allocation of funds towards various attributes of the built environment can become transparent in public process using the PED-Learner as a demonstration of citizen input and economic trade-offs. Being able to predict health outcomes from design decisions with some level of accuracy is a final future application. Expansion of the database to include specialized groups of experts or health care groups will help designers and health practitioners determine how the variables in the environment might affect the health outcomes or delivery. If the PED-Learner and a pedestrian simulation tool are tied together in future research, we would expect that designers, practitioners and researchers would have a valuable design tool for addressing the challenges of urbanization.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.aei.2005.08.002

References

[1] Reich Y, Konda S, Levy SN, Monarch I, Subrahmanian E. New roles for machine learning in design. Artif Intell Eng 1993;8(3):165–81.
[2] Maher ML, Garza AG. Case-based reasoning in design. IEEE Expert 1997;12(2):34–41.
[3] Rosenman MA, Gero JS. Evolving designs by generating useful complex gene structures. In: Bentley P, editor. Evolutionary design by computers. San Francisco: Morgan Kaufmann; 1999. p. 345–64.
[4] Arciszewski T, Mustafa M, Ziarko W. The formation and use of abstract concepts in design. In: Fisher DHJ, Pazzani MJ, Langley P, editors. Concept formation: knowledge and experience in unsupervise learning. Los Altos, CA: Morgan Kauffman; 1991. p. 323–53.
[5] Coyne RD, Newton S, Sudweeks F. A connectionist view of creative design reasoning. In: Gero JS, Maher ML, editors. Modelling creativity and knowledge-based creative design. New Jersey: Lawrence Erlbaum; 1993. p. 177–209.
[6] Reich R, Fenves SJ. A system that learns to design cable-stayed bridges. J Struct Eng 1995;121(7):1090–100.
[7] Wang W, Gero JS. Sequence-based prediction in the conceptual design of bridges. ASCE J Comput Civ Eng 1997;11(1):37–43.
[8] Duffy A. What and how of learning in design. IEEE Expert 1997;12(3):71–6.
[9] Kaplan Rachel. The analysis of perception via preference: a strategy for studying how the environment is experienced. Landscape Plan 1985;12:161–76.
[10] Buhyoff GJ, Miller PA, Roach JW, Zhou D, Fuller LG. An AI methodology for landscape visual assessments. AI Appl 1994;8:1–13.
[11] Naderi JR. On the nature of walking and learning pedestrian environment. In: The architectural research centers consortium (ARCC), Montreal, Canada; 2002.
[12] Jacobs AB. Great streets. Cambridge, MA: MIT Press; 1993.
[13] Naderi JR, Raman B. Design considerations in simulating pedestrian environments Transportation research board annual meeting, Washington 2002.
[14] Mitchell T. Machine learning. New York: Mcgraw Hill Press; 1997.
[15] Quinlan JR. C4.5: programs for machine learning. San Mateo, CA: Morgan Kaufmann; 1993.
[16] Cardie C. Using decision trees to improve case-based reasoning. In: Proceeding of 10th international conference on machine learning. Amherst, MA: Morgan Kaufmann. 1993, p. 25–32.
[17] Lynch K. Images of a city. Cambridge, MA: MIT Press; 1960.
[18] Appleton Jon. The experience of landscape. London: Wiley; 1975.
[19] Forman Richard. Some general principle of landscape and regional ecology. Landscape Ecol 1995;10(3):133–42.
[20] Kweon Byoung-Suk, Naderi Jody Rosenblatt, Maghelal Praveen, Shin Woohwa. Correlates of environmental constraucts and perceived safety enhancements in pedestrian corridors adjacent to urban streets. Report 167722-1, Southwest University Transportation Centers, Texas Transportation Institute; 2004.