Steven P. French
Professor
City and Regional Planning Department
California Polytechnic State University
San Luis Obispo, California 93407
Tel. (805) 756-2496
FAX (805) 756-5986

A KNOWLEDGE-BASED APPROACH TO USING EXISTING DATA FOR SEISMIC RISK ASSESSMENT

Abstract: Techniques for seismic risk assessment have improved rapidly in the past decade. Much more is known about how different types of structures respond to an earthquake. To make assessments of seismic risk, this earthquake hazard information must be combined with an inventory of buildings by structure type. The cost of obtaining such inventory information presents a key stumbling block to implementing risk assessment techniques for large urban areas. Structure inventories based on field surveys are too expensive and become outdated as soon as they are completed. No single secondary data source contains the detailed structural information required by today's risk assessment techniques. This paper describes a knowledge-based system that makes inferences about structure types based on secondary data. Existing land use, structure age, size, number of stories and improvement are used to classify buildings by structure type. The knowledge-based system described here uses a set of rules to make inferences about structure type based on these characteristics. In this way the knowledge-based system creates new data not available from secondary sources.

INTRODUCTION

This paper describes a prototype knowledge-based system that generates a building structure inventory for use in seismic risk assessment. The approach begins with data extracted from an assessor's .land parcel database. The system then uses a set of rules to classify individual buildings into seven structural classes. The rules operate on a number of facts drawn for the assessor's database, such as building's age, size use and number of floors. The prototype knowledge base was developed using VP-Expert which operates on data stored in a DBase III database format. This paper discusses the system and its application to a subset of buildings in Palo Alto, California.

The state of seismic risk assessment has improved markedly over the past decade. While the nature and size of individual earthquakes still cannot be predicted with a high degree of reliability, probabilistic risk analysis provides a way to estimate the amount of damage that is likely to result from events of various sizes. By understanding the likely results that could be caused by earthquakes of different sizes, federal, state and local governments can develop mitigation policies and plan to respond to emergencies. .

Seismic risk is a function of the degree of earthquake hazard present in a locality (both its probability and its severity) and the value of the property at risk (e.g., a city's building inventory). Risk analysis is a procedure through which these two factors are combined to estimate the potential losses likely to occur from earthquakes of different sizes. Understanding the potential losses that can be expected is an important step in rationally planning to deal with the earthquake hazard.

Risk analysis has been used in a variety of circumstances over the past 20 years. The methods used generally include three components: (1) information about the hazard itself (e.g., the probability and intensity of likely events), (2) an inventory of the property at risk to the hazard, and (3) estimates of the response of various classes of property to different levels of hazard. This paper focus exclusively on the second of these factors: creating an inventory of buildings at risk. See Scawthorn (1986), Jaffe et al., (1984) or Bolton et al.(1986) for more general discussions of the risk analysis technique.

To understand why a building inventory is necessary, it is important to understand that different types of buildings behave differently when subjected to an identical seismic load. For example, unreinforced masonry buildings are notoriously bad for their lack of seismic resistance. They will sustain much greater damage than wood or steel frame structures. The composition of the building inventory is then an important factor in determining the overall damage that results from an earthquake. In addition the value of the structures is necessary if the damage is to be expressed in dollar terms.

BUILDING INVENTORIES

While considerable scientific attention has been given to improving our understanding of the geology and seismicity of earthquake-prone areas, less attention has been given to the task of conducting the inventory of property at risk. Since the characteristics of the property at risk are a key determinant of expected damage, it is necessary to know the number and value of the buildings exposed to an earthquake hazard by structural type. A typical building inventory classifies structures into a finite number of types that are known to experience similar vulnerability to the hazard.

Insurance companies use a structure classification scheme for determining earthquake (and fire) premiums (See Insurance Services Office, 1983). Other classification schemes range from as few as four classes (French and Isaacson, 1984), to more elaborate classifications, such as the twelve classes used in the Rapid Visual Screening Method (Applied Technology Council, 1988), or the forty building classes used by the Earthquake Engineering Facility Classification (Applied Technology Council, 1985). Most of these schemes classify structures based on their construction material, type of framing system, and the age of the structure (which determines the stringency of the building code in force at the time of construction).

Table 1 lists a set of building types should be adequate for most urban risk assessments. Once the number and value of structures in each class is known, this information can be fed into a risk analysis model. In the risk analysis process a separate vulnerability function is applied to each structure type.

There is little theoretical uncertainty associated with an inventory of structures (after all, it is possible to inspect each structure and count the number in each class). However, no one source maintains the kind of data needed to support the sophisticated damage assessment models that have been developed over the past decade. When a risk assessment for a large urban region is desired, the problem of obtaining a suitable inventory is not trivial.

TABLE 1
TYPICAL BUILDING CLASSIFICATION SCHEME

Wood Frame
Steel Moment Resisting Frame
Braced Steel Frame
Light Metal
Steel Frame with Concrete Shear Wall
Concrete Tilt-Up
Precast Concrete Frame
Reinforced Concrete with Moment Resisting Frame
Reinforced Concrete Shear Wall, no MRF
Reinforced Masonry
Unreinforced Masonry Infill
Unreinforced Masonry

The cost of conducting a complete field inspection is prohibitive. Furthermore, once the inventory is complete, the data begins to decay. As buildings are added and demolished the inventory changes, therefore the "half life" of any data collected by field inspection is relatively short. This makes the expense all the more prohibitive.

Using secondary data poses a different set of problems. While much of the needed data is available from secondary sources, it is scattered among assessor's files, building permit files, real-estate files, census data, and other databases. None of these sources has all the data required, and it is not in a form readily used by most risk analysis models. The most effective approach seems to be to use data from one or more sources to generate the needed structural information. This suggests finding factors such as age and size that are strongly correlated with structure type. If such factors exist, the structure inventory can be produced from secondary data.

ASSESSOR'S LAND PARCEL DATA

Tax assessors maintain land parcel files that provide information about each parcel within a jurisdiction. In many jurisdictions these files are stored in a database form. Assessors parcel files are generally thought to be the single most important source of existing data for conducting a building inventory (see Jones et al., 1987). These files are particularly useful because they (1) provide nearly complete coverage of all parcels in a jurisdiction, (2) have at least a minimal level of information such as size and value about structures and (3) are widely available in machine readable form. Unfortunately, the content and format of these files is likely to vary considerably between jurisdictions. For example, California law requires only that the county tax assessor maintain a file which includes the following items:

1) a unique identification number,
2) an address for tax billing,
3) the assessed value of improvements,
4) the assessed value of the land, and
5) any exemptions which apply to the property.

Most other states require tax assessor's to maintain similar information. Thus, the assessor's parcel file seems to be the logical starting point for any structure inventory.

For tax purposes the assessor maintains the value of the land and any improvements separately. For the purpose of conducting a structure inventory, it is helpful that the improvement (structure) value is separated from the land value. In California these values are, however, assessed values which may not accurately reflect replacement cost due to the valuation methods instituted as a result of Proposition 13. Under Proposition 13 assessed values were set at their 1975 level and can only be increased by 2% annually, unless the property is sold. When a property is sold the assessed valuation is set at its market value. The increase in valuation is then limited to 2% annually from the new value until another change in ownership occurs. Thus, assessed improvement costs will only equal market value for those properties that have changed ownership in a given year. It is possible to use recent values to estimate market value for properties with outdated valuations.

Another useful piece of information available from the assessor's parcel file is the type and amount of any exemptions on the property. The most common type of exemption is the homeowner exemption. This is useful in that it is an indicator that the structure is a residence and that the structure is owner-occupied.

Nearly all tax assessors, however, augment the basic required information considerably to support their appraisal activities. Most assessors maintain additional information that includes the square footage of structures on the parcel, the number of stories, year of construction, indicators of construction quality and an existing land use category for the property (e.g., single family residential or retail commercial). Many of these land use coding schemes are derived for the SIC code.

The Santa Clara County Assessor maintains a database of 450,000 records. In Santa Clara County the assessor's data is divided into two systems the Parcel Data System (or

Assessor's Master File) and the Appraisal Data System. The Parcel Data System provides the basic information required by the state Board of Equalization. The Appraisal Data System provides more detailed information, about the attributes of the structure including number of square feet, number of stories and year of construction. For commercial and industrial structures this database also includes a categorical indicator describing the parcel's topography and an index of construction quality. For single family residential properties the database includes a categorical indicator of construction quality (primarily used to identify custom finishes, walls and flooring materials, etc.) and a topographic flag that indicates hillside construction.

To support development and testing of our knowledge-based system, we obtained the assessor's data for 20,420 parcels in and around Palo Alto, California. This data was provided Santa Clara County Center for Urban Analysis in ASCII format on a 9 track tape. Table 2 shows the data fields obtained from the Santa Clara Assessor's files.

TABLE 2
SANTA CLARA COUNTY ASSESSOR'S DATA ACQUIRED

Assessor's Parcel Number
Tax Rate Area
Owner Name
Land Use Code
Zoning Classification
Land Value
Improvement Value
Personnel Property Value
Exemption Value
Site Address
Postal City
Zip Code
Topographic Flag
Construction Class
Construction Quality
Shape
Total Building Area
Number of Floors

The assessor's data was loaded into an Oracle database on a DecStation 3100 running Ultrix 2.1. Having the data in Oracle allows for easy queries, report writing and relational linking to other database tables as they become available.

The Santa Clara Assessor's database includes one field that is rather unique. The Construction Class field describes each building's construction class. This is a rather general classification that puts buildings into five classes. While this field provides a crude basis for a risk assessment, it was considered unsuitable for two reasons. First, it does not categorize the structures finely enough. For example, it does not draw a distinction between unreinforced masonry and concrete tilt-ups. Second, most jurisdictions are not likely to have this level of structural information in their assessor's files. Given that we want to develop a knowledge-based system that can be generalized to a range of different jurisdictions, we decided not to use this field directly. This field was useful in developing the rules of the knowledge base and for checking system validity.

A KNOWLEDGE BASED APPROACH

Clearly, a useable building inventory is critical for any regional seismic risk assessment. The cost of field inspection is just too expensive for large urban areas, especially considering that the data begins to become obsolete almost as soon as it has been collected. While a number of institutions maintain partial inventory files, these files do not include the type of structural information needed to categorize buildings properly for seismic risk assessment. A knowledge-based system that can make judgements about structural types based on the available information provides the best solution to the inventory problem. This is the type of strictly bounded domain in which knowledge based systems have been successfully applied in the past.

A knowledge-based system to support seismic risk assessment should use existing data to classify structures into categories that can be input into a typical risk analysis model. Furthermore, the knowledge-based system should be able to integrate data from multiple sources. The assessor's parcel file provides a good initial set of data, but information from additional special purpose databases should be integrated where appropriate.

A knowledge-based system that classifies structures based on available data has additional benefits associated with how handles updating. If the secondary data source (e.g., the assessor's parcel file) is regularly updated by its owner, getting an updated structure inventory only requires rerunning the knowledge-base system. Thus, the costs of updating the inventory are considerable smaller than if the inventory were based on field inspection.

We use the term knowledge-based system rather than the commonly used "expert system." Since there is no human expert whose rules-of-thumb can be used to develop the knowledge base, the rules must be developed from analysis of the existing data. Relationships between structural type and available characteristics such and age and number of stories must be derived to form the knowledge base.

DEVELOPING THE SYSTEM

A knowledge base consists of a collection of "if-then" rules. These rules act upon a set of facts that are supplied to the system in an interactive consultation or from some machine-readable source, such as a database. In a knowledge-based system the rules and the facts are completely separate. As a result, each can be modified separately. Most knowledge-based systems also have a separate actions module that contains instructions that are activated when particular rules are found to be true or false. Reading and writing to a database or seeking the value of a particular variable are examples of typical actions. While each component is necessary, the critical factor is the rules that make up the knowledge base.

To develop the knowledge base for this project we conducted rather extensive statistical analyses of the Santa Clara Assessor's data file. We used a combination of selection logic and contingency tables to isolate the characteristics of various structural types Using the Chi-Square statistic, we were able to identify those characteristics that were most closely associated with various structural categories. We used the structural class variable from the assessor's data base as the dependent variable to be explained by combinations of other variables.

The expert system shell VP-Expert was used to prototype the knowledge-based system. This is a relatively simple expert system development tool that has good database links. Several other software approaches were considered, such as the Nexpert expert system shell and the Prolog language. The additional power of these systems may be required as the system is refined, but VP-Expert was found to be quite adequate for initial prototyping.

Since most residential construction in the study area consists of wood frame buildings, we decided to focus the prototype development on non-residential structures. Using a standard SQL query, we extracted the 1373 non-residential records from our assessor's database. These records were then transferred to DBase on a microcomputer for compatibility with VP-Expert.

Based on the relationships uncovered by the statistical analysis, we developed a knowledge base consisting of sixteen rules. Each rule contains on or more premises (if clauses) and a single conclusion (then clause). The rules are tested sequentially using the facts that describe a particular record taken from the assessor's parcel database. The rules evaluate one database record (land parcel) at a time. When the evaluation of a particular record is complete, the system assigns it to a building class and moves to the next record.

In the prototype system, the rules act upon the following fields: land use class, improvement value, year of construction, building area, number of floors and building shape. Based on the values of these variables for a particular record, the rule base assigns the parcel to one of seven structure categories. The assigned category is then written back into a blank field in the database and the next record is processed.

Table 3 shows the structure of a typical rule. While a rule can have only a single conclusion, most have numerous premises. The use of multiple premises allows the system to identify that combination of factors most often likely to be associated with a particular building type. In this case the land use class 5 includes most retail commercial buildings and the shape value A indicates that the building is nearly square.

TABLE 3
TYPICAL RULE IN THE KNOWLEDGE BASE

```
IF    YEAR < 1946 AND
      LAND USE CLASS = 5 AND
      SHAPE = A  AND
      NUMBER OF FLOORS <= 2  AND
      BUILDING AREA < 10000

THEN BUILDING CLASS = UNREINFORCED MASONRY
```

Since the assessor's database includes land parcels that may not have buildings, the first order of business is to determine if a given parcel has a building on it. The land use coding scheme assigns values over 90 to unimproved parcels. In addition the system examines the improvement value on the parcel. If the land use class is over 90 or the improvement value is less than $2500, the parcel is classed as no building. This is the easiest classification to make. The model has more difficulty in distinguishing between wood frame and concrete block structures.

Table 4 shows the classification results produced by the initial prototype model for nonresidential buildings of Palo Alto. Obviously the number of "Unknown structures is too high. At this point the model cannot classify roughly a third of the structures. The number of unreinforced masonry structures seems a bit high, while the proportion of wood frame structures seems rather low.

TABLE 4
MODEL'S CLASSIFICATION OF NONRESIDENTIAL STRUCTURES

Structure Type	Number	Percentage
Wood Frame	503	36.6
Unreinforced Masonry	195	14.2
Concrete Tilt-up	23	1.7
Steel Frame	17	1.2
Concrete Wall	0	0.0
No Building	144	10.5
Unknown	491	35.8

The process of improving the accuracy of the judgements made by the knowledge base is an iterative one. The structure class variable from the assessors database was used to check validity of the system's classifications. Comparing the models classifications to the assessor's structure variable suggests that of those buildings the model was able to classify, it is about 80% accurate. Most of the classification error comes from assigning concrete wall structures to the wood frame category. This indicates that the rules associated with wood frame and concrete wall structures need clarification. The statistical analysis of the original data does show that these two structures share many similar features. Assessing how possible it is to develop rules that clearly separate structure types with very similar characteristics is the major challenge in refining this prototype system.

While these results are somewhat disappointing, it is important to remember that this prototype was developed using only nonresidential structures. These buildings were chosen because they are the most heterogeneous and difficult to classify. When we run the model on the full assessors database, the percentage of error will be much lower as the model easily classifies all the one and two story residential structures as wood frame. They structures analyzed here are the most difficult to classify.

FURTHER DEVELOPMENT

The model described here is the initial prototype. We feel it is sufficient for proof of concept, but we intend to elaborate the model considerably to improve its accuracy and to lower the proportion of "Unknown" buildings. Further improvements require balancing two type of error: that associated with being unable to classify a building and that associated with placing the building in the wrong class.

As the knowledge base becomes more complex, we will probably want to move its development to the Nexpert system. Using this system will allow the knowledge base to work directly with the entire Oracle database.

Being able to link to directly to Oracle, a fully relational database, will also allow the model to integrate multiple databases into the model. Several special purpose databases can provide additional detailed information regarding specific types of structures. For example, the California Seismic Safety Commission maintains a database of unreinforced buildings. This data base can be used to cross check the buildings classified as unreinforced masonry.

While we have used the available assessor data for an initial validity checking, a more thorough validity check based on field inspections needs to be undertaken. A sample of 100-200 buildings should be adequate to this purpose. This will allow us to assess the accuracy of the system's predictions more fully.

While a knowledge-based approach to classifying structures appears sound, improving the performance of the model is the major task at hand.

ACKNOWLEDGEMENTS

The work discussed in this paper was supported in part by a grant from the National Science Foundation (BCS-8822125). The author wishes to acknowledge the valuable collaboration of Dr. Craig Howard, the project director, and Dr. Anne Kiremidjian of the Civil Engineering Department at Stanford University. A number of graduate students also contributed to various aspects of this project: Sarah Keown and John Castenon of the City and Regional Planning Department at California Polytechnic State University and John Luce, Rajan Vasudevan and Dimitris Rentzis at Stanford University. We would also like to thank Frank Lockfeld adn Carol Hunt of the Santa Clara County Center for Urban Analysis for their assistance in acquiring the test data.

REFERENCES

Applied Technology Council. 1985. Earthquake Damage Evaluation Data for California . Report ATC-13. Redwood City, California: Applied Technology Council.

Applied Technology Council. 1988. Rapid Visual Screening of Buildings for Potential Seismic Hazards. Report ATC-21. Redwood City, CA: Applied Technology Council.

Association of Bay Area Governments. 1986. Building Stock and Earthquake Losses: The San Francisco BAy Area Example. OAkland, CA: ABAG.

Bolton, Patricia A. Susan G. Heikkala, Marjorie M. Greene and Peter J. May. 1986. Land Use Planning for Earthquake Hazard Mitigation: A Handbook for Planners. Boulder, CO: Natural Hazards Research and Applications Information Center, University of Colorado.

California Seismic Safety Commission. 1987. Guidebook to Identify and Mitigate Seismic Hazards in Buildings. Sacramento, CA: Seismic Safety Commission.

French, Steven P. and Mark S. Isaacson. 1984. "Applying Earthquake Risk Analysis Techniques to Land Use Planning." Journal of the American Planning Association. 50:4 (Autumn) pp. 509-522.

Insurance Services Office. 1983. Guide for Determination of Earthquake Classifications. New York: Insurance Services Office.

Jaffe, Martin, Joann Butler and Charles Thurow. 1984. Reducing Earthquake Risks: A Planners Guide. Chicago: American Planning Association.

Jones, Barclay G., Donald M. Manson, Charles M. Hotchkiss, Micheal J. Savonis and Kimberley A. Johnson. 1987. Estimating Building Stocks and their Characteristics. Ithaca, NY: Cornell Institute for Socioual and Economic Research.

National Research Council. 1989.Estimating Losses from Future Earthquakes. Washington, D.C.: National Academy Press.

Scawthorn, Charles. 1986. "Rapid Assessment of Seismic Vulnerability" in Techniques for Rapid Assessment of Seismic Vulnerability, Charles Scawthorn, ed., New York: American Society of Civil Engineers.