Financial Applications of Fuzzy Case-Based Reasoning to Residential Property Valuation

William Cheetham
Piero Bonissone
GE Research and Development Center

Discussion of homework

- evaluate data
  what technique did you use?
  what did you find?
- retrieval
  how did you select similar cases?
- reuse
  how did you determine a final price?
  did you determine confidence?
- what was you average error?

Outline

- **Problem Description & Motivation**
- **Related Work**
- **APVT Architecture**
- **Comparable-based Approach (Fuzzy CBR)**
- **CBR Design**
- **CBR Validation Stage**
- **Confidence Assessment**
- **Result Analysis**
- **Conclusions**

Introduction

Residential property valuation is the process of determining a monetary estimate of the value of a single family residence.
- at a given location
- at a given time

•Needed to determine:
- –Collateral value for mortgage origination
- –Asset value for mortgage insurance
- –Portfolio value for mortgage packages, etc.

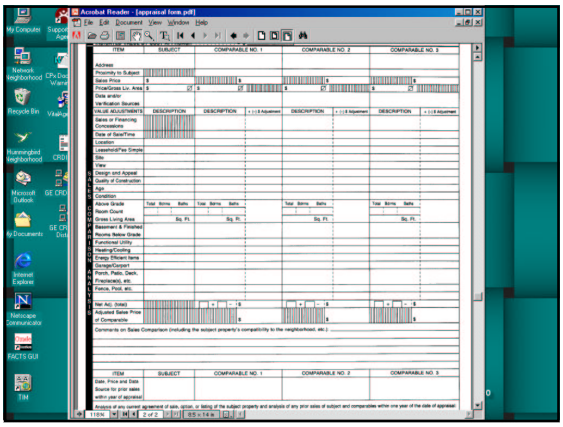
The estimate is called an appraisal = $110,000

Problem Description and Motivation

Appraisals are needed to:
- grant most new mortgages
- evaluate the value of packages of mortgages that may be purchased

The current manual process for appraising properties usually:
- requires an on-site visit by a human appraiser
- costs about $500
- lasts three to four days
- subject to human variability

The most common method used by human appraisers is the "sales comparison approach" which involves:
- finding recent sales comparable to the subject property
- adjusting the comparables' sales price to reflect differences Vs subject
- reconciling the comparables' adjusted prices to create an estimate



Related Work (External)

- Commercial Vendors
  - GA-trained NN for each Census Track model
    - Very computational Intensive
    - Not scaleable nationwide
  - Non-fuzzy Case-based retrieval
    - Brittle Retrieval Process
  - Statistical regressions
    - Difficult to maintain
    - Inconsistent sign in coefficients
    - Not transparent

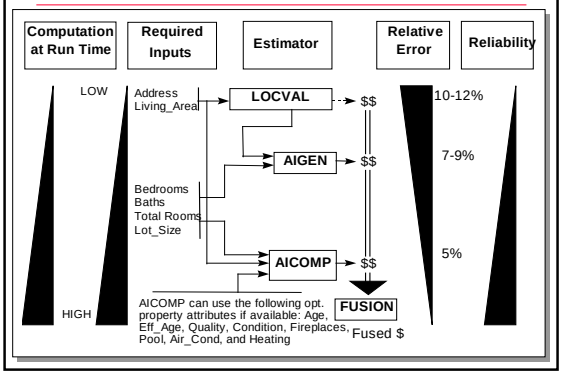
Related Work (Internal)

Multiple methods were used by different GE researchers to develop property value estimators.

- –LocVal estimator
- –StatGen estimator
- –Index based estimator

Method Used	Data Needed	Error (median)
Location Value	Address & Liv. Area	10%
Statistical Formula	10 attributes	8%
Fuzzy-Neural Net	10 attributes	7%
Fuzzy CBR	11-30 attributes	5%
Human Appraiser	Site Inspection	3%

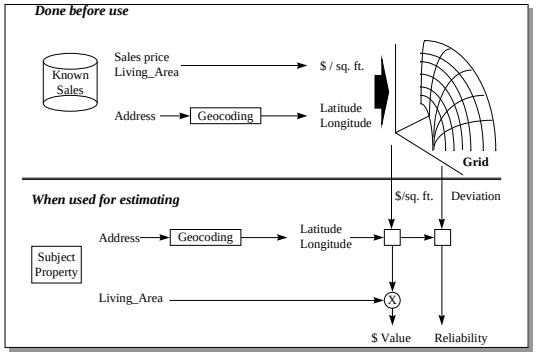
AI Models For Valuation



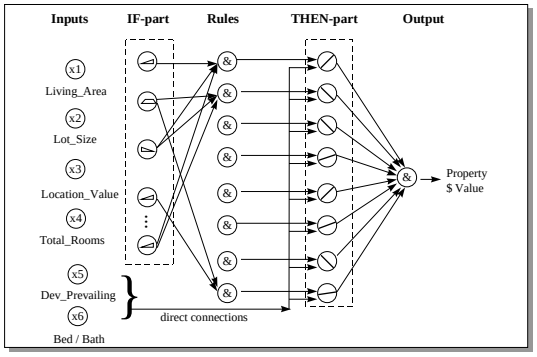
APVT Architecture

- Components:
  - 2 Cascaded Generative Estimators (LocVal & AIGEN)
  - 1 Fuzzy CBR Estimator
  - 1 Fusion module
  - Output of each estimator and of fusion module
    - Value
    - Confidence
    - Justification (only for CBR and fusion)
- See Architecture Diagram

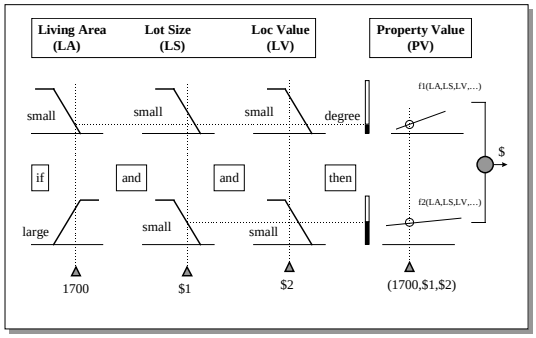
Locational Value Estimator



Generative Estimator (AIGEN)



Generative Estimator (AIGEN) - cont.



Comparable-based Approach (AICOMP)

- AICOMP is a Comparable-based Approach built with Fuzzy CBR technology
- AICOMP Description
  - Process Flow
  - Preference Criteria & Similarity Measures
  - Comparable Adjustments
  - Comparable Filtering
  - Comparable Final Selection and Aggregation

PROFIT Overview

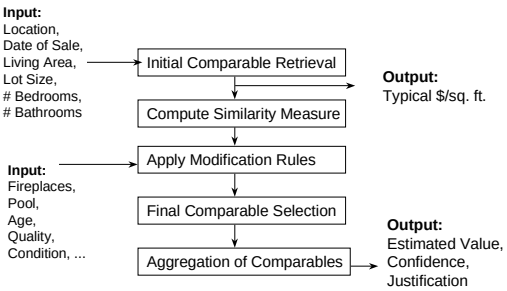
The PROperty Financial Information Technology (PROFIT) system uses:
- case-based reasoning techniques with
- fuzzy predicates and fuzzy-logic based similarity measures

The case-base consisted of over 250,000 properties that had sold in California during the last five years.

Each case consisted of up to 166 property attributes

Attribute	Value
Sale Price	$185,000
Address	2 Bronco Ln.
Living Area	2000 sq. ft
Bedrooms	3
Bathrooms	2.5
etc.	etc.

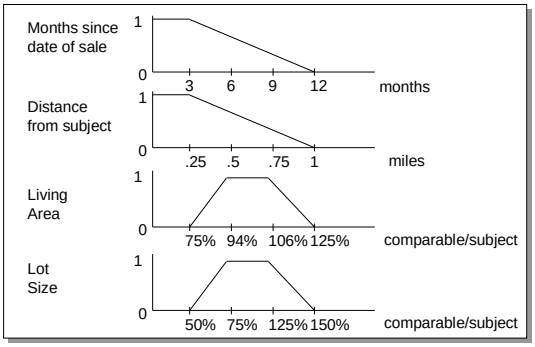
AICOMP Process Flow



Initial Case Retrieval

- Initial retrieval is done by a standard SQL query against a DB.
- The query uses the following attributes & their corresponding maximum allowable deviations
  - Date of sale (within 12 months)
  - Distance (within 1 mile)
  - Living area (+ / - 25%)
  - Lot size (+ 100% / - 50%)
  - Number of bedrooms (+/- 3)
  - Number of bathrooms (+/- 3)
- Retrieved cases are ranked according to (fuzzy membership) preference functions

Preference Functions



Asymmetric Preference Functions

Subject	Comparable								
	1	1.5	2	2.5	3	3.5	4	4.5	5+
1	1.00	0.75	0.20	0.05	0.01	**0.00**	**0.00**	**0.00**	**0.00**
1.5	0.60	1.00	0.60	0.25	0.10	0.05	**0.00**	**0.00**	**0.00**
2	0.10	0.70	1.00	0.70	0.25	0.05	**0.00**	**0.00**	**0.00**
2.5	0.05	0.20	0.75	1.00	0.75	0.20	0.05	**0.00**	**0.00**
3	0.01	0.10	0.40	0.80	1.00	0.80	0.40	0.10	0.05
3.5	**0.00**	0.05	0.15	0.45	0.85	1.00	0.85	0.45	0.30
4	**0.00**	**0.00**	0.05	0.20	0.50	0.90	1.00	0.90	0.70
4.5	**0.00**	**0.00**	**0.00**	0.10	0.30	0.70	0.95	1.00	0.95
5+	**0.00**	**0.00**	**0.00**	0.05	0.15	0.35	0.75	0.95	1.00

Example:

Subject Bathrooms = 2
Comparable Bathrooms = 2.5
Preference = 0.7

Preference Function for Number of Bedrooms

Comparable's # Bedrooms		1	2	3	4	5
	1	1.00	0.50	0.05	0.00	0.00
Subject's	2	0.20	1.00	0.50	0.05	0.00
# Bedrooms	3	0.05	0.30	1.00	0.60	0.05
	4	0.00	0.05	0.50	1.00	0.60
	5	0.00	0.00	0.05	0.60	1.00
	6+	0.00	0.00	0.00	0.20	0.80

Example:

Subject Bedrooms = 5
Comparable Bedrooms = 4
Preference = 0.6

Similarity Measure Computation

- Preference Weighting and Aggregation
  - Weights obtained using experts guesses then optimized
  - The subject property is compared against each comparable along the six variables used in the initial retrieval
  - The preference functions are used to evaluate each variable
  - The similarity measure is the weighted sum of the the preferences

Similarity Measurement Computation

Attribute	Subject	Comparable	Comparison	Preference	Weight	Weighted Preference
Months since date of	X	6 months	6 months	0.67	0.222	0.1489
Distance	X	0.2 miles	0.2 miles	1.00	0.222	0.2222
Living Area	2000	1800	90%	0.79	0.333	0.2633
Lot Size	20000	35000	175%	0.33	0.111	0.0367
# Bedroom	3	3	0%	1.00	0.056	0.0556
# Bathrooms	2.5	2	2.5 -> 2	0.75	0.056	0.0417
Similarity Measure (Sum of Weighted Preference/Sum of Weights) =						**0.768333**

Adjustment Rules

Living Area (subject - comp) * (22 + (Sales_Price_of_comp * .00003))
Lot Area (subject - comp) * 1
Fireplaces (subject - comp) * 2000
Effective Year Built
w * (Age_comp-Age_subject) * (Sale_Price_comp/1000)
if (Age_subject + Age_comp) / 2 < 4 then w = 4 else
if (Age_subject + Age_comp) / 2 < 6 then w = 3 else
if (Age_subject + Age_comp) / 2 < 8 then w = 2 else
if (Age_subject + Age_comp) / 2 < 15 then w = 1 else
w = .5
max of 10% of salePrice
Quality (.02 * sale price) for each level of difference:
(Luxury > Excellent > Good> Average > Fair > Poor)
Pool $10000 for a pool

Adjustment Table for Bathrooms

Subject	Comp								
	1	1.5	2	2.5	3	3.5	4	4.5	5+
1	**0.00**	*-1.50*	-3.00	*-5.00*	*-8.00*	N/A	N/A	N/A	N/A
1.5	1.00	**0.00**	-1.00	*-3.50*	*-6.00*	*-9.00*	N/A	N/A	N/A
2	*4.00*	*1.50*	**0.00**	-2.25	*-4.00*	*-6.50*	N/A	N/A	N/A
2.5	*7.00*	*4.50*	2.00	**0.00**	-2.00	*-4.50*	*-7.00*	N/A	N/A
3	*9.00*	*6.50*	3.00	2.00	**0.00**	-2.50	-5.00	*-7.50*	'@*5
3.5	N/A	*8.50*	*6.50*	*4.50*	*2.50*	**0.00**	*-3.00*	*-5.50*	'@*5
4	N/A	N/A	*8.50*	*7.00*	*5.50*	*3.00*	**0.00**	*-3.00*	'@*5
4.5	N/A	N/A	N/A	*10.00*	*8.00*	*6.00*	*3.00*	**0.00**	'@*5
5+	N/A	N/A	N/A	'@*5	'@*5	'@*5	'@*5	'@*5	**0.00**

Adjustment Example

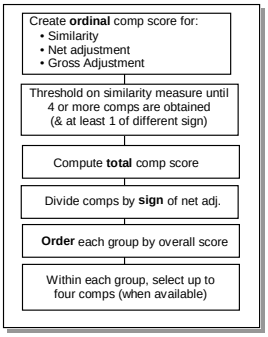
Attribute	Subject	Comparable	Adjustment
SalePrice	?	175000	175000
LivingArea	2000	1800	5450
LotArea	20000	25000	-5000
TotalBaths	2.5	2	2000
Bedrooms	3	3	
Fireplaces	1	0	2000
EffYearBuilt	93	89	2800
Quality	Good	Average	3500
Condition	Average	Average	
Pool	Yes	No	10000
Adjusted Price =			**195750**

Comparable Filtering

We would like the selected comparables to have the following properties:

- No single adjustment should be larger (in absolute value) than 10% of sales price
- Net adjustment should not exceed 15% of sales price
- Gross adjustment should not exceed 25% of sales price
- The unit price for living area of the comparables should not vary more than 15% from each other and should bracket that of the subject
- Comparables should be as close as possible to the subject
- The value estimated for the subject should be bracketed by the sales price of the comparables

Comparable Selection



Example of Comparable Ranking

Comparable	Score Value	Score Rank	Net Adjust Value	N. A. Rank	Gross Adjust Value	G. A. Rank	Total Rank
113-012	0.95	1	1344	2	5924	4	7
306-018	0.88	2	3586	5	4186	1	8
093-011	0.78	3	5686	7	8191	7	17
305-006	0.67	4	6150	8	6160	6	18
685-046	0.64	5	3139	3	6099	5	13
847-984	0.58	6	-948	1	5670	3	10
873-005	0.53	7	-5261	6	9261	8	21
431-023	0.48	8	3546	4	4410	2	14
331-018	0.44	9	9310	9	11300	9	27

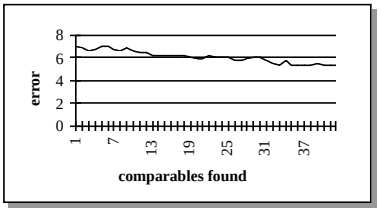
Example of Final Aggregation

Comparable	Adjusted Price	Score	Weighted Price
113-012	197000	0.95	187150
306-008	202000	0.88	177760
093-011	196500	0.78	153270
685-046	192000	0.64	122880
847-984	201000	0.58	116580
Total		3.83	757640
Final estimate = 757640/ 3.83 =			**199900**

Validation Stage

- **System tested on 7,293 properties from Contra Costa county in California, USA.**
- **For each property, we computed:**
  - the predicted sales price of each property & compared it with its actual sales price to derive the estimate's error.
  - the percentage error and its five confidence characteristics
- **With these new data, we:**
  - analyzed the conditional distributions of the estimate's error, given each of its five confidence characteristics
  - used C4.5 to create rules predicting the error from the estimator's characteristics
  - validated these rules via data visualization
  - manually transformed the rules into the membership functions for confidence assessment

Example of P(error |# comps found)

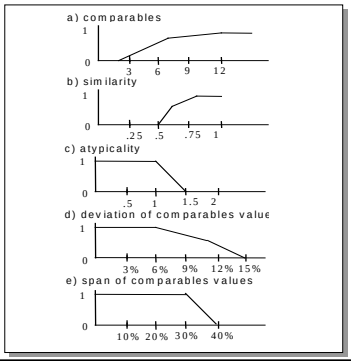


- For instance, the figure above shows that the estimate error (in percentage) decreases as the number of comparables found in the initial retrieval increases.
- Therefore, we can use this number as one of the filters to predict the expected error.

Confidence Assessment

- The confidence value is obtained from the conjunctive evaluation of five soft constraints defined on the estimator's internal parameters:
- Number of cases found in the initial retrieval
  - Average of the similarity values for the best four cases
  - Typicality of problem with respect to the case-base (i.e. if the attributes of the subject fall within typical ranges for the subjects five digit zip code region)
  - Span of adjusted sales prices of highest confidence solutions
  - Distribution of adjusted sales prices of highest confidence solutions

Membership Function for Confidence Assessment



Random Sample of 10 of 7293 Validation Tests

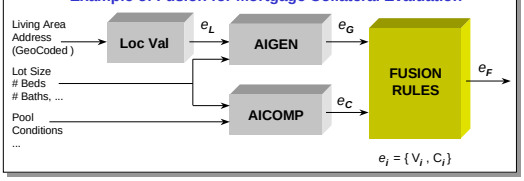
Error	Comp Found	Simil.	Atyp.	Comp Dev.	Comp Span	Conf. Value	
-9.8	3	0.63	1.42	2.02	6.32	**0.15**	
-2	35	0.94	0.38	2.24	8.57	**1.00**	
17.3	11	0.71	0.94	5.67	19	**0.70**	
0.5	24	0.85	0.66	2.05	7.24	**1.00**	
-1.6	14	0.95	0.29	2.89	9.33	**1.00**	
5.2	15	0.90	0.73	3.24	12	**1.00**	
5.2	12	0.74	0.17	4.5	18	**0.80**	
3.1	19	0.74	0.81	2.83	8.11	**0.80**	
-14	12	0.82	1.97	3.85	15	**0.00**	
7.8	11	0.77	1.34	4.24	13	**0.32**	

- Each row is a different property.
- Columns show the estimate's error, its five characteristics, and its confidence value (CV)
- CV is the conjunctive (minimum) evaluation of the membership functions used for confidence assessment

Fusion of Reasoning Models

- **Develop Collection of Quasi-independent Models**
- **Each Model Generates:**
  - Output Value (v_i) - **Prediction**
  - Confidence parameter (c_i) derived from training stats. - **Introspection**
- **Intelligent Fusion Rules**
  - Consider discrepancies among Output values (v)
  - Consider dynamic confidence parameter (c) associated with each output

Example of Fusion for Mortgage Collateral Evaluation



Analysis of Results (cont.)

- The confidence value was subdivided into three groups (*good, fair, and poor*) to identify the largest good set with the lowest error.
- Results of testing 7,293 properties:

Label	Group Size [% of Test size]	Median Absolute Relative Error
Good (E)	*63%*	*5.4%*
Fair (I)	*24%*	*7.7%*
Poor (U)	*13%*	*11.8%*

Conclusions

- Developed a CBR system for residential property valuation, which generates an estimate and a confidence value
- The system uses Fuzzy Logic to translate current appraisers practices into:
  - Retrieval preference criteria
  - Similarity computation
  - Solution adaptation
  - Confidence value generation
- The confidence value determine the CBR estimate suitability for decision making.

Conclusions (cont.)

- The system scalability was proven by thousands of transactions used in validation stage.
- The system can also be used to validate a property value provided by an external source.
  - The system identifies the best set of comparables to justify the given value and provides an associated confidence value.
- Possible Future work:
  - automatic case-base maintenance and update (determination of whether the selection or adaptation rules need to be changed, due to changing market conditions)
  - automatic generation of the new selection & adaptation rules.