# Soft Computing Models in Online Real Estate

Jozo Dujmović[1], Guy De Tré[2], Navchetan Singh[1], Daniel Tomasevich[1], and Ryoichi Yokoohji[1]

[1] Department of Computer Science, San Francisco State University, San Francisco, CA, USA
jozo@sfsu.edu
[2] Dept. of Telecommunications and Information Processing, Ghent University, Ghent, Belgium
Guy.DeTre@telin.ugent.be

**Abstract.** In this paper we present a decision support system that uses soft computing models for evaluation, selection and pricing of homes. The system (called LSPhome) is based on the Logic Scoring of Preference (LSP) evaluation method and implemented in the context of online real estate. The goal of this system is to use weighted compensative logic models that can precisely express user needs, and help both buyers and sellers of homes. The design of such a system creates specific logic and computational challenges. Soft computing logic problems include the use of verbalized importance scales for derivation of andness, penalty-controlled missingness-tolerant logic aggregation, detailed and verbalized presentation of evaluation results, and development of optimum pricing models. Computational problems include fast and parallel collection of heterogeneous information from the Internet, and development of user interface for fast and simple creation of customized soft computing decision criteria by nonprofessional decision makers.

**Keywords:** Evaluation, selection, real estate, missing data, verbalization.

## 1 Introduction

Real estate is an area that includes a spectrum of soft computing decision problems. In this paper we present a survey of the most important soft computing models that are used in online real estate (ORE). The first such a problem is the development of criteria for evaluation and selection of homes. The home evaluation criteria are based on weighted compensative logic functions that can model adjustable degrees of simultaneity and replaceability, mandatory, sufficient, and optional requirements, as well as adjustable degrees of importance of various home attributes. The aggregation of home quality and home affordability is also a soft computing logic problem. Similarly, the problem of optimum home pricing can also be solved using soft computing models. In ORE we frequently encounter problems of decision making with incomplete (missing) inputs, and the need to expand aggregation models with missingness-tolerant aggregators. Finally, the users of ORE decision models are not decision experts, but nonprofessionals who need simple verbalized approach to specifying soft computing decision models. These seemingly heterogeneous problems are closely related in the

M. Jamshidi et al. (eds.), *Advance Trends in Soft Computing WCSC 2013*,
Studies in Fuzziness and Soft Computing 312,
DOI: 10.1007/978-3-319-03674-8_8, 

context of ORE. Thus, the goal of this paper is to show all fundamental components of the soft computing decision infrastructure in ORE.

In the USA the real estate market data and procedures are governed by the National Association of Realtors [11]. Full information about homes on sale and other marketed properties can be found in the Multiple Listing Service (MLS) [14]. ORE web sites (e.g. [13],[16]) use MLS data and provide application programming interfaces (API) that can be used to access data about available homes and their characteristics. These data can be used as inputs for evaluation and selection process based on soft computing criteria.

The paper is organized in three main sections. Section 2 describes soft computing models for home evaluation in the context of buying and selling a home. Section 3 surveys the penalty-controlled missingness-tolerant aggregation, and the verbalization problems. Section 4 presents experimental results generated by the LSPhome system.

## 2 LSP criterion Function for Home Evaluation

The LSP method [5] provides soft computing evaluation criteria built in three basic steps. The first step develops a list of attributes $a_1,...,a_n$ , $a_i \in R,\ i = 1,...,n$ that characterize relevant properties of evaluated homes. The second step is to provide requirements for each attribute in the form of elementary criteria functions $g_i : R \to I,\ I = [0,1]$; they assign degrees of satisfaction to attribute values $x_i = g_i(a_i),\ i = 1,...,n$. The third step generates an overall degree of satisfaction (overall suitability) as an aggregate of attributes' degrees of satisfaction: $S = L(x_1,...,x_n)$. The mapping $L : I^n \to I$ is based on weighted compensative logic functions [1],[10],[3],[12] that are implemented as specific forms of means [9], [2].

### 2.1 Attribute Tree

The home evaluation attribute tree based on data that can be retrieved from the Internet is shown in Fig. 1. The attributes are grouped in two main groups: the quality of home location, and the quality of the home. The quality of home location is based on an analysis of points of interest that are available from Google. The attributes that affect the home quality come from ORE web sites.

### 2.2 Elementary Criteria

The number of home evaluation attributes in the attribute tree in Fig 1. is 36. For each of these attributes we provide an attribute criterion function that reflects the requirements of a specific user. Some of attribute criteria are specific for each user and others can be shared by all users. Two such examples are shown in Fig. 2. The criterion #112 uses data obtained from the LSPhome user interface shown in Fig.3.

```
1 HOME SUITABILITY

   11 QUALITY OF LOCATION
      111 Suitability of neighborhood
         1111 Walkability
         1112 Shopping and dining
         1113 Health support
         1114 Suitability for children
         1115 Suitability for seniors
      112 Walking distance from the ideal location

   12 QUALITY OF HOME
      121 Available space
         1211 Area belonging to home
            12111 Total internal living area of home
            12112 Outer usable area belonging to home
         1212 Rooms and other designated areas
            12121 Primary rooms
               121211 Number of bedrooms
               121212 Number of bathrooms
               121213 Kitchen
               121214 Dining room/area
               121215 Living/family room
            12122 Additional space and storage
               121221 Additional space
                  1212211 Breakfast room/area
                  1212212 Home office
                  1212213 Laundry
               121222 Storage and auxiliary areas
                  1212221 Walk-in closets
                  1212222 Pantry
                  1212223 Auxiliary utility areas
         1213 Parking space
            12131 Reserved parking
               121311 Garage
                  1213111 Private garage
                  1213112 Shared garage
               121312 Reserved uncovered parking space
            12132 Public parking (first-come, first-served)
               121321 Free public parking
                  1213211 Street parking next to home
                  1213212 Street parking close to home
               121322 Paid public parking
      122 Home features
         1221 Home organization/layout
            12211 Type of home
            12212 Number of floors
         1222 Home construction features
            12221 External wall material
            12222 Type of floor
            12223 Type of roof
         1223 Home energy supply
         1224 Home temperature regulation
            12241 Source of energy for heating
            12242 Type of heating system
            12243 Type of cooling system
         1225 Home age and maintenance
            12251 Home age
            12252 Last modification/improvement
```

**Fig. 1.** The home evaluation attribute tree

The presented interface provides a limited capability for homebuyers to specify their requirements. This is necessary to avoid too much detail that would discourage the majority of general population users. In all cases the users are expected to specify the ideal location of their desired home and the maximum allowable distance $D_{max}$ from the ideal location. The evaluation of homes using the attribute criterion #112 (Fig. 2) is based on the relative distance $100D/D_{max}$ . The presented attribute criterion shows a relatively high tolerance for all distances except those close to $D_{max}$ . By selecting $D_{max}$ (see Fig. 3) the users can customize the attribute criterion function.

| 112 | | Walking distance from the ideal location |
|---|---|---|
| Value | % | The ideal location is a user-specified location selected as a point that completely satisfies all user requirements. The distance can be expressed as (1) walking, (2) car, (3) public transport, or (4) bicycle distance. We use the normalized relative walking distance x = 100D/Dmax, where<br>D = walking distance between an evaluated home and the ideal location (miles or km)<br>Dmax = The maximum acceptable walking distance from the ideal location (miles or km). *Dmax must be selected by each user* |
| 0 | 100 | |
| 40 | 86 | |
| 70 | 70 | |
| 90 | 50.6 | |
| 100 | 0 | |
| **12222** | | **Type of floor** |
| Value | % | The type or material of the walking surface of the primary living areas of the home. The main options are:<br>ST = stone, HW = hardwood, SW = softwood<br>L = laminate floor, V = vinyl/linoleum, P = parquet<br>SL = slate, T = tile (ceramic), C = carpet<br>Evaluation method:<br>1 = ST/SL/T, 2 = V, 3 = SW/C, 4 = L, 5 = HW, 6 = P |
| 1 | 35 | |
| 2 | 50 | |
| 3 | 70 | |
| 4 | 75 | |
| 5 | 85 | |
| 6 | 100 | |

**Fig. 2.** Sample elementary criteria

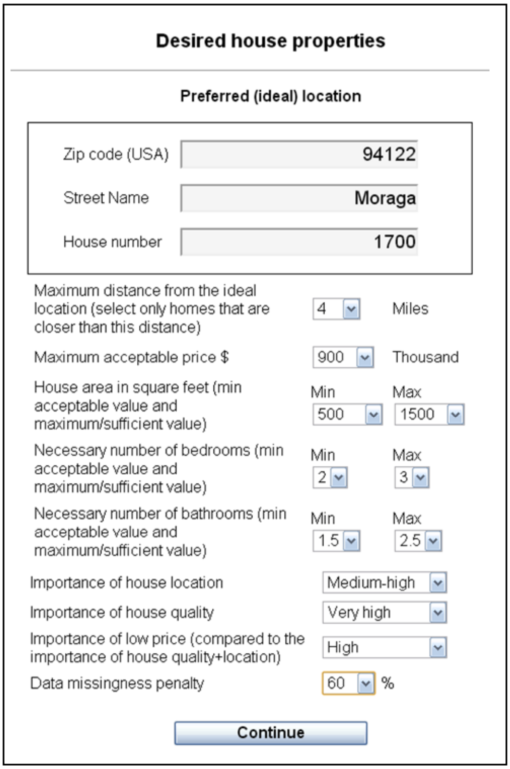


**Fig. 3.** LSPhome interface for specifying user requirements

Other user supplied elementary criteria are the available area, the number of bedrooms and the number of bathrooms. All of them are specified in the range from the minimum acceptable value $a_{\min}$ to the maximum (sufficient) value $a_{\max}$. The simplest form of such elementary criteria is the following: $x = g(a) = \min[1, (a / a_{\max})]$.

An alternative more flexible version can be obtained by assigning the minimum default suitability $x_{\min}$ to the minimum acceptable value $a_{\min}$ as follows:

$$x = g(a) = \begin{cases} 0, & a < a_{\min} \\ \min\left(1, \left(a - a_{\min} + x_{\min}(a_{\max} - a)\right) / \left(a_{\max} - a_{\min}\right)\right), & a \geq a_{\min} \end{cases}$$

To simplify the use of LSPhome, only the essential user requirements are customizable. All user-shareable and less specific elementary criteria are not customizable and one such example is the criterion #12222 shown in Fig.2. That criterion is a fixed scoring system that reflects an average standpoint acceptable for the majority of users. E.g., if the ORE web site provides a home with hardwood floor, then, for all homebuyers, the corresponding floor satisfaction degree is 85%. The use of fixed elementary criteria significantly reduces the number of necessary user inputs and simplifies the communication with users.

### 2.3 Logic Aggregation Structure

Aggregation of all attribute suitability degrees yields the overall suitability of the evaluated home. The aggregation is based on the superposition of several basic aggregators that are implemented using the generalized conjunction/disjunction function (GCD) [4]. The soft computing *suitability aggregation structure* (SAS), in the form of a "shade diagram," [5] is shown in Figs. 4 and 5. The suitability aggregation structure uses a spectrum of weighted compensative logic functions. In the case of GCD we use the system of 17 distinct degrees of orness $\omega = 0, 1/16, ..., 1$, (or andness $\alpha = 1 - \omega$) symbolically denoted C, C++, C+, C+-, CA, C-+, C-, C--, A, D--, D-, D-+, DA, D+-, D+, D++, D, described in [3]. The aggregators starting with letter C denote various forms of conjunction (pure and hard or soft partial) and aggregators starting with letter D denote various forms of disjunction (pure and hard or soft partial) [4]. The hard partial conjunction function is a model of mandatory requirements ($f_c(x_1, ..., x_k) = 0,\ x_i = 0,\ i \in \{1, ..., k\},\ k > 1$) and the hard partial disjunction is a model of sufficient requirements ($f_d(x_1, ..., x_k) = 1,\ x_i = 1,\ i \in \{1, ..., k\},\ k > 1$). Soft versions provide a positive output if a single input is positive. The aggregator A denotes the neutrality (the arithmetic mean). E.g., to evaluate the suitability of neighborhood (#111) we first identify the locations of all relevant points of interest for evaluation of walkability, shopping and dining, health support, and suitability for children and seniors, and then we aggregate these suitability degrees using a weighted soft partial conjunction C- (andness $\alpha = 5/8$), with the highest relative importance (weight) assigned to walkability and to suitability for children. Such weights reflect the fact that most homebuyers are young families. In Figs. 4 and 5 "+" in the first column denotes mandatory attributes and "-" denotes optional attributes.

The SAS shown in Fig. 4 includes a user-supplied final aggregator $F$ of location suitability and home quality. The user is requested to specify in verbal form the overall importance of location suitability and house quality (Fig. 3) and the parameters of the resulting GCD aggregator are then computed as shown in Section 3 and [7]. The resulting overall suitability scores for $N$ competitive homes are $S_i, i = 1, ..., N$.

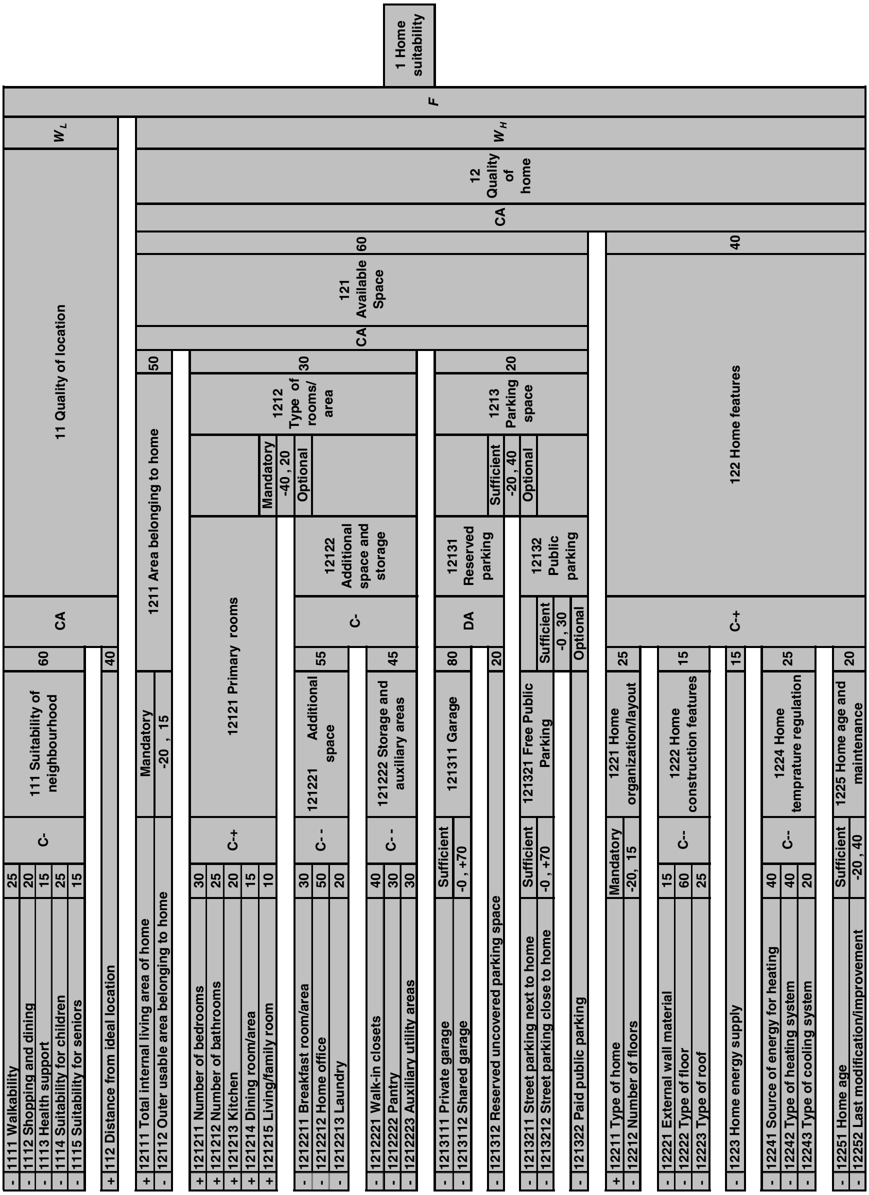


**Fig. 4.** The home criterion suitability aggregation structure (SAS)

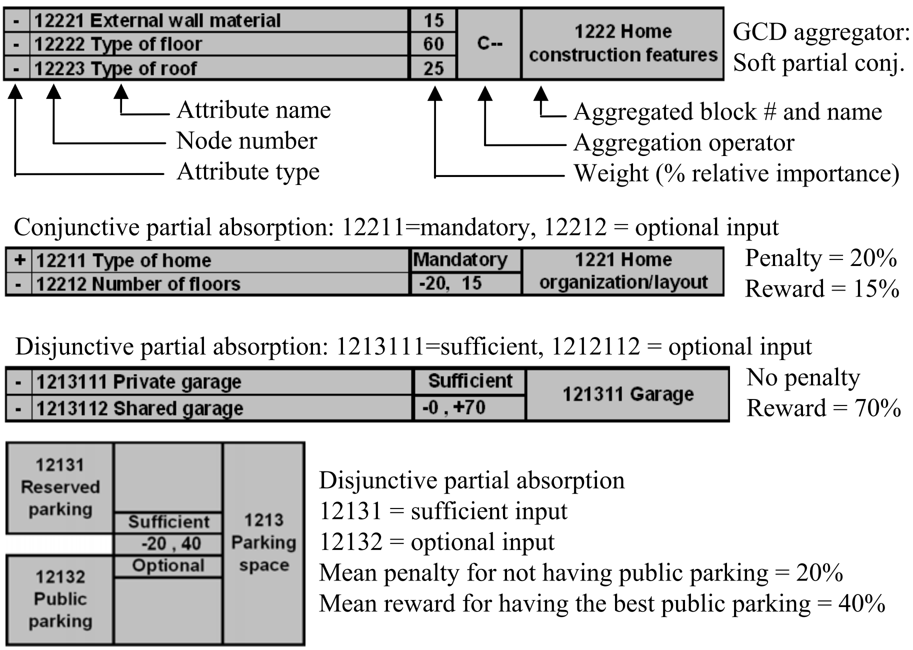


**Fig. 5.** Explanation of fields in the shade diagram

We also express the soft computing logic relationships by using the conjunctive partial absorption aggregators that aggregate mandatory and optional inputs and the disjunctive partial absorption aggregators that aggregate sufficient and optional inputs [3],[5]. In both cases the properties of these aggregators are determined using the desired level of penalty (decrease of output in the case of unsatisfied optional input) and reward (increase of output in the case of perfectly satisfied optional input). E.g., in the case of parking space (#1213) it is sufficient to have a reserved parking place and the availability of public parking is optional with the mean penalty of -20% and the mean reward of +40%. An obvious advantage of shade diagrams is their rectangular form: similarly to Nassi-Shneiderman structured flow charts, a new shade diagram can be inserted in each rectangular space, making easily readable aggregation structures. Shading of diagrams facilitates the perception of grouping of inputs.

If home costs are $C_1,...,C_N$ then logic aggregation also includes a hard partial conjunction ($\bar{\Delta}$) for aggregating the overall home quality $Q_i = S_i / \max(S_1,...,S_N) \in [0,1]$ and the home affordability $A_i = \min(C_1,...,C_N)/C_i \in [0,1]$ yielding (in the case of equal weights) the overall home values $V_i = Q_i \bar{\Delta} A_i,\ i = 1,...,N$. In the case of home sale this model can be used to find the *maximum price of our home* $C_i^*$, so that (even with that price) the home is still the most attractive in a selected area (attains the maximum value $V_i = Q_i \bar{\Delta} \min(C_1,...,C_N)/C_i^* \geq Q_j \bar{\Delta} \min(C_1,...,C_N)/C_j,\quad \forall j \neq i$).

## 3 Missingness-Tolerant Aggregation and Verbalization

The LSP criterion function consists of attribute criteria and the suitability aggregation structure and assumes the availability of all $n$ input attribute values. In reality, however, the ORE web sites regularly offer incomplete data about available homes. For example, our experiments with homes available through ORE API in San Francisco show on the average the availability in the range from 50% to 70% of input attributes, as illustrated in Fig. 6. For each of ten zip codes we averaged the availability of attributes for all marketed homes providing reliable insight into the missingness problem. The home attribute data come from various sources: home owner/seller, county records, and broker listing feeds, and some of them are frequently incomplete. So, we have two options: to abandon the idea of home evaluation and selection using ORE data, or to use techniques for penalty-controlled missingness-tolerant aggregation. We use the method presented in [8] where the user can select the degree of penalty $P \in [0,1]$ (or $P \in [0,100\%]$) for missing data, as shown in Fig. 3. Then nonnegative inputs $x_i \geq 0$ correspond to known attributes, and negative inputs denote unknown attributes defined as $x_i = P - 1$. So, $x_i = 0$ denotes either no satisfaction of the corresponding elementary criterion or the maximum penalty assigned to an unknown attribute. In the case of negative suitability we have $0 \leq P < 1$, $-1 \leq x_i < 0$ and the zero penalty yields $x_i = -1$. Our missingness-tolerant aggregation structure maps $[-1,1]^n \rightarrow [0,1]$; for details, see [8].

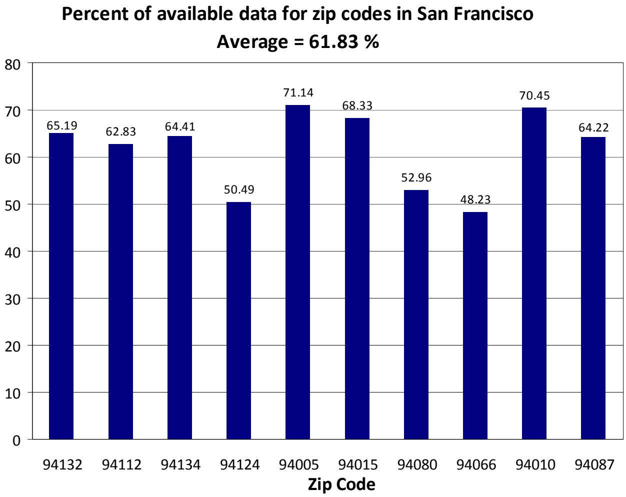


**Fig. 6.** ORE data availability for ten zip codes in San Francisco

The aggregation of suitability degrees is related to the perception of the importance of inputs. For example, if a homebuyer requires a high degree of simultaneity of the home quality and the home location quality, that requirement necessarily yields a perception that both the home quality and the location quality are (for that specific homebuyer) very important. Thus, a high andness is a consequence of high overall importance of inputs. Similar situation also holds in the case of high orness. However, while the concepts of andness and orness are familiar to professional decision-makers, the concept of overall importance is familiar to everybody. This fact can be used to derive the andness/orness and other parameters of partial conjunction and partial disjunction from the verbalized perception of the overall importance of inputs. This idea was introduced in [7] and implemented in the LSPhome interface shown in Fig. 3 where users can select verbalized degrees of importance of home location, home quality, and home price. Using the method presented in [7] the selected degrees of importance of home location and home quality are used to derive the andness and weights for the final suitability aggregation block ($W_L, W_H, F$, Fig. 4). Then, the mean importance of location and home quality and the importance of price are used to derive the andness and the weights of the aggregator that aggregates the overall suitability and the overall affordability and provides the overall home value.

If the user wants a simultaneous satisfaction of $k$ inputs and has the perception that their levels of overall importance (selected from the verbalized importance scale with $L+1$ levels) are $S_1,...S_k$, $S_i \in \{0,...,L\}$, $i = 1,...,k$, $k > 1$, then, according to [7], the corresponding andness is interpreted as the mean relative overall importance: $\alpha = (S_1 + ... + S_k)/kL$. Indeed, the perception of importance and the value of andness increase simultaneously and the above model is based on the linear relationship between the andness and the mean overall importance. The verbalized overall importance scale can also be used to derive the degrees of relative importance of inputs. Among three linear models proposed in [7] for computing weights $W_1,...W_k$, the simplest is the proportional scaling model $W_i = S_i / (S_1 + ... + S_k)$, $i = 1,...,k$.

## 4 Experimental Results

Using the LSP methodology presented in previous sections we developed a web application called LSPhome that helps users to find the most suitable home according to their specific criteria. Traditional ORE web sites offer searches of available real estate inventory in the style of the traditional SQL SELECT-FROM-WHERE statement. The user is only allowed to specify a few crisp conditions in the WHERE clause. Such conditions are used as a filter, i.e. as a strictly binary selector that rejects all homes that do not satisfy any of the filtering conditions selected by the user. In a typical case the ORE web sites offer the filter conditions from the following list: (1) home type, (2) price range, (3) minimum number of bedrooms, (4) minimum number of bathrooms, (5) square feet range, (6) lot size range, (7) home age range, (8) time on market, (9) keywords used to select desired features (e.g. pool, patio), and (10) desired location/neighborhood. The filtering method considers all selected filtering conditions as the binary mandatory requirements. For example, if the home type is specified as condo/apartment, then all single family and multi family homes will be rejected. Obviously, the filtering process is useful, but it is not the home evaluation.

The home filtering is merely the partitioning of inventory in two basic groups: homes that are not acceptable and homes that might be acceptable. At this time the customers of ORE web sites do not have the possibility to determine the degree of acceptability (suitability) according to their specific needs. Evaluation and ranking of potentially suitable homes is left to the user and it is done intuitively. Of course, the number of attributes is too big for easy intuitive evaluation, and the process of home selection is usually stressful and time consuming.

The primary advantages of the soft computing approach and the LSP method with respect to the traditional filtering process are the evaluation and ranking of homes according to user needs, the reduction of search/decision time, and the justification of proposed decisions; that increases the confidence and improves the experience of homebuyer (and/or home seller). Of course, the central problem is how to define the user needs. The number of home attributes that we used (36) is a typical value and it is difficult to reduce it without losing the credibility of evaluation results. On the other hand, it is not reasonable to ask an average homebuyer to specify 36 elementary criteria (or fuzzy set membership functions) followed by an advanced aggregation structure. Thus, we proposed a hybrid approach: the user specifies 9 crucial requirements using the LSPhome interface shown in Fig. 3 and the remaining parts of the LSP criterion are prefabricated (fixed, reflecting average general requirements). In this way we combine the simplicity of specification of requirements and the breadth of covering relevant attributes. In particular, a significant advantage of our method is the integration of the home quality and the location quality attributes, with the possibility to conveniently adjust the relative importance of home quality versus the location quality. The location quality analysis is based on data about all points of interest provided by Google using techniques developed in [6] and [15].

**LSPhome has found the following houses for you:**

| Rank | Score [%] | Price | Value | Address |
|---|---|---|---|---|
| 1 | 76.94 | 890000 | 100.00 | 1746 20th Ave, San Francisco, CA 94122 |
| 2 | 71.41 | 850000 | 98.58 | 1675 26th Ave, San Francisco, CA 94122 |
| 3 | 71.01 | 860000 | 97.73 | 1621 31st Ave, San Francisco, CA 94122 |
| 4 | 67.80 | 827000 | 97.38 | 1242 12th Ave, San Francisco, CA 94122 |
| 5 | 29.20 | 649999 | 72.08 | 1350 41st Ave, San Francisco, CA 94122 |
| 6 | 36.01 | 900000 | 68.03 | 1699 19th Ave, San Francisco, CA 94122 |
| 7 | 35.33 | 900000 | 67.38 | 1683 27th Ave, San Francisco, CA 94122 |
| 8 | 31.31 | 900000 | 63.44 | 1526 45th Ave, San Francisco, CA 94122 |
| 9 | 26.49 | 780000 | 62.68 | 1879 48th Ave, San Francisco, CA 94122 |
| 10 | 23.47 | 800000 | 58.25 | 1207 30th Ave, San Francisco, CA 94122 |

Show all competitive systems    Show all evaluation results

Make LSP suitability maps    Analyze the quality of urban location

New home search    Done

**Fig. 7.** An example of typical LSPhome evaluation results

A typical example of the summarized home evaluation and selection results generated by the LSPhome system is shown in Fig. 7 ("score" denotes the overall suitability). The user looking for a home in the vicinity of the 19th Avenue in San Francisco is given the ranking of 10 homes selected by LSPhome. The first four homes satisfy more than 2/3 of user requirements and other have too low suitability scores. The overall value is computed as a hard partial conjunction of the normalized suitability and normalized affordability: $V_i = W(\min(C_1,...,C_N)/C_i)\ \overline{\Delta}\ (1-W)(S_i/\max(S_1,...,S_N)) \in [0,1]$, $i = 1,...,N$ . The weight $W$ denotes the relative importance of affordability compared to the relative importance of home quality. It is computed from the importance of low price selected by the homebuyer using the LSPhome interface (Fig. 3). The results in Fig. 7 show the normalized relative value $V_i^{(norm)} = 100V_i / \max(V_1,...,V_N))$, $i = 1,...,N$, so that the top ranking home is rated 100%. In our example the four leading homes differ for less than 3%, while others have significantly lower values. Consequently, in this case the user is expected to focus on the four best options, compare homes using the values and the suitability of attributes (Fig. 8 and Fig. 9) and expand the investigation using suitability maps or a detailed analysis of the quality of urban location.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Walkability | 47.0 | 54.0 | 49.0 | 49.0 | 23.0 | 44.0 | 32.0 | 82.0 | 53.0 | 40.0 |
| Shopping and dining | 76.0 | 79.0 | 79.0 | 84.0 | 55.0 | 83.0 | 66.0 | 87.0 | 80.0 | 81.0 |
| Health support | 38.0 | 42.0 | 64.0 | 66.0 | 41.0 | 59.0 | 35.0 | 74.0 | 64.0 | 63.0 |
| Suitability for children | 64.0 | 34.0 | 56.0 | 59.0 | 36.0 | 65.0 | 38.0 | 89.0 | 56.0 | 69.0 |
| Suitability for seniors | 40.0 | 39.0 | 61.0 | 59.0 | 28.0 | 58.0 | 51.0 | 79.0 | 63.0 | 54.0 |
| Walking distance from the ideal location | 1.76 | 1.75 | 0.02 | 0.65 | 1.29 | 0.72 | 1.92 | 1.01 | 0.26 | 0.88 |
| Total internal living area of home | 1350.0 | 982.0 | 1700.0 | 1385.0 | 1152.0 | 2266.0 | 1647.0 | 1046.0 | 2190.0 | 1686.0 |
| Outer usable area belonging to home | 1650.0 | 2014.0 | 1700.0 | 1611.0 | 1500.0 | 206.0 | 474.0 | 1046.0 | 2190.0 | 1912.0 |
| Number of bedrooms | 2.0 | 2.0 | 3.0 | 2.0 | 2.0 | 5.0 | 4.0 | 3.0 | 4.0 | 2.0 |
| Number of bathrooms | 2.0 | 1.0 | 2.0 | 1.0 | 1.0 | 4.0 | 2.0 | 1.5 | 3.0 | 1.1 |
| Kitchen | 50.0 | 40.0 | 60.0 | ***** | 0.0 | 0.0 | 40.0 | ***** | ***** | ***** |
| Dining room/area | 0.0 | 1.0 | 0.0 | ***** | 1.0 | 0.0 | 1.0 | ***** | ***** | ***** |
| Living/family room | 1.0 | 0.0 | 1.0 | ***** | 0.0 | 1.0 | 0.0 | ***** | ***** | ***** |
| Breakfast room/area | 1.0 | 1.0 | 1.0 | ***** | 0.0 | 0.0 | 0.0 | ***** | ***** | ***** |
| Home office | 1.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 1.0 | ***** | ***** | ***** |
| Laundry | 70.0 | 70.0 | 100.0 | ***** | 0.0 | 0.0 | 35.0 | ***** | ***** | ***** |
| Walk-in closets | 1.0 | 0.0 | 1.0 | ***** | 0.0 | 0.0 | 0.0 | ***** | ***** | ***** |
| Pantry | 0.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | ***** | ***** | ***** |
| Auxiliary utility areas | 0.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | ***** | ***** | ***** |
| Private garage | 0.0 | 3.0 | 3.0 | ***** | 3.0 | 3.0 | 3.0 | 3.0 | ***** | ***** |
| Shared garage | 0.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | 0.0 | ***** | ***** |
| Reserved uncovered parking space | 0.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | 0.0 | ***** | ***** |
| Street parking next to home | 1.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | 0.0 | ***** | ***** |
| Street parking close to home | 0.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | 0.0 | ***** | ***** |
| Paid public parking | 0.0 | 0.0 | 0.0 | ***** | 0.0 | 0.0 | 0.0 | 0.0 | ***** | ***** |
| Type of home | 4.0 | 4.0 | 1.0 | 4.0 | 4.0 | 4.0 | 4.0 | 1.0 | 1.0 | 4.0 |
| Number of floors | 1.0 | 1.0 | 2.0 | ***** | ***** | 1.0 | 2.0 | ***** | ***** | ***** |
| External wall material | 1.0 | 1.0 | ***** | ***** | ***** | 2.0 | 1.0 | ***** | ***** | ***** |
| Type of floor | 5.0 | 3.0 | 2.0 | ***** | 1.0 | 5.0 | 1.0 | ***** | ***** | ***** |
| Type of roof | 1.0 | 1.0 | ***** | ***** | 0.0 | 0.0 | 0.0 | 1.0 | ***** | ***** |
| Home energy supply | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |
| Source of energy for heating | 3.0 | ***** | ***** | ***** | ***** | 3.0 | 3.0 | ***** | ***** | ***** |
| Type of heating system | 2.0 | ***** | ***** | ***** | ***** | 4.0 | 4.0 | ***** | ***** | ***** |
| Type of cooling system | 0.0 | 0.0 | ***** | ***** | 0.0 | 0.0 | 0.0 | 0.0 | ***** | ***** |
| Home age | 97.0 | 109.0 | 12.0 | 77.0 | 90.0 | 84.0 | 65.0 | 108.0 | 12.0 | 81.0 |
| Last modification/improvement | 33.0 | ***** | ***** | ***** | ***** | 7.0 | ***** | 5.0 | ***** | ***** |

**Fig. 8.** Attributes of ten competitive homes showing typical cases of missing data

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1213 | Parking space | 33.60 | 75.50 | 75.50 | ***** | 75.50 | 75.50 | 75.50 | ***** | 75.50 | ***** |
| 12132 | Public parking (first-come, first-served) | 75.00 | 0.00 | 0.00 | ***** | 0.00 | 0.00 | 0.00 | ***** | 0.00 | ***** |
| 122 | Home features | 43.37 | 30.30 | 62.74 | 63.18 | 32.64 | 53.16 | 48.79 | 61.72 | 30.48 | 62.15 |
| 1225 | Home age and maintenance | 22.92 | 9.17 | 90.00 | 35.83 | 25.00 | 49.26 | 45.83 | 32.50 | 42.49 | 90.00 |
| 1221 | Home organization/layout | 94.28 | 94.28 | 58.87 | 100.00 | 100.00 | 94.28 | 100.00 | 100.00 | 50.00 | 50.00 |
| 12131 | Reserved parking | 0.00 | 94.48 | 94.48 | ***** | 94.48 | 94.48 | 94.48 | ***** | 94.48 | ***** |
| 121311 | Garage | 0.00 | 100.00 | 100.00 | ***** | 100.00 | 100.00 | 100.00 | ***** | 100.00 | ***** |
| 1212 | Type of rooms/area | 0.00 | 0.00 | 0.00 | 58.35 | 0.00 | 0.00 | 0.00 | 61.02 | 87.53 | 100.00 |
| 1211 | Area belonging to home | 91.86 | 71.75 | 100.00 | 93.76 | 81.07 | 83.91 | 88.77 | 100.00 | 74.01 | 100.00 |
| 121321 | Free public parking | 75.00 | 0.00 | 0.00 | ***** | 0.00 | 0.00 | 0.00 | ***** | 0.00 | ***** |
| 1224 | Home temperature regulation | 50.37 | 0.00 | ***** | ***** | 0.00 | 64.32 | 64.32 | ***** | 0.00 | ***** |
| 1222 | Home construction features | 77.23 | 68.46 | 50.00 | ***** | 44.22 | 81.13 | 46.46 | ***** | 70.00 | ***** |
| 12121 | Primary rooms | 0.00 | 0.00 | 0.00 | 58.35 | 0.00 | 0.00 | 0.00 | 61.02 | 87.53 | 100.00 |
| 12122 | Additional space and storage | 58.34 | 2.64 | 29.83 | ***** | 0.00 | 0.00 | 4.27 | ***** | ***** | ***** |
| 121222 | Storage and auxiliary areas | 29.83 | 0.00 | 29.83 | ***** | 0.00 | 0.00 | 0.00 | ***** | ***** | ***** |
| 121221 | Additional space | 93.64 | 26.09 | 29.83 | ***** | 0.00 | 0.00 | 42.16 | ***** | ***** | ***** |
| 11 | QUALITY OF LOCATION | 65.60 | 60.69 | 72.02 | 71.94 | 46.80 | 70.65 | 53.58 | 69.67 | 86.08 | 72.84 |
| 111 | Suitability of neighborhood | 56.89 | 50.65 | 60.06 | 61.61 | 34.48 | 60.17 | 42.84 | 59.36 | 82.96 | 61.68 |
| 12252 | Last modification/improvement | 29.75 | ***** | ***** | ***** | ***** | 79.00 | ***** | ***** | 85.00 | ***** |
| 12251 | Home age | 19.17 | 9.17 | 90.00 | 35.83 | 25.00 | 30.00 | 45.83 | 32.50 | 10.00 | 90.00 |
| 12243 | Type of cooling system | 0.00 | 0.00 | ***** | ***** | 0.00 | 0.00 | 0.00 | ***** | 0.00 | ***** |
| 12242 | Type of heating system | 60.00 | ***** | ***** | ***** | ***** | 100.00 | 100.00 | ***** | ***** | ***** |
| 12241 | Source of energy for heating | 90.00 | ***** | ***** | ***** | ***** | 90.00 | 90.00 | ***** | ***** | ***** |

**Fig. 9.** A fragment of evaluation results showing the missing attribute and subsystem data

A typical problem of missing data is visible in Fig. 8 (obtained using the "show all competitive systems" option in Fig. 7) and in Fig 9 (obtained using the "show all evaluation results" option in Fig. 7). Out of ten competitive homes only two homes have the complete attribute data. All other data are incomplete. Furthermore, the missing attribute data propagate through the aggregation tree and some subsystems (e.g. garage and reserved parking) have missing values shown in Fig. 9.

In order to deal with missing data evaluators must decide about the most suitable value of the missingness penalty parameter. The effects of missingness penalty are shown in Fig. 10. In all cases increasing the missingness penalty causes a decrease of the overall suitability. For the maximum penalty the overall suitability for missing nonmandatory attributes is positive, and for missing mandatory attributes it is zero.

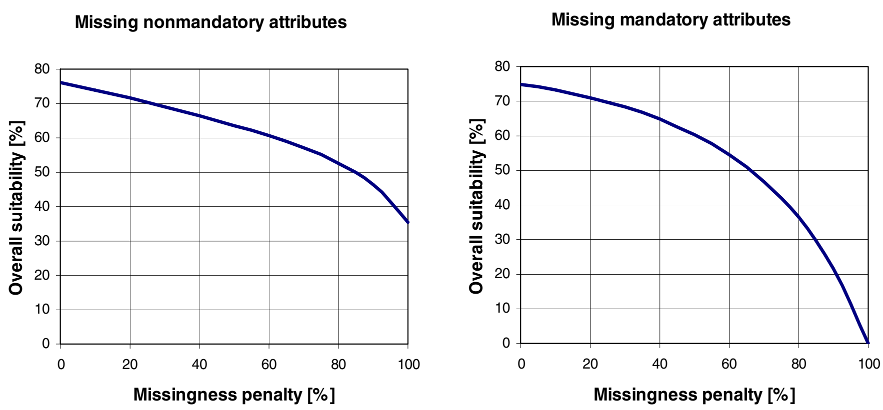


**Fig. 10.** Overall suitability as a function of missingness penalty

The selection of missingness penalty is based on the decision maker missingness tolerance level. Indeed, the missing data can be intentionally hidden because they are inconvenient, or they can be unknown to all data providers. In the case of suspected inconvenient data it is justifiable to apply the highest penalty. In the case where we have reasons to believe that the unknown attributes are satisfied (e.g. the house with missing parking data is in a residential district that is known to have free public parking space) we may select a lower penalty value. To decide about the most appropriate missingness penalty it is suitable to first plot and analyze the overall suitability curves similar to those shown in Fig. 10. The suitability functions in Fig. 10 are strictly concave and the penalty of 80% should be applied if we want to get the overall suitability that is approximately halfway between the extreme values.

The evaluation results (Fig. 7) offer the possibility for detailed investigation of the suitability of home location and its neighborhood. The evaluation tools [6], [15] (also available at www.seas.com) provide suitability maps, which are geographic maps with an overlay showing the distribution of suitability degrees. Fig. 11 shows the suitability map for walkability (possibility to access selected points of interest by walking) where the suitability degrees are presented as numeric values on top of a Google map with selected points of interest. We define walkability as a conjunctive partial absorption of a set of mandatory points of interest and a set of optional points of interest. For the best home proposed in Fig. 7, the walkability is 53%. The potential homebuyer can also investigate the suitability of selected neighborhood for business, children, entertainment, shopping, etc., or make his/her own suitability map.

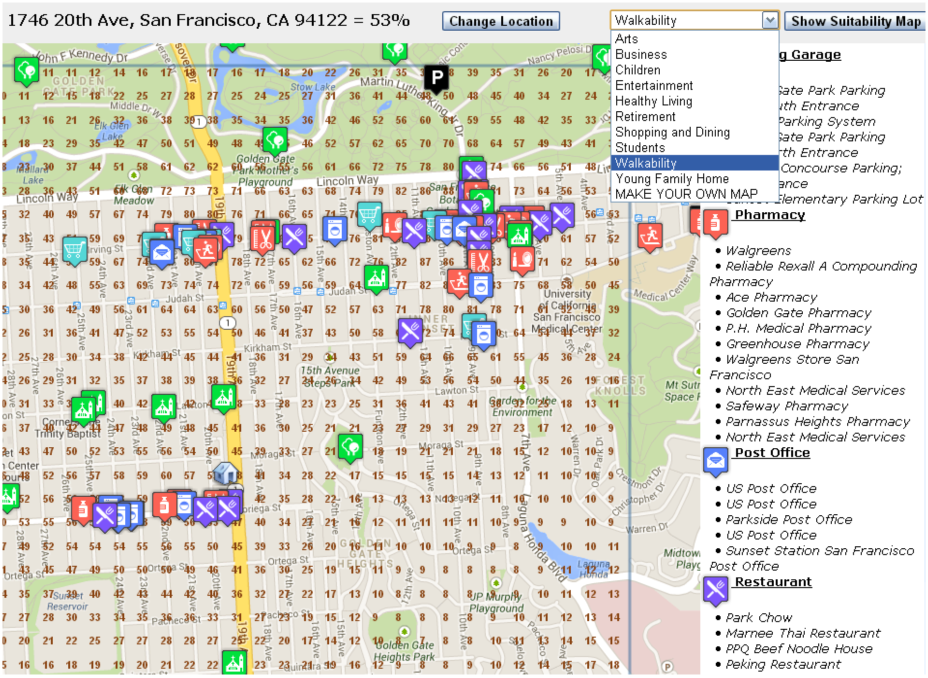


**Fig. 11.** Suitability map on top of a Google map in the vicinity of the selected home

Suitability maps based on points of interest provided by Google give useful information about the suitability of neighborhood, but they do not include physical, environmental, and safety aspects of the neighborhood. Such information can be collected from other sources (e.g. government) and used for additional analysis of the suitability of location. Fig. 12 shows a sample of such an analysis (based on the analyzer of the quality of urban locations developed in [15] and activated as an option in Fig. 7). The quality of urban locations is analyzed using suitability maps based on 11 diverse attributes presented in Fig. 12. In the given point of the best home reported by LSPhome the quality of location is 64%. This value can be compared with the presented distribution of the location quality in the whole city (the best values around 70% and the mean value of 55.41%). Thus, the neighborhood of the selected home is notably above the city average and not too far from the best locations in the city.

| Suitability of the home location | | |
|---|---|---|
| Total quality of location | : | 64 % |

| Ranking of Components | | |
|---|---|---|
| 1 | Elevation | 100 % |
| 2 | Safety | 94 % |
| 3 | Accessibility - public transportation | 83 % |
| 4 | Shopping and dining | 80 % |
| 5 | Health support | 64 % |
| 6 | Requirements of seniors | 63 % |
| 7 | Pollution caused by traffic | 59 % |
| 8 | Requirements of children | 56 % |
| 9 | Walkability | 53 % |
| 10 | Proximity to park | 34 % |
| 11 | Availability of urban forest | 13 % |

| Distribution of Suitability for Quality of Location | |
|---|---|
| Mean value | 55.41 % |
| Sigma | 12.39 |
| Coefficient of variation | 22.36 % |

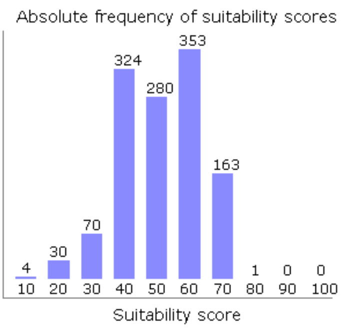


**Fig. 12.** Quality of urban locations based on 11 diverse attributes

## 5 Conclusions

Evaluation and selection of homes is essentially a soft computing logic problem. ORE web sites offer data that enable the use of customized compensative logic criteria for fast ranking of available homes. This paper shows a way such criteria can be designed using the LSP method, and implemented in a software tool available over the Internet. Specific problems related to online buying and selling of homes include the missingness-tolerant aggregation and the use of the verbalized concept of overall importance to derive the andness/orness and weights of partial conjunction and partial disjunction aggregators. Soft computing decision methods are a way to significantly improve both the efficiency and the customer experience in online real estate.

## References

1. Beliakov, G., Pradera, A., Calvo, T. (eds.): Aggregation Functions: A Guide for Practitioners. STUDFUZZ, vol. 221. Springer, Heidelberg (2007)
2. Bullen, P.S.: Handbook of means and their inequalities. Kluwer (2003)
3. Dujmović, J.J.: Preference logic for system evaluation. IEEE Transactions on Fuzzy Systems 15(6), 1082–1099 (2007)
4. Dujmović, J.J., Larsen, H.L.: Generalized conjunction/disjunction. International J. Approx. Reas. 46, 423–446 (2007)
5. Dujmović, J.J., Nagashima, H.: LSP method and its use for evaluation of Java IDE's. International J. Approx. Reas. 41(1), 3–22 (2006)
6. Dujmović, J.J., De Tré, G.: Multicriteria Methods and Logic Aggregation in Suitability Maps. International Journal of Intelligent Systems 26(10), 971–1001 (2011)
7. Dujmović, J.J.: Andness and Orness as a Mean of Overall Importance. In: Proceedings of the IEEE World Congress on Computational Intelligence,-Brisbane, Brisbane, Australia, June 10-15, pp. 83–88 (2012)
8. Dujmović, J.: The Problem of Missing Data in LSP Aggregation. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R., et al. (eds.) IPMU 2012, Part III. CCIS, vol. 299, pp. 336–346. Springer, Heidelberg (2012)
9. Gini, C., Barbensi, G., Galvani, L., Gatti, S., Pizzetti, E.: Le Medie. Unione Tipografico-Editrice Torinese, Torino (1958)
10. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation Functions. Cambridge Univ. Press (2009)
11. National Association of Realtors, NAR Fact Sheet (2013), `http://www.realtor.org/for-the-media/nar-fact-sheet`
12. Torra, V., Narukawa, Y.: Modeling Decisions. Springer (2007)
13. Trulia (2013), `http://www.trulia.com/`
14. Wikipedia, Multiple listing service (2013), `http://en.wikipedia.org/wiki/Multiple_Listing_Service`
15. Yokoohji, R.: LSP Suitability Maps for the Quality of Urban Locations. MS thesis. San Francisco State University, Department of Computer Science, Culminating Experience Report SFSU-CS-CE-12.09 (2012)
16. Zillow (2013), `http://www.zillow.com/`