

A System For Generating A Value Index For Properties And A Method Thereof**Cross-Reference to Related Applications**

[0001] The present application claims the benefit of Singapore Patent Application No. 10202102385W filed March 9, 2021 which is incorporated by reference herein.

Technical Field

[0002] The present invention relates to a system for generating a value index for properties and a method thereof. Preferably, the present invention relates to a system for generating an estimated value of a property based on a value index.

Background

[0003] Real estate price indexes are often a high-level summation of a large geographic area's housing price transactions. These indexes are often published on a monthly or quarterly basis and thus there is a delay between the index numbers and the actual transactions happening daily and thus greatly reduces the benefits of the information.

[0004] Additionally, as the indexes cover large predefined geographic areas, these indexes "average" out heterogeneous properties and does not provide sufficient insight into individual properties, sub-sets of the market composed of similar properties of a certain kind, projects, geographical locations, etc.

[0005] Commonly, the primary methods used to calculate real estate price index are: (A) the repeat sales approach or (B) the hedonic approach. While each approach has advantages and disadvantages, both seek to provide a housing price index that closely follows the changes in housing prices.

[0006] At an individual property level, prices can vary greatly based on the preferences of the individual buyer and seller as well as the characteristics of the property. While representative of the unique nature of real estate transactions, these variations can lead to extended

negotiation periods and an increase in failed transactions. Thus, there exist a need for a method to remove some of the subjectivity and biases from the property valuation process.

[0007] When a person wishes to purchase a property, it would be useful to provide the person with current value of the property based on factors relevant to the property. However, conventional real estate price indexes are not able to provide such a value.

Summary

[0008] According to various embodiments, a system for generating a value index for properties is provided. The system includes a processor, and a memory in communication to the processor for storing instructions executable by the processor, such that the processor is configured to categorise the properties into one or more categories, determine an optimal number of pre-clusters of properties based on the one or more categories, cluster the properties into a plurality of pre-clusters based on the optimal number of clusters, determine a set of search radii, generate a cluster of properties based on one of the plurality of pre-clusters and the set of search radii, and generate a value index for the cluster of properties.

[0009] According to various embodiments, each of the plurality of pre-clusters may include of properties in geographically separated areas.

[0010] According to various embodiments, the process may be further configured to receive a filtering criteria for a target property, determine one or more categories of the target property, identify a pre-cluster from the plurality of pre-clusters that matches the one or more categories of the target property, generate the cluster of properties based on the pre-cluster and the set of search radii, and generate the value index for the cluster of properties.

[0011] According to various embodiments, each search radius of the set of search radii may be longer than another, such that each of the search radius may be determined based on information of the territory where the properties are located in.

[0012] According to various embodiments, the processor may be configured to expand the search radius and generate another cluster of properties when the number of data points in the

cluster based on the search radius is lower than a threshold number of data points to generate the value index.

[0013] According to various embodiments, the processor may be configured generate estimated values of the properties based on the value index.

[0014] According to various embodiments, the processor may be configured to de-cluster a transaction value of a cluster of property units into estimated values of each of the property units.

[0015] According to various embodiments, a method of generating a value index for properties is provided. The method includes categorising the properties into one or more categories, determining an optimal number of pre-clusters of the properties based on the one or more categories, clustering the properties into a plurality of pre-clusters based on the optimal number of clusters, determining a set of search radii, generating a cluster of properties based one of the plurality of pre-clusters and the set of search radii, and generating a value index for the cluster of properties.

[0016] According to various embodiments, each of the plurality of pre-clusters may include of properties in geographically separated areas.

[0017] According to various embodiments, the method may further include receiving a filtering criteria for a target property, determining one or more categories of the target property, identifying a pre-cluster from the plurality of pre-clusters that matches the one or more categories of the target property, generating the cluster of properties based on the pre-cluster and the set of search radii, and generating the value index for the cluster of properties.

[0018] According to various embodiments, each search radius of the set of search radii may be longer than another, such that the method may further include determining each of the search radius based on information of the territory where the properties are located in.

[0019] According to various embodiments, the method may further include expanding the search radius and generate another cluster of properties when the number of data points in the

cluster based on the search radius may be lower than a threshold number of data points to generate the value index.

[0020] According to various embodiments, the method may further include generating estimated values of the properties based on the value index.

[0021] According to various embodiments, the method may further include de-clustering a transaction value of a cluster of property units into estimated values of each of the property units.

[0022] According to various embodiments, a non-transitory computer readable storage medium comprising instructions is provided, such that the instructions, when executed by a processor in a system, cause the system to categorise the properties into one or more categories, determine an optimal number of pre-clusters of properties based on the one or more categories, cluster the properties into a plurality of pre-clusters based on the optimal number of clusters, determine a set of search radii, generate a cluster of properties based on one of the plurality of pre-clusters and the set of search radii, and generate a value index for the cluster of properties.

Brief Description of Drawings

[0023] Fig. 1 shows an exemplary embodiment of a system for generating a value index for properties.

[0024] Fig. 2 shows a flow diagram of an exemplary method of generating a value index for properties.

[0025] Fig. 3 shows a flow diagram of an exemplary method of generating a value index and a method of generating an estimated value for a property.

[0026] Fig. 4 shows a graph displaying the relationship between a silhouette score and the number of clusters.

[0027] Fig. 5 shows an example of the plurality of pre-clusters on a map of a territory after pre-clustering.

[0028] Fig. 6 shows an extract of the map in Fig. 5 and an example of the cluster.

[0029] Fig. 7 shows a flow diagram of an exemplary method of generating the value index.

[0030] Fig. 8 shows a flow diagram of an exemplary method of generating a value index based on a target property.

[0031] Fig. 9 shows an exemplary embodiment of the property data resources accessed by the system when generating the value index.

[0032] Fig. 10 shows a flow diagram of a method of collecting and filling property data of the properties in the property data resources.

[0033] Fig. 11 shows an exemplary embodiment of a property value matrix.

[0034] Fig. 12 shows a flow diagram of a method of generating an estimated value of the properties.

[0035] Fig. 13 shows a flow diagram of an exemplary method of generating estimated values for a cluster of property units within the cluster.

Detailed Description

[0036] Fig. 1 shows an exemplary embodiment of a system 100 for generating a value index for properties. Value Index may include a price index of the property. System 100 may include a processor 110, a memory 120 in communication with the processor 110 for storing instructions executable by the processor 110. Processor 110 is configured to categorise the

properties into one or more categories, determine an optimal number of pre-clusters of properties based on the one or more categories, cluster the properties into a plurality of pre-clusters based on the optimal number of clusters, determine a set of search radii, generate a cluster of properties based on one of the plurality of pre-clusters and the set of search radii, generate a value index for the cluster of properties. System 100 may further include an I/O interface 130, a display 140, a power source 150, a communication module 160. System 100 may be in communication with user devices, e.g. computer devices, via a network. System 100 may be in communication with property data resources 180, which are databases for storing property data of the properties.

[0037] Fig. 2 shows a flow diagram of an exemplary method 200 of generating a value index for properties. Method 200 includes categorising the properties into one or more categories in block 210, determining an optimal number of pre-clusters of the properties based on the one or more categories in block 220, clustering the properties into a plurality of pre-clusters based on the optimal number of clusters in block 230, determining a set of search radii in block 240, generating a cluster of properties based one of the plurality of pre-clusters and the set of search radii in block 250 and generating a value index for the cluster of properties in block 260. By generating a plurality of pre-clusters, it is possible for method 200 to provide an accurate and efficient way of generating a value index for properties. In addition, the method 200 provides an accurate and current value index for properties.

[0038] To generate a value index for the properties, the system 100 may retrieve the transaction data of past transactions from the property data resources. In order to generate a quality value index, i.e. an index that accurately reflects the value of the property, the system 100 would require a high density of data points of the transaction values of the properties. However, as transaction values of properties are “scattered” all over an area or territory and sporadic, it is advantageous to use transaction values from similar properties to fill a data gap, i.e. in parts of the areas without transaction values, so as to provide a “complete” picture” of the transaction values of the properties as much as possible. In this way, the system would be able to generate a higher quality value index with more data points, which can be used to generate a useful index value of a property. Based on the information of the properties in the property data resource, the system 100 is configured to generate an estimated value for each

of the properties in the property data resource. Based on the characteristics of the properties, the system 100 is configured to cluster the properties into a plurality of clusters and generate a value index for each of the clusters from the transaction values and estimated values of the properties in the each of the clusters. Plurality of clusters may be formed based on types of properties, e.g. condominium block, detached housing estate, etc. Before the system 100 generates the clusters, the system 100 may be configured to pre-cluster the properties, which will be explained in detail later. Pre-clustering enables the system 100 to generate value index more accurately and efficiently.

[0039] Fig. 3 shows a flow diagram of an exemplary method 300 of generating a value index. Fig. 3 also shows the method 300 may include a method 300 of generating an estimated value of a property. Once the property data and transaction values of the properties in the databases are populated, the system 100 is able to generate a more accurate value index from the data. Method 300 may include pre-clustering the property data at block 310. System 100 may include a pre-cluster module configured to pre-cluster the property data. To pre-cluster the property data, the pre-cluster module is configured to determine an optimal number of pre-clusters. System 100 may include a pre-clustering module configured to pre-cluster the properties into a plurality of pre-clusters based on the optimal number of pre-clusters. Method 300 may be generating a set of search radius during the pre-clustering step in block 310. Method 300 may include clustering the properties into a plurality of clusters in block 320 based on the optimal number of pre-clusters and a search radius of the set of search radii. System 100 may include a clustering module 322 configured to cluster the properties into the plurality of clusters. While Fig. 3 shows a flow diagram, for easy reference, the relevant modules for the step may be shown at the step. Method 300 may include generating the value index for each of the plurality of clusters in block 330. System 100 may include an index module 332 configured to generate the value index. Method 300 may include generating estimated values of the properties at block 340. System 100 may include a valuation module 342 configured to generate estimated values of the properties based on the value index. In this way, it is possible to generate an accurate estimated value at an individual property level. Method 300 may include de-clustering the value generated for a cluster into individual values of the properties in the cluster in block 350. System 100 may include a de-clustering module configured to de-cluster a transaction value of a cluster of property units into estimated values of each of the property units.

[0040] Fig. 4 shows a graph 400 displaying the relationship between a silhouette score and the number of clusters. To pre-cluster the property data, the system 100 may be configured to determine the optimal number of pre-clusters to pre-cluster the properties. Pre-cluster module may be configured to run a pre-clustering search to determine the optimal number of pre-clusters. In order to obtain a quality index, the number of data points to form the value index is important. Hence, pre-clustering the properties enables the system 100 to identify an optimal number of pre-clusters with sufficient data points to form a quality index. Pre-cluster module may use the K-Silhouette analysis to generate a Silhouette score for each of the pre-clusters to determine the number of pre-clusters with the highest Silhouette score. Referring to Fig. 8, for example, the number of pre-clusters with the highest Silhouette score (0.275) is 48. Hence, the number of pre-clusters is identified as the optimal number of pre-clusters.

[0041] Fig. 5 shows an example of the plurality of pre-clusters on a map of a territory after pre-clustering. As shown in Fig. 5, the properties in the territory are pre-clustered by the pre-clustering step. Pre-cluster module may be configured to categorize the properties into categories. Categories may include property type, e.g. condominium, single detached houses, etc., condition of properties, etc. System 100 may categorize the properties as much as possible depending on the data available. For example, system 100 may categorize the properties according to size of property, number of rooms, etc. Pre-clustering is useful to ensure that there are sufficient property data points in each category. Each of the plurality of pre-clusters may include properties in the same geographical area or properties in geographically separated areas. For example, the number of single detached houses are rarely transacted and hence the transaction values of such a type of property are not easily available. After pre-clustering, the properties of the same category, which may be located at separated geographical areas of the territory, e.g. Cluster A, are pre-clustered into the same pre-cluster to provide sufficient data points for the value index. Pre-cluster module may be configured to store the optimal number of clusters and cluster data into a cluster database.

[0042] Pre-cluster module may also be configured to determine a set of search radii for generating a value index. Each search radius of the set of search radii is longer than another. Each of the search radius is determined based on information of the territory where the

properties are located in. Set of search radii is used for radius index expansion during a search. Radius index expansion is an iterative process where the search radius for properties for an index is automatically increased until a maximum radius is reached, or a suitable number of data points is obtained within the search radius. System 100 may be configured to determine if the number of data points is sufficient or beyond a threshold for each search radius from the shortest search radius to the longest search radius. When generating the index, the system 100 may be configured to expand the current search radius to the next search radius when the number of data points, i.e. transaction values, in the cluster based on the current search radius is not sufficient or lower than a threshold number of datapoints to generate the value index. When the system 100 is generating the index for a cluster for every time period, e.g. month, quarter, the system 100 identifies the data points within the corresponding pre-cluster and within a search radius, e.g. shortest search radius, and/or within a time period, e.g. month. However, the number of data points, i.e. transaction values, within the search radius may not be sufficient to generate the index. In this situation, the system 100 is configured to expand the search radius to the next search radius in the same pre-cluster, until there is a sufficient number of data points in the cluster for generating the index. The set of search radii is determined based on the information of the territory during the pre-clustering step. For example, if the system 100 identifies the territory to be a densely populated area, e.g. a town or city area, the radius increments for the set of search radii may be set at standard increments of distance taken by a person walking, e.g. 5 mins walking distance interval from 5 mins up to an hour. However, if the system 100 identifies the territory to be in a less densely populated area, e.g. rural area, the radius increment for the set of radii may be set at standard increments of distance taken by a person driving, e.g. 5 mins driving distance interval from 5 mins to 2 hours. Set of radii may represent the time a person is willing to travel to a location of interest, e.g. location of the property, when looking for a property to buy. If a person is looking at a property near a work location and is willing to travel 1 hour by driving from the property location, the system 100 would be able to generate an index based on data points of properties in the plurality of clusters and within the search radius of 1 hour drive from the property location, i.e. the location of interest.

[0043] Fig. 6 shows an extract of the map in Fig. 5 and an example of the cluster. To generate the value index, the system 100 may cluster the data into the plurality of clusters. Based on the category of a target property, e.g. condominium, the cluster module generates a

plurality of clusters based on the plurality of pre-clusters. For example, referring to Fig. 5, if a value index for a cluster 602C comprising a condominium block is to be generated, the system 100 may identify the category, i.e. condominium, and retrieve the pre-cluster 602P of the same category and cluster the properties in the pre-cluster 602P relevant to the cluster 602C into the cluster 602C. Pre-cluster module may cluster the properties within a search radius of the set of search radii from the location of the target property. If the number of data points in the cluster is not sufficient, the cluster module may expand the search by searching for properties in the same pre-cluster in the next expanded search radius. Cluster module may cluster the property data using K-prototype clustering if dataset contains categorical attributes, or K-means clustering if dataset contains only numerical attributes.

[0044] Fig. 7 shows a flow diagram of an exemplary method 700 of generating the value index. Method 700 may include generating the value index for each of the plurality of clusters at block 710. System 100 may include an index generating module configured to generate and analyse an index based on the transaction values and estimated values of the properties. Method 700 may further include validating the value index in block 720 after the index is generated. System 100 may include a validating module configured to validate the index. Method 700 may include smoothening the value index in block 730. Value index may be smoothened if the index is validated. System 100 may include an index smoothening module configured to smoothen the index. Method 700 may include expanding the search radius in block 740 to form a new cluster to re-generate the index.

[0045] Once the plurality of clusters are formed, the index generating module generates a value index for each of the plurality of clusters. Index generating module may retrieve data, e.g. transaction date, transaction amount, transaction type, etc. from a transaction database in the property data resources and data containing features and attributes related to the project of a property, e.g. geographic area, the building structure, and amenities, etc. Based on the data retrieved, the index generating module may be configured to generate a value index for each of the plurality of clusters. Value index may be based on a monthly or quarterly timeline from the date the cluster has a transaction until a targeted end date, e.g. current date, month or quarter.

[0046] Index generating module may be configured to retrieve the transaction values and estimated values from a cluster within a search radius, e.g. the smallest search radius, and generate the value index from the retrieved values of the properties within the search radius. Validating module may be configured to validate if the index can be generated or if there are sufficient data points or above a threshold to generate the value index. If the validating module determines that the index is not able to be generated, the index generating module is configured to activate the cluster module to expand the search radius to the next expanded search radius to form a new cluster based on the expanded search radius. Once the new cluster is formed, the index generating module generates index. Validating module may again validate the re-generated index again. Validating module may be configured to validate the index by determining if the index is generated for every pre-determined period of time, e.g. monthly, quarterly and/or if the value in the index is within a specified range, e.g. 0.5-2.0 (log scale -0.7-0.7). System 100 may repeat the process iteratively until the validating module validates that the index can be generated. Index generating module may be configured to produce a monthly index. If the monthly index is not valid, a quarterly index may be generated. However, if a valid quarterly index could not be generated, the index generating module may be configured not to generate an index for the cluster. Index generating module may be configured to store the index in an index database.

[0047] Index generating module may be configured to use the hedonic model and repeated sales model simultaneously to generate the index. Index generating module may be configured to determine the index to be used by determining the model that produces valid index and has smaller mean square error, i.e. fits the transaction subset better. Hedonic model uses attributes like weighted least square model, dependent variable - transacted, weight – 1/distance to current index in meters, independent variables – various transaction attributes AND month index (as categorical attributes). Repeated sales model uses attributes like Case-Shiller index, definition of repeated sales, e.g. when two transactions have same addresses, and the area size/room type are the same. Smoothing module may apply cubic spline smoothing factor may be applied to smoothen the index values. System 100 may be configured to automate the generation of the value index. System 100 may be configured to store an automated index model to automate the generation of the value index.

[0048] Fig. 8 shows a flow diagram of an exemplary method 800 of generating a value index based on a target property. A user may enter a desired filtering criteria, i.e. a search query, into the system 100 via a graphical user interface on the user device 20. System 100 may be configured to display a value index based on the filtering criteria on the display 140. System 100 may also display an estimated value or index value of the target property on the display 140. The user would be able to receive an estimated value of the target property upon entering the desired filtering criteria. Method 800 may include receiving the filtering criteria and configured to define and determine the filtering criteria in block 810. When the system 100 receives the filter criteria for a target property, the system 100 determines one or more categories of the target property and identifies the category of the pre-cluster that is relevant to the one or more categories. System 100 may be configured to identify a pre-cluster from the plurality of pre-clusters that matches the one or more categories of the target property. System 100 may be configured to generate the cluster of properties based on the pre-cluster and the set of search radii. System 100 may be configured to gather data on transaction values and estimated values of the properties in the cluster corresponding to the target property based on the defined criteria in block 320. Thereafter, the system 100 may apply the automated index model to the data in block 330 to generate the value index for the cluster of properties relevant to the target property. System 100 may generate value indexes for a plurality of clusters beforehand and retrieve the value index relevant to the target property if the value index has been generated. Otherwise, the system 100 may generate the value index for the target property. When the value index is generated, it may be displayed on the display via a graph, a map, a chart, a list, etc.

[0049] Fig. 9 shows an exemplary embodiment of the property data resources 180 accessed by the system 100 when generating the value index. Property data resources may include at least the databases as shown in Fig. 9. Data in the databases may go through a series of cleaning and normalization process before being used. Listing database 180A may include records of properties for sale and rent. Transaction database 180B may include records of transacted sales or rentals of the properties, e.g. transaction date, transaction amount, transaction type, etc. Property database 180C may include property details/information about the properties, e.g. size of rooms, no. of rooms, etc. Address database 180D may include address, points of interest, nearby amenities, e.g. schools, malls, etc, transportation information, distance between points, geo-location of property, e.g. latitude and longitude.

Building database 180E may include information about a building, e.g. height, no. of floors, no. of elevators, construction material, etc. Policy database 180F may include information about government policies, land, etc. Economic database 180G may include mortgage rates, population growth, income growth, etc. Jurisdiction database 180H may include breakdown of geography by city area, state/province, unique housing policies of specific countries that need to be incorporated into the algorithm. Project database 180J may include information about specific developments, new developments being built, geographical area, building structure, amenities, etc. Land database 180K may include information about the legal land lot, zoning information, etc.

[0050] Fig. 10 shows a flow diagram of a method 1000 of collecting and filling property data of the properties in the property data resources. In order to generate a value index that is as accurate and relevant as possible, it is necessary to generate the value index based on sufficient number of data points, i.e. transaction values and estimated values of the properties. As it is not possible to obtain transaction values of all the properties, it is useful to estimate the value of the property based on transacted prices of other corresponding properties that are similar or substantially similar to the property.

[0051] Method 1000 may include collecting data by the system 100 from data sources at block 1010. Data sources may include maps, websites, property databases, etc. System 100 may include a data extraction module 1012 configured to extract data from data sources. Data extraction module 1012 may include a map analysis module configured to analyse a map to extract information of the properties. Data collection may include extracting data from the data sources. Method 1000 may include normalizing the property data at block 1020. System 100 may normalize and clean the data to ensure that the data collected from various sources conforms to the data requirement and needs of the system 100. Method 1000 may include checking the consistency of the data in block 1030. At this step, the system 100 may be configured to run an inference engine to deduce property data of a property. Method 1000 may include filling in missing data in the property data resources, e.g. address database, building database, land database, etc. in block 1040. System 100 may be configured to collect data periodically, e.g. daily, weekly, monthly. Dynamic property data, e.g. transaction prices, that are available on a frequent basis, may need to be collected and updated into the relevant databases so that the index generated is current.

[0052] When extracting property data from data sources in block 1012, the data extraction module 1012 may be configured to extract the information from the data sources that may not be text-based. The non-text-based information needs to be extracted before it can be used by the system 100. For example, the data extraction module may include a floor plan processing module configured to extract data from documents using text recognition and computer vision processing. Floor plan processing module may use Optical Character Recognition (OCR) to extract the words in the floor plan and clustering methods, e.g. DBSCAN, to group the information into blocks and lines of text. Floor plan processing module may use Lexical analysis and parsing (Yacc + Lex) to extract features, e.g. area size, number of rooms, unit type, stack, etc. from the text data. For the graphical information included in the floor plan, the data extraction module 1012 may use an object recognition module to extract the different parts of the floor plan, e.g. unit floor plan, development project plan, tower plan, and the position of the different parts may be used to improve the text extraction of the floor plan processing module.

[0053] Data extraction module 1012 may include a contour detection and analysis module to identify the wall positions of the walls in the floor plan. Based on the wall positions, a spatial information-based image segmentation may be used to identify the individual rooms in the floor plan and estimate the size of the rooms. Based on the identified rooms, the room position may be used to extract room features from the text information. For example, the text of a room named “Master Bedroom” may not be clear and recognizable. Based on the position and the size of the room, the system 100 is able to identify the room to be a “Master Bedroom” and fill in the data accordingly.

[0054] Property data is primarily based on address information. Accurate address data provides accuracy in the property database as without an accurate address search, there could be duplicate or bad records in the database. System 100 may be configured to process address information from the data sources to improve the accuracy of the address data. Using similarity between entities based on a mixture of geographical, named entities and other heterogeneous features, the system 100 may be configured to provide an improved address and geo-location searching method.

[0055] System 100 may include an address processing module configured to process the address of each of the properties from various data sources. Address processing module may be configured to handle misspelled words, different address formats, missing address parts and positional-dependent address part meanings. Address processing module may use word encoding for pre-processing. Address processing module may use Levenshtein distance and bag of ngrams encoding methods to identify misspelled parts. Address processing module may use Named Entity Recognition (NER) for address part identification and meaning assignment. Address processing module may use cosine similarity to determine the distance between two addresses and TF-IDF weights and the total entropy of present address components may be used for the confidence score. After the processing, the address part may be handled as another standard numerical features. For example, the address part may be associated with the geo-location, i.e. latitude and longitude co-ordinates, of the property. The address part may be used in combination with other features to find similarity between properties.

[0056] Normalizing the data in block 520 may include matching the address of the properties from various sources, including the address in the address database of the system 100. For example, the system 100 may retrieve an address from a map with a street name “Albert Ave”. System 100 recognises “Ave” to be “Avenue” and normalize the street name of the property to “Albert Avenue”. In another example, the system 100 may receive a condominium name “Blue2 Apt @ LakeHouse”. System 100 may clean and normalise the name to “Blue Two Condo At LakeHouse”. This process of standardizing and normalizing enables the machine learning and statistical models to perform much better.

[0057] As the system 100 fills the missing data of the property, the system 100 may execute consistency checks to ensure that there are no conflicts in the data in block 1030. System 100 may include a consistency checking engine configure to check the data consistency. Consistency checking module may use new data sources and information to augment existing data. For example, if the exact floor size of an apartment unit A is not known but a similar unit in the same building unit B was recently sold, then it is possible that floor size of B can be used to populate the floor size of A. In another example, if the extracted number of room data of a property is 100 rooms, the consistency checking module would be able to identify

that the data is inaccurate as a similar unit in the same building has only 3 rooms. System 100 may include a property analysis module configured to analyse the property data of properties.

[0058] As mentioned, one of the obstacles to generating quality index prices is the lack of complete information across the various databases and data sources. System 100 may include a data filling module configured to fill in the missing data of the properties in block 1040. Data filling module may be configured to use an iterative fill method to fill in the missing data. Data filling module is capable of handling both numerical and categorical values and any combination of missing values for a property. Data filling module may also make use of domain knowledge to fill in the missing data. For example, the data filling module is configured to assume that properties with the same postal code would likely share the same geographical features, like construction year or nearby bus stations, while properties in the same high-rise stack would share floor plan features, like size, number of bedrooms, number of bathrooms, and location views. Data filling module may then fill in the missing data based on the known property data of the properties. Data filling module may apply either KNNImputer (K nearest neighbor imputer) or Iterative Decision Tree Regressor fill based on the problem's characteristics to iteratively fill the missing data of the property.

[0059] Fig. 11 shows an exemplary embodiment of a property value matrix 1100. System 100 may be configured to normalize the property using a grading system 100. System 100 may include a grading system 100 for grading the properties. By providing a grading system 100, the system 100 is able to normalize the property based on the quality and condition of the properties in addition to the collected data. This would enable the system 100 to provide a more realistic and accurate assessment of the properties and therefore able to better normalize the properties. Grading system 100 may include a quality variable and/or a condition variable. Quality variable may be a measure of the standard of design and construction of the property based on parameters such as type of materials used, type of fixtures and fittings, dimensions of rooms and common areas, floor to ceiling heights, etc. Quality variable may include at least the following categories: luxury, high-end, standard, affordable, basic or unfinished. Condition variable may be a measure of how well preserved or deteriorated the property is and may take into consideration the normal wear and tear of the property as well as any damages and the number of years since the property was built or renovated. Condition

variable may include at least the following categories: new, maintained as new, well-maintained, slightly deteriorated, deteriorated. After grading the properties, the system 100 may be able to generate a property value matrix 1100 to use it to generate a relative value of a property based on the condition of the property. To generate the matrix 1100, the system 100 may retrieve the transacted prices of the properties of the respective categories and calculate the relative value of the property in each category with respect to the highest category so as to normalize the categories in the property value matrix 1100. For example, referring to Fig. 11, the “New Luxury Category” may be identified as the highest priced category and the transacted prices may be used to normalize the values of the properties in the other categories. Based on the transacted prices of the “New High-end” category, the relative value of the “New High-end” category may be calculated to be 85% of the “New Luxury” category. The “New Standard” category may be calculated to be 60% so on and so forth. Accordingly, the relative value of the categories may be generated. Property value matrix 1100 may be used as one of the factors to estimate the value of a property.

[0060] Fig. 12 shows a flow diagram of an exemplary method 1200 of generating an estimated value of the properties. Method 1200 may be an iteration method. Valuation module 342 may be configured to generate the estimated value. Valuation module 342 may be configured to retrieve the value index from the index database, the transaction data from the transaction database, the property data from the property database. Based on the data extracted, the valuation module 342 may be configured to generate an estimated value for each of the properties. Valuation module 342 may be configured to store the estimated value in an estimated value database.

[0061] Referring to Fig. 12, for each cluster, the method 1200 may include adjusting past transactions values based on the corresponding index in block 1210. Method 1200 may include generating a floor coefficient, i.e. effect of floor on transacted price, for each building or cluster in block 1220. Method 1200 may include valuating the properties in block 1230.

[0062] System 100 may include a regression module 1222 configured to generate the floor coefficient. Regression module 1222 may be configured to run a linear regression model using independent variable and dependent variables. Independent variable may include transacted price after adjustment in block 1210. Dependent variables may include other

attributes associated with each transaction and floor number (as a numeric attribute), and data including transactions of target project of target property. Regression module may pre-set a minimum floor coefficient value of 0.001.

[0063] Valuation module 342 may be configured to generate an estimated value for each of the properties by applying the following valuation models for all the properties:

- if the properties are in the same building and within a stack, retrieve the latest transaction or transactions that took place in the same building and same stack, adjust the value by the floor coefficient using the following formula,

$$e^{(\text{floor difference} \times \text{floor coefficient})}$$

and take mean values if there are multiple latest transactions.

- if the properties are in the same area and same building, retrieve the latest transition or transactions that took place in the same area and building, adjust the value by the floor coefficient using the following formula,

$$e^{(\text{floor difference} \times \text{floor coefficient})}$$

and take mean values if there are multiple latest transactions.

- if the properties are in the same building and have an area difference within 10%, retrieve the latest transaction or transactions that took place in the same building and have an area difference within 10%, adjust the value by the by the floor coefficient using the following formula,

$$e^{(\text{floor difference} \times \text{floor coefficient})}$$

and take mean values if there are multiple latest transactions.

- if the properties are in same project and same floor, retrieve the latest transaction or transactions that took place in the same project and same floor, and take mean values if there are multiple latest transactions.
- if the properties are in the same project and same area, retrieve the latest transaction or transactions that took place in the same project and have the same area, adjust the value by the by the floor coefficient using the following formula,

$$e^{(\text{floor difference} \times \text{floor coefficient})}$$

and take mean values if there are multiple latest transactions.

[0064] Valuation module 342 may apply machine learning method to generate the estimated values of the properties. Valuation module 342 may run machine learning models simultaneously. Machine learning models may include ridge, gradient boost, random forest, etc. Valuation module 342 may be based on dependent variable, e.g. unit per square feet, and independent variable, e.g. transacted properties' attributes, data including transactions that belong to same cluster and within a specified radius. Valuation module 342 may compare the mean square errors on fitting the training dataset, select the model with smallest mean square error, and use the model to predict the unit square feet price of the current property.

[0065] Referring to Fig. 12, the method 1200 may include validating the estimated value in block 1240. The estimated value may be validated to ensure that there are no negative values. If the estimated value is not valid, the method 1200 may include iterating the valuation module 342 to generate another estimated value.

[0066] Referring to Fig. 12, the method 1200 may include de-clustering cluster value, i.e. a transaction value of a cluster of property units, or the estimated value generated for a cluster of property units, into estimated values of each of the property units in the cluster of property units in block 1250. System 100 may include a de-clustering module configured to de-cluster the cluster value into estimated values of each of the property units in the cluster of property units.

[0067] Fig. 13 shows a flow diagram of an exemplary method 1300 of generating estimated values for a cluster of property units within the cluster. Method 1300 may include de-clustering of the cluster value for the cluster of property units into individual estimated values for each of the property units in the cluster. System 100 may include a de-clustering module configured to de-cluster the cluster value of the cluster of property units. De-clustering is useful to generate an estimated value for the property units when the transaction information does not uniquely identify a particular property or unit. De-clustering module may be configured to de-cluster the cluster value into individual estimated values and assign the generated individual estimated values to the appropriate property units in the cluster of property units.

[0068] Method 1300 may include clustering of property units in block 1310, generating a cluster representative unit in block 1320 and generating individual estimated values for each of the property units of the cluster in block 1330.

[0069] Clustering of property units may include creating a cluster containing the property units to which a transaction can refer to. This step is applicable to transaction types that do not specify an individual unit. For example, if a transaction value specifies a sale in a high-rise building on a specific stack and floor range, then a cluster of property units is created containing all property units in the specified stack and floor range.

[0070] Generating a cluster representative unit may include creating a cluster representative unit for the non-specific transaction. For non-specific features, a median value is computed so the cluster representative unit can be treated as a regular property for the purpose of index generation and value estimation or valuation.

[0071] De-clustering the cluster value into individual estimated values may include applying a regression model to estimate the values of all the property units in the cluster. The regression model may be trained on the cluster representative unit. The regression model utilizes the differentiating features inside the cluster, e.g. floor, area size, build up area, view, property improvement, etc. to generate a distinct estimated value for each unit. Based on the same example above, assuming that the property units within the floor range are of the same size, the de-clustering module is configured to utilize the floor level as the differentiating feature and generate the estimated value of the unit at each floor level from the cluster value.

[0072] Generating individual estimated value may include dimensionality reduction at block 1332 and regression modelling at block 1334. Dimensionality reduction techniques are used to reduce the differentiating features to the features most relevant for the estimation and to improve the learning of these features using techniques like Principal Component Analysis (PCA), Factor Analysis of Mixed Data (FAMD), target encoding, etc. Regression modelling may include either Neural Networks or Random Forest Regression depending on the type of features used during clustering.

[0073] System 100 may include a server, a laptop, a computer, etc. Processor 110 typically controls overall operations of the system 100, such as the operations associated with display, data communications. Processor 110 may include one or more processors to execute instructions in the above-described modules to perform all or part of the steps in the above-described methods. Moreover, the processor 110 may include one or more modules which facilitate the interaction between the processor 110 and other modules. System 100 may be in communication with the user devices via the network.

[0074] Memory 120 may be configured to store various types of data to support the operation of the system 100. For example, the data may include instructions for any applications or above methods operated on the system 100, programmes, applications, modules etc. Memory 120 may be implemented using any type of volatile or non-volatile memory devices, or a combination thereof, such as a static random access memory (SRAM), an electrically erasable programmable read-only memory (EEPROM), an erasable programmable read-only memory (EPROM), a programmable read-only memory (PROM), a read-only memory (ROM), a magnetic memory, a flash memory, a magnetic or optical disk.

[0075] Power source 150 provides power to various modules of the system 100. Power source 150 may include a power management system 100, one or more power sources, and any other modules associated with the generation, management, and distribution of power in the system 100.

[0076] Display 140 may include a screen providing an output interface for the system 100 and the user. In some embodiments, the screen may include a liquid crystal display (LCD), organic light-emitting diode (OLED), a touch panel, etc. If the screen includes the touch panel, the screen may be implemented as a touch screen to receive input signals from the user. Touch panel may include one or more touch sensors to sense touches, swipes, and gestures on the touch panel. The touch sensors may not only sense a boundary of a touch or swipe action, but also sense a period of time and a pressure associated with the touch or swipe action.

[0077] I/O interface 130 provides an interface between the processor 110 and peripheral interface modules, such as a keyboard, a click wheel, buttons, and the like. The buttons may

include, but are not limited to, a home button, a volume button, a starting button, and a locking button.

[0078] Communication module 160 may be configured to facilitate communication, wired or wirelessly, between the system 100 and other devices. System 100 may access a wireless network based on a communication standard, such as WiFi, 2G, or 3G, LTE, and 4G cellular technologies or a combination thereof. In one exemplary embodiment, the communication module may receive a broadcast signal or broadcast associated information from an external broadcast management system 100 via a broadcast channel. In one exemplary embodiment, the communication module may further include a near field communication (NFC) module to facilitate short-range communications. For example, the NFC module may be implemented based on a radio frequency identification (RFID) technology, an infrared data association (IrDA) technology, an ultra-wideband (UWB) technology, a Bluetooth (BT) technology, and other technologies.

[0079] In exemplary embodiments, the system 100 may be implemented with one or more application specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), controllers, micro-controllers, microprocessors, or other electronic modules, for performing the above described methods.

[0080] A skilled person would appreciate that the features described in one example may not be restricted to that example and may be combined with any one of the other examples.

[0081] The present invention relates to a system for generating a value index for properties and a method thereof generally as herein described, with reference to and/or illustrated in the accompanying drawings.

Claim

1. A system for generating a value index for properties, the system comprising:
 - a processor, and
 - a memory in communication to the processor for storing instructions executable by the processor,wherein the processor is configured to:
 - categorise the properties into one or more categories,
 - determine an optimal number of pre-clusters of properties based on the one or more categories,
 - cluster the properties into a plurality of pre-clusters based on the optimal number of clusters,
 - determine a set of search radii,
 - generate a cluster of properties based on one of the plurality of pre-clusters and the set of search radii, and
 - generate a value index for the cluster of properties.
2. The system according to claim 1, wherein each of the plurality of pre-clusters comprises of properties in geographically separated areas.
3. The system according to claim 1 or 2, wherein the process is further configured to:
 - receive a filtering criteria for a target property,
 - determine one or more categories of the target property,
 - identify a pre-cluster from the plurality of pre-clusters that matches the one or more categories of the target property,
 - generate the cluster of properties based on the pre-cluster and the set of search radii, and
 - generate the value index for the cluster of properties.
4. The system according to any one of claims 1 to 3, wherein each search radius of the set of search radii is longer than another, wherein each of the search radius is determined based on information of the territory where the properties are located in.

5. The system according to any one of claims 1 to 4, wherein the processor is configured to expand the search radius and generate another cluster of properties when the number of data points in the cluster based on the search radius is lower than a threshold number of data points to generate the value index.
6. The system according to any one of claims 1 to 5, wherein the processor is configured generate estimated values of the properties based on the value index.
7. The system according to any one of claims 1 to 6, wherein the processor is configured to de-cluster a transaction value of a cluster of property units into estimated values of each of the property units.
8. A method of generating a value index for properties, the method comprising:
 - categorising the properties into one or more categories,
 - determining an optimal number of pre-clusters of the properties based on the one or more categories,
 - clustering the properties into a plurality of pre-clusters based on the optimal number of clusters,
 - determining a set of search radii,
 - generating a cluster of properties based one of the plurality of pre-clusters and the set of search radii, and
 - generating a value index for the cluster of properties.
9. The method according to claim 8, wherein each of the plurality of pre-clusters comprises of properties in geographically separated areas.
10. The method according to claim 8 or 9, further comprising:
 - receiving a filtering criteria for a target property,
 - determining one or more categories of the target property,
 - identifying a pre-cluster from the plurality of pre-clusters that matches the one or more categories of the target property,
 - generating the cluster of properties based on the pre-cluster and the set of search radii, and

generating the value index for the cluster of properties.

11. The method according to any one of claims 8 to 10, wherein each search radius of the set of search radii is longer than another, wherein the method further comprises determining each of the search radius based on information of the territory where the properties are located in.

12. The method according to any one of claims 8 to 11, further comprising expanding the search radius and generate another cluster of properties when the number of data points in the cluster based on the search radius is lower than a threshold number of data points to generate the value index.

13. The method according to any one of claims 8 to 12, further comprising generating estimated values of the properties based on the value index.

14. The method according to any one of claims 8 to 13, further comprising de-clustering a transaction value of a cluster of property units into estimated values of each of the property units.

15. A non-transitory computer readable storage medium comprising instructions, such that the instructions, when executed by a processor in a system, cause the system to:

- categorise the properties into one or more categories,
- determine an optimal number of pre-clusters of properties based on the one or more categories,
- cluster the properties into a plurality of pre-clusters based on the optimal number of clusters,
- determine a set of search radii,
- generate a cluster of properties based on one of the plurality of pre-clusters and the set of search radii, and
- generate a value index for the cluster of properties.