May 2020

Multi-resolution Semantic Area Search

Yan Mayster

Brian Shucker

Multi-resolution Semantic Area Search

ABSTRACT

There are situations where people want to find geographic locations or areas that are similar to a given area, e.g., in terms of its urban or natural features, types of buildings, parks, views, suitability for particular lifestyles or purposes, etc. This disclosure utilizes machine learning techniques to provide answers to the general question “where can I experience life similarly (e.g., to where I am now)?” by providing strong clues about nearby areas that may be of interest to the person conducting the search.

KEYWORDS

- Semantic search
- Image search
- Multi-resolution search
- Geographical search
- Spatial index
- Machine learning
- Embeddings

BACKGROUND

There are situations where people want to find geographic locations or areas that are similar to a given area, e.g., in terms of its urban or natural features, types of buildings, parks, views, general suitability for particular lifestyles or purposes, etc. For example, a person shopping for a home may like a given area that has the kind of homes and city layout that they like, e.g., the proximity of a dog park, the relative remoteness of busy highways and thoroughfares; yet that area may not fit their other criteria of being close to their work location.

A natural question they may have is “are there neighborhoods like this one within X miles of my work location?”

As another example, a person may seek neighborhoods that support their lifestyle, e.g., one oriented towards rowing sports, and thus seek neighborhoods of a particular profile, e.g., must have parks with lakes, a river, etc. A tourist may enjoy a location due to the views it affords of nature or buildings, and may pose a question about whether there are other areas nearby that are similar and thus worth visiting too. Most generally, the question may be formulated as “where can I experience life similarly (e.g., to where I am now)?”

On embeddings

In the context of machine learning, embeddings are learned continuous vector representations of discrete variables. An image is just such a discrete variable represented in the high dimensional space of width×height×number_of_channels dimensions. A 64-dimensional or 128-dimensional embedding of an image is a low-dimensional representation of the image, in which each dimension represents some machine-learned feature. The machine-learned feature is not typically related to any specific semantics. Thus, an embedding is a vector in an *N*-dimensional space, usually with the values between 0.0 and 1.0 in each dimension, the entire embedding space being the surface or the volume of a unit sphere in that space.

A machine-learning algorithm trains the embedding model such that when all dimensions of the resulting vectors are taken together, any two sufficiently close embedding vectors will exhibit similarity in terms of shapes, colors, spacing between shapes, and more general visual patterns in the corresponding images. Additionally, the model designer can enumerate image features of interest and train the model to cluster these features together in the embedding space.

The distance between a pair of embedding vectors is usually measured as the cosine of the angle between them (the normalized dot product).

DESCRIPTION

This disclosure describes machine learning techniques to provide answers to the above-mentioned question "where can I experience life similarly (e.g., to where I am now)?" by providing strong clues about nearby areas that may be of interest to the person conducting the search.

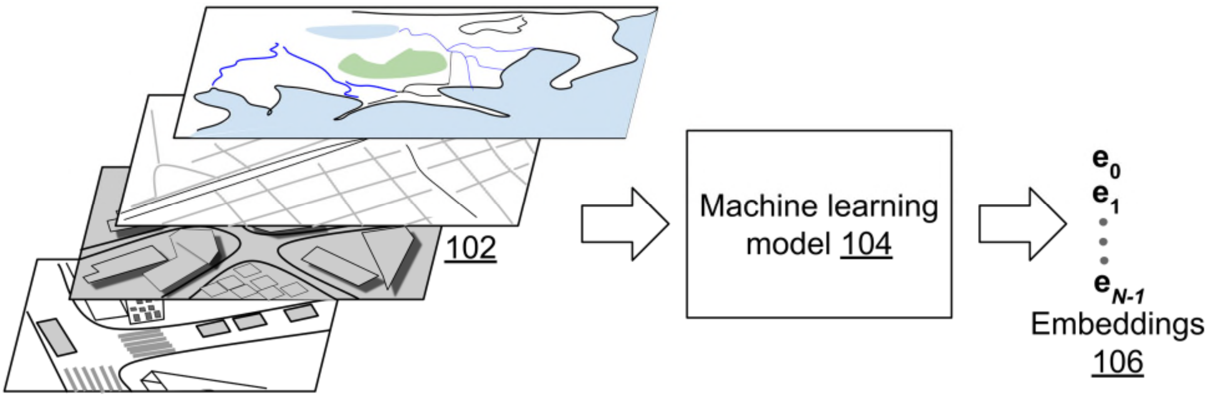

Fig. 1: Generating embeddings of multi-resolution images of a point on the earth's surface

Per the techniques, illustrated in Fig. 1, aerial and satellite imagery (102) of the earth's surface is resampled at N different zoom levels and is divided into sub-images or tiles on grids at each zoom level. In this manner, a point on the earth's surface is associated with a stack of nested tiles of differing sizes, e.g., with the lowest level corresponding to the size of an individual building and the highest level to that of a county or equivalent. For each zoom level, a machine learning model (104) is trained to output an embedding vector (106) for each tile such that tiles that depict areas with feature similarities are clustered together.

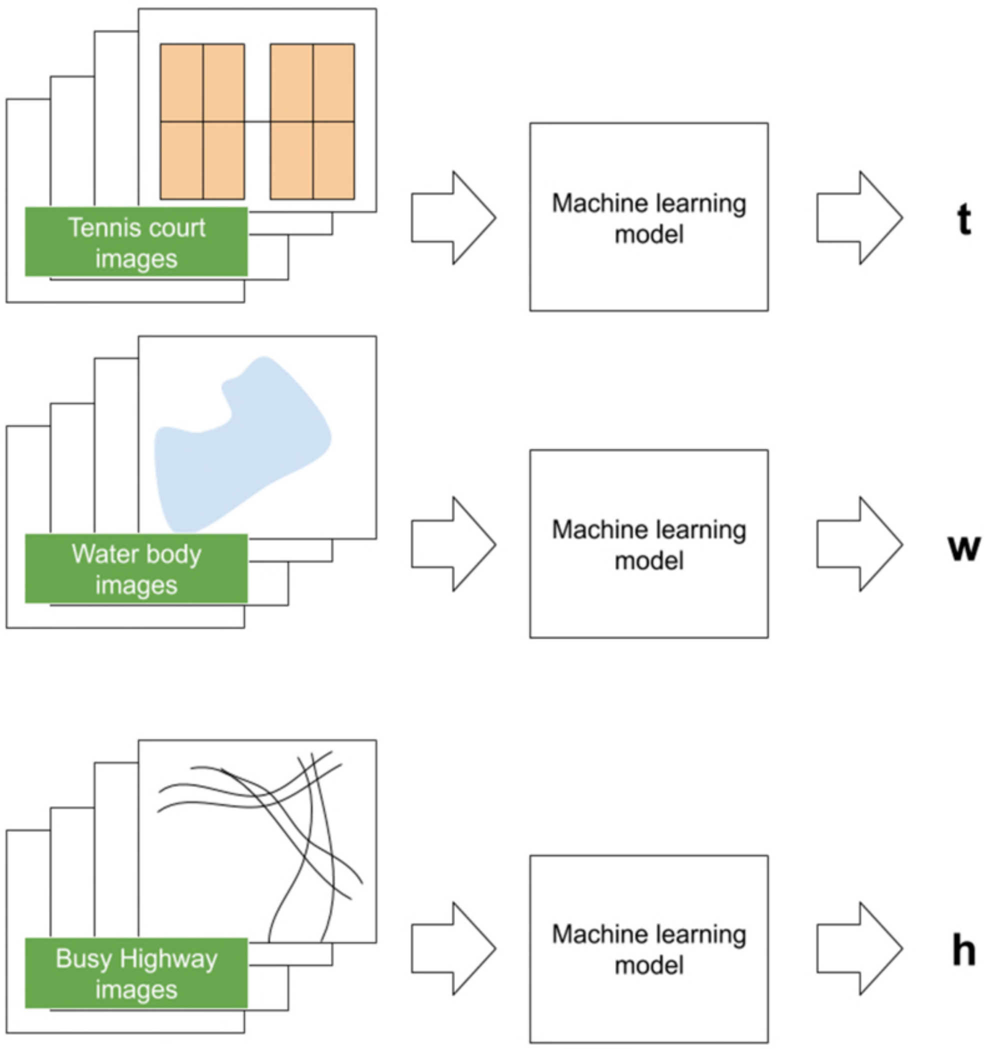

Fig. 2: Creating embeddings of features of interest

As illustrated in Fig. 2, large numbers of images that include particular features of interest, e.g., tennis courts, water bodies, shopping malls, busy highways, etc. are obtained and their embedding values computed for tiles at each zoom level (or the zoom level that is the most appropriate for that type of semantic feature). Embeddings of images of a given feature, e.g., tennis court (**t**), water body (**w**), highway (**h**), etc., of interest are clustered in embeddings space, and the regions of clustering are identified as such.

The embedding space $\mathbf{e}_0, \mathbf{e}_1, \ldots \mathbf{e}_{N-1}$ (Fig. 1) is analyzed by identifying neighborhoods (epsilon-balls in mathematical terms) of the points that correspond to images with known semantic features of interest. Other images are likely to depict these features, e.g., tennis court, if their embeddings fall into one of these regions. Effectively, a dictionary of typically desired features is created in embedding space that maps to different lifestyles or experiences.

$\mathbf{e}_0$	`{ t(0.9); w(0.5); h(0.1), ... }`
$\mathbf{e}_1$	`{ t(0.95); w(0.09); h(0.05), … }`
$\mathbf{e}_{N-1}$	`{ t(0.95); w(0); h(0), … }`

Fig. 3: Assigning lifestyle scores at each zoom level

Each tile at each zoom level is assigned a “lifestyle” score for each specific lifestyle, based on the number of different features detected and the degree of similarity (e.g., as measured by cosine distance) versus known examples of such features in the contained tiles at the same or lower zoom levels. For example, as illustrated in Fig. 3, at a relatively zoomed-out level ($\mathbf{e}_0$), the image matches for tennis court (**t**) with a score of 0.9, water-body (**w**) with a score of 0.5, busy highway (**h**) with a score of 0.1, etc. As the zoom level increases, the tennis court feature gains prominence. Note that a single tile, though it has a single embedding value at its zoom level, may match multiple features (because of overlaps in embedding space) at different distances/costs.

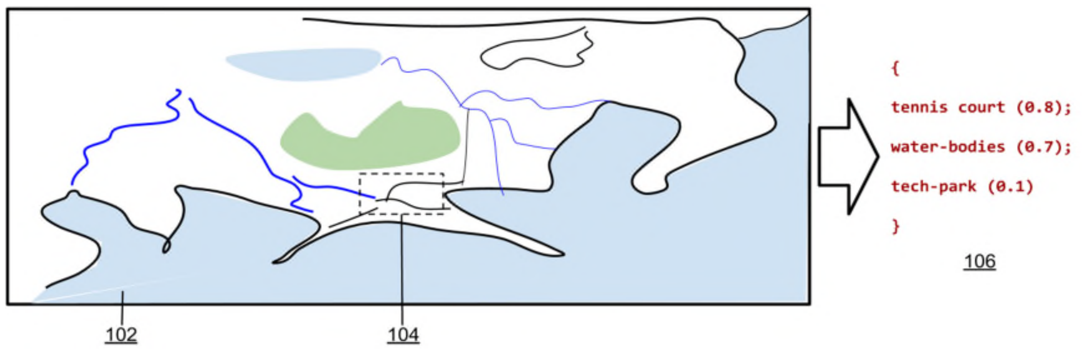

Fig. 4: A user search for features within an area

As illustrated in Fig. 4, when a user searches for areas (perhaps by drawing a certain shape (104) on a map (102), or within a certain radius of a specific point) that can support a specific lifestyle or matches a particular interest (semantic query), the space of the tiles at the zoom level corresponding to area size and overlapping the stipulated region is searched, and the top-scoring tiles are picked (106) based on that lifestyle, e.g., the "combined feature" score for that semantic query.

Likewise, queries of the form "show me a settled area similar to the one I'm in" (with the implication of a match in, e.g., urban environment features), or "I like this neighborhood in city X, show me neighborhoods in city Y like it" (with the implication of a real-estate search), or "show me an area with views similar to the one I'm in" (with the implication of a match in, e.g., features that tend to dominate the skyline) can be answered by consulting the dictionary for features that match these queries and looking for, across zoom levels, tiles or collections of tiles that have similar feature scores for the same features from these semantic groups as the ones in the area where the user is located or has specified as the source area for the query.

The dictionary of features or, alternatively, the dimensions of the embedding space (which, as stated, are not in a one-to-one map to semantic features, but rather each dimension a combination of semantic features) may be explicitly exposed to the search user as tunable parameters when modeling their search criteria, e.g., a user seeking travel advice, a real estate agent conducting a house search, an insurance agent analyzing certain risks that are dependent on the presence of particular urban features or terrain type, etc., with statistical data linking change in each feature or embedding space dimension to price changes (from descriptions “more/less” expensive to predictive quantitative changes in today’s and expected future prices). Thus, the human agent can formulate the criteria by selecting from the features or by simply changing the boundaries of the embedding space region(s), with these being automatically translated into embedding search settings to obtain search results as described above.

In this manner, the machine learning techniques described herein enable a user to search for geographic areas by specifying multiple features or ranges/boundaries, using GPS as needed, in some (possibly pictorial) representations of space with multiple dimensions/axes. The techniques yield geographic areas that match the user search (have a similarity score with the query specified) and enable the user to answer the question “where can I experience life similarly (e.g., to where I am now),” or some variant thereof. The techniques can be implemented as part of a mapping or navigation application or in other applications that include features that require neighborhood search, e.g., real estate, insurance, travel advice, or other types of applications.

CONCLUSION

There are situations where people want to find geographic locations or areas that are similar to a given area, e.g., in terms of its urban or natural features, types of buildings, parks, views, suitability for particular lifestyles or purposes, etc. This disclosure utilizes machine learning techniques to provide answers to the general question “where can I experience life similarly (e.g., to where I am now)?” by providing strong clues about nearby areas that may be of interest to the person conducting the search.