



(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number: PCT/US2017/059006

(22) **International Filing Date:** 30 October 2017 (30.10.2017)

(25) **Filing Language:** English

(26) **Publication Language:** English

(71) **Applicant: HEWLETT-PACKARD DEVELOPMENT COMPANY, L. P.** [US/US]; 10300 Energy Drive, Spring, Texas 77389 (US).

(72) **Inventors:** **BALINSKY, Helen**; Filton Road Stoke Gifford, Pt., Bristol Bristol BS34 8QZ (GB). **BALINSKY, Alexander**; Prof Alexander Balinsky, School of Mathematics, Cardiff University, Senghennydd Road, Cardiff South Wales CF24 4AG (GB). **PERCY, Isabelle**; Oak Lodge, Mill Lane, Blackpill, Swansea South Glamorgan SA3 5BD (GB). **SIMSKE, Steve, J**; 3390 E Harmony Rd, Ft. Collins, Colorado 80528 (US).

(74) **Agent:** **BURROWS, Sarah** et al.; HP Inc, 3390 E. Harmony Road, Mail Stop 35, Fort Collins, Colorado 80528 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: CRIME PROFILES

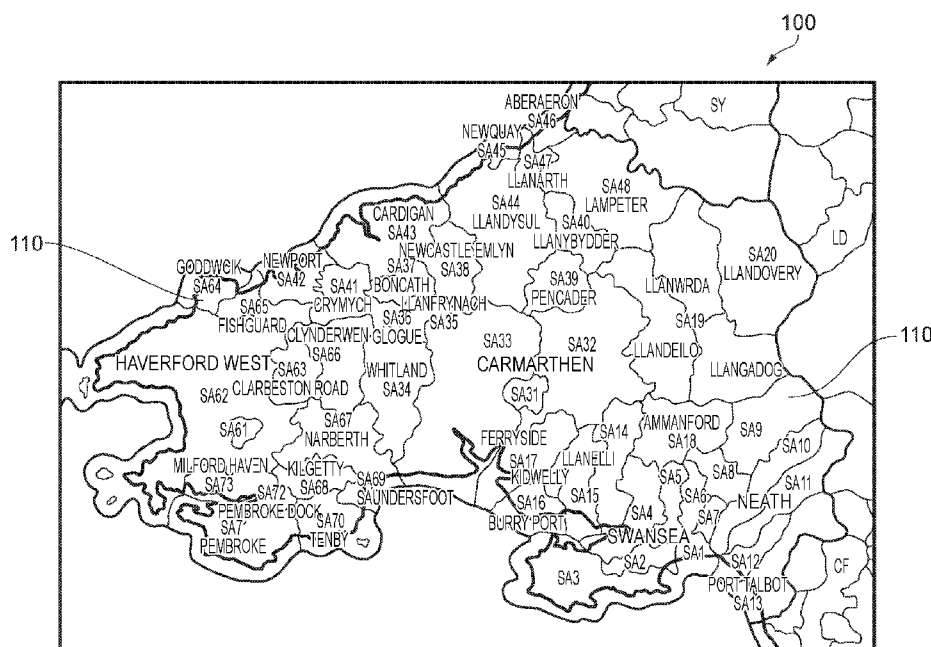


FIG. 1

(57) Abstract: A method for identifying regions with similar, temporally stable crime profiles, the method comprises partitioning a geographical area into multiple non-overlapping regions, for each region, generating a set of tuples representing a regional crime record comprising a sequence of offence data, and generating a measure of similarity between respective pairs of the multiple regions.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

Published:

— *with international search report (Art. 21(3))*

CRIME PROFILES

BACKGROUND

[0001] Urban population growth, one of the main trends of today's world, is likely to result in vast, yet sparsely populated areas outside of large or mega-cities. Such sparsely populated areas can be challenging to police. Regions in the same area may themselves vary in population density, occupancy and types of crime committed. This can impede the creation of a uniform crime prediction model for the area.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Various features of certain examples will be apparent from the detailed description which follows, taken in conjunction with the accompanying drawings, which together illustrate, by way of example only, a number of features, and wherein:

[0003] Figure 1 is a schematic showing a police area, partitioned on relatively small regions (UK postcodes);

[0004] Figure 2A is a flow chart for identifying regions with similar, temporally stable crime profiles according to an example;

[0005] Figure 2B is a flow chart for identifying regions with similar, temporally stable crime profiles according to an example;

[0006] Figure 3 is a schematic showing sequential crime records for two different regions and different time intervals;

[0007] Figure 4 is a schematic showing a sample measure similarity matrix computed for three regions: R1, R2 and R3;

[0008] Figure 5 is a schematic showing a visualization of the measure similarity for over one hundred regions in different periods of time, where denser areas correspond to higher similarity;

[0009] Figure 6A is a schematic showing a visual analytic of clusters having similar crime statistics and where regions in the area are represented by labels on the curve; and

[0010] Figure 6B is a schematic showing clusters of similarity that are marked by circles and where regions in the area are represented by labels on the curve.

DETAILED DESCRIPTION

[0011] In the following description, for purposes of explanation, numerous specific details of certain examples are set forth. Reference in the specification to "an example" or similar language means that a particular feature, structure, or characteristic described in connection with the example is included in at least that one example, but not necessarily in other examples.

[0012] For the purposes of crime prediction and prevention, some methods use a very specific location analysis (latitude and longitude) of historic crimes to try and predict the location of future crime hotspots. For example, a method may use an assignment of binary values to generate a list of crime hotspots. These hotspots are then used to cluster together areas that are closer than a threshold distance to determine regions where a crime is likely to occur in future. Other research into re-offending can be based on various psychological assessments and tested on those who have already been incarcerated.

[0013] According to an example, there is provided a method for using crime similarities between areas for crime analysis and prediction. A measure of similarity for different regions is determined, thus allowing a group of similar regions to be used to provide a larger crime data set. This information can be used to build reliable models for crime prediction in regions with relatively low counts for particular crimes. The ability to combine crime data from regions thus enables a predictive crime model to be built with better accuracy. For areas with a similar pattern of crimes over different periods of time, information from one area can be used to predict crime in a similar area or areas. This provides a more efficient approach to managing resources, policies, and/or strategies.

[0014] In contrast to cities, in rural areas there is a sparse supply of data and large policing areas may only be covered by a few police personnel. This means that in many areas there are insufficient crime numbers for crime analysis and predictions. In an example, a measure of similarity between regions is used to identify exploitable patterns between similar areas. This “cold start” problem is aided by the fact that crime similarity for, for example, zip/postcode areas in sequential years remains largely unchanged, hence providing the basis for crime predictions.

[0015] The methods described herein use known data (i.e. memory based system) of where and when a crime was committed to enable prediction of where and when a crime is likely to happen in the future. As such, a recommendation or level of risk of crimes that are likely to occur in the same location or within a period of time can be provided, i.e. a reoffending and crime prediction model based on crime similarity can be developed. There may be provided a calculation of an offender's likelihood of re-offending and the production of a list of the most likely individuals within a given location to re-offend.

[0016] The production of a measure of similarity between one region and another can provide a list of the most likely areas in which similar crimes occur. Such a list of locations is helpful in reactive situations. Since the similarity between two different locations is calculated using the similarity of the crime types and times at each location, the police will be able to know both when and in which areas crimes are likely to occur, given that a crime has or is likely to occur in a similar area. In an example, these locations can be grouped into clusters of similar locations. A list of likely offences within a location can be provided and predicted, including when these offenses are most likely to occur. To find this list of most likely residents to re-offend in a given location, a model has been built, based on a Random Forests algorithm, with the ability to accurately predict the re-offending behaviour of an individual offender.

[0017] Figure 1 is a schematic showing a police area, partitioned on relatively small regions (UK postcodes) according to an example. A police area is the area for which a territorial police force is responsible for policing. In the present example this area is assumed to be large, with high variability of

population and types of crimes across the area. It is further assumed that the area is partitioned on relatively small regions (e.g. zip/postcodes), which do not overlap and which jointly cover the entire area.

[0018] For example, the first four letters of UK postcodes (e.g. BS34, CF23, SA48) can be considered as regions. Figure 1 shows West Wales's partition area 100 on postcode regions 110. Depending on a task, area properties, population density, types of crimes, etc. different partitions of the area onto regions can be considered, where larger/smaller regions are considered for different partitions. In an example, multiple partitions can be employed in parallel. The results can then be combined in a voting, boosting, or ensemble approach. As such, it is possible to find an "optimal" partitioning that optimizes a cost function (e.g. one maximizing the similarity measure between closely alike regions, etc.).

[0019] According to an example, a policing area A can be partitioned into regions R , using, for example, a processor of a system, such that:

$$A = \cup_i R_i \text{ and } R_i \cap R_j = \emptyset \text{ for any } i \neq j$$

[0020] It is possible to identify regions that have similar crime statistics and which remain stable over time, i.e. stable over time crime similarity regions, such that data for these regions have internal similarity and thus can be effectively combined to create a reliable prediction model (for these regions).

[0021] According to an example, regions do not need to be spatially adjacent or be similar at the same time period. The same policing area can be partitioned on different regions and produce similarity clusters for different groups of crimes, where a cluster relates to regions having similar crime statistics. For regions with similarity over different periods of time: the information about earlier regions can be used to anticipate the development in later regions. Thus, if some types of crimes increased in one of the similarity regions, this information can be used for successful crime management in similar regions. If crime similarity of regions stays the same for a while but then changes, it can be reported and investigated further, i.e. why was one area/region subject to a rise/decline of a particular type(s) of crime(s), whilst the other(s) was not?

[0022] In an example, crime similarity can be valid for a subset of crimes (of interest) and not for the entire set of crimes for regions. This information provides a closer look at former similar regions and allows for the identification of influencing factors, which can be leveraged from a more successful region(s). In an example, one or a few regions from similarity clusters can be used for one or more: testing/evaluation of the effects of new policies, new area development, or deploying new policing methods. If proven to be effective, the results can be rolled out to the entire cluster.

[0023] Figure 2A is a flow chart for identifying regions with similar, temporally stable crime profiles according to an example. At block 210 a geographical area is partitioned into multiple non-overlapping regions. At block 220 a set of tuples representing a regional crime record comprising a sequence of offence data is generated for each region. At block 230 a measure of similarity between respective pairs of the multiple regions is generated.

[0024] Further aspects are described below with reference to Figure 2B. In this figure, blocks 210, 220 and 230 are the same as those described above with reference to Figure 2A. Those aspects will not therefore be discussed again. Rather, the additional aspects will be described.

[0025] Figure 2B also shows a flow chart for identifying regions with similar, temporally stable crime profiles according to an example. The similarities between locations are calculated and clustered. A location can be selected and other locations can be recommended that may experience similar crimes at similar times. At block 240 data from selected regions is combined to form an inter-regional crime record, where regions may be selected according to a measure of similarity. For example, regions may be selected that have a measure of similarity above a predetermined threshold value. At block 250 the measure of similarity between respective pairs of the multiple regions may be generated using a parameter representing a regional characteristic and/or using crime records.

[0026] Figure 3 is a schematic showing sequential crime records for two different regions (region i, R_i 310 and region j, R_j 320) and different time intervals according to an example. In an example, for each region R_i there

exists a crime record C_i (a sequence of crimes) ordered in time 330 as they occurred. In each area there is a finite number of defined types of crimes. The column "OffenceDesc" (offence description) 340 contains specific code words for crime, e.g. "THEFT", "CRIMDAM", "VIOLENCE". The total number of different crime types recorded by police is well defined by local laws and regulations and may be up to a hundred or a few hundred different crime types.

[0027] A dictionary of crimes, D , can provide an exhaustive list of all crimes for an area, e.g. types of crimes that occur within various locations within a dataset and the times at which they occur. A criminal record for a region from this area during some time interval is represented as a sequence of elements from D . In an example, compound crimes (analogous to compound words) can also be used. For example, "grand larceny auto" + "drug dealing" can be a compound crime as there is a causal link between the act of stealing a car and drug dealing.

[0028] The measure of similarity is indicated by Sym for such sequences to identify similar regions, and thus cluster the area. As such, constructed measures of similarity can be used to perform clustering (unsupervised learning) or to group regions with similar crimes together. For example, TF-IDF vector representation can be used directly to perform K-mean clustering. Other examples include Spectral Clustering and Affinity Propagation Algorithms, which require only similarity measures.

[0029] In an example, "users" in the dataset can be considered to be various locations within a police dataset, with "items" representing information about each crime, and a "rating" representing the number of crimes of that particular combination that happen within that location. Crimes in a particular location can be envisioned as a text, with words being crimes in their sequential order: techniques from text classifications can then be applied to group the locations into similar "topics". According to an example, to design a "recommender system", collaborative filtering can be used. Collaborative filtering is a method by which information representing the interests of several different users can be collaborated together to form automatic predictions about the interests of another user. In this case, the distributions of crime within different locations

can be grouped together to form automatic predictions about the crime distributions of another area.

[0030] According to an example, a numerical measure of similarity for two regions can be determined, which remains stable over some period of time for different regions, and which thus follow similar crime patterns.

[0031] For example, each crime record C_i may be considered as a bag of crimes from D . The "Jaccard" similarity (or TF-IDF similarity) of two bags A and B is defined by:

$$Sym(A, B) = \frac{\sum_x \min(n_x(A), n_x(B))}{\sum_x (n_x(A) + n_x(B))}$$

where $n_x(A)$ is a number of occurrences of crime x in bag A .

[0032] Collaborative filtering is based on the assumption that if two users, A and B , rate a series of n items similarly, or have similar behavior (e.g. buying items on an e-Commerce site, listening to songs on a digital music listening site), they will continue to rate and act on those items similarly in the future. For each pair of regions R_i and R_j , a corresponding similarity value can be computed using a processor of the system and organized into a matrix $M_{ij} = Sym(R_i, R_j)$. The matrix is symmetrical as the measure is commutative $Sym(A, B) = Sym(B, A)$.

[0033] Figure 4 is a schematic showing a sample measure similarity matrix 400 computed for three regions: R_1 , R_2 and R_3 . The maximum value is reached on the second diagonal (shaded boxes 410), where each region is compared to itself $Sym(A, A) = 0.5$.

[0034] Jaccard similarity is not the only measure of similarity that can be introduced to compare regions. In an example, TF-IDF vector representation and cosine similarity may be used. In an example, every region can be represented by vector where the length equals the total number of crimes. For example, for a crime x and region A , the A_x component equals:

$$\frac{n_x(A)}{\max_y(n_y(A))} \cdot \left(\log \frac{N}{N_x + 1} + 1 \right)$$

where N is the total number of regions and N_x is the number of regions where crime x occurs.

[0035] After defining vector representation of each region, it is possible to use a processor of the system to calculate the similarity as a cosine similarity of vectors:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Here A and B are the TF-IDF vectors for any two separate locations within the dataset. A_i and B_j are, therefore, the components of these vectors A and B respectively.

[0036] Figure 5 is a schematic showing a visualization of the measure similarity (as described above with reference to Figure 4) for over one hundred regions in different periods of time (2010, 2012, 2014), where denser areas correspond to higher similarity. Along the x and y axes are the many regions (postcodes) and the maximum value diagonal 520 is shown. It can be seen that crime similarity for regions in sequential years remains largely unchanged providing the basis for prediction.

[0037] The Jaccard Similarity for bags is defined for two bags A and B by counting an element n times in the intersection, where n is the minimum number of times that element appears in either A or B . The union is then defined by counting for the element the sum of the number of times that element appears in both A and B . This results in a visualization or “heat map” of a similarity matrix for a dataset. In this example, the similarity matrix is for crime data from 2010 and locations are divided by postcode sector. Darker areas indicate greater similarity, while brighter areas indicate less similarity between the two locations in question.

[0038] In Figure 5, there is a pattern that can be seen within the similarity matrix and whether this pattern holds over time must be established in order to determine whether or not collaborative filtering techniques are appropriate for this particular data set (of around two hundred pairs of regions/postcodes). In order to illustrate the validity of the assumptions that individuals who held similar opinions in the past will continue to do so in the future and that individuals will like similar kinds of items in the future as they did in the past, plots of the similarities for various years and months can be produced within the dataset (2010, 2012, 2014). As shown, the pairwise similarities between the locations in this dataset do not change from year to year, indicating that there is no significant seasonal or yearly trend in the pattern of similarities. As such, the assumption that a "user's preferences" (i.e. a location's crime patterns) do not change over time appears to hold. To make the precomputed similarity matrix usable, clustering algorithms are run on the matrix to discover the hidden groupings within the dataset. An example of two suitable algorithms are the Spectral Clustering and Affinity Propagation algorithms. It is noted that it has been found that the clusters shown cannot be seen based on a simple rural/urban classification of locations.

[0039] Alternatively, or additionally, the similarities between postcodes may be visualized using a t-distributed stochastic neighbour embedding (t-SNE) visualization, as shown in Figure 6A. Figure 6A is a schematic showing a visual analytic of clusters having similar crime statistics according to an example, where regions in the area are represented by labels on the curve. In this example, it can be seen that there is no well-defined strong clustering but postcode areas of similar crimes are grouped along the one-parametric curve 610. This visualization allows for visual analytics to be built for police officers providing a better understanding of crime similarities. In an example, visual analytics could use Tanimoto's of similarity and distance as well as Jaccard if bitmap similarity is considered.

[0040] Figure 6B is a schematic showing clusters of similarity that are marked by circles and where regions in the area are represented by labels on the curve. Figure 6B shows circles 620 around points (regions or postcodes) which are

closer on the parametric curve and hence have a more similar crime structure, i.e. darker edges between postcodes tell us that regions are more similar.

[0041] By using k-mean or hierarchical clustering, we can combine many regions in several groups and train machine learning models with more data and better modelling of specific crime behaviours. The curves shown in Figures 6A and 6B give an estimation on how many clusters there should be in k-mean or hierarchical clustering (where the number of clusters is a parameter that should be fixed before using clustering algorithms).

[0042] The present disclosure presents a method for generating a temporally stable measure of the similarity of crime profiles between regions in an area. The similarity measure can be used to cluster regions by crime profile, the aim being to better enable law enforcement to direct services or predict future trends. The methods described herein assists the police in the allocation of increasingly limited resources, especially in the face of budget freezes and government cuts. It provides a way to predict types of crime and reoffenders.

[0043] The present disclosure addresses one of the key problems of policing in rural areas: dealing with statistically low numbers of different crimes in some vast, low populated areas. Numbers are not small from a policing and crime prevention point of view, but insufficient for building Machine Learning models for individual crimes in particular areas.

[0044] Given historical records of crimes in policing areas and their partitions (regions) the methods described herein determine a stable measure of similarity, which allows identification of areas with similar patterns of crimes, which occur over the same period of time or over separated similar periods of time. Combining regions of similar crime patterns over a period of time, it is possible to collect accurate and reliable data for areas of crime pattern similarity, thus building bespoke prediction models for different group regions or clusters. This can provide clues to questions such as what are the factors for such behaviour, for example average/median salary, depreciation index or something else.

[0045] The methods described predicts crime on a crime by crime basis for all crimes within the dataset. This removes the need to carry out costly psychological assessments on offenders, or to wait until the individual has carried out enough crimes to be brought into custody, i.e. from the very first crime that the offender commits, it is possible to predict what their next move will be, based on the profile developed for the subset of offenders matching this individual.

[0046] The approaches described may not be limited to crime predication but may be extended to predict other trends based on data analysis of other log-files, including those associated with equipment and system monitoring and diagnostics.

[0047] Examples in the present disclosure can be provided as methods, systems or machine-readable instructions. Such machine-readable instructions may be included on a computer readable storage medium (including but not limited to disc storage, CD-ROM, optical storage, etc.) having computer readable program codes therein or thereon.

[0048] The present disclosure is described with reference to flow charts and/or block diagrams of the method, devices and systems according to examples of the present disclosure. Although the flow diagrams described above show a specific order of execution, the order of execution may differ from that which is depicted. Blocks described in relation to one flow chart may be combined with those of another flow chart. In some examples, some blocks of the flow diagrams may not be necessary and/or additional blocks may be added. It shall be understood that each flow and/or block in the flow charts and/or block diagrams, as well as combinations of the flows and/or diagrams in the flow charts and/or block diagrams can be realized by machine readable instructions.

[0049] The machine-readable instructions may, for example, be executed by a general-purpose computer, a special purpose computer, an embedded processor or processors of other programmable data processing devices to realize the functions described in the description and diagrams. In particular, a processor or processing apparatus may execute the machine-readable instructions. Thus, modules of apparatus may be implemented by a processor

executing machine readable instructions stored in a memory, or a processor operating in accordance with instructions embedded in logic circuitry. The term 'processor' is to be interpreted broadly to include a CPU, processing unit, ASIC, logic unit, or programmable gate set etc. The methods and modules may all be performed by a single processor or divided amongst several processors.

[0050] Such machine-readable instructions may also be stored in a computer readable storage that can guide the computer or other programmable data processing devices to operate in a specific mode.

[0051] For example, the instructions may be provided on a non-transitory computer readable storage medium encoded with instructions, executable by a processor.

[0052] Figure 7 shows an example of a processor 710 associated with a memory 720. The memory 720 comprises computer readable instructions 730 which are executable by the processor 710. The instructions 730 may comprise:

Instructions to partition a geographical area into multiple non-overlapping regions;

Instruction to, for each region, generate a set of tuples representing a regional crime record comprising a sequence of offence data; and

Instructions to generate a measure of similarity between respective pairs of the multiple regions.

[0053] Such machine-readable instructions may also be loaded onto a computer or other programmable data processing devices, so that the computer or other programmable data processing devices perform a series of operations to produce computer-implemented processing, thus the instructions executed on the computer or other programmable devices provide a operation for realizing functions specified by flow(s) in the flow charts and/or block(s) in the block diagrams.

[0054] Further, the teachings herein may be implemented in the form of a computer software product, the computer software product being stored in a storage medium and comprising a plurality of instructions for making a

computer device implement the methods recited in the examples of the present disclosure.

[0055] While the method, apparatus and related aspects have been described with reference to certain examples, various modifications, changes, omissions, and substitutions can be made without departing from the spirit of the present disclosure. In particular, a feature or block from one example may be combined with or substituted by a feature/block of another example.

[0056] The word "comprising" does not exclude the presence of elements other than those listed in a claim, "a" or "an" does not exclude a plurality, and a single processor or other unit may fulfil the functions of several units recited in the claims.

[0057] The features of any dependent claim may be combined with the features of any of the independent claims or other dependent claims.

CLAIMS

1. A method for identifying regions with similar, temporally stable crime profiles, the method comprising:
 - partitioning a geographical area into multiple non-overlapping regions;
 - for each region, generating a set of tuples representing a regional crime record comprising a sequence of offence data; and
 - generating a measure of similarity between respective pairs of the multiple regions.
2. A method as claimed in claim 1, further comprising:
 - generating a measure of similarity between respective pairs of the multiple regions using a parameter representing a regional characteristic.
3. A method as claimed in claim 1, further comprising:
 - generating a measure of similarity between respective pairs of the multiple regions using crime records.
4. A method as claimed in claim 1, further comprising:
 - combining data from selected regions to form an inter-regional crime record.
5. A method as claimed in claim 4, wherein regions are selected according to a measure of similarity.
6. A method as claimed in claim 5, wherein regions are selected that have a measure of similarity above a predetermined threshold value.
7. A method as claimed in claim 1, wherein generating a measure of similarity comprises calculating one of a Jaccard similarity coefficient and a term-frequency inverse document frequency vector representation and cosine similarity between respective pairs of the multiple regions.

8. A non-transitory machine-readable storage medium encoded with instructions executable by a processor for identifying regions with similar, temporally stable crime profiles, the machine-readable storage medium comprising instructions to:

generate a data structure comprising multiple parts representing a regional crime record comprising a sequence of offence data for respective ones of multiple regions subdividing an area of interest; and

comparing a measure of the similarity between respective pairs of the multiple regions.

9. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

instructions to generate a measure of similarity between respective pairs of the multiple regions using a parameter representing a regional characteristic.

10. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

instructions to generate a measure of similarity between respective pairs of the multiple regions using crime records.

11. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

instructions to combine data from selected regions to form an inter-regional crime record.

12. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

instructions to select regions according to a measure of similarity.

13. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

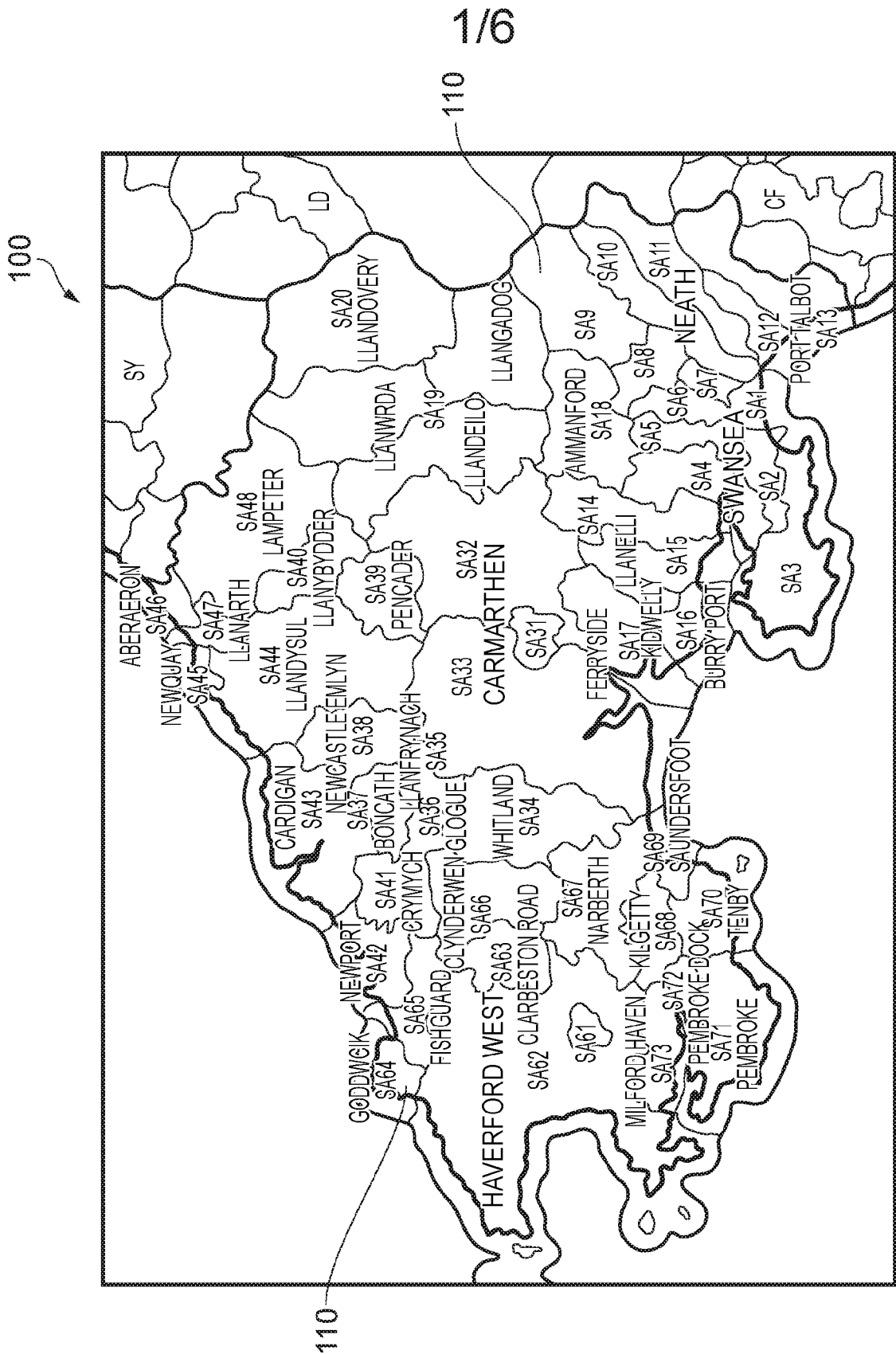
instructions to select regions with a measure of similarity above a predetermined threshold value.

14. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

instructions to calculate one of a Jaccard similarity coefficient and a term-frequency inverse document frequency vector representation and cosine similarity between respective pairs of the multiple regions.

15. A non-transitory machine-readable storage medium as claimed in claim 8, further encoded with:

instructions to cluster multiple regions into groups.



2/6

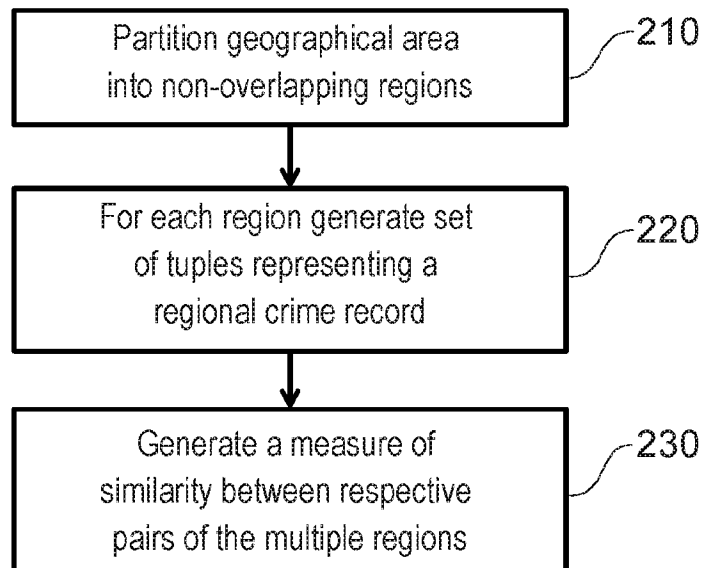


FIG. 2A

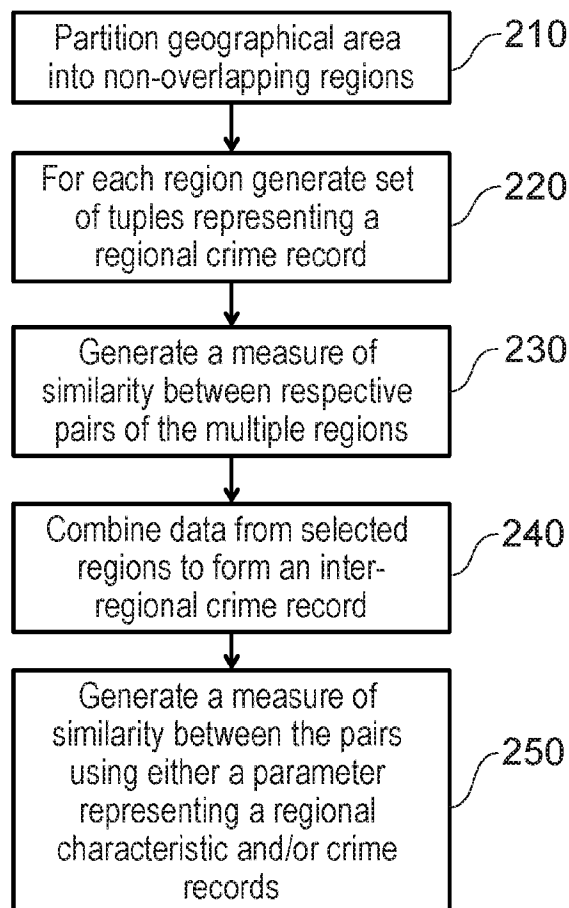


FIG. 2B

310	330	Region 1 (R ₁)	ReportedDate	OffenceDesc
			02/01/2014 00:00	SEX
			02/01/2014 13:12	THEFT
			03/01/2014 15:54	THEFT
			07/01/2014 17:32	VIOLENCE2
			09/01/2014 09:28	BURGDW
			13/01/2014 10:43	THEFT
			19/01/2014 17:28	THEFT
			19/01/2014 17:28	THEFT
			24/01/2014 10:36	THEFT
			30/01/2014 09:43	BURGDW
			02/02/2014 01:15	CRIMDAM
			02/02/2014 11:58	THEFT
			02/02/2014 23:37	CRIMDAM
			02/02/2014 23:37	VIOLENCE
320	330	Region 2 (R ₂)	Reported Date	OffenceDesc
			06/01/2010 11:00	CRIMDAM
			30/01/2010 01:21	VIOLENCE
			05/02/2010 08:43	BURGDW
			05/02/2010 08:43	BURGDW
			05/02/2010 08:43	BURGDW
			14/02/2010 16:56	VIOLENCE2
			15/02/2010 15:09	SEX
			20/02/2010 18:00	VIOLENCE2
			29/03/2010 19:22	THEFT
			08/04/2010 00:59	VIOLENCE2
			24/04/2010 23:02	VIOLENCE
			15/05/2010 10:56	THEFT
			11/06/2010 08:51	CRIMDAM
			11/06/2010 21:50	VIOLENCE2

FIG. 3

4/6

410

400

R ₃	0.1	0.01	0.5
R ₂	0.3	0.5	0.01
R ₁	0.5	0.3	0.1
Region (postcode)	R ₁	R ₂	R ₃

FIG. 4

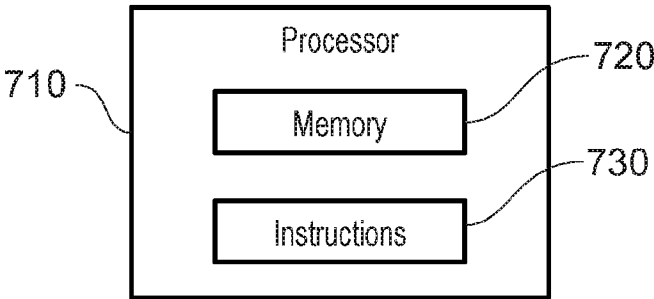


FIG. 7

5/6

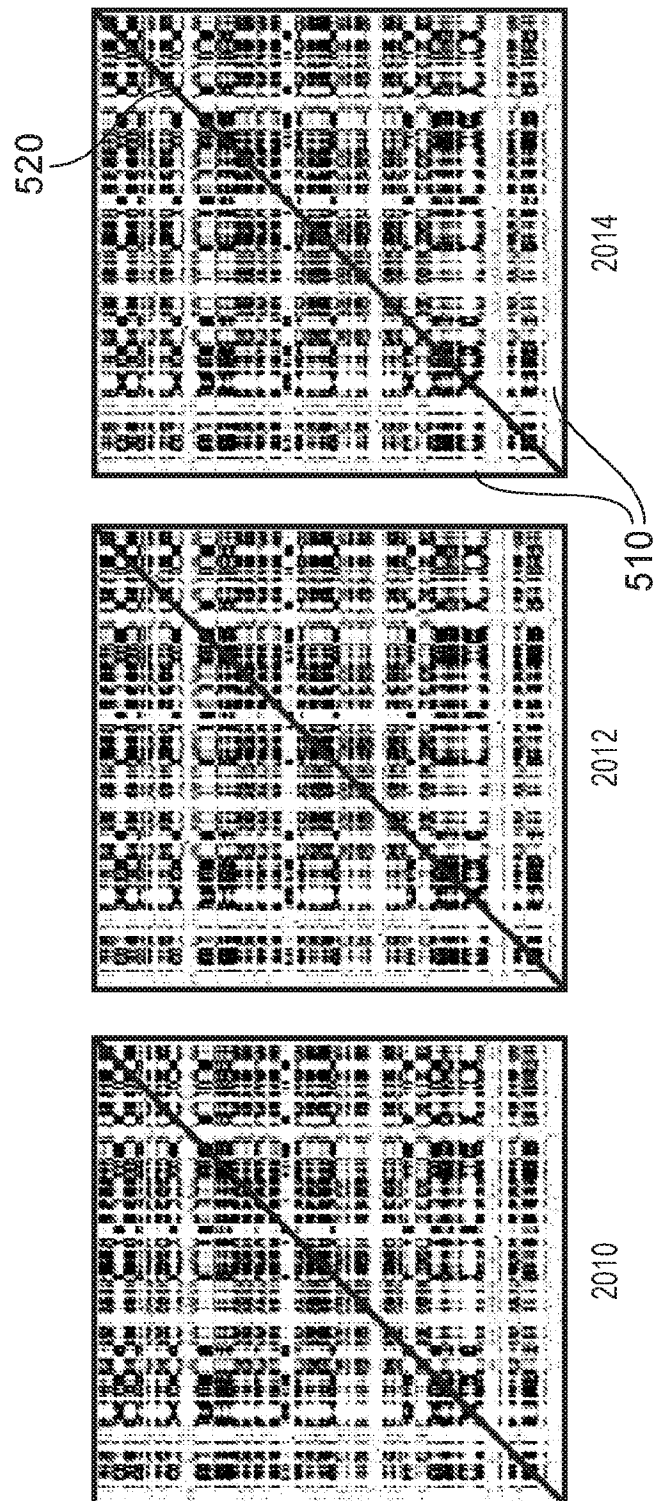


FIG. 5

6/6

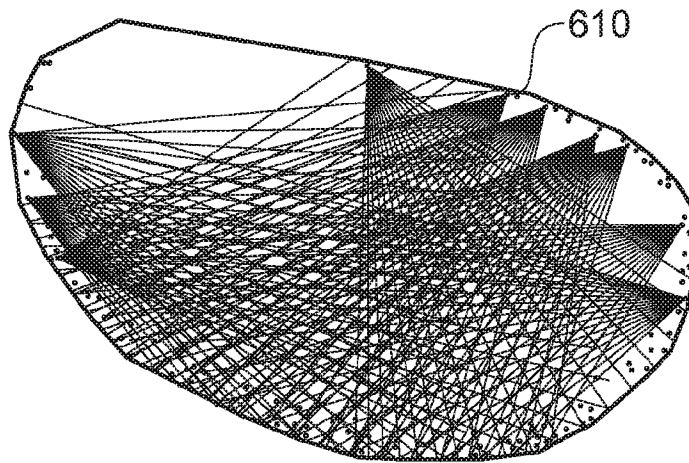


FIG. 6A

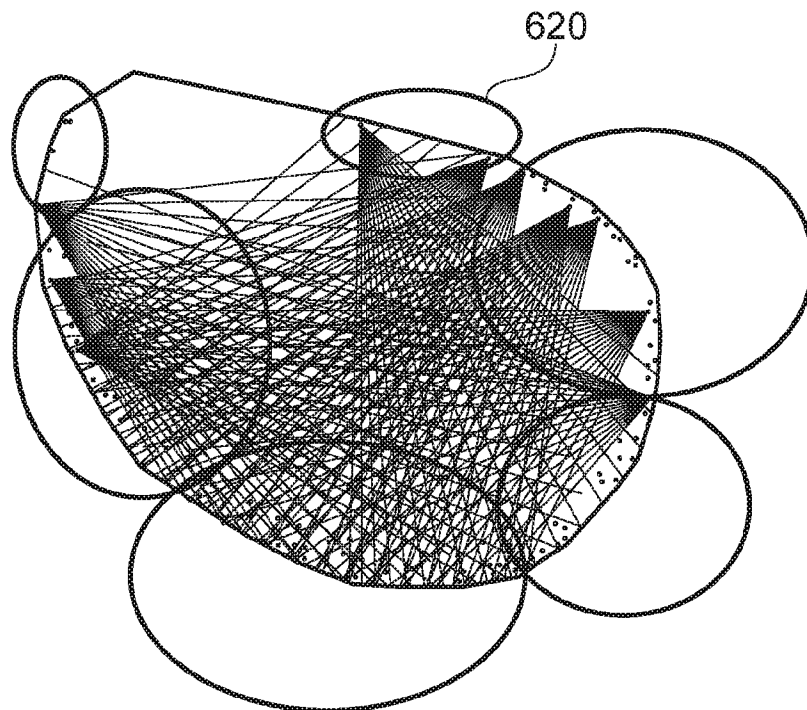


FIG. 6B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2017/059006

A. CLASSIFICATION OF SUBJECT MATTER <i>G06F 17/30 (2006.01)</i> According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06F Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatSearch (RUPTO internal), USPTO, PAJ, K-PION, Esp@cenet, Information Retrieval System of FIPS		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010/0100319 A1 (THOMAS TRINKO et al.) 22.04.2010, abstract, paragraphs [0002], [0005]-[0007], [0009]-[0013], [0023]-[0024], [0042]-[0051], [0054]-[0055], [0058]-[0065], [0069]-[0072], claims 1, 19-20, fig.1	1-6, 8-13, 15
A		7, 14
Y	US 2003/0140040 A1 (ANDREW SCHILLER) 24.07.2003, paragraphs [0002]-[0006], [0008], [0011], [0013]-[0014], [0017], [0030], [0035], claim 1	1-6, 8-13, 15
A		7, 14
A	US 2015/0379413 A1 (PALANTIN TECHNOLOGIES, INC.) 31.12.2015	1-15
A	CN 106096623 A (WANG MEIHUA et al.) 09.11.2016	1-15
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
*	Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family
“A”	document defining the general state of the art which is not considered to be of particular relevance	
“E”	earlier document but published on or after the international filing date	
“L”	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	
“O”	document referring to an oral disclosure, use, exhibition or other means	
“P”	document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search 30 August 2018 (30.08.2018)		Date of mailing of the international search report 27 September 2018 (27.09.2018)
Name and mailing address of the ISA/RU: Federal Institute of Industrial Property, Berezhkovskaya nab., 30-1, Moscow, G-59, GSP-3, Russia, 125993 Facsimile No: (8-495) 531-63-18, (8-499) 243-33-37		Authorized officer M. Demchenko Telephone No. (499) 240-25-91