

Análisis de datos de calidad del aire de la Zona Metropolitana del Valle de México mediante técnicas de agrupamiento

Pablo Camarillo-Ramírez, Abraham Sánchez-López, Luis J. Calva-Rosales,
Ivan Pérez-Vázquez

Facultad de Ciencias de la Computación,
Benemérita Universidad Autónoma de Puebla, México

{pablo.camarillo, asanchez, luis.calva}@cs.buap.mx, navip_1301@hotmail.com

Resumen. El cambio climático representa uno de los problemas a los que la sociedad actual se enfrenta, debido a sus impactos en la salud de los seres vivos. Es por ello que las autoridades necesitan de herramientas que les brinden la información necesaria para tomar decisiones que disminuyan el impacto de dicho cambio. En este trabajo, se propone una estrategia para agrupar los datos de la calidad del aire de la zona metropolitana de la Ciudad de México de los años 1999 al 2002, para reconocer los patrones de mediciones en los contaminantes del ambiente que detonan en precontingencias y contingencias ambientales en la zona de la Ciudad de México.

Palabras clave: Calidad del aire, clustering, K-means.

1. Introducción

El ser humano a lo largo de su existencia ha ido cambiando su entorno para vivir de manera cómoda y segura, prueba de ello son los grandes alcances para transportarse por cielo, mar, tierra y el espacio. Los avances tecnológicos han facilitado los hábitos cotidianos, los negocios, la fabricación de grandes cantidades de productos, etc. Sin embargo, estos avances han llevado a un deterioro ambiental que amenaza seriamente el desarrollo actual y futuro de las naciones [14], [9].

La contaminación del aire o contaminación atmosférica es un problema que produce cambios climáticos en todo el mundo y afecta a la salud de millones de personas. Es por esta razón que las herramientas tecnológicas que contribuyan en el estudio de esta contaminación son de vital importancia en la elaboración de políticas que erradiquen la contaminación o que mitiguen los efectos de la misma.

Actualmente, diversas organizaciones y gobiernos han implementado mecanismos de medición de contaminantes en el aire para conocer los índices de la

calidad del aire de las diferentes regiones del planeta. Los índices de calidad del aire (ICA) son número usados por agencias del gobierno para determinar la calidad del aire. En la ciudad de México y en la zona metropolitana del valle de México (ZMVM) la contaminación del aire es medida con el índice metropolitano de calidad del aire (IMECA). El IMECA es usado para mostrar el nivel de contaminación y el nivel de riesgo que representa a la salud humana en un tiempo determinado y así poder tomar medidas de protección.

En [14] la autora propone una aplicación de la Inteligencia de Negocios (Business Intelligence) para analizar datos de cambio climático para la zona sur del valle de Puebla. Los resultados de los procesos de Inteligencia de Negocios aplicados a los datos de la calidad del aire de la zona sur del valle de Puebla, que presenta la autora, señalan una relación muy fuerte entre la calidad del aire y las variables climáticas, también muestran que la calidad del aire respecto de los niveles de concentración de los contaminantes atmosféricos, se encuentra determinada por la presencia de partículas menores a 10 micrómetros (PM_{10}) y químicos de ozono (O_3). En la ciudad de Puebla se cuenta con 3 estaciones de medición de contaminantes, pero no se cuenta con un modelo propio que brinde información sobre la calidad del aire. En este trabajo, buscamos obtener conclusiones similares a las obtenidas por la autora del trabajo previamente descrito, aunque el enfoque de nuestra estrategia solo toma en cuenta la información de la calidad del aire para la Zona Metropolitana de la Ciudad de México.

Otro trabajo relacionado con el análisis de datos de cambio climático, es el presentado por Reyes Salazar et al. [10], en el cual se propone un algoritmo genético para agrupar los datos de cambio climático de la Zona Metropolitana del Valle de México. En este trabajo, los autores crean los *patrones* a partir de las mediciones de diversas estaciones de la región estudiada y las agrupan para determinar el tipo de contaminantes que son claves para la activación de una contingencia ambiental, según las normas de calidad del aire. Esta estrategia de agrupación resultó en la obtención de 10 grupos, en los que se pueden agrupar los datos de cambio climático, concluyendo que los patrones que representaban mayores niveles de medición en ciertos contaminantes, tales como (PM_{10}) y el ozono coinciden con mediciones altas de IMECAs.

En la estrategia propuesta en este trabajo, cada medición de contaminantes conocida es tomada como un *patrón* en el cual, los atributos del patrón, son los valores de cada contaminante y de esta forma su agrupamiento nos conducirá a conclusiones sobre la relación entre los valores de los contaminantes y la calidad del aire. La literatura ofrece diversas técnicas de agrupamiento de datos [9], sin embargo en nuestra estrategia hemos usado el método *K-means* para agrupar los datos de calidad del aire. En la sección 2 se presenta una breve reseña de los conceptos más importantes referentes al cambio climático y a las técnicas de agrupamiento, después, en la sección 3 se describe la estrategia propuesta, para que en la sección 4 se presenten e interpreten los resultados obtenidos y, finalmente 5 se presentan las conclusiones obtenidas y el trabajo futuro de esta propuesta.

2. Marco teórico

2.1. Cambio climático

El cambio climático es uno de los procesos más relevante del deterioro ambiental mencionado anteriormente. Para entender qué es el cambio climático, debemos mencionar que la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMUCC) dice que “*cambio climático*” se entiende un cambio de clima atribuido directa o indirectamente a la actividad humana que altera la composición de la atmósfera mundial y que se suma a la variabilidad natural del clima observada durante periodos de tiempo comparables. El clima es una descripción estadística de las condiciones de tiempo y sus variaciones, incluyendo condiciones promedio y extremas. Los gases de efecto invernadero juegan un rol importante en el deterioro del clima y causan el cambio climático. Los gases de efecto invernadero incluyen vapor de agua, dióxido de carbono (CO_2), metano (CH_4), óxido nitroso (N_2O) y algunos gases industriales tales como clorofluorocarbonos (CFC_s). Estos gases actúan como una manta aislante, manteniendo la superficie de la tierra más caliente al no reflejar los rayos del sol al espacio, manteniendo el calor en la superficie de la tierra más caliente de lo que debería estar si estos gases no se presentaran en la atmósfera. Una vez liberados a la atmósfera, muchos de estos gases permanecerán ahí por un largo periodo de tiempo [6,8,13].

2.2. Calidad del aire

Se entiende por contaminación atmosférica a la presencia, en la atmósfera, de sustancias en una cantidad que implique molestias o riesgo para la salud de las personas y de los demás seres vivos. La calidad del aire es medida por la concentración de los contaminantes aéreos, es decir a mayor presencia en el aire, menos es la calidad y consecuentemente mayor es el impacto en la naturaleza y los seres vivos [7].

De acuerdo a las normas mexicanas de calidad del aire, para reportar la calidad del aire, el índice emplea cinco categorías [12]. La Tabla resume las categorías de calidad del aire, a partir de las mediciones de IMECA.

Tabla 1. Categorías de calidad del aire.

Buena	Regular	Mala	Muy mala	Extremadamente mala
0 - 50	51 - 100	101 - 150	151 - 200	201 - 500

2.3. Agrupamiento de datos

Cuando se trabaja con grandes cantidades de información, es necesario agruparla para que ésta sea más fácil de manejar. Agrupar esta información incrementa la eficiencia de su administración y su interpretación [15]. Existen dos

grupos de técnicas de agrupamiento: Agrupamiento jerárquico y agrupamiento no jerárquico.

Agrupamiento jerárquico Los métodos jerárquicos generan una sucesión de particiones, donde cada partición se obtiene uniendo o dividiendo grupos. Dentro de los métodos jerárquicos se distinguen dos tipos:

1. Métodos aglomerativos. En estos, los nuevo grupos se forman uniendo los grupos existentes. La ventaja de estos métodos es su rapidez.
2. Métodos divisivos. A diferencia de los métodos aglomerativos, en estos métodos, los nuevo grupos se forman dividiendo los grupos existentes. La ventaja de estos métodos es que parten de la información global que hay en los datos que además el proceso de división no tiene por que seguir hasta que cada elemento forme un grupo.

Sin embargo, de acuerdo a [1], estos métodos suelen ser muy lentos y en general aplicables solo para datos con pocos casos.

Agrupamiento no jerárquico El agrupamiento no jerárquicos los datos se dividen en k particiones o grupos donde cada partición representa un grupo. El funcionamiento básico de los métodos de agrupamiento no jerárquico son:

1. Seleccionar K centroides iniciales, siendo K el número de grupos deseados.
2. Asignar cada patrón al grupo que le sea más cercano.
3. Reasignar o relocalizar cada observación a uno de los K grupos de acuerdo con alguna regla de paro.
4. Termina si no hay reasignaciones de los puntos o si la reasignación satisface la regla de parada. En otro caso se regresa al paso dos.

Validación de clusters: silueta de los clusters La silueta de un cluster se refiere a un método de interpretación y validación del agrupamiento de datos. La técnica provee una representación gráfica de que tan bien está situado cada objeto en su cluster. Es un método propuesto por [11]. El ancho de la silueta (silhouette width) de la i -ésima observación es definida por: $sil_i = \frac{(b_i - a_i)}{\max(a_i, b_i)}$ Donde, a_i denota la distancia promedio entre la observación i y todas las otras que están en el mismo cluster de i , y b_i denota la distancia promedio mínima de i a las observaciones que están en otros clusters.

Las observaciones con ancho de silueta grande están bien agrupadas mientras aquellas con ancho de silueta baja tienden a estar ubicada en medio de dos clusters. Para un número de clusters dado K , el ancho de silueta promedio de la configuración de conglomerados será simplemente el promedio de sil_i sobre todas las observaciones. Es decir, $\bar{s} = \frac{\sum_i sil_i}{n}$ Kaufman y Rousseeuw [3] sugirieron estimar el número óptimo de cluster K para el cual el ancho de silueta promedio sea la mayor posible.

Después de haber hecho una breve revisión de los conceptos básicos de nuestro trabajo, en la siguiente sección se presenta nuestra propuesta de análisis de datos.

3. Estrategia de agrupamiento propuesta

La estrategia de agrupamiento propuesta para los datos de cambio climático consiste en emplear el método K-means. Las instancias se forman a partir de la información de las mediciones de contaminantes de la Zona Metropolitana del Valle de México, cada instancia consiste en un vector que contiene las mediciones de seis contaminantes criterio de cada hora de cierto año (desde 1995 hasta 2002), resultando en un total de 8760 instancias por año. Estas mediciones se agrupan empleando la técnica mencionada anteriormente y se validan los grupos generados a partir de la silueta de los clusters resultantes.

3.1. Preparación de los datos

La primera fase de nuestra estrategia consiste en preparar los datos para su agrupamiento. Los datos originales consisten en un conjunto de hojas de cálculo que contienen las mediciones de diversas estaciones, sin embargo muchas de las estaciones reportan valores negativos, lo cual es imposible e indican un fallo en la estación, es por ello que en este estudio solo nos basamos en una estación, que es la que reporta menos valores incongruentes. La presencia de estos valores incorrectos se corrigió sustituyendo esos valores incongruentes por la media aritmética de los valores correctos.

3.2. Agrupamiento con K-means

El algoritmo empleado para agrupar los datos de calidad del aire, es el algoritmo K-means, que es uno de los algoritmos de agrupamiento no jerárquico más utilizado para encontrar grupos (*clusters*). Lo anterior, debido a su fácil implementación y a su rápida ejecución [2]. Este algoritmo se presentó desde los años sesentas [4] [5]. Parte un problema de m atributos, cada instancia se traslada a un espacio m -dimensional. El centroide del cluster describe cada cluster y es un punto en el espacio m -dimensional alrededor del cual cada instancia es agrupada. La distancia más usada de la instancia al centroide del cluster es la distancia Euclidiana. El algoritmo de K-means consiste en dos principales pasos:

1. El paso de **asignación** consiste en mover cada instancia a la clase más cercana.
2. El paso de **re-estimación** consiste en recalcular los centroides de las clases a partir de las instancias asignadas a cada clase(cluster).

Se repiten los dos pasos del algoritmo hasta que el paso de re-estimación produce un cambio mínimo de los centroides de las clases.

Una vez corregidos los datos, se procede a construir, de tal forma que se pueda usar el algoritmo K-means para su agrupamiento. Las instancias quedan formadas por las mediciones de los contaminantes criterio. Contaminantes criterio, es un termino usado internacionalmente para describir contaminantes aéreos que han sido regulados y son usados como indicadores de la calidad del aire. Los

contaminantes criterio son: El Ozono (O_3), el dióxido de sulfuro (SO_2), dióxido de nitrógeno (NO_2), monóxido de carbono (CO) y las partículas suspendidas (PM). Con esa información, las instancias se forman como vectores como a continuación se describe:

$$X = (X_1, X_2, X_3, X_4, X_5, X_6)$$

Donde:

- X_1 corresponde a la medición del Monóxido de Carbono (CO)
- X_2 corresponde a la medición del Dióxido de Nitrógeno (NO_2)
- X_3 corresponde a la medición del Oxido de Nitrógeno (NOX)
- X_4 corresponde a la medición del Ozono (O_3)
- X_5 corresponde a la medición de partículas suspendidas menores a 10 micrómetros (PM_{10})
- X_6 corresponde a la medición del Dióxido de Sulfuro (SO_2)

Es necesario mencionar que se agrega la medición del óxido de nitrógeno (NOX), ya que el ozono es creado por reacciones químicas entre este compuesto.

4. Experimentos y resultados

Como se menciono anteriormente, se tienen 8760 instancias por año, y son 8 años sobre los cuales se hizo el análisis de datos reportado en este trabajo. Esta información fue obtenida de la página oficial de calidad del aire del gobierno del Distrito Federal (<http://www.calidadaire.df.gob.mx/calidadaire/>). Se aplicó el algoritmo de agrupamiento K-means a cada conjunto de datos anuales, y para verificar el número de clusters encontrados el óptimo, se empleo la técnica de validación con las siluetas de los clusters. El resultado de la aplicación del algoritmo K-means a los 8 conjuntos de datos que representan las mediciones anuales de los datos de cambio climático, arrojan la información reportada en la Tabla 2.

Tabla 2. Número óptimo de grupos para los datos de calidad del aire.

Año	Número óptimo de grupos
1995	9
1996	9
1997	6
1998	6
1999	6
2000	6
2001	6
2002	6

4.1. Validación con las siluetas de los clusters

El número óptimo de grupos fue calculado a partir de la información obtenida de la silueta de cada ejecución del método K-means con diferente número de grupos. En la Tabla 3 se presentan las pruebas que se hicieron con un diferente número de grupos y el error que arroja la silueta de cada uno de ellos, para los 8 años de mediciones que se están estudiando en este trabajo. Se remarcan los resultados que presentan menores errores en las siluetas de cada prueba, con ello se justifica el número óptimo de grupos para cada medición anual. Los resultados de las pruebas muestran agrupaciones con mínimo 6 grupos y máximo 11 grupos, ya que con un número de grupos menor que 6 y mayor que 11 el error se incrementa y por fines prácticos se decidió omitir esos resultados.

Tabla 3. Errores arrojados por la silueta de cada agrupación.

Año \ Número de grupos	6	7	8	9	10	11
1995	0.5037	0.5050	0.5050	0.4943	0.4943	0.5292
1996	0.4292	0.3922	0.4190	0.3883	0.4180	0.3922
1997	0.3841	0.4585	0.4083	0.4379	0.4585	0.4319
1998	0.4372	0.4415	0.4612	0.4612	0.4612	0.4386
1999	0.4155	0.4603	0.4674	0.4793	0.5228	0.5249
2000	0.4218	0.7771	0.8226	0.8712	0.8705	0.8608
2001	0.4520	0.4628	0.5340	0.4628	0.4694	0.5340
2002	0.3124	0.3928	0.3517	0.3795	0.3131	0.3517

Una vez que se tienen agrupados los datos de la calidad del aire de cada año, desde 1995 hasta el 2002, la Figura 1, muestra las siluetas de estos clusters.

Con los grupos obtenidos por el método de K-means para cada conjunto de datos anuales, se ha relacionado la información de los contaminantes asociada a cada cluster de datos. En las Tablas 4–11 se concentra la información sobre los valores máximos y mínimos para cada contaminante criterio en cada grupo. Con esta información se puede observar una clara tendencia que la calidad del aire esta fuertemente influenciada por las partículas suspendidas menores a 10 micrómetros (PM_{10}).

5. Conclusiones

El interés por realizar la presente investigación es para contestar las siguientes preguntas: ¿Existe un patrón en los registros de cada año?, ¿Sólo un contaminante criterio se dispara por medición?, ¿Cuáles son los contaminantes que se disparan con mayor frecuencia?. Estas preguntas no pueden ser contestadas con solo tener el registro de la calidad del aire en un momento determinado, si no

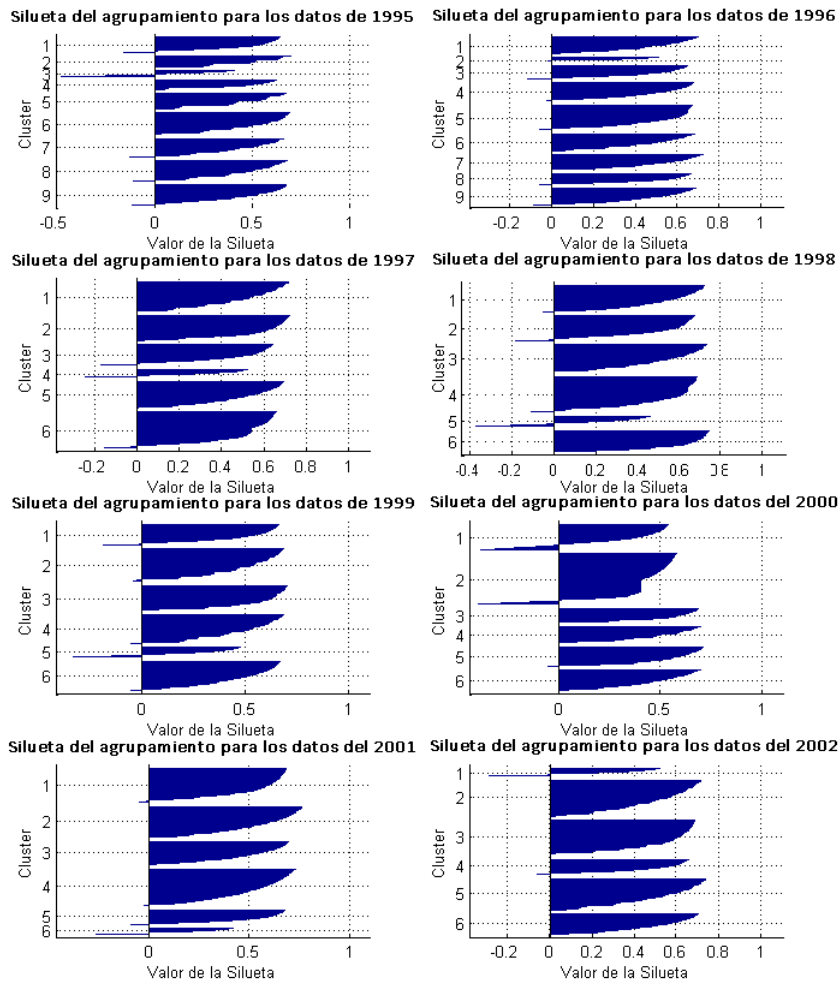


Fig. 1. Siluetas de los agrupamientos de los datos de calidad del aire anuales

Tabla 4. Valores máximos y mínimos para cada cluster I.

Datos del agrupamiento de los datos de calidad del aire de 1995							
Número de grupo \ Contaminante		CO	NO ₂	NOX	O ₃	PM ₁₀	SO ₂
1	Máximos	15.7000	0.2390	0.4860	0.2690	134.0000	0.2400
	Mínimos	0.1000	0.0030	0.0040	0.0010	91.0000	0.0010
2	Máximos	12.0000	0.1950	0.4110	0.1950	24.0000	0.1730
	Mínimos	0.1000	0.0020	0.0030	0.0010	18.0000	0.0010
3	Máximos	12.4000	0.2870	0.4800	0.2920	782.0000	0.1280
	Mínimos	0.2000	0.0100	0.0050	0.0030	134.0000	0.0010
4	Máximos	10.6000	0.1540	0.3440	0.1630	17.0000	0.1620
	Mínimos	0.1000	0.0010	0.0010	0.0010	1.0000	0.0010
5	Máximos	7.3000	0.0840	0.2310	0.1720	32.0000	0.1450
	Mínimos	0.1000	0.0020	0.0010	0.0010	25.0000	0.0010
6	Máximos	13.8000	0.1480	0.4910	0.2460	67.0000	0.1760
	Mínimos	0.1000	0.0040	0.0030	0.0010	53.0000	0.0010
7	Máximos	8.7000	0.2360	0.3560	0.2070	41.0000	0.1950
	Mínimos	0.1000	0.0010	0.0020	0.0010	32.0000	0.0010
8	Máximos	13.3000	0.1310	0.4340	0.2090	52.0000	0.1190
	Mínimos	0.1000	0.0050	0.0020	0.0010	41.0000	0.0010
9	Máximos	13.1000	0.1730	0.4110	0.2350	91.0000	0.1300
	Mínimos	0.1000	0.0020	0.0030	0.0010	68.0000	0.0010

Tabla 5. Valores máximos y mínimos para cada cluster I.

Datos del agrupamiento de los datos de calidad del aire de 1996							
Número de grupo \ Contaminante		CO	NO ₂	NOX	O ₃	PM ₁₀	SO ₂
1	Máximos	10.1000	0.1390	0.4210	0.1780	42.0000	0.1970
	Mínimos	0.1000	0.0040	0.0020	0.0010	30.0000	0.0010
2	Máximos	20.0000	0.2460	0.4250	0.2040	863.0000	0.1560
	Mínimos	0.3000	0.0110	0.0120	0.0040	239.0000	0.0010
3	Máximos	19.3000	0.3010	0.5000	0.2270	239.0000	0.1870
	Mínimos	0.2000	0.0090	0.0050	0.0020	151.0000	0.0010
4	Máximos	24.4000	0.2620	0.4920	0.2390	151.0000	0.1690
	Mínimos	0.1000	0.0090	0.0040	0.0010	104.0000	0.0010
5	Máximos	21.8000	0.2270	0.4930	0.2620	103.0000	0.1350
	Mínimos	0.1000	0.0030	0.0020	0.0010	75.0000	0.0010
6	Máximos	12.5000	0.1440	0.4440	0.2200	57.0000	0.1420
	Mínimos	0.1000	0.0050	0.0020	0.0010	42.0000	0.0010
7	Máximos	7.3000	0.1070	0.4050	0.2210	29.0000	0.1410
	Mínimos	0.1000	0.0050	0.0040	0.0020	18.0000	0.0010
8	Máximos	8.3000	0.0740	0.2340	0.1730	18.0000	0.1380
	Mínimos	0.1000	0.0020	0.0010	0.0020	1.0000	0.0010
9	Máximos	11.7000	0.1650	0.4880	0.2440	74.0000	0.1500
	Mínimos	0.1000	0.0010	0.0010	0.0010	57.0000	0.0010

Tabla 6. Valores máximos y mínimos para cada cluster I.

Datos del agrupamiento de los datos de calidad del aire de 1997								
Número de grupo	Contaminante		<i>CO</i>	<i>NO₂</i>	<i>NOX</i>	<i>O₃</i>	<i>PM₁₀</i>	<i>SO₂</i>
	1	Máximos		10.6000	0.1170	0.2110	0.2250	37.0000
Mínimos			0.1000	0.0010	0.0030	0.0010	21.0000	0.0010
2	Máximos		10.9000	0.0910	0.2890	0.2090	21.0000	0.1420
	Mínimos		0.1000	0.0010	0.0010	0.0020	1.0000	0.0010
3	Máximos		12.8000	0.1540	0.4690	0.2920	168.0000	0.2260
	Mínimos		0.2000	0.0030	0.0130	0.0010	93.0000	0.0010
4	Máximos		16.9000	0.1960	0.4960	0.2130	912.0000	0.1430
	Mínimos		0.1000	0.0040	0.0180	0.0020	169.0000	0.0020
5	Máximos		8.9000	0.1210	0.2840	0.2600	57.0000	0.1630
	Mínimos		0.1000	0.0010	0.0030	0.0010	38.0000	0.0010
6	Máximos		12.7000	0.1490	0.3790	0.2710	92.0000	0.2300
	Mínimos		0.1000	0.0010	0.0020	0.0010	57.0000	0.0010

Tabla 7. Valores máximos y mínimos para cada cluster II.

Datos del agrupamiento de los datos de calidad del aire de 1998								
Número de grupo	Contaminante		<i>CO</i>	<i>NO₂</i>	<i>NOX</i>	<i>O₃</i>	<i>PM₁₀</i>	<i>SO₂</i>
	1	Máximos		11.3000	0.1980	0.4990	0.2500	108.0000
Mínimos			0.1000	0.0020	0.0030	0.0030	67.0000	0.0020
2	Máximos		8.2000	0.1170	0.4350	0.1960	43.0000	0.1580
	Mínimos		0.1000	0.0030	0.0010	0.0030	28.0000	0.0010
3	Máximos		8.1000	0.1800	0.2530	0.1490	16.0000	0.0940
	Mínimos		0.1000	0.0030	0.0010	0.0020	0.0000	0.0010
4	Máximos		9.6000	0.1230	0.3540	0.1650	27.0000	0.1880
	Mínimos		0.1000	0.0030	0.0010	0.0030	16.0000	0.0010
5	Máximos		12.9000	0.2160	0.4960	0.2230	686.0000	0.1530
	Mínimos		0.2000	0.0040	0.0040	0.0050	109.0000	0.0020
6	Máximos		9.9000	0.1920	0.4320	0.2160	67.0000	0.1990
	Mínimos		0.1000	0.0040	0.0010	0.0030	43.0000	0.0010

Tabla 8. Valores máximos y mínimos para cada cluster II.

Datos del agrupamiento de los datos de calidad del aire de 1999							
Número de grupo	Contaminante	<i>CO</i>	<i>NO₂</i>	<i>NOX</i>	<i>O₃</i>	<i>PM₁₀</i>	<i>SO₂</i>
	1	Máximos	11.3000	0.1980	0.4990	0.2500	108.0000
Mínimos		0.1000	0.0020	0.0030	0.0030	67.0000	0.0020
2	Máximos	8.2000	0.1170	0.4350	0.1960	43.0000	0.1580
	Mínimos	0.1000	0.0030	0.0010	0.0030	28.0000	0.0010
3	Máximos	8.1000	0.1800	0.2530	0.1490	16.0000	0.0940
	Mínimos	0.1000	0.0030	0.0010	0.0020	0.0000	0.0010
4	Máximos	9.6000	0.1230	0.3540	0.1650	27.0000	0.1880
	Mínimos	0.1000	0.0030	0.0010	0.0030	16.0000	0.0010
5	Máximos	12.9000	0.2160	0.4960	0.2230	686.0000	0.1530
	Mínimos	0.2000	0.0040	0.0040	0.0050	109.0000	0.0020
6	Máximos	9.9000	0.1920	0.4320	0.2160	67.0000	0.1990
	Mínimos	0.1000	0.0040	0.0010	0.0030	43.0000	0.0010

Tabla 9. Valores máximos y mínimos para cada cluster II.

Datos del agrupamiento de los datos de calidad del aire del 2000							
Número de grupo	Contaminante	<i>CO</i>	<i>NO₂</i>	<i>NOX</i>	<i>O₃</i>	<i>PM₁₀</i>	<i>SO₂</i>
	1	Máximos	13.4000	0.2260	0.4910	0.2100	623.0000
Mínimos		0.4000	0.0040	0.0020	0.0020	78.0000	0.0030
2	Máximos	10.5000	0.1780	0.4310	0.2200	78.0000	0.3890
	Mínimos	0.1000	0.0030	0.0030	0.0010	46.0000	0.0010
3	Máximos	5.6000	0.0720	0.1520	0.1560	13.0000	0.1510
	Mínimos	0.1000	0.0010	0.0040	0.0030	0.0000	0.0010
4	Máximos	5.9000	0.0810	0.2390	0.1720	23.0000	0.1680
	Mínimos	0.1000	0.0030	0.0040	0.0020	13.0000	0.0010
5	Máximos	8.7000	0.1230	0.3760	0.1910	46.0000	0.2260
	Mínimos	0.2000	0.0030	0.0010	0.0020	34.0000	0.0010
6	Máximos	8.1000	0.0990	0.3460	0.1800	34.0000	0.3380
	Mínimos	0.1000	0.0020	0.0040	0.0010	23.0000	0.0010

Tabla 10. Valores máximos y mínimos para cada cluster II.

Datos del agrupamiento de los datos de calidad del aire del 2001								
Número de grupo	Contaminante		<i>CO</i>	<i>NO₂</i>	<i>NOX</i>	<i>O₃</i>	<i>PM₁₀</i>	<i>SO₂</i>
	1	Máximos		8.5000	0.0520	0.1750	0.1370	32.0000
Mínimos			0.1000	0.0040	0.0030	0.0010	1.0000	0.0010
2	Máximos		11.0000	0.0820	0.3210	0.1670	73.0000	0.4930
	Mínimos		0.1000	0.0040	0.0050	0.0010	53.0000	0.0010
3	Máximos		13.6000	0.0970	0.4020	0.1890	100.0000	0.4980
	Mínimos		0.6000	0.0060	0.0060	0.0020	74.0000	0.0010
4	Máximos		7.3000	0.0820	0.1860	0.1750	53.0000	0.4920
	Mínimos		0.1000	0.0040	0.0030	0.0010	33.0000	0.0010
5	Máximos		13.9000	0.1130	0.4610	0.1960	141.0000	0.4980
	Mínimos		0.9000	0.0070	0.0070	0.0020	100.0000	0.0030
6	Máximos		11.6000	0.1600	0.4130	0.1880	721.0000	0.4690
	Mínimos		1.2000	0.0100	0.0150	0.0020	142.0000	0.0040

Tabla 11. Valores máximos y mínimos para cada cluster III.

Datos del agrupamiento de los datos de calidad del aire del 2002								
Número de grupo	Contaminante		<i>CO</i>	<i>NO₂</i>	<i>NOX</i>	<i>O₃</i>	<i>PM₁₀</i>	<i>SO₂</i>
	1	Máximos		11.9000	0.1640	0.3930	0.1680	368.0000
Mínimos			0.1000	0.0050	0.0120	0.0020	97.0000	0.0020
2	Máximos		7.7000	0.1230	0.3620	0.1590	36.0000	0.1880
	Mínimos		0.1000	0.0050	0.0100	0.0010	22.0000	0.0010
3	Máximos		6.5000	0.1340	0.2980	0.1450	21.0000	0.2240
	Mínimos		0.1000	0.0040	0.0050	0.0010	0.0000	0.0010
4	Máximos		11.3000	0.1470	0.3460	0.1980	97.0000	0.2680
	Mínimos		0.2000	0.0070	0.0140	0.0010	69.0000	0.0010
5	Máximos		6.9000	0.1320	0.3110	0.1640	51.0000	0.2230
	Mínimos		0.1000	0.0020	0.0050	0.0010	37.0000	0.0010
6	Máximos		8.0000	0.1180	0.3290	0.1760	69.0000	0.1810
	Mínimos		0.1000	0.0060	0.0120	0.0010	52.0000	0.0010

que se requiere del análisis de las mediciones de calidad del aire para observar cual es el comportamiento de los datos. Y obtener las conclusiones pertinentes.

La estrategia de agrupamiento presentada en este trabajo brinda una herramienta para el análisis de datos de calidad del aire, en específico se realizaron las pruebas con la información de la calidad del aire de la Zona Metropolitana del Valle de México. El estudio realizado ha demostrado que existe una variación importante en los datos de calidad del aire de un año a otro., ya que el número de grupos varía de un año a otro, lo que con ayuda de expertos, se puede interpretar como una causa del cambio climático.

La interpretación exacta del agrupamiento obtenido en nuestro estudio no es una tarea trivial, debido a la falta de un modelo que identifique los contaminantes criterio que influyan en el aumento de los niveles de IMECAs y por consecuencia, en la declaración de una contingencia ambiental. Nuestra estrategia propone una forma de agrupar estos valores y puede ser empleada para cualquier otra región que tenga estaciones que midan los contaminantes criterio.

Referencias

1. Araujo, B.: Aprendizaje automático: conceptos básicos y avanzados : aspectos prácticos utilizando el software Weka. Pearson Prentice Hall (2006)
2. Davidson, I.: Understanding k-means non-hierarchical clustering. Tech. Rep. 02-2, State University of New York, 1400 Washington Ave., Albany 12205 (February 2002)
3. Kaufman, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. John Wiley and Sons, New York (1990)
4. Max, J.: Quantizing for minimum distortion. Information Theory, IRE Transactions on 6(1), 7–12 (March 1960)
5. McQueen, J.: Some methods for classifications and analysis of multiattribute instances. In: Proceedings of the Fifty Berkley Symposium on Mathematics, Statics and Probability Vol. 1. pp. 284–269 (1967)
6. ONU: Convención marco de las naciones unidas (1992)
7. Osorio, M.A., Torrijos, T., Sánchez, A., Arroyo, O.: Preliminary analysis for an air quality management dss in the metropolitan valley of puebla, mexico. In: Proceedings of the 13th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems, and 10th WSEAS International Conference on Non-linear Analysis, Non-linear Systems and Chaos, and 7th WSEAS International Conference on Dynamical Systems and Control, and 11th WSEAS International Conference on Wavelet Analysis and Multirate Systems: Recent Researches in Computational Techniques, Non-linear Systems and Control. pp. 210–215. MAMECTIS/NOLASC/CONTROL/WAMUS'11, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA (2011)
8. Pachauri, R., Reisinger, A.: Contribución de los grupos de trabajo i, ii y iii al cuarto informe de evaluación del grupo intergubernamental de expertos sobre el cambio climático. Cambio climático 2007 p. 104 (2007)
9. Reyes-Salazar, J.E.: Descubriendo conocimiento en bases de datos de cambio climático con técnicas de cómputo suave. Master's thesis, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico (2012)

10. Reyes-Salazar, J.E., Sanchez, A.: Analysis of air quality data in Mexico City with clustering techniques based on genetic algorithms. In: Electronics, Communications and Computing (CONIELECOMP), 2013 International Conference on. pp. 27–31 (March 2013)
11. Rousseeuw, P.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20(1), 53–65 (Nov 1987)
12. Secretaria del Medio Ambiente: Imeca (2011), disponible en: <http://www.calidadaire.df.gob.mx/calidadaire/index.php?opcion=2&opcioninfoproductos=22>
13. Semarnat: Estrategia de México ante el cambio climático global (2008)
14. Torrijos-Muñoz, M.T.: Aplicación de la tecnología de inteligencia de negocios como una estrategia de análisis que contribuya a llevar un control de la calidad del aire y de las variables climáticas, así como a proporcionar instrumentos para mitigar el cambio climático en la zona sur del valle de Puebla. Ph.D. thesis, Universidad Popular Autónoma del Estado de Puebla, Puebla, Mexico (2012)
15. Ward, J.H.: Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301), 236–244 (1963)