**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Programas de Maestría y Doctorado en Ingeniería Telemática**

**Seminario de Investigación**

**Detección de productos volcánicos basado en aprendizaje incremental**

**José Eduardo Gómez Daza**

Estudiante de Maestría

4 de septiembre de 2015

1. **Introducción**

En esta sección se presentó las principales definiciones de los temas del dominio de aplicación (vulcanología), sistemas dinámicos y de la parte ingenieril, como los es las definiciones de aprendizaje incremental y finalmente el aprendizaje supervisado.

1.1 Dominio de aplicación

**VOLCÁN**

Es un sistema natural que traslada energíadesde el interior de la tierra hacia el exterior, por medio de la transferencia de magma. La creación del magma ocurre en profundidad, por fusión de rocas que luego se mueven hacia la superficie hasta hacer erupción [1]

Se presentan además los productos volcánicos generados en un volcán activo, los cuales se muestran a continuación:

* Erupciones
* Flujos piroclásticos
* Caída de ceniza
* Emisión de gases volcánicos.
* Lluvias acidas
* Sismicidad.
* Flujos de lodo.

**Sistema dinámico**

 Se considera un sistema como una entidad formada por un conjunto de elementos, que son los componentes básicos del sistema, y por las relaciones existentes entre sí y con su entorno. Cuando el sistema involucra la variable tiempo, se le denomina sistema dinámico**,** ya que este evolucionará a lo largo del tiempo [2].

En la figura 1 se explica lo que es un sistema, y luego se describe algunos tipos de sistemas existentes, y se presentan algunos ejemplos de cada sistema, finalmente se concluye que un sistema dinámico es a su vez un sistema estocástico y un sistema abierto que varía en el tiempo y genera datos no estacionarios.



Figura 1. Tipos de sistemas.

1.2 Informática.

**Aprendizaje supervisado**

Es una técnica para deducir una funcióna partir de datos de entrenamiento. Los datos de entrenamiento consisten de pares de objetos (datos de entrada y resultados deseados). El objetivo del aprendizaje supervisado es el de crear una función capaz de predecir el valor correspondiente a cualquier objeto de entrada [3].

En la figura 2 se muestra como a partir de un dataset, que contiene atributos y una clase, es posible llegar a una función objetivo pasando el dataset a un algoritmo de aprendizaje supervisado.



Figura 2. Aprendizaje supervisado.

**Aprendizaje incremental**

Es una técnica del aprendizaje automático que permite analizar flujo de datoscontinuos o de gran tamaño, adaptándose adecuadamente a entornos dinámicos permitiendo al clasificador emitir una respuesta en cualquier momento, sin necesidad de reentrenamiento previo de los clasificadores [6].

1. **Motivación**

Se presenta un escenario de motivación, en el cual se toma en cuenta que los volcanes son monitoreados por sensores instalados en cercanías a los volcanes, dichos sensores son pertenecientes a distintas metodologías de vigilancia, tales como geoquímica, los cuales miden concentraciones de gas radón, Co2 y So2; geofísica: los cuales son sensores que miden el movimiento de la tierra a través de sismómetros, actividad superficial: miden los cambios visibles que ocurren y son monitoreados a través de cámaras digitales y sensores climáticos.

Ahora bien, estos sensores son almacenados en repositorios de información ubicados en un observatorio vulcanológico. Si se quisiera hacer uso de las técnicas de minería de datos, específicamente de aprendizaje supervisado sobre el conjunto de datos que se tiene a través del repositorio de información volcánica, y teniendo en cuenta que los datos son adquiridos en tiempo real. Lo primero que se debería hacerse es seleccionar el conjunto de datos de un lapso de tiempo, por ejemplo de los años 2005 a 2014, luego pasar todo el conjunto de datos a un algoritmo de aprendizaje supervisado, este algoritmo nos generaría un clasificador, el cual claramente estaría entrenado con datos de los años seleccionados. Si quisiéramos entonces clasificar el tipo de ceniza que está arrojando el volcán, de acuerdo a atributos como por ejemplo tipo de sismicidad generada, duración del sismo e inflación del cono volcánico, el clasificador nos retornaría como respuesta el tipo de sismicidad que está cayendo (alta, media o baja). Sin embargo, al pasar el tiempo y como se mencionó anteriormente el volcán es un sistema dinámico que varía en el tiempo, y sus condiciones físicas cambian. El clasificador pasado este tiempo no arrojara resultados precisos, obteniendo falsos positivos en sus clasificaciones, lo cual lo convertiría en obsoleto.

1. **Trabajos relacionados.**

Debido a que no se encontró trabajos relacionados con el dominio de aplicación (vulcanología) que hagan uso de aprendizaje incremental, se decidió agrupar la sección de trabajos relacionados en vulcanología (geofísica, geodesia y geoquímica) con aprendizaje supervisado y aprendizaje incremental con otros dominios de aplicación como por ejemplo salud, óptica, robótica, redes, y aplicaciones de voz y texto. Los trabajos relacionados fueron consultados de los motores de búsqueda Springger, IEEE y google scholar tal y como se observa en la figura 3.



Figura 3. Motores de búsqueda consultados.

Las figuras 4 y 5 muestran que la tendencia en aprendizaje supervisado en dominio de vulcanología está encaminada a la sismología y en cuanto algoritmos a Maquinas de vector soporte, quizá por la precisión que esta ofrece.



Figura 4. Aprendizaje supervisado en volcanología.



Figura 5. Algoritmos de aprendizaje supervisado en volcanología.

Las brechas existentes en esta sección son:

* Obsolescencia en los clasificadores.
* Conjunto de datos muy grande para reentrenar clasificadores.

Ahora bien, desde el lado del aprendizaje incremental se encuentran las gráficas 6 y 7. Donde se observa que la tendencia se encuentra en los artículos de voz y texto, sin embargo en los últimos años los trabajos relacionados con óptica están aumentando. Mientras que el algoritmo ISVM es el más utilizado para los dominios de aplicación mencionados anteriormente.



Figura 6. Otros dominios de aplicación en aprendizaje incremental.



Figura 7. Algoritmos de aprendizaje incremental en otros dominios de aplicación.

En esta sección no se encuentran brechas como tal sino que se toma como punto de partida lo encontrado en estas investigaciones, las cuales son las siguientes.

* Trabajos investigación con dominios de aplicación más populares en aprendizaje incremental.
* Familias de algoritmos más usados.
* Tomar como punto de partida las familias de algoritmos más usadas.
1. **conclusiones.**

Se presenta, como conclusiones una comparación entre aprendizaje supervisado y aprendizaje incremental.

|  |  |
| --- | --- |
| **Aprendizaje supervisado** | **Aprendizaje incremental** |
| En el sector volcánico los trabajos recientes se enfocan en: Geofísica (32/54) ya que dentro se encuentra la sismología que permite obtener un estado del nivel del volcán de manera inmediata. | Los dominios de aplicación más populares en aprendizaje incremental y los que más se está trabajando en la actualidad son en el área de óptica (9/48) y robótica (10/48). |
| Se encuentra que los algoritmos más usados en el área volcánica son:* Redes neuronales (24/54).
* Máquinas de vector soporte (12/54).

 Puede estar asociado a que estos algoritmos son de caja negra y mantienen una mejor **precisión** en **comparación a otros** algoritmos (Corrales. C ,2014). | * ISVM (23/48) el más usado junto con las redes neuronales (8/48) adaptadas para aprendizaje incremental.
* Los multi - clasificadores (7/48) incrementales son una buena opción ya que estos mejoran **la precisión de un solo** clasificador (Corrales. C ,2014).
 |
| Necesita un reentrenamiento del nuevo conjunto de datos para no ser obsoleto. |  |

1. **pregunta de investigación.**

¿Cómo detectar productos volcánicos haciendo uso de algoritmos de aprendizaje incremental?

1. **Hipótesis.**

Finalmente se presentan dos hipótesis:

* El uso de técnicas de aprendizaje incremental soluciona los problemas de obsolescencia de clasificadores en los sistemas dinámicos.
* Las redes neuronales y las máquinas de vector soporte arrojarán resultados más precisos que otros clasificadores.
1. **Referencias**
2. V. Zobin, Introduction to volcanic seismology, Mexico, 2000.
3. A. C. Orozco Mauricio, Earthquake Research and Analysis - Seismology, Seismotectonic and Earthquake Geology, 2012.
4. Van Wyk Benjamin, evaluación y prevesion de riesgos ambientales en centro america. 1996.
5. Chouet, B.A, precursory swarm of long period events at redoubt volcano, 1989
6. Kotsiantis C, machine learning, 2007.
7. Suarez. A, Estudio e implementación en MOA de nuevos algoritmos de aprendizaje incremental basados en support vector machine, 2013.