**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

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

**Relator: Ing. Jose Eduardo Gomez, estudiante de Maestría**

**Co-relator: Mgtr. Camilo Corrales,**

**Protocolante: Diego Fabian Duran Dorado, estudiante de Doctorado**

**Fecha: 30 de Octubre de 2015**

**Hora inicio: 10:10 a. m.**

**Hora fin: 11:02 a. m.**

**Lugar:** Salón de posgrado, FIET, Universidad del Cauca, Popayán

**Asistentes:**

Dr. Juan Carlos Corrales, coordinador del seminario

Mgtr. Camilo Corrales, co-relator

Estudiantes de Maestría y Doctorado en Ingeniería Telemática

Estudiantes de pregrado de la FIET

**Orden del día:**

1. Presentación a cargo del relator
2. Intervención del co-relator
3. Discusión
4. Conclusiones
5. **PRESENTACIÓN A CARGO DEL RELATOR**

**Hora de inicio: 10:10 a.m.**

**Hora de finalización: 10:35 a.m.**

El Ing. Gómez se presentó como estudiante de Maestría en Ingeniería Telemática. Además presentando al director de su trabajo, el Mgtr. Camilo Corrales.

A continuación presenta la agenda de su presentación, la cual se compone de: Contexto, Escenario de motivación, Trabajos relacionados, Pregunta de investigación, Hipótesis y Objetivos.

1. **CONTEXTO**

El Ing. Gómez inició su relatoría indicando la definición de volcán, haciendo énfasis en los productos volcánicos. Posteriormente mencionó las principales características de un sistema dinámico: sistema que físico cuyo comportamiento evoluciona con el tiempo.

Otro concepto descrito fue el de aprendizaje supervisado. Finalmente, se conceptualizó el aprendizaje incremental, el cual es una técnica del aprendizaje automático que permite analizar flujo de datos continuos 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. Además se describieron los diferentes cambios del contexto: abruptos, graduales, recurrentes y ruido.

1. **ESCENARIO DE MOTIVACIÓN**

El ing. Gómez describe el escenario como:

En la dinámica de las detecciones volcánicas, los clasificadores serán obsoletos con el transcurso del tiempo, causando precisiones erróneas en las detecciones o predicciones.

1. **TRABAJOS RELACIONADOS**

La revisión de trabajos se realizó de acuerdo a una clasificación en dos grandes grupos: aprendizaje supervisado y aprendizaje incremental.

En cuanto al aprendizaje supervisado, se indicaron las siguientes conclusiones:

* La sismología es el área en la cual se está trabajando más en la actualidad.
* Específicamente la clasificación de eventos sísmicos.
* Los núcleos temáticos de geoquímica y geodesia no han sido usados actualmente.
* Los algoritmos SVM y RNA han sido los más usados en el dominio de aplicación volcánica.

En cuanto al aprendizaje incremental, se presentó un cuadro con las características de los algoritmos incrementales. Con ayuda de expertos se definieron tres áreas de vigilancia volcánica, y éstas fueron clasificadas según las características de los algoritmos incrementales.

Entre las conclusiones se destacaron:

* Tomar como punto de partida las familias de algoritmos más usadas.
* Identificar las características más relevantes de cada algoritmo.

1. **PREGUNTA DE INVESTIGACIÓN**

La pregunta de investigación planteada fue:

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

1. **HIPÓTESIS**

El uso de técnicas de aprendizaje incremental soluciona los problemas de obsolescencia de clasificadores en los sistemas dinámicos.

1. **OBJETIVOS**

General

Construir un mecanismo que permita clasificar productos volcánicos haciendo uso de técnicas de aprendizaje incremental.

Específicos:

* Definir un conjunto de datos de entrenamiento para la clasificación de productos volcánicos.
* Adaptar algoritmo(s) incremental (es) que permitan clasificar productos volcánicos.
* Desarrollar y evaluar experimentalmente un prototipo que implemente las capacidades del sistema.

1. **CO-RELATORÍA**

**Hora de inicio: 10:35 a.m.**

**Hora de finalización: 10:40 a.m.**

El Mgtr. Corrales expresó lo siguiente:

La idea es aprovechar el vínculo que tiene Jose con el Servicio Geológico Colombiano y mirar de alguna forma cómo es el problema que va del lado del dominio de aplicación y adaptarlo al aprendizaje supervisado. Entonces usar el aprendizaje supervisado tradicional no es la mejor opción, porque por lo general estos datos se transmiten en tiempo real y la idea es tomar decisiones en la medida que ocurran cambios en los sistemas volcánicos. Entonces a partir de eso, abordamos el trabajo desde el aprendizaje incremental. La idea en esta segunda presentación es que Jose fuera definiendo cuales son los algoritmos de aprendizaje incremental más adecuado de acuerdo a las áreas de vigilancia volcánica. Con el experto de Ingeominas, se definieron cuáles son las características de las áreas de vigilancia que pueden servir. Con base en eso, se tuvo un punto de partida para definir el dataset. Aparte, lo que Jose ha hecho es realizar una revisión del estado del arte sobre el aprendizaje incremental y ha encontrado un autor el cual tiene en cuenta un grupo de criterios para clasificar los algoritmos de aprendizaje incremental. Entonces el siguiente paso era definir esas áreas de vigilancia volcánica y cuáles eran los algoritmos que nos podían servir. Entonces esa era la tarea de Jose, combinar las áreas de aprendizaje incremental con las del área de vulcanología. Y al final, él presenta los posibles algoritmos para cada una de éstas áreas. Ya cual sería el paso a seguir, como la idea es utilizar aprendizaje incremental y métodos de conjuntos, la idea es decir voy a utilizar los algoritmos en por lo menos un área de vigilancia y de alguna forma poderlos combinar. De pronto faltó aclarar más el trabajo de Jose porque se dieron muchas vueltas en lo de los algoritmos de aprendizaje incremental y relacionarlos con la parte de la vulcanología.

1. **DISCUSIÓN**

**Hora de inicio: 10:40 a.m.**

**Hora de finalización: 11:02 a.m.**

Se abre la discusión:

* Pregunta 1: en el primer objetivo dices que vas a definir un conjunto de entrenamiento. Vas a tener en cuenta ahí el aspecto dinámico de los datos? No sé si es por redacción o no lo vas a tener en cuenta.

Respuesta: no. No lo voy a tener en cuenta. Se supone que en esa parte voy a sacar los parámetros necesarios en cada área para definir mi conjunto.

* Pregunta 2. Si los datos son estáticos, cómo van a ver el funcionamiento del algoritmo incremental?

Respuesta: se supone que el primer paso del aprendizaje incremental es entrenarlo normal, como si se tratase de aprendizaje supervisado. Después habría que mirar esta parte del segundo objetivo para que tenga en cuenta el flujo de datos en tiempo real de cada una de las áreas.

La cuestión es que el flujo de datos va a ser dinámico, pero la estructura de los datos no. Entonces lo que podemos hacer es cambiar una palabra en el primer objetivo por “estructura”.

* Pregunta 2: en la motivación tienes una entradas, y en los algoritmos incrementales esa caída de ceniza se volvería una entrada? Es decir, cómo hace el sistema para irse acomodando a los cambios?

Respuesta: no. El sistema se retroalimentaría, él almacena los datos que van llegando y los va almacenando. Entonces cuando encuentra que hay un cambio entonces lo que hace es entrenar el algoritmo.

Pero cómo hace el algoritmo incremental para detectar los cambios si no hay retroalimentación.

Pero es que la ceniza no lee con un sensor, sino con los mismos tres parámetros de entrada.

* Pregunta 3: en el objetivo 2 hablas de adaptar el algoritmo, a qué se refieren con adaptar el algoritmo? La parte matemática hay que cambiarla, o se refiere a la estraucutra de los datos, que sería el primer objetivo.

Respuesta: tenemos una serie de algoritmos que pueden servir. Hay unos que sirven para determinadas cosas. La idea es que podríamos combinar clasificadores, lo que se llaman multiclasificadores, para generar un nuevo algoritmo que nos permita realizar lo que necesitamos.

Ese híbrido es algo que ya está en la literatura o creen que hay una brecha.

De hecho en las clasificaciones tenemos sus clasificadores, y no todos cumplen con todo lo que queremos.

* Pregunta 4: para este tipo de presentación falta definir cual es el problema, que marque el derrotero y defina qué se busca. Cómo está la presentación no sabía para dónde iba hasta cuando Camilo aclaró en su co-rrelatoría que se está buscando.
* Pregunta 5: con respecto a la evaluación. Se habla de unos datos del 2014 y después del 2020. No es un tiempo muy largo para llevar a cabo la evaluación?

Afortunadamente Jose tiene los expertos a la mano. Entonces ha sido fácil ir de la caracterización del problema a los algoritmos. Y luego a una captura de requisitos que nos ha permitido tener una correspondencia entre los algoritmos y la caracterización del problema. Lo interesante aquí es que es posible llegar a un estudio de caso. Lo que queremos es llegar a tener un prototipo para empezar la validación y comparar los resultados con las apreciaciones de los expertos semana tras semana según lleguen los datos. Para ir afinando el sistema.

* Pregunta 6: ya decidieron un producto o un volcán seleccionado?

Respuesta: pensamos en definirlos. Pero por ahora creemos que nos hace falta cosas por leer, pero el volcán Puracé ya está definido. Según los expertos, les interesa ese volcán.

El Dr. Juan Carlos Corrales finaliza la sesión.