

RESUMEN

El presente trabajo de investigación trata acerca del análisis de una base de datos por medio de algoritmos de aprendizaje de máquina y la selección de prototipos de los datos obtenidos de un guante electrónico traductor de señas básicas enfocado especialmente para las personas que presentan una capacidad especial, en este caso personas con discapacidad auditiva y de lenguaje; se utilizará un sistema electrónico inteligente con la capacidad de detectar un número de signos del idioma. En el sistema electrónico se usará un sensor flexible en cada dedo, mismos que se utilizan para recolectar datos (se coloca en la mano derecha de la persona para obtener información de los números del 0 al 9 en el lenguaje de señas). Los datos obtenidos se analizan a través de un esquema que involucra las siguientes etapas: Balanceo de datos con Kennard-Stone (KS), selección de prototipos con algoritmo evolutivo (CHC) y procedimiento de optimización para reducción decremental (DROP3). Consecutivamente, el algoritmo K-Nearest Neighbours (KNN) se utiliza para la clasificación numérica. Este trabajo presenta un análisis de los considerados mejores algoritmos de clasificación de prototipos (DROP3 y CHC) y así poder determinar la adecuada para nuestro conjunto de datos. El principal objetivo del trabajo es usar las nuevas tecnologías del ambiente del Big Data para desarrollar y probar un algoritmo que consiga reducir un dataset de clasificación compuesto por muchos ejemplos a pocos prototipos que los representen sin perder calidad para ayudar a los sistemas de clasificación a enfrentarse a datasets grandes.

PALABRAS CLAVES:

- **SELECCIÓN DE PROTOTIPOS**
- **LENGUAJE DE SEÑAS**
- **KNN**
- **DROP3**
- **CHC**

ABSTRACT

The present investigation deals with the analysis of a database by means of machine learning algorithms and the selection of prototypes of the data obtained from an electronic translator glove basic sign language focused especially for people who have a special ability, in this case people with hearing disabilities and language; it will be used an intelligent electronic system with the ability to detect a number of language signs. In the electronic system will be used a flexible sensor on each finger, which is used to collect data (placed on the right hand of the person to obtain information from numbers 0 to 9 in sign language). The data obtained are analyzed through a scheme that involves the following stages: Data balancing with the Kennard-Stone (KS), selection of prototypes with evolutionary algorithm (CHC) and decremental reduction optimization procedure (DROP3). Consecutively, the algorithm K-Nearest Neighbors (KNN) is used for the numerical rating. This paper presents an analysis of the considered to be the best classification algorithms of prototypes (DROP3 and CHC) and thus be able to determine the appropriate for our data set. The main objective of this work is to use new technologies of the Big Data environment to develop and test an algorithm that get reduce a dataset classification composed of many examples to a few prototypes that represent them without losing quality to help the classification systems to face large datasets.

KEYWORDS:

- **SELECTION OF PROTOTYPES**
- **SIGN LANGUAGE**
- **KNN**
- **DROP3**
- **CHC**