

Chapter 7

Classification Techniques in Data Mining: Classical and Fuzzy Classifiers

Ali Hosseinzadeh

Comprehensive Imam Hossein University, Iran

S. A. Edalatpanah

Ayandegan Institute of Higher Education, Tonekabon, Iran

ABSTRACT

Learning is the ability to improve behavior based on former experiences and observations. Nowadays, mankind continuously attempts to train computers for his purpose, and make them smarter through trainings and experiments. Learning machines are a branch of artificial intelligence with the aim of reaching machines able to extract knowledge (learning) from the environment. Classical, fuzzy classification, as a subcategory of machine learning, has an important role in reaching these goals in this area. In the present chapter, we undertake to elaborate and explain some useful and efficient methods of classical versus fuzzy classification. Moreover, we compare them, investigating their advantages and disadvantages.

INTRODUCTION

Our level of learning depends on the perfection degree of our former knowledge (Alpayden, 2010). Learning is an important human behavior and very close to artificial intelligence which makes mankind able to boost his knowledge in relation to the environment. Received information through senses, are processed by brain, extracting knowledge from the received information and further preserving them (Nilsson, 2005). Learning is the ability to improve behavior based on former experiences and observations. Thus, learning ability must be regarded as a potential tool (Mitchell, 1997; Bishop 2006; Natarajan 1991). Development of computer technologies and automatic learning techniques can result in easier, more efficient. Numerous approaches of decision making techniques exist in machine learning domain where computers decide or make suggestions for the decision. The aim of machine learning is producing smart systems with high levels of flexibility and intelligence able to extract knowledge (learning) from

DOI: 10.4018/978-1-5225-0914-1.ch007

the environment, and simulating human behavior, use former experiences in solving a problem. Machine learning is in three forms of supervised, unsupervised, and semi supervised of which we have focused on the supervised form (Nilsson, 2005; Mitchell, 2006). In recent decades, advances in data collection and conserving capabilities have resulted in high sum of information in many sciences. Advanced-databases management technology can contain different types of data. Therefore, statistical techniques and traditional management tools are not sufficient for the analysis of these data, and extracting knowledge from this amount is a true difficulty (Bishop, 2006). Data mining is an attempt to obtain useful information out of these data which has become even more significant with overdue growth of data (Jing He, 2009).

Nowadays, data mining, as a subcategory of machine learning, plays a vital role in retrieving information for the classification of large collections of textual or non-textual documents. Basically, the most important knowledge mankind has achieved is classification. Classification is a process which divides data sets into determined parts, making organizations able to discover patterns in particular problems in complex, large sets (Heikki, 1996; Han & Kamber, 2001).

Data classification is a 2-phase process. First, a model is generated based on train data sets existing in the database. Train data sets consist of records, samples, examples, and or objects which include a set of features or aspects. Second, every sample has a label of predetermined class verified in one feature named 'class level'. When 'class level' is determined, the learning phase is called supervised learning (Nilsson, 2005; Bishop, 2006). In classification process, objects are assigned to distinct classes with distinguished attributes, and are introduced as a model. Subsequently, having considered features of each category, the new object is dedicated to them, and its label and type is predicted. Some of common, important methods used nowadays for data mining in classification, supervised problems are listed below (Han & Kamber, 2001):

- Support vector machine,
- Linear regression,
- Decision tree,
- K- nearest neighbor.

In real-world environment data are irresolute and vague, and in these classifiers, error-value importance is the same for different train sample, while it should not logically be so (Zadeh, 1965; Zadeh, 1968; Baldwin, 1981; Zadeh, 1984). (Since some data are defected by noise confusions of filter,) fuzzy classification methods were proposed for this difficulty (Joachims, 2002). Since quantitative data can be classified, and these techniques can be used for detecting and eliminating noise and intruder, these methods are useful for contracting amount of data. These methods include:

- Fuzzy support-vector machine,
- Fuzzy linear regression,
- Fuzzy decision tree,
- Fuzzy nearest neighbor.

In this chapter, we aim at clarification, description, and comparison of different classical versus fuzzy classification methods of data as an important branch of machine learning.

34 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/classification-techniques-in-data-mining/171905

Related Content

Data-Driven Strategies for Enhancing Customer Retention in Moroccan Telecoms

Adil Garoheand Rachid Zammar (2024). *AI and Data Engineering Solutions for Effective Marketing* (pp. 284-298).

www.irma-international.org/chapter/data-driven-strategies-for-enhancing-customer-retention-in-moroccan-telecoms/350758

Beyond Ambient Display: A Contextual Taxonomy of Alternative Information Display

Andrew Vande Moereand Dietmar Offenhuber (2009). *International Journal of Ambient Computing and Intelligence* (pp. 39-46).

www.irma-international.org/article/beyond-ambient-display/3877

Autonomous Unmanned Aerial Vehicle for Post-Disaster Management With Cognitive Radio Communication

Raja Guru R.and Naresh Kumar P. (2021). *International Journal of Ambient Computing and Intelligence* (pp. 29-52).

www.irma-international.org/article/autonomous-unmanned-aerial-vehicle-for-post-disaster-management-with-cognitive-radio-communication/272038

IronyTR: Irony Detection in Turkish Informal Texts

Asli Umay Ozturk, Yesim Cemekand Pinar Karagoz (2021). *International Journal of Intelligent Information Technologies* (pp. 1-18).

www.irma-international.org/article/ironytr/289965

Hybrid Intelligent Diagnosis: Differentiating Oligodendroglioma and Astrocytoma Through Combined Radiology and Pathology Using DL

Bhargavi Maridu, Renugadevi R., Shareefunnisa Syedand Sajida Sultana S. K. (2024). *Artificial Intelligence Transformations for Healthcare Applications: Medical Diagnosis, Treatment, and Patient Care* (pp. 233-257).

www.irma-international.org/chapter/hybrid-intelligent-diagnosis/351608