**A Fuzzy Approach to Text Classiﬁcation With Two-Stage Training for Ambiguous Instances**

**Abstract:**

Sentiment analysis is a very popular application area of text mining and machine learning. The popular methods include support vector machine, naive bayes, decision trees, and deep neural networks. However, these methods generally belong to discriminative learning, which aims to distinguish one class from others with a clear-cut outcome, under the presence of ground truth. In the context of text classiﬁcation, instances are naturally fuzzy (can be multilabeled in some application areas) and thus are not considered clear-cut, especially given the fact that labels assigned to sentiment in text represent an agreed level of subjective opinion for multiple human annotators rather than indisputable ground truth. This has motivated researchers to develop fuzzy methods, which typically train classiﬁers through generative learning, i.e., a fuzzy classiﬁer is used to measure the degree to which an instance belongs to each class. Traditional fuzzy methods typically involve generation of a single fuzzy classiﬁer and employ a ﬁxed rule of defuzziﬁcation outputting the class with the maximum membership degree. The use of a single fuzzy classiﬁer with theabove-ﬁxed ruleof defuzziﬁcation is likely to get the classiﬁer encountering the text ambiguity situation on sentiment data, i.e., an instance may obtain equal membership degrees to both the positive and negative classes. In this paper, we focus on cyberhate classiﬁcation, since the spread of hate speech via social media can have disruptive impacts on social cohesion and lead to regional and community tensions. Automatic detection of cyberhate has thus become a priority research area. In particular, we propose a modiﬁed fuzzy approach with twostage training for dealing with text ambiguity and classifying four types of hate speech, namely, religion, race, disability, and sexual orientation—and compare its performance with those popular methods as well as some existing fuzzy approaches, while the features are prepared through the bag-of-words and word embedding feature extraction methods alongside the correlationbased feature subset selection method. The experimental results show that the proposed fuzzy method outperforms the other methods in most cases.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows XP/UBUNTU.
* Implementation : NS2
* NS2 Version : 2.28
* Front End : OTCL (Object Oriented Tool Command  Language)
* Tool : Cygwin (To simulate in Windows OS)