Support Fuzzy-Set Machines
From Kernels on Fuzzy Sets to Machine Learning Applications

Jorge. Guevara
IBM Research
Sao Paulo, Brazil
jorgegd@br.ibm.com

Roberto Hirata Jr.
University of Sao Paulo
Sao Paulo, Brazil
hirata@ime.usp.br

Stéphane Canu
INSA de Rouen
Rouen, France
scanu@insa-rouen.fr

1 Abstract

This work introduces a new class of kernel machines: the *Support fuzzy-sets machines*. This machines can be used to solve machine learning tasks, like classification, regression or clustering, on data with point-wise uncertainty. We advocate the use of fuzzy set for modeling the uncertainty around the vicinity of observations, and for incorporating those uncertainties into the learning machine. Support fuzzy-sets machines are defined by kernel gram (covariance) matrices defined by kernel on fuzzy sets, which are special kind of real-valued (kernel) functions whose domain is the set of all fuzzy sets, i.e., $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, where $\mathcal{X}$ is a fuzzy set. Under fuzzy set modeling, such kernels can be used to estimate covariance matrices for observations with point-wise uncertainty and for estimating similarity measures for that kind of data. Previous research showed in fact, an improved performance in learning machines when is considered the information given by the neighborhood around observations, see for example local learning [Bottou and Vapnik, 1992], Vicinal kernels [Vapnik, 1995], Vicinal risk minimization [Chapelle et al., 2001] and the RBF network. More recent approaches consider kernel machines defined by kernels on probability measures [Muandet et al., 2012]. Support Fuzzy-set Machines thanks to the reproducing theorem of kernel methods learn $f = \sum_i \alpha_i k(X,)$ in a high dimensional Hilbert space called Reproducing Kernel Hilbert Space, where the *support fuzzy-sets* is the set given by all the fuzzy sets such the correspondent $\alpha_i > 0$. Several positive definite kernels on fuzzy sets can be used for training the proposed kernel machines, as for example: **The cross product kernel on fuzzy sets** [Guevara et al., 2017], **the intersection kernel on fuzzy sets**, [Guevara et al., 2014], **the non-singleton kernel on fuzzy sets**, [Guevara et al., 2013] or **the Distance-based kernels on fuzzy sets** ([Guevara et al., 2015]).

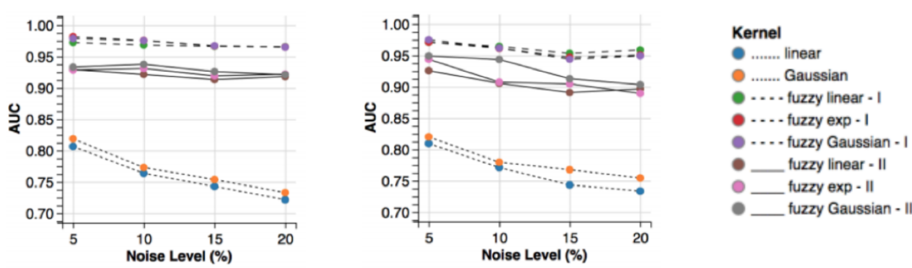
Ontic and epistemic interpretation of uncertainty by fuzzy sets Fuzzy sets (FS) were introduced by Lotfi A. Zadeh in 1965 [Zadeh, 1965], those sets differentiate from classical sets because they have a L-valued characteristic function, where L is a complete lattice. For instance, if L is the unit interval, one can define a *degree of membership* between that interval for each element in the set. In that sense, a fuzzy set X is completely characterized by its membership function:$X : \Omega \rightarrow [0, 1]$, where $X(x)$ for some $x \in \Omega$ is interpreted as the *degree of membership* of x to the fuzzy set with that membership function. Fuzzy sets are used to model *uncertainty* in observational data, based on either *ontic* or *epistemic* interpretation. Ontic, in the sense that point-wise uncertainties can be treated as entities, e.g. modeling set-valued predictors in data. Epistemic, in the sense that a fuzzy set is a model for incomplete information on single-valued predictors, i.e., a model for non-precise data. Ontic interpretation enables to say that fuzzy sets are elements with some underlying probabilistic law and that they are the realization of fuzzy-valued random variables [Kwakernaak, 1978]. Examples of ontic point of view fuzzy sets modeling are a region within a gray-scale image, a frequency profile, fuzzy clusters, a convolutional kernel on deep learning, etc. From the epistemic point of view, FS can be used to model a region within images describing the no well-known location of an object, for example, a statement describing the (unknown) age of a person, a nested set of intervals containing some unknown deterministic value [Hüllermeier, 2005, Dubois, 2011].

Methodology for constructing a Support Fuzzy-Set Machine In practical ML tasks like classification or regression, membership functions for predictors can be constructed from expert's knowledge or using data-driven approaches (from histograms or quantile functions, for example) without assuming any probabilistic law for the data generation process. Then, a kernel on fuzzy sets can be selected for estimating the covariance matrix among the fuzzified version of the data. Finally, a kernel machine algorithm, like Gaussian process regression or support vector machine can be used without modification with the kernel on fuzzy sets.

Dataset	%Noise
pima-5an-nn	5%
pima-10an-nn	10%
pima-15an-nn	15%
pima-20an-nn	20%
pima-5an-nc	5%
pima-10an-nc	10%
pima-15an-nc	15%
pima-20an-nc	20%

Figure 1: Experimental datasets

Supervised classification experiments on data with noise in the predictors We used eight datasets, listed in Figure 1, from Keel repository, a percentage of the predictor values of those datasets were corrupted by noise, the percentage is indicated by the substring "5", "10", "15" or "20" in the dataset's name. the first four datasets with suffix "nn" indicate that the noise corruption occurs in for both training and testing values. the last four datasets with suffix "nc" indicate that only the training data was corrupted by noise. We used two "fuzzification" approaches for constructing fuzzy sets for this data: the first one was given by estimating Gaussian membership functions for each predictor, we learn the parameter of those fuzzy sets from the values of each predictor. The second approach defines a fuzzy set for each predictor given by a function that interpolates the empirical distribution of the values of each predictor. Using the estimated membership functions, we estimated a second data set of membership degrees, such a dataset gives information about the certainty of each predictor value on the original dataset. As the nature of the problem was a classification problem, we define a support fuzzy-set machine as being a support vector machine with a cross product kernel on fuzzy sets. Figures 1 shows the experimental results in terms of the Area Under the Curve (AUC) metric estimated by nested cross-validation. Figure a) shows the results for datasets with the suffix "nn" and Figure b) shows results for datasets with suffix "nc". The two worst performers are an SVM with linear kernel (light blue dots) and an SVM with RBF kernel (orange dots). It is possible to observe that when the noise level increases the AUC decrement for those SVMs. On the other side, support fuzzy-set machines are the best performers (see [Guevara et al., 2017] for a detailed description of the kernels used on this machines). The suffix "I" or "II" on the legend of Figure 1 indicates whether the first or second fuzzification was used for estimating the fuzzy sets used by a particular kernel on fuzzy sets. It is possible to observe



(a) (b)

that a support fuzzy-set machine is more noise resistant in contrast with a classical support vector machine for this task. All the kernels on fuzzy sets used in this experiment show a noise resistance property.

Discussion and conclusion Support fuzzy-set machines bring a way to include and model point-wise uncertainty into kernel machines. Those learning machines use a kernel on fuzzy sets for estimating a (covariance) similarity matrix between fuzzy samples. Experimental results on noisy datasets show the potential use of those learning machines for dealing with noisy corrupted datasets.

References

Léon Bottou and Vladimir Vapnik. Local learning algorithms. *Neural computation*, 4(6):888–900, 1992.

Olivier Chapelle, Jason Weston, Léon Bottou, and Vladimir Vapnik. Vicinal risk minimization. In *Advances in neural information processing systems*, pages 416–422, 2001.

Didier Dubois. Ontic vs. epistemic fuzzy sets in modeling and data processing tasks. In *IJCCI (NCTA)*, page 13, 2011.

J. Guevara, R. Hirata, and S. Canu. Kernel functions in takagi-sugeno-kang fuzzy system with nonsingleton fuzzy input. In *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pages 1–8, 2013. doi: 10.1109/FUZZ-IEEE.2013.6622409.

J. Guevara, R. Hirata, and S. Canu. Positive definite kernel functions on fuzzy sets. In *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pages 439–446, July 2014. doi: 10.1109/FUZZ-IEEE.2014.6891628.

J. Guevara, R. Hirata, and S. Canu. *Fuzzy Set Similarity using a Distance-Based Kernel on Fuzzy Sets*. 2015.

J. Guevara, R. Hirata, and S. Canu. Cross product kernels for fuzzy set similarity. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, July 2017. doi: 10.1109/FUZZ-IEEE.2017.8015459.

Eyke Hüllermeier. Fuzzy methods in machine learning and data mining: Status and prospects. *Fuzzy sets and Systems*, 156(3):387–406, 2005.

Huibert Kwakernaak. Fuzzy random variables—i. definitions and theorems. *Information Sciences*, 15(1):1–29, 1978.

Krikamol Muandet, Kenji Fukumizu, Francesco Dinuzzo, and Bernhard Schölkopf. Learning from distributions via support measure machines. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 10–18. 2012.

Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995. ISBN 0-387-94559-8.

Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.