Modification of soft connectives in machine learning models

V. S. Kalnitsky

St.-Petersburg state university

Motivations

In 1970, Bellman and Zadeh (Bellman, Zadeh, 1970) suggested using fuzzy set theory to solve multicriteria choice problems. Criteria reduction is considered as an operation with fuzzy sets using analogues of set-theoretic operations on them. The decision maker chooses an operation on fuzzy sets after evaluating his potential actions. In this work, Bellman and Zadeh use the minimum and maximum functions as the AND and OR connectives corresponding to the intersection and union of goals. These connectives later became known as rigid. However, experiments have shown that this preset does not always reflect the behavior of the decision maker (Thole, Zimmermann, Zysno, 1979). This circumstance gave impetus to the search for alternative connectives. The use of the latter has fully justified itself, for example, when describing complex linguistic categories (Zimmermann, Zysno, 1983). It was shown that logical connectives with a compensation effect are more effective from the point of view of control. Such connectives are the average values between the maximum and minimum and are called soft connectives.

Summing up the study of averaging operations, Zimmerman lists the most studied generalized averages in terms of applications (Zimmerman, 2010): arithmetic and geometric means, symmetric sum and difference, “fuzzy and” and “fuzzy or”, compensatory AND and OR, convex combinations of maximum and minimum, of products and sums, OWA-operators.

Bobyr (Bobyr, 2018) justified the use of soft connectives in machine learning models and presented the results of computer experiments. The use of a hard minimum in traditional fuzzy inference algorithms does not allow obtaining satisfactory indicators of model accuracy. The main reason is that the minimum has a specific nature of the value dependence on the argument. When the critical value is crossed, the uniform increase in values is abruptly replaced by a loss of dependence. Due to this, dead zones of the ML model are formed.

There are two standard tricks to solve this problem. The first one is to decrease the angle between the tangents at the break point (for the minimum function, this angle is 45^o). The second one is the smoothing of the function. However, there are two other circumstances that negatively affect the quality of the model using soft connectives: possible violation of monotonicity in arguments and non-associativity. The latter is inevitable, since all smooth averages are non-associative (Aczel, 1969). So, the main problem of soft computing is that there is no monotonic, smooth, and associative soft connectives. The classical approach to solving this problem is to use a one-parameter family of soft connectives approximating the minimum. The optimal parameter for the accuracy of the ML model is selected experimentally.

The general disadvantage of the used approximations is the impossibility of independent influence on the behavior near the smoothing point and on the measure of non-associativity. We suggest using a spline to smooth the minimum with the smallest possible deviation from associativity.

Method

Definition. A soft signum is a family of functions $\varphi_\delta(z)$ converging pointwise to the function $sgn(z)$ as $\delta \to 0$ on the interval $[-1, 1]$.

Definition. A soft minimum is a function of the form

$$min_\delta(x, y) = \frac{x + y - (x - y)\varphi_\delta(x - y)}{2}$$

When choosing the optimal parameter, two trends collide. On the one hand, softening eliminates a sharp change in the growth rate of values, but on the other hand, a soft minimum is increasingly different from an associative operation. The latter becomes a negative factor for the learning process. The smaller the angle between the tangents, the greater the difference. The optimal parameter δ is the compromise between two tendencies.

In an attempt to get closer to the signum, reducing the non-associativity, one has to increase the curvature of the mean value at the diagonal point. The reason is the particular choice of approximation in the form of smooth functions of a fixed type. In these classes of functions, there is no possibility to improve one factor by fixing another. For this to be possible, the value of the derivative at zero should not affect the behavior of the generalized signum outside some neighborhood of zero. We suggest using the spline method. The soft signum must be an odd C^1-function and consist of two splines on the interval $[0, 1]$. The first spline is responsible for the curvature, the second one generates a connective that differs little from the associative one. As a parameter, we choose the value of the derivative of the soft signum at zero. In addition, it is necessary to guarantee the commutativity, monotonicity, and pointwise convergence to signum of the generated connectives.

Definition. A function φ_k is k-admitted, $k > 0$, if the properties hold

1. $\varphi_k \in C^1[-1, 1]$.
2. There exists $z_0 \in [0, 1]$ such that $\varphi(z) = kz$ for $z \in [-z_0, z_0]$.
3. For all $z \in [-1, 1]$ holds $\varphi_k(z) + z \cdot \varphi'_k(z) \le 1$.

Problem. To find k-admitted function φ_{k_0} for which the measure of non-associativity

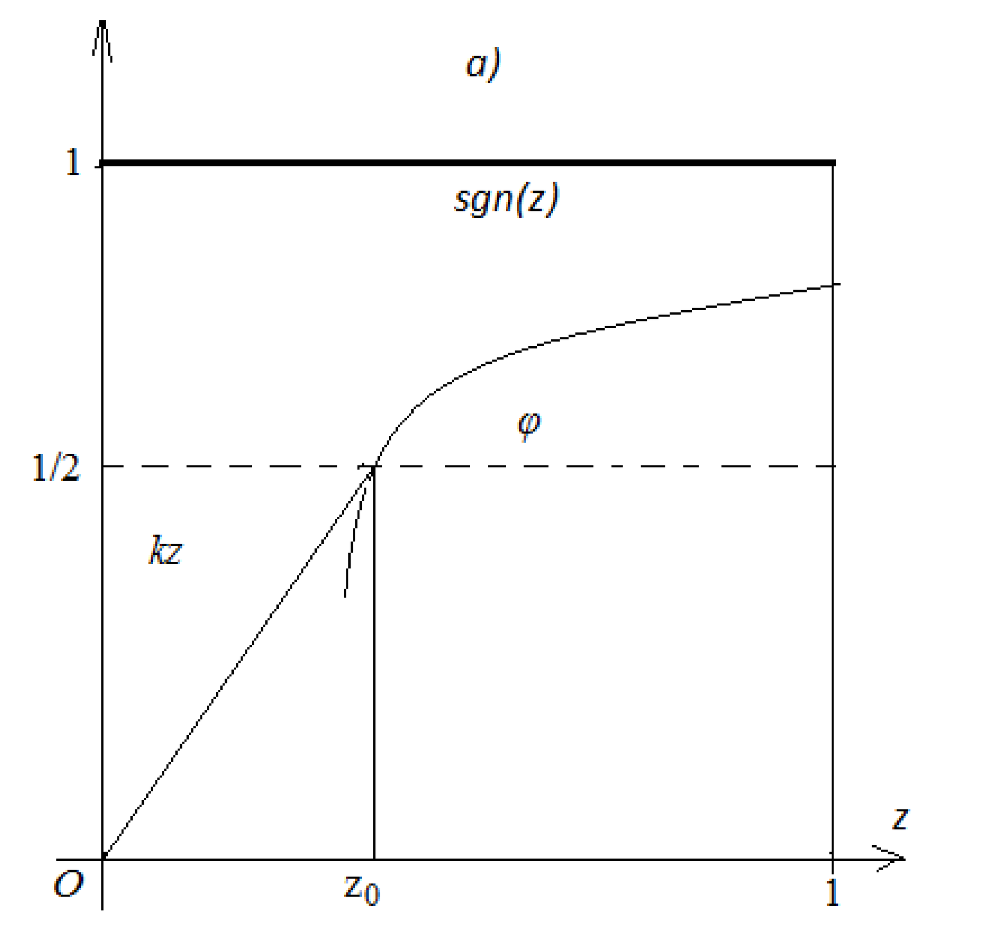
$$\Delta(\varphi_k) = \int_0^1 |1 - \varphi_k(z)| dz$$

has minimal rate. As k increases, the function φ_{k_0} must converge pointwise to the signum.

In this formulation, the problem has an exact solution.

$$sgn_k^0(z) = \begin{cases} kz, z \in \left[0, \frac{1}{2k}\right], \\ 1 - \frac{1}{4kz}, z \in \left[\frac{1}{2k}, 1\right]. \end{cases}$$

Spline-solution



Comparison

Soft sgn type	Rate of $\Delta(\varphi)$	Non-associativity	Monotonicity
Pegat, 2009	$0.5/k + o(1/k)$	almost all triples	yes
Bobyr, 2015	$1/k + o(1/k)$	almost all triples	no
Berenji, 1992	$0.69/k + o(1/k)$	almost all triples	no
Spline-solution	$0.37/k + o(1/k)$	almost 2/3 of triples	yes

Conclusion

In the presented study, a new approach was proposed for constructing connective approximations for soft computations in machine learning. Instead of using smooth functions, we suggested using spline approximations. Thanks to this, the problem of separating the negative factors that worsen the accuracy of ML-models was solved. Such factors include curvature, non-monotonicity and non-associativity of soft connectives. We have built a spline-connective that has a vast area of exact associativity, which was not previously noticed by specialists. The described model can be improved in several ways. Due to the fact that the weight of the curvature factor at the critical point is not known, one can try to design the experiment with zero curvature. To do this, it suffices to specify on the segment $[-z_0, z_0]$ not a linear dependence, but a monomial of odd degree, for example, kz^3. In this case, for any k the curvature will remain zero as the rate of decrease of deviation will drop. What will be the optimal value of the parameter in this case, needs to be set experimentally. We can forego the smoothness of a continuous soft signum outside zero, which will improve the deviation, but form the original problem of a sharp jump in the dependence on the argument in the neighborhood of the diagonal. Alternatively, you can keep the break at the critical point by allowing the soft spline-signum to break at zero, greatly reducing the deflection. In all proposed variants of modifications, spline-signums have more advantageous positions in comparison with regular functional classes of high smoothness.

References

Zadeh, L. A. (1965). Fuzzy sets. Information Control, 8, 338-353.
Zimmermann H.J., Zysno P. (1983). Decisions and evaluations by hierarchical aggregation of information. Fuzzy Sets and Systems, 10(3), 243-260.
Bellman R.E., Zadeh L.A. (1970) Decision-making in a fuzzy environment. Management Science, 17, B141-B164.
Thole U., Zimmermann H.J., Zysno P. (1979). On the suitability of minimum and product operators for the intersection of fuzzy sets. Fuzzy Sets and Systems, 2, 167-180.
Aczel J. (1969). Lectures on Functional Equations and Their Applications. Academic Press, New York.
Bobyr M.V., Milostnaya N.A. (2015). Analysis of the use of soft arithmetic operations in the structure of fuzzy logic inference. Bulletin of Computer and Information Technologies.
Bobyr M.V. (2018). Nonlinear learning method for a neuro-fuzzy inference system. Artificial intelligence and decision making, 1, 66-74.
Pegat A. (2009). Fuzzy modeling and control. M.: Binom. Knowledge Lab.
Berenji H. R., Khedkar P. (1992) Learning and tuning fuzzy logic controllers through reinforcements. IEEE Transactions on Neural Networks, 3(5), 724–740.

Contact: st006987@spbu.ru