

On obtaining intuitionistic fuzzy estimates for similarity of strings

Peter Vassilev and Todor Stoyanov

Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences, Acad. G. Bonchev Str.
Bl. 105, Sofia-1113, Bulgaria
e-mail: peter.vassilev@gmail.com

Received: 28 June 2016

Revised: 21 July 2016

Accepted: 25 July 2016

Abstract: In this paper we introduce a conceptual approach for utilizing the apparatus of intuitionistic fuzzy sets in the problem of approximate string matching. We briefly discuss the general ideas and end with a particular example of implementation which utilizes orderings between intuitionistic fuzzy pairs.

Keywords: Intuitionistic fuzzy sets, strings, approximate matching, similarity.

AMS Classification: 03E72.

1 Introduction

Given an alphabet Σ and two strings of *similar* length x and y , defined over it, further we will try to provide an intuitionistic fuzzy estimate of how similar they are. We will stay clear of the typical use of Levenshtein distance, instead preferring another approach to define similarity. In order to do so we require two things. The first is a surjective mapping s , defined for all strings over Σ . There are in fact many ways to choose such mapping and further we plan to investigate the viability of these that preserve information about the ordering of the symbols in the strings. However, for simplicity in this paper we will consider only mappings invariant with respect to the ordering. Next we require an intuitionistic fuzzy estimate which will tell us how similar and how different are the surjective images of the string. In other words we transform the couple (x, y)

to $(s(x), s(y))$ and we give an intuitionistic fuzzy estimate of the type $(\mu_{s(x),s(y)}, \nu_{s(x),s(y)})$ which we assume can be understood as $(\mu_{x,y}, \nu_{x,y})$. Having outlined the essence of the approach we will continue with some preliminary definitions and notions we will require further.

2 Preliminaries

Further we remind some basic definitions.

Definition 2.1 (cf. [2]). *An intuitionistic fuzzy pair (IFP) is an ordered couple of real non-negative numbers $\langle a, b \rangle$, with the constraint:*

$$a + b \leq 1. \quad (1)$$

This concept is very important in practice since many methods implementing IFSs lead to estimates in the form of IFPs. One way to measure which of two IFPs is better is by using an ordering:

Definition 2.2 (cf. [1, 2]). *Given two IFPs: $u = \langle u_1, u_2 \rangle$ and $v = \langle v_1, v_2 \rangle$, we say that u is less or equal to v , and we write:*

$$u \leq v,$$

iff

$$\begin{cases} u_1 \leq v_1 \\ u_2 \geq v_2 \end{cases}. \quad (2)$$

Remark 2.1. *It is obvious that the above is a partial ordering, since it is transitive, reflexive and antisymmetric but there exist u and v , for which conditions (2) are not satisfied.*

Remark 2.2. *An equivalent form of (2) is:*

$$\begin{cases} u_1 \leq v_1 \\ 1 - u_2 \leq 1 - v_2 \end{cases}. \quad (3)$$

A formulation of the necessary and sufficient condition for the fulfillment of (2) is made obvious by (3) (due to (1)), namely:

$$\begin{cases} \min(u_1, v_1, 1 - u_2, 1 - v_2) = u_1 \\ \max(u_1, v_1, 1 - u_2, 1 - v_2) = 1 - v_2 \end{cases}.$$

Another partial ordering is the following:

Definition 2.3 (cf. [3]). *Given two IFPs $u = \langle u_1, u_2 \rangle$ and $v = \langle v_1, v_2 \rangle$, we say that u is (definiteness-based) less or equal to v , and we write:*

$$u \preceq_{\sigma} v,$$

iff

$$\begin{cases} (u_1 + u_2)u_1 \leq (v_1 + v_2)v_1 \\ (u_1 + u_2)u_2 \geq (v_1 + v_2)v_2. \end{cases} \quad (4)$$

Further we will introduce some definitions about strings, which will be used later.

Definition 2.4. We say that two non-empty strings x and y are of similar length, iff

$$|l(x) - l(y)| \leq k, \quad (5)$$

where k is a preliminary chosen integer constant, and $l(\cdot)$ denotes the length of a string.

Definition 2.5. For a given string x over Σ , we say that Σ_x is an alphabet induced by the string if it contains exactly the different symbols occurring in x .

Remark 2.3. We note that Σ_x may coincide with Σ_y , when $x \neq y$. Indeed, if we take the two strings “aaabbbbc” and “abc” – both strings induce the same alphabet, namely $\{a, b, c\}$.

3 Construction of intuitionistic fuzzy estimate for the similarity of two strings

Let us be given two strings x and y , such that they are similar in length. If the alphabets induced by them are also similar in size, we can define the following intuitionistic fuzzy estimate for the similarity of the two strings.

Definition 3.1. For any non-empty two strings x and y , we define the intuitionistic fuzzy estimate

$$\left\langle \frac{|\Sigma_{xy}|}{\max(|\Sigma_x|, |\Sigma_y|)}, \frac{\min(|\Sigma_x|, |\Sigma_y|) - |\Sigma_{xy}|}{\max(|\Sigma_x|, |\Sigma_y|)} \right\rangle, \quad (6)$$

where $\Sigma_{xy} = \Sigma_x \cap \Sigma_y$, and $|\cdot|$ denotes the number of elements of the set.

We will show that the above definition is correct, i.e., that it produces an IFP (see Definition 2.1). Indeed, it is obvious that the sum of the two components is always less than or equal to 1, since

$$\min(|\Sigma_x|, |\Sigma_y|) \leq \max(|\Sigma_x|, |\Sigma_y|).$$

We only have to show that $\min(|\Sigma_x|, |\Sigma_y|) - |\Sigma_{xy}| \geq 0$. Since, we have $\Sigma_{xy} \subseteq \Sigma_x$ and $\Sigma_{xy} \subseteq \Sigma_y$, this implies $|\Sigma_{xy}| \leq |\Sigma_x|$ and $|\Sigma_{xy}| \leq |\Sigma_y|$. Hence, $|\Sigma_{xy}| \leq \min(|\Sigma_x|, |\Sigma_y|)$, i.e., the definition is correct.

Corollary 3.1. For any non-empty two strings x and y , we can define the intuitionistic fuzzy estimate

$$\langle \mu(x, y), \nu(x, y) \rangle = \left\langle \frac{|\Sigma_{xy}|}{\max(l(x), l(y))}, \frac{\min(l(x), l(y)) - |\Sigma_{xy}|}{\max(l(x), l(y))} \right\rangle, \quad (7)$$

where $\Sigma_{xy} = \Sigma_x \cap \Sigma_y$, and $|\cdot|$ denotes the number of elements of the set.

Theorem 3.1. Given three strings over Σ , x, y, z , such that $l(y) = l(z)$, we have that either

$$\langle \mu(x, y), \nu(x, y) \rangle \preceq_{\sigma} \langle \mu(x, z), \nu(x, z) \rangle,$$

or

$$\langle \mu(x, z), \nu(x, z) \rangle \preceq_{\sigma} \langle \mu(x, y), \nu(x, y) \rangle.$$

Proof. Since $l(y) = l(z)$, we have $\max(l(x), l(y)) = \max(l(x), l(z)) = M$ and $\min(l(x), l(y)) = \min(l(x), l(z)) = m$. Hence, we obtain:

$$(1 - \mu(x, y), \nu(x, y)) = (1 - \mu(x, z), \nu(x, z)) = \frac{M - m}{m}. \quad (8)$$

Therefore, following Definition 2.3,

$$\langle \mu(x, y), \nu(x, y) \rangle \preceq_{\sigma} \langle \mu(x, z), \nu(x, z) \rangle$$

is equivalent to $|\Sigma_{xy}| \leq |\Sigma_{xz}|$ and

$$\langle \mu(x, z), \nu(x, z) \rangle \preceq_{\sigma} \langle \mu(x, y), \nu(x, y) \rangle$$

is equivalent to $|\Sigma_{xy}| \geq |\Sigma_{xz}|$, which proves the Theorem. \square

Corollary 3.2. *Given a string x and equal in length strings y_1, \dots, y_n , we can order them with respect to the definiteness ordering (which implies the classical ordering between IFPs), so that the greatest estimate will give us the best possible match with x .*

4 Conclusion

We have provided a preliminary overview of a possible way for constructing intuitionistic fuzzy estimates showing the similarity between two strings. In future work the methods will undergo further improvement and refinement (in particular, with respect to the choice of surjective mapping and inclusion of the information regarding the ordering of the symbols in the strings).

Acknowledgment

Work presented here is supported by the Young Scientists Support Program of the Bulgarian Academy of Sciences, Grant No. DFNP-123/2016 “Finding approximate matching of string or parts of biomedical signals with defined models with the help of intuitionistic fuzzy estimates”.

References

- [1] Atanassov, K. T. (2012) *On Intuitionistic Fuzzy Sets Theory*. Springer, Berlin.
- [2] Atanassov, K., Szmidt, E., & Kacprzyk, J. (2013) On Intuitionistic Fuzzy Pairs, *Notes on Intuitionistic Fuzzy Sets*, 19(3), 1–13.
- [3] Vassilev, P. A note on new partial ordering over intuitionistic fuzzy sets. *Annual of “Informatics” Section, Union of Scientists in Bulgaria*, 8, 2015 (to appear).