Distribution-Based Behavioural Distance for Nondeterministic Fuzzy Transition Systems

Hengyang Wu and Yuxin Deng

***Abstract*—Modal logics and behavioural equivalences play an important role in the specification and verification of concurrent systems. In this paper, we first present a new notion of bisimulation for nondeterministic fuzzy transition systems, which is distribution-based and coarser than state-based bisimulation appeared in the literature. Then, we define a distribution-based bisimilarity metric as the least fixed point of a suitable monotonic function on a complete lattice, which is a behavioural distance and is a more robust way of formalising behavioural similarity between states than bisimulations. We also propose an on-the-fly algorithm for computing the bisimilarity metric. Moreover, we present a fuzzy modal logic and provide a sound and complete characterization of the bisimilarity metric. Interestingly, this characterization holds for a class of fuzzy modal logics. In addition, we show the non-expansiveness of a typical parallel composition operator with respect to the bisimilarity metric, which makes compositional verification possible.**

***Index Terms*—Fuzzy transition system, Bisimulation, Behavioural distance, Modal logic, Non-expansiveness.**

I. Introduction

Modal logics and behavioural equivalences are very important for the specification and verification of concurrent systems. The former can be used for model checking, particularly for specifying the properties to be verified. The latter can be used for state-aggregation algorithms that compress models by merging bisimilar states but guarantee that the required properties are preserved.

Recently, bisimulations have been investigated in fuzzy systems. For example, Cao et al. [1], [2] considered bisimulations for deterministic (resp. nondeterministic) fuzzy transition systems, abbreviated as **FTS**s (resp. **NFTSs**); Ćirić et al. [3] investigated bisimulations for fuzzy automata; Qiu and Deng [4], and Xing et al. [5] studied (bi)simulations for fuzzy discrete event systems; Fan [6] defined bisimulations for fuzzy Kripke structures. For more information about fuzzy (bi)simulations, we also refer to [7]–[14].

The bisimulation proposed by Cao et al. [2] is state-based and represents a branching-time semantics. For instance, it can distinguish states s and t in Fig.1. However, if we are mainly interested in the maximal possibilities of the occurrences of some events, then the two states should be intuitively identified. After performing the same action a, both s and t can reach some state with maximal possibility 0.8; after performing the action a immediately followed by b, both s and t can reach some state with maximal possibility 0.6. Based on this observation, we introduce a notion of distribution-based bisimulation that represents a linear-time semantics. This bisimulation is close to the classical bisimulation [20], which is based on transitions between (bare) states. We view distributions as *generalized* states and thus define transitions between distributions. Two bisimilar distributions are required to have the same height, which shows the same maximal possibility of reaching some states, and after some matching transitions, the two successor distributions still preserve the same height.

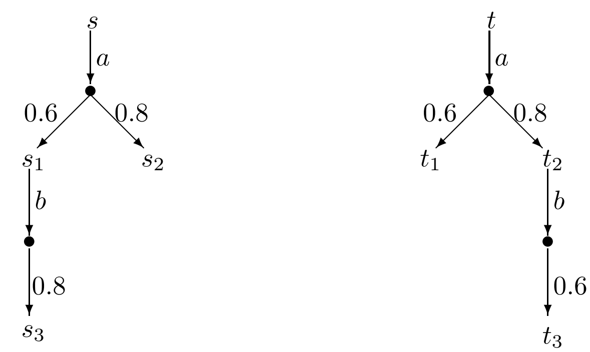

Fig.1 s and t are distribution-bisimilar but not state-bisimilar

Built on bisimulations, a notion of behavioural ultrametric [2] has been proposed to measure the similarity of states in an **NFTS**. It is a more robust way of formalising behavioural similarity between fuzzy systems than bisimulations. The smaller the behavioural distance, the more similarly the states behave. In particular, the behavioural distance between two states is 0 if and only if they are exactly bisimilar. In this paper, we generalize behavioural ultrametrics and propose a notion of triangular-conorm-based metric, called $\oplus$-metric, to measure the distance between possibility distributions in an **NFTS**. Technically speaking, we define a distribution-based bisimilarity metric as the least fixed point of a suitable monotonic function, which captures the similarity of the behaviour of possibility distributions. We also show that a parallel composition in the style of communicating sequential processes (CSP) [15] is non-expansive with respect to distribution-based bisimilarity metric, which makes compositional verification possible. Our metric differs from the ultrametric of Cao et al. [2] in two aspects: (1) the former is defined on $\mathcal{F}(S)$, while the latter is defined on S; (2) the former is based on a class of

This work is partially supported by the National Natural Science Foundation of China (61321064, 61370100, 61672229, 61261130589), Shanghai Municipal Natural Science Foundation (16ZR1409100), and ANR 12IS02001 PACE.

H. Wu is with MoE Engineering Center for Software/Hardware Co-Design Technology and Application, East China Normal University, Shanghai 200062, P.R. China. Email: wuhengy_1974@aliyun.com.

Y. Deng (corresponding author) is with Shanghai Key Laboratory of Trustworthy Computing, MOE International Joint Lab of Trustworthy Software and International Research Center of Trustworthy Software, East China Normal University, Shanghai 200062, P.R. China. Email: yxdeng@sei.ecnu.edu.cn.

triangular-conorms, while the later is only based on $\vee$-conorm, which is a special case of ours. This generalization allows us, according to the properties of different systems, to compute behavioural distances by choosing appropriate operators.

Modal logics and behavioural equivalences are closely related. Whenever a new equivalence is proposed, the quest starts for the associated logic such that two systems (or states) are behaviourally equivalent if and only if they satisfy the same set of modal formulae. Along this line, a great amount of work has appeared that characterizes various kinds of classical (or probabilistic) bisimulations by appropriate logics, e.g. [16]–[24]. Although bisimulations have been investigated extensively in fuzzy systems, there is little work about the connection between bisimulations and the corresponding modal logics. Fan [6] characterized (fuzzy) bisimulations for fuzzy Kripke structures in terms of Gödel modal logic; Wu and Deng [14] characterized bisimulations for **FTS**s by using a fuzzy style Hennessy-Milner logic. In this paper, we present a fuzzy modal logic, which is directly interpreted on distributions. This logic characterizes not only distribution-based bisimilarity but also distribution-based bisimilarity metric for **NFTS**s soundly and completely. We show that the distance between two distributions is 0 if and only if they have the same value on each formula. Moreover, this characterization holds for a class of fuzzy modal logics, by instantiating a triangular-norm (t-norm) as Łukasiewicz t-norm, Gödel t-norm, product t-norm, or nilpotent minimum t-norm.

The rest of this paper is structured as follows. We briefly review some basic concepts on fuzzy sets in Section II and introduce a fuzzy logic used in this paper. In Section III, we define a new bisimulation for **NFTS**s, which is distribution-based. Section IV embarks upon the development of distribution-based bisimilarity metric. We first define triangular-conorm-based metric, called $\oplus$-metric. Then we give the notion of distribution-based bisimulation metric and discuss the monotonicity of a function on behavioural metrics. Thanks to Tarski's fixed point theorem, we can obtain the least fixed point of the function and define it as the distribution-based bisimilarity metric. Moreover, we present an on-the-fly algorithm to compute the bisimilarity metric. In Section V, we give a fuzzy modal logic, which characterizes distribution-based bisimilarity metric soundly and completely. In Section VI, we define a parallel composition and show that it is non-expansive with respect to the distribution-based bisimilarity metric and preserves distribution-based bisimilarity. As an example, we discuss the potential application of our results in Section VII and review some related work in Section VIII. Finally, we conclude in Section IX with some future work.

II. Preliminaries

A. Fuzzy Set

We first briefly recall some basic concepts on fuzzy sets. Let S be an ordinary set. A *fuzzy set* μ of S is a function that assigns each element s of S with a value $\mu(s)$ in the unit interval $[0,1]$. The *support* of μ, written $supp(\mu)$, is the set $\{s \in S \mid \mu(s) > 0\}$. We denote by $\mathcal{F}(S)$ the set of all fuzzy sets in S. For convenience, whenever supp(μ) is a finite set, say $\{s_1, s_2, \cdots, s_n\}$, then a fuzzy set μ is written as $\mu = \frac{r_1}{s_1} + \frac{r_2}{s_2} + \cdots + \frac{r_n}{s_n}$, where $r_i \in (0,1]$ and $r_i = \mu(s_i)$ with $1 \le i \le n$. With a slight abuse of notations, we sometimes write a possibility distribution to mean a fuzzy set[1].

For any $\mu, \nu \in \mathcal{F}(S)$, we say that μ is contained in ν (or ν contains μ), denoted by $\mu \subseteq \nu$, if $\mu(s) \le \nu(s)$ for all $s \in S$. Notice that $\mu = \nu$ if both $\mu \subseteq \nu$ and $\nu \subseteq \mu$. We use $\emptyset$ to denote the empty fuzzy set with $\emptyset(s) = 0$ for all $s \in S$. For any $s \in S$, we write $\bar{s}$ for the point distribution, satisfying $\bar{s}(s') = 1$ if $s' = s$ and 0 otherwise.

For any $p, q \in [0,1]$, we write $p \vee q$ and $p \wedge q$ to mean $\max(p,q)$ and $\min(p,q)$, respectively, write $\bigvee_{i \in I} p_i$ for the supremum of $\{p_i \mid i \in I\}$, and $\bigwedge_{i\in I} p_i$ for the infimum, where $\{p_i \mid i \in I\}$ is a family of elements in $[0,1]$. In particular, $\bigvee \emptyset = 0$ and $\bigwedge \emptyset = 1$. For any $c \in [0,1]$ and $\mu \in \mathcal{F}(S)$, the *scalar multiplication* $c \cdot \mu$ of c and μ is defined by letting $(c \cdot \mu)(s) = c \wedge \mu(s)$, for each $s \in S$. Let $\{\mu_i \mid i \in I\}$ be a family of elements of $\mathcal{F}(S)$, the *union* of μ_i, denoted $\bigcup \mu_i$, is given by $(\bigcup_{i\in I} \mu_i)(s) = \bigvee_{i \in I} \mu_i(s)$. The *height* of possibility distribution μ, written $ht(\mu)$, is defined as $\bigvee_{s \in S} \mu(s)$.

Throughout this paper, we assume an ordinary finite set S and let it denote the set of states of a fuzzy transition system.

B. Fuzzy Logic

Let us introduce an algebraic structure, which will play a key role in defining the bisimilarity metric and in its logical characterization.

Definition 1. [26] Let $(C, \le, 0, 1)$ be a bounded linearly ordered set. A function $n : C \to C$ is called an *involution* or *strong negation* function if it is order-reversing with $n(1) = 0$ and $n(n(x)) = x$ for all $x \in C$.

When C is taken as the unit interval $[0,1]$, the involutive negation is usually interpreted by the *standard negation*: $n(x) = 1 - x$ for any $x \in [0,1]$.

Definition 2. We consider the unit interval $[0,1]$ as an algebra endowed with the operations min and max, a fixed *left-continuous* t-norm $\otimes$ and its *residum* $\Rightarrow$, as well as the standard negation n. The ordering $\le$ is as usual. This algebra will be denoted by $\mathcal{L}_\otimes$.

Notice that left-continuity is a necessary and sufficient condition for a t-norm $\otimes$ and its residuum $\Rightarrow$, defined as $x \Rightarrow y = \sup\{z \in [0,1] \mid x \otimes z \le y\}$, to verify the so-called residuation property: $x \le y \Rightarrow z$ iff $x \otimes y \le z$. In this case, we say $(\otimes, \Rightarrow)$ is an *adjoint pair* [26].

For the basic properties of t-norms we refer to [27], [28]. There are four commonly used t-norms: Łukasiewicz t-norm ($x \otimes y = \max(x+y-1, 0)$), Gödel t-norm ($x \otimes y = \min(x,y)$), product t-norm ($x \otimes y = x \times y$), and nilpotent minimum t-norm [29] ($x \otimes y = \min(x,y)$ if $x+y > 1$ and 0 otherwise). The first three are continuous, whereas the fourth is left-continuous. The corresponding algebras are denoted by $\mathcal{L}^L_\otimes$, $\mathcal{L}^G_\otimes$, $\mathcal{L}^P_\otimes$, and $\mathcal{L}^N_\otimes$, respectively. For convenience, we sometimes use the symbols

[1] Strictly speaking a possibility distribution is different from a fuzzy set, though the former can be viewed as the generalized characteristic function of the latter. See [25] for more detailed discussion.

$\Rightarrow_L$, $\Rightarrow_G$, $\Rightarrow_P$ and $\Rightarrow_N$ to denote the implications in the algebras $\mathcal{L}^L_\otimes$, $\mathcal{L}^G_\otimes$, $\mathcal{L}^P_\otimes$, $\mathcal{L}^N_\otimes$, respectively, which can be found in [28], [29].

In general, an important aim of introducing an algebraic structure is to establish the completeness of a formal system. Examples include the **BL**-algebra for the basic logic introduced by Hájek [28] and the **MTL**-algebra for the monoidal t-norm based logic introduced by Esteva and Godo [26]. Nevertheless, our motivation is to establish an abstract metric (see Section IV) and to characterize it (see Section V). Below we discuss the reasons of choosing $\mathcal{L}_\otimes$ by comparing it with some relevant algebras.

- The standard semantics of **BL** (resp. **MTL**) uses the unit interval $[0,1]$ as the set of truth values and a continuous (resp. left-continuous) t-norm $\otimes$ as an interpretation of the strong conjunction $\&$. The implication $\rightarrow$ is interpreted by the residuum of $\otimes$. It is well known that $([0,1], \min, \max, \otimes, \Rightarrow, 0, 1)$ is an **MTL**-algebra [26], and it is a **BL**-algebra provided that $\otimes$ is continuous [28].
- In **BL** and **MTL**-algebras, the negation $\neg x = x \Rightarrow 0$ is in general not involutive. Thus, **BL** and **MTL** as well as some of their axiomatic extensions lack a strong disjunction dual to the strong conjunction. Attempts have been made to generalize them. For example, an **IMTL**-algebra [26], where **I** stands for involution, is obtained by adding $\neg\neg x = x$ to an **MTL**-algebra. However, in **IMTL** the involution depends on the t-norm. As a consequence, **IMTL** admits only those left-continuous t-norms which yield an involutive negation, but rules out operators like Gödel and product t-norms. In order to fill such a gap, an $\mathbf{MTL_n}$-algebra[2] is obtained by adding an independent involutive negation n and a δ operation to **MTL**; see [30] for more details. The extensions of **BL**-algebra can be found in, for example, [31], [32]. Obviously, our $\mathcal{L}_\otimes$ is like the $\mathbf{MTL_n}$-algebra giving up the operator δ. In $\mathcal{L}_\otimes$, we can obtain a t-conorm $\oplus$ by letting $x \oplus y = 1 - [(1-x) \otimes (1-y)]$, which is dual to $\otimes$ and $(\otimes, \oplus, n)$ is called a *De Morgan triple*.
- In this paper, the operator n is used for defining $\oplus$; the operator $\otimes$ is used for producing the adjoint pair $(\otimes, \Rightarrow)$ and the De Morgan triple $(\otimes, \oplus, n)$. It is not used to interpret the strong conjunction $\&$ as **BL** and **MTL**-algebras do, since $\varphi \& \psi$ in general is not a formula in our fuzzy modal logic; the induced operator $\oplus$ is used for defining an abstract metric and for modeling the strong disjunction (see Definition 6). Note that it is not used to interpret the strong disjunctive connective $\veebar$ between formulae, since $\varphi \veebar \psi$ in general is not a formula in our fuzzy modal logic. This point is very different from [30], [32]. The induced biimplication operator $\Leftrightarrow$ helps us to model equivalence $\equiv$ and is given by $x \Leftrightarrow y = (x \Rightarrow y) \wedge (y \Rightarrow x)$ (see [3], [27]), which also deviates from **BL** and **MTL**-algebras, where it is defined as $x \Leftrightarrow y = (x \Rightarrow y) \otimes (y \Rightarrow x)$. For more differences between $\mathcal{L}_\otimes$ and the above algebras, please see Section V.
- The De Morgan triple $(\otimes, \oplus, n)$ and the adjoint pair $(\otimes, \Rightarrow)$ guarantee the main results of this paper to hold.

[2]We use n instead of $\sim$ because this notion will be reserved for bisimilarity.

The following properties will be useful.

Proposition 1. Under $\mathcal{L}_\otimes$, for any $x, y, z, x_1, x_2, y_1, y_2, y_i$ in the interval $[0,1]$, where $i \in I$. The following properties hold.

(1) $x \Rightarrow y \ = \ 1$ if and only if $x \leq y$,
(2) $1 \Rightarrow x \ = \ x$,
(3) $x \otimes (x \Rightarrow y) \ \leq \ y$,
(4) $(x \Leftrightarrow y) \otimes (y \Leftrightarrow z) \ \leq \ x \Leftrightarrow z$,
(5) $(x_1 \Leftrightarrow y_1) \wedge (x_2 \Leftrightarrow y_2) \ \leq \ (x_1 \wedge x_2) \Leftrightarrow (y_1 \wedge y_2)$,
(6) $x_1 \Leftrightarrow x_2 \ \leq \ (x_1 \Rightarrow y) \Leftrightarrow (x_2 \Rightarrow y)$,
(7) $x_1 \Leftrightarrow x_2 \ \leq \ (y \Rightarrow x_1) \Leftrightarrow (y \Rightarrow x_2)$,
(8) $x \oplus (\bigvee_{i\in I} y_i) \ \geq \ \bigvee_{i\in I}(x \oplus y_i)$,
(9) $x \oplus (\bigwedge_{i\in I} y_i) \ = \ \bigwedge_{i\in I}(x \oplus y_i)$.

Proof. The first five items can be found in [27], [28], [33]. Items (6) and (7) can be proven by adjoint condition. As an example, we choose to prove (6). Let $t \leq x_1 \Leftrightarrow x_2$. Then $t \leq x_1 \Rightarrow x_2$ and $t \leq x_2 \Rightarrow x_1$. It follows by adjoint condition that $t \otimes x_1 \leq x_2$ and $t \otimes x_2 \leq x_1$. Further, we have that $t \otimes x_2 \otimes (x_1 \Rightarrow y) \leq x_1 \otimes (x_1 \Rightarrow y) \leq y$ by (3). Again, by adjoint condition, we have that $t \otimes (x_1 \Rightarrow y) \leq x_2 \Rightarrow y$, and then $t \leq (x_1 \Rightarrow y) \Rightarrow (x_2 \Rightarrow y)$. In a similar way, we can get $t \leq (x_2 \Rightarrow y) \Rightarrow (x_1 \Rightarrow y)$. So $t \leq (x_1 \Rightarrow y) \Leftrightarrow (x_2 \Rightarrow y)$. Thus, we obtain $(x_1 \Leftrightarrow x_2) \leq (x_1 \Rightarrow y) \Leftrightarrow (x_2 \Rightarrow y)$.

Item (8) is trivial since $\oplus$ is monotonic. Item (9) can be obtained by using the duality between $\oplus$ and $\otimes$ and the left-continuity of $\otimes$. □

III. Distribution-based bisimulation

In this section, we introduce the notion of distribution-based bisimulation for nondeterministic fuzzy transition systems.

Definition 3. [34] A *nondeterministic fuzzy transition system* (**NFTS**) is a triple $\mathcal{M} = (S, A, \rightarrow)$, where S is a set of states, A is a set of actions, and $\rightarrow \subseteq S \times A \times \mathcal{F}(S)$ is the transition relation.

For each state s, more than one possibility distribution may be reached by performing action a because of nondeterminism. We often write $s \xrightarrow{a} \mu$ for $(s, a, \mu) \in \rightarrow$. We write $s \not\xrightarrow{a}$ if there exists no μ such that $s \xrightarrow{a} \mu$. Let us define a transition relation between distributions in an **NFTS**.

Definition 4. Given an **NFTS** $\mathcal{M} = (S, A, \rightarrow)$, the transition relation $\rightarrow$ is lifted to $\rightarrow_d \subseteq \mathcal{F}(S) \times A \times \mathcal{F}(S)$ by taking the smallest relation satisfying:

- if $s \xrightarrow{a} \mu$ then $\bar{s} \xrightarrow{a}_d \mu$;
- if $s \not\xrightarrow{a}$ then $\bar{s} \xrightarrow{a}_d \emptyset$;
- if $\bar{s} \xrightarrow{a}_d \nu_s$ for each $s \in supp(\mu)$, then

$$\mu \xrightarrow{a}_d \bigcup_{s \in supp(\mu)} \mu(s) \cdot \nu_s.$$

We often abuse the notation and $\rightarrow_d$ is also denoted by $\rightarrow$. We define $T_a(\mu)$ as the set $\{\nu \in \mathcal{F}(S) \mid \mu \xrightarrow{a} \nu\}$. An **NFTS** $(S, A, \rightarrow)$ is *image-finite* if for each distribution μ and each action a, the set $T_a(\mu)$ is finite. The **NFTS** is *finitary* if it is image-finite and has finitely many states.

Note that if $\mu \xrightarrow{a} \nu$ then some (not necessarily all) states in the support of μ can perform action a. For example, consider the two states s_1 and s_2 in Figure 1. Since $\bar{s}_1 \xrightarrow{b} \frac{0.8}{s_3}$ and s_2 cannot perform action b, the distribution $\mu = \frac{0.6}{s_1} + \frac{0.8}{s_2}$ can make the transition $\mu \xrightarrow{b} \frac{0.6}{s_3}$ to reach the distribution $\frac{0.6}{s_3}$.

Definition 5 (Distribution-based bisimulation)**.** A symmetric relation $\mathcal{R} \subseteq \mathcal{F}(S)\times\mathcal{F}(S)$ is a distribution-based bisimulation on an **NFTS** $(S, A, \rightarrow)$ if for all $\mu, \nu \in \mathcal{F}(S)$, whenever $\mu\mathcal{R}\nu$ then:

(i) $ht(\mu) = ht(\nu)$ and

(ii) if $\mu \xrightarrow{a} \mu'$, then there exists some $\nu' \in \mathcal{F}(S)$ such that $\nu \xrightarrow{a} \nu'$ and $\mu'\mathcal{R}\nu'$.

Two distributions $\mu, \nu \in \mathcal{F}(S)$ are bisimilar, written $\mu \sim \nu$, if there exists a distribution-based bisimulation $\mathcal{R}$ with $\mu\mathcal{R}\nu$.

It is easy to see that $\sim$ is the greatest distribution-based bisimulation since it is the union of all distribution-based bisimulations. Moreover, it is an equivalence relation on $\mathcal{F}(S)$.

IV. Distribution-based Bisimilarity Metric

Before defining the distribution-based bisimilarity metric, we first introduce a triangular conorm-based metric.

A. Triangular Conorm-Based Metric

Recall that the *similarity relation* on S proposed by Zadeh in [35] is a fuzzy relation $\mathcal{S}$ on S, i.e., a function of type $S \times S \to [0, 1]$, that satisfies the following three conditions:

(1) reflexive, i.e., $\mathcal{S}(s, s) = 1$ for all $s \in S$,

(2) symmetric, i.e., $\mathcal{S}(s, t) = \mathcal{S}(t, s)$ for all $s, t \in S$,

(3) max-star transitive, i.e., $\mathcal{S} * \mathcal{S} \subseteq \mathcal{S}$, or, more explicitly,

$$\bigvee_{s' \in S} \mathcal{S}(s, s') * \mathcal{S}(s', t) \leq \mathcal{S}(s, t)$$

where $*$ is associative and monotone non-decreasing in each of its arguments.

Now, we take $*$ to be a t-norm $\otimes$ and define, for all $s, t \in S$, $d(s, t) = 1 - \mathcal{S}(s, t)$. It turns out that, for all $s, s', t \in S$, we have $d(s, t) \leq d(s, s') \oplus d(s', t)$. Moreover, $d(s, s) = 0$ and $d(s, t) = d(t, s)$. We call the distance function d a $\oplus$*-metric*. Next, we generalize the $\oplus$-metric from S to $\mathcal{F}(S)$.

Definition 6. Under $\mathcal{L}_\otimes$, a function $d : \mathcal{F}(S)\times\mathcal{F}(S) \to [0, 1]$ is called a $\oplus$*-metric* on $\mathcal{F}(S)$ if for all $\mu, \nu, \eta \in \mathcal{F}(S)$,

$(P1)$ $d(\mu, \mu) \ = \ 0$,

$(P2)$ $d(\mu, \nu) \ = \ d(\nu, \mu)$,

$(P3)$ $d(\mu, \eta) \ \leq \ d(\mu, \nu) \oplus d(\nu, \eta)$.

We call $(\mathcal{F}(S), d)$ a $\oplus$-metric space.

This definition generalizes the notion of the pseudo-ultrametric given by Cao et al. in [2] in two aspects: (1) our metric is directly defined on distributions, whereas Cao's is defined on states; (2) we use $\oplus$ instead of $\vee$ in Cao's definition, i.e, the $\max$-disjunction is replaced by the strong disjunction. The following proposition indicates that $\oplus$-metric always exists.

Proposition 2. Given a t-conorm $\oplus$ from $\mathcal{L}_\otimes$, the $\oplus$-metric always exists.

Proof. Let $\perp(\mu, \nu) = 1 - (ht(\mu) \Leftrightarrow ht(\nu))$ for any distributions $\mu, \nu \in \mathcal{F}(S)$. Then it is easy to see that $\perp$ satisfies $P1$ and $P2$. Now, we verify $P3$. For any $\mu, \nu, \eta \in \mathcal{F}(S)$, we have

$$\begin{aligned}
&\perp(\mu, \nu) \oplus \perp(\nu, \eta) \\
&= [1 - (ht(\mu) \Leftrightarrow ht(\nu))] \oplus [1 - (ht(\nu) \Leftrightarrow ht(\eta))] \\
&= 1 - (ht(\mu) \Leftrightarrow ht(\nu)) \otimes (ht(\nu) \Leftrightarrow ht(\eta)) \\
&\geq 1 - (ht(\mu) \Leftrightarrow ht(\eta)) \quad \text{(by Proposition 1 (4))} \\
&= \perp(\mu, \eta)
\end{aligned}$$

as expected. $\square$

We can understand $\perp(\mu, \nu)$ as how far distributions μ and ν are from the same height based on the $\oplus$-metric. If $\oplus$ is taken from $\mathcal{L}^L_\otimes$, then we have $\perp(\mu, \nu) = |ht(\mu) - ht(\nu)|$, which gives a pseudometric bounded by 1 often used in probabilistic systems. If $\oplus$ is taken from $\mathcal{L}^G_\otimes$, then $\perp(\mu, \nu) = 0$ when $ht(\mu) = ht(\nu)$, and $\max(1 - ht(\mu), 1 - ht(\nu))$ otherwise, which is a pseudo-ultrametric. In the same way, one can get $\perp(\mu, \nu)$ when $\oplus$ is taken from $\mathcal{L}^N_\otimes$ and $\mathcal{L}^P_\otimes$, respectively. A pseudo-ultrametric is a $\oplus$-metric because $\vee$ is the minimum t-conorm, but the inverse is not necessarily true, as witnessed by $\perp$ when $\oplus$ is taken from $\mathcal{L}^L_\otimes$, $\mathcal{L}^N_\otimes$, and $\mathcal{L}^P_\otimes$, respectively. Note that $\perp$ is important for defining our bisimilarity metric and we use it as a special notation in the sequel.

Let $\mathcal{D}_{\mathcal{F}(S)}$ be the set of all $\oplus$-metrics on $\mathcal{F}(S)$. The order $\preceq$ on $\mathcal{D}_{\mathcal{F}(S)}$ is defined by letting $d_1 \preceq d_2$ if for all $\mu, \nu \in \mathcal{F}(S)$ we have $d_1(\mu, \nu) \leq d_2(\mu, \nu)$.

It is easy to check that $\preceq$ is indeed a partial order on $\mathcal{D}_{\mathcal{F}(S)}$. Recall that a partially ordered set $(X, \leq)$ is called a *complete lattice* if every subset of X has a supremum and an infimum in X. We now show that $\mathcal{D}_{\mathcal{F}(S)}$ endowed with the above order forms a complete lattice.

Lemma 1. Under $\mathcal{L}_\otimes$, $(\mathcal{D}_{\mathcal{F}(S)}, \preceq)$ is a complete lattice.

Proof. For any $X \subseteq \mathcal{D}_{\mathcal{F}(S)}$, we define $\bigsqcup X$ by letting $(\bigsqcup X)(\mu, \nu) = \bigvee\{d(\mu, \nu) \mid d \in X\}$ and $\bigsqcap X$ by letting $\bigsqcap X = \bigsqcup\{d \in \mathcal{D}_{\mathcal{F}(S)} \mid \forall d' \in X, d \preceq d'\}$. We need to verify that $\bigsqcup X$ and $\bigsqcap X$ are the supremum and infimum of X, respectively. We only prove that $\bigsqcup X$ is the supremum; the infimum can be proven similarly. We first show that $\bigsqcup X \in \mathcal{D}_{\mathcal{F}(S)}$. It is obvious that $(\bigsqcup X)(\mu, \mu) = 0$ and $(\bigsqcup X)(\mu, \nu) = (\bigsqcup X)(\nu, \mu)$. For $(P3)$, we have that

$$\begin{aligned}
& (\textstyle\bigsqcup X)(\mu, \nu) \\
= \ & \textstyle\bigvee_{d \in \mathcal{D}_{\mathcal{F}(S)}} d(\mu, \nu) \\
\leq \ & \textstyle\bigvee_{d \in \mathcal{D}_{\mathcal{F}(S)}} (d(\mu, \eta) \oplus d(\eta, \nu)) \\
\leq \ & \textstyle\bigvee_{d \in \mathcal{D}_{\mathcal{F}(S)}} d(\mu, \eta) \oplus \bigvee_{d \in \mathcal{D}_{\mathcal{F}(S)}} d(\eta, \nu) \\
= \ & [(\textstyle\bigsqcup X)(\mu, \eta)] \oplus [(\bigsqcup X)(\eta, \nu)]
\end{aligned}$$

for any $\mu, \nu, \eta \in \mathcal{F}(S)$. Hence $\bigsqcup X$ is a $\oplus$-metric. In addition, it is not hard to prove that $\bigsqcup X$ is the supremum of X. Hence $(\mathcal{D}_{\mathcal{F}(S)}, \preceq)$ is a complete lattice. $\square$

Hausdorff distance measures how far two subsets of a metric space are from each other, which allows us to lift a $\oplus$-metric on $\mathcal{F}(S)$ to a $\oplus$-metric on $\mathcal{P}(\mathcal{F}(\mathcal{S}))$.

Definition 7. Let $(\mathcal{F}(S), d)$ be a $\oplus$-metric space. For any $\mu \in \mathcal{F}(S)$ and a finite subset A of $\mathcal{F}(S)$, define

$$d(\mu, A) = \bigwedge_{\nu \in A} d(\mu, \nu).$$

Furthermore, given a pair of finite subsets $A, B \subseteq \mathcal{F}(S)$, the Hausdorff distance induced by d is defined as

$$H_d(A, B) = [\bigvee_{\mu \in A} d(\mu, B)] \vee [\bigvee_{\nu \in B} d(\nu, A)].$$

As expected, H_d has the following property.

Lemma 2. Under $\mathcal{L}_\otimes$, if d is a $\oplus$-metric on $\mathcal{F}(S)$, then H_d is a $\oplus$-metric on $\mathcal{P}_f(\mathcal{F}(S))$, where $\mathcal{P}_f(\mathcal{F}(S))$ stands for the set of all finite subsets of $\mathcal{F}(S)$.

Proof. Firstly, it is easy to see that $H_d(A, A) = 0$ and $H_d(A, B) = H_d(B, A)$. Secondly,

$$\begin{array}{ll} & H_d(A,B) \oplus H_d(B,C) \\ \geq & [\bigvee_{\mu \in A} d(\mu, B)] \oplus [\bigvee_{\nu \in B} d(\nu, C)] \\ \geq & \bigvee_{\mu \in A} \bigvee_{\nu \in B} [d(\mu, B) \oplus d(\nu, C)] \quad \text{(by Proposition 1 (8))} \\ \geq & \bigvee_{\mu \in A} \bigvee_{\nu \in B} [d(\mu, \nu') \oplus d(\nu, C)] \ \text{(for some } \nu' \in B) \\ \geq & \bigvee_{\mu \in A} [d(\mu, \nu') \oplus d(\nu', C)] \\ = & \bigvee_{\mu \in A} [d(\mu, \nu') \oplus (\bigwedge_{\eta \in C} d(\nu', \eta))] \\ = & \bigvee_{\mu \in A} [\bigwedge_{\eta \in C} (d(\mu, \nu') \oplus d(\nu', \eta))] \\ \geq & \bigvee_{\mu \in A} \bigwedge_{\eta \in C} d(\mu, \eta) \\ = & \bigvee_{\mu \in A} d(\mu, C). \end{array}$$

Similarly, we have $H_d(A, B) \oplus H_d(B, C) \geq \bigvee_{\eta \in C} d(\eta, A)$. Hence,

$$\begin{array}{ll} & H_d(A,B) \oplus H_d(B,C) \\ \geq & [\bigvee_{\mu \in A} d(\mu, C)] \vee [\bigvee_{\eta \in C} d(\eta, A)] \\ = & H_d(A, C). \end{array}$$

This completes the proof. □

The following lemma shows that the distance between a distribution and a finite set of distributions is preserved by the sup of an increasing $\oplus$-metric chain.

Lemma 3. Let $(\mathcal{F}(S), d)$ be a $\oplus$-metric space, $\{d_n \mid n \in \mathbb{N}\}$ be an increasing chain, i.e., $d_0 \preceq d_1 \preceq d_2 \cdots$, and B be a finite subset of $\mathcal{F}(S)$. Then for any $\mu \in \mathcal{F}(S)$, we have

$$(\bigsqcup_{n \in \mathbb{N}} d_n)(\mu, B) \ = \ \bigvee_{n \in \mathbb{N}} d_n(\mu, B).$$

Proof. It is sufficient to prove that

$$\inf_{\nu \in B} \sup_{n \in \mathbb{N}} d_n(\mu, \nu) = \sup_{n \in \mathbb{N}} \inf_{\nu \in B} d_n(\mu, \nu).$$

First, for any $n \in \mathbb{N}$, it is clear that

$$\inf_{\nu \in B} d_n(\mu, \nu) \leq \inf_{\nu \in B} \sup_{n \in \mathbb{N}} d_n(\mu, \nu).$$

This shows that $\inf_{\nu \in B} \sup_{n \in \mathbb{N}} d_n(\mu, \nu)$ is an upper bound of the set $\{\inf_{\nu \in B} d_n(\mu, \nu) \mid n \in \mathbb{N}\}$.

Second, for any $\nu \in B$, we consider $\sup_{n \in \mathbb{N}} d_n(\mu, \nu)$. For any $\epsilon > 0$, there exists some n_ν such that

$$d_{n_\nu}(\mu, \nu) > \sup_{n \in \mathbb{N}} d_n(\mu, \nu) - \epsilon.$$

Since B is finite and $d_n (n \in \mathbb{N})$ is an increasing chain, there exists a d such that $d(\mu, \nu) > \sup_{n \in \mathbb{N}} d_n(\mu, \nu) - \epsilon$ for any $\nu \in B$. Hence,

$$\inf_{\nu \in B} d(\mu, \nu) \ > \ \inf_{\nu \in B} (\sup_{n \in \mathbb{N}} d_n(\mu, \nu) - \epsilon) \ = \ \inf_{\nu \in B} \sup_{n \in \mathbb{N}} d_n(\mu, \nu) - \epsilon.$$

Thus, $\inf_{\nu \in B} \sup_{n \in \mathbb{N}} d_n(\mu, \nu)$ is the supremum of the set $\{\inf_{\nu \in B} d_n(\mu, \nu) \mid n \in \mathbb{N}\}$. We have completed the proof. □

An immediate corollary is that the Hausdorff distance between two finite sets of possibility distributions is preserved by the sup of an increasing $\oplus$-metric chain.

Corollary 1. Let $(\mathcal{F}(S), d)$ be a $\oplus$-metric space, A, B be finite subsets of $\mathcal{F}(S)$, and $\{d_n \mid n \in \mathbb{N}\}$ be an increasing chain. Then $H_{\bigsqcup_{n \in \mathbb{N}} d_n}(A, B) \ = \ \bigvee_{n \in \mathbb{N}} H_{d_n}(A, B)$.

Proof. Straightforward by using Lemma 3. □

B. Distribution-Based Bisimulation Metric

Inspired by [36], we define a bisimilarity metric that can measure the behavioural distance between two distributions.

Definition 8. Given an **NFTS** $(S, A, \rightarrow)$ and $\mathcal{L}_\otimes$, a $\oplus$-metric d on $\mathcal{F}(S)$ is called a *distribution-based bisimulation metric* if for all $\mu, \nu \in \mathcal{F}(S)$,

(i) $\perp \preceq d$ and
(ii) $d(\mu, \nu) < 1$ implies that whenever $\mu \xrightarrow{a} \mu'$ then there exists a transition $\nu \xrightarrow{a} \nu'$ with $d(\mu', \nu') \leq d(\mu, \nu)$.

The first condition says that d should be at least $\perp$. The second condition requires the matching of transitions and the distance between the two successor distributions should not exceed the distance between the original distributions.

Below we aim to give a fixed-point characterization of bisimulation metric.

Definition 9. Given an image-finite **NFTS** $(S, A, \rightarrow)$, the functional $\Delta : \mathcal{D}_{\mathcal{F}(S)} \rightarrow \mathcal{D}_{\mathcal{F}(S)}$ is defined as follows: for any $d \in \mathcal{D}_{\mathcal{F}(S)}$ and $\mu, \nu \in \mathcal{F}(S)$,

$$\Delta(d)(\mu, \nu) \ = \ \max(\bigvee_{a \in A} H_d(T_a(\mu), T_a(\nu)), \perp(\mu, \nu)).$$

Now, we check that Δ is well-defined.

Proposition 3. Given an image-finite **NFTS** $(S, A, \rightarrow)$, then for any $d \in \mathcal{D}_{\mathcal{F}(S)}$, we have $\Delta(d) \in \mathcal{D}_{\mathcal{F}(S)}$.

Proof. It is easy to see that $\Delta(d)$ satisfies $P1$ and $P2$. Now, we verify $P3$. For any $\mu, \nu, \eta \in \mathcal{F}(S)$, it is obvious that, for any $a \in A$, $\Delta(d)(\mu, \nu) \geq H_d(T_a(\mu), T_a(\nu))$ and symmetrically $\Delta(d)(\nu, \eta) \geq H_d(T_a(\nu), T_a(\eta))$. Hence, for any $a \in A$,

$$\begin{array}{ll} & \Delta(d)(\mu, \nu) \oplus \Delta(d)(\nu, \eta) \\ \geq & H_d(T_a(\mu), T_a(\nu)) \oplus H_d(T_a(\nu), T_a(\eta)) \\ \geq & H_d(T_a(\mu), T_a(\eta)) \quad \text{(by Lemma 2).} \end{array}$$

Consequently,

$$\Delta(d)(\mu, \nu) \oplus \Delta(d)(\nu, \eta) \geq \bigvee_{a \in A} H_d(T_a(\mu), T_a(\eta)).$$

On the other hand, we also have

$$\Delta(d)(\mu, \nu) \geq \perp(\mu, \nu) \text{ and } \Delta(d)(\nu, \eta) \geq \perp(\nu, \eta).$$

Hence,

$$\Delta(d)(\mu,\nu)\oplus\Delta(d)(\nu,\eta)\geq\bot(\mu,\nu)\oplus\bot(\nu,\eta)\geq\bot(\mu,\eta)$$

by Proposition 2. It follows that

$$\begin{array}{rl} & \Delta(d)(\mu,\nu)\oplus\Delta(d)(\nu,\eta) \\ \geq & \max(\bigvee_{a\in A}H_d(T_a(\mu),T_a(\eta)),\bot(\mu,\eta)). \end{array}$$

That is, $\Delta(d)(\mu,\nu)\oplus\Delta(d)(\nu,\eta)\geq\Delta(d)(\mu,\eta)$ as desired. □

Moreover, Δ is monotonic with respect to the order $\preceq$, i.e., if $d_1\preceq d_2$, then we have $\Delta(d_1)\preceq\Delta(d_2)$. Recall that the Knaster-Tarski fixed-point theorem [37] says that each monotonic function on a complete lattice has a least fixed point. Hence the following proposition holds.

Proposition 4. Given an image-finite **NFTS** $(S,A,\rightarrow)$, the functional $\Delta:\mathcal{D}_{\mathcal{F}(S)}\rightarrow\mathcal{D}_{\mathcal{F}(S)}$ has a least fixed point given by

$$\Delta_{\min}=\bigsqcap\{d:\ \Delta(d)\preceq d\}.$$

Let us present an explicit characterization of bisimulation metrics in terms of pre-fixed points of Δ.

Lemma 4. Given an image-finite **NFTS** $(S,A,\rightarrow)$, for any $d\in\mathcal{D}_{\mathcal{F}(S)}$, d is a pre-fixed points of Δ if and only if it is a distribution-based bisimulation metric.

Proof. Straightforward by the definitions of distribution-based bisimulation metric and Δ. □

Proposition 4 and Lemma 4 imply the following theorem.

Theorem 1. Given an image-finite **NFTS** $(S,A,\rightarrow)$. The smallest distribution-based bisimulation metric exists, moreover it is just $\Delta_{\min}$. Let $d_b\overset{def}{=}\Delta_{\min}$ and call it *distribution-based bisimilarity metric*.

In what follows, we use d_b^L, d_b^G, d_b^P, and d_b^N to denote distribution-based bisimilarity metrics based on $\mathcal{L}^L_\otimes$, $\mathcal{L}^G_\otimes$, $\mathcal{L}^P_\otimes$ and $\mathcal{L}^N_\otimes$, respectively. Under the assumption of image-finiteness, Corollary 1 implies that Δ preserves the sup of an increasing chain of $\mathcal{D}_{\mathcal{F}(S)}$.

Lemma 5. Given an image-finite **NFTS** $(S,A,\rightarrow)$, we have

$$\Delta(\bigsqcup_{n\in\mathbb{N}}d_n)\ =\ \bigsqcup_{n\in\mathbb{N}}\Delta(d_n)$$

where $d_n(n\in\mathbb{N})$ is an increasing chain of $\mathcal{D}_{\mathcal{F}(S)}$.

Combining Theorem 1 and Lemma 5 gives a way to calculate the distribution-based bisimilarity metric for image-finite **NFTS**s.

Corollary 2. Let $(S,A,\rightarrow)$ be an image-finite **NFTS**. We define $\Delta^0(\bot)\overset{def}{=}\bot$ and $\Delta^{n+1}(\bot)\overset{def}{=}\Delta(\Delta^n(\bot))$. Then

$$d_b=\Delta_{\min}=\bigsqcup_{n\in\mathbb{N}}\Delta^n(\bot).$$

Proof. The proof is standard; see e.g. Page 183 in [37]. □

We give a simple example below to illustrate the above notions and results.

Example 1. Consider Figure 1 again. Let $\mu_1=\frac{0.6}{s_1}+\frac{0.8}{s_2}$, $\mu_2=\frac{0.8}{s_3}$, $\nu_1=\frac{0.6}{t_1}+\frac{0.8}{t_2}$ and $\nu_2=\frac{0.6}{t_3}$. We write d_n for $\Delta^n(\bot)$, clearly $d_i\preceq d_{i+1}$ for any $i\in\mathbb{N}$. Now, we compute $d_b^G(\bar{s},\bar{t})$ and $d_b^G(\bar{s}_1,\bar{t}_2)$. For any $n\geq 3$,

$$\begin{array}{rcl} d_n(\bar{s},\bar{t}) & = & \Delta(d_{n-1})(\bar{s},\bar{t}) \\ & = & \max(d_{n-1}(\mu_1,\nu_1),1-(ht(\bar{s})\Leftrightarrow_G ht(\bar{t}))) \\ & = & d_{n-1}(\mu_1,\nu_1) \\ & = & \Delta(d_{n-2})(\mu_1,\nu_1) \\ & = & \max(d_{n-2}(\frac{0.6}{s_3},\frac{0.6}{t_3}),1-(ht(\mu_1)\Leftrightarrow_G ht(\nu_1))) \\ & = & d_{n-2}(\frac{0.6}{s_3},\frac{0.6}{t_3}) \\ & = & \Delta(d_{n-3})(\frac{0.6}{s_3},\frac{0.6}{t_3}) \\ & = & \max(d_{n-3}(\emptyset,\emptyset),1-(ht(\frac{0.6}{s_3})\Leftrightarrow_G ht(\frac{0.6}{t_3}))) \\ & = & 0. \end{array}$$

Hence $d_b^G(\bar{s},\bar{t})=0$. In the same way, we can calculate that $d_b^G(\bar{s}_1,\bar{t}_2)=0.4$.

For any pair of possibility distributions μ and ν, we observe that the smaller the value of $d_b(\mu,\nu)$, the more similar the distributions.

Corollary 3. Given an image-finite **NFTS** $(S,A,\rightarrow)$. For any $s,t\in S$, let $\mathcal{S}(s,t)\overset{def}{=}1-d_b(\bar{s},\bar{t})$. Then $\mathcal{S}$ is a similarity relation on S.

The above corollary relates distribution-based bisimilarity metric to Zadeh's similarity relation.

We can see that $\bot$ plays a key role in computing the distribution-based bisimilarity metric, which is obtained by an iteration of Δ starting from $\bot$. This paper uses the least fixed point to establish the distribution-based bisimilarity metric, while Cao et al. [2] used the greatest fixed point to establish the behavioural distance. However, there is no essential difference because in [2] the partial order on $\mathcal{D}_{\mathcal{F}(S)}$ is reverse, i.e., $d_1\preceq d_2$ if $d_1(s,t)\geq d_2(s,t)$ for all $s,t\in S$.

C. An Algorithm for Computing d_b

First of all, let us fix a finitary **NFTS** $\mathcal{M}=(S,A,\rightarrow)$. Given any pair of possibility distributions $\mu,\nu\in\mathcal{F}(S)$, we present an on-the-fly algorithm to compute their distance.

In order to compute the distance $d_b(\mu,\nu)$, we need to successively compute $d_n(\mu,\nu)=\Delta^n(\bot)(\mu,\nu)$ by Corollary 2, until an N appears such that $d_N(\mu,\nu)=d_{N+1}(\mu,\nu)$. Then $d_b(\mu,\nu)=d_N(\mu,\nu)$. The tricky point is that there may be several n satisfying $d_n(\mu,\nu)=d_{n+1}(\mu,\nu)$ and we cannot simply choose any of them to be N. We have to make sure that similar condition $d_N(\mu',\nu')=d_{N+1}(\mu',\nu')$ also holds for any μ' and ν' reachable from μ and ν, respectively.

Algorithm 1 gives the details. The main procedure is **distance**(μ,ν), which first uses the procedure **generate** to generate all the generalized states reachable from μ and ν, then repeatedly calls **dist**(n,μ',ν') for computing $d_n(\mu',\nu')$. The procedure **dist**(n,μ,ν) faithfully implements the inductive definition $\Delta^n(\bot)(\mu,\nu)$ for $n\geq 0$. When a sufficiently large number n is obtained such that $d_n(\mu',\nu')=d_b(\mu',\nu')$ for any μ' and ν' reachable from μ and ν, respectively, the algorithm terminates. Therefore, the correctness of the algorithm is ensured by the fixed point characterization of the bisimilarity metric given in Corollary 2.

Proposition 5. For any infinite sequence of transitions

$$\mu_0 \xrightarrow{a_1} \mu_1 \xrightarrow{a_2} \mu_2 \xrightarrow{a_3} \cdots$$

there exists some $i \geq 0$ such that for all $j > i$ we have $ht(\mu_j) = ht(\mu_i)$.

Proof. Suppose for a contraction that $\forall i$, $\exists j > i$ with $ht(\mu_j) \neq ht(\mu_i)$. So in particular, for $i = 0$, $\exists j_1 > 0$ such that $ht(\mu_{j_1}) \neq ht(\mu_0)$. Then for j_1, $\exists j_2 > j_1$ such that $ht(\mu_{j_2}) \neq ht(\mu_{j_1})$. Continuing this process, we will get that $ht(\mu_{j_n}) \neq ht(\mu_{j_{n-1}})$ for any $n \geq 2$. Note that Definition 4 implies that $ht(\mu) \geq ht(\nu)$ for any $\mu \xrightarrow{a} \nu$ with $a \in A$. Hence, we get that $ht(\mu_{j_1}) > ht(\mu_{j_2}) > \cdots$, i.e., an infinite descending chain of non-zero values of $ht(\mu_{j_n})$. However, this is impossible. The reason is as follows. Let $X = \{\mu_0(s') \mid s' \in S\}$ and Y be the set defined by $\{\nu(s') \mid s' \in S, \exists s \in S, s \xrightarrow{a} \nu\}$. Note that the two sets are finite because S is finite and the transitions are finite. Now, by Definition 4, $\mu_1 = \bigcup_{s \in supp(\mu_0)} \mu_0(s) \cdot \nu_s$, where the distribution ν_s is obtained by s in supp(μ_0) performing the action a. For any $s' \in S$, $\mu_1(s')$ only takes values in the set X or Y, implying that $ht(\mu_1)$ only takes values in these two sets. In the same way, each $ht(\mu_{j_n})(n \geq 1)$ only takes value in these two sets, which is in contradiction with the fact that $\{ht(\mu_{j_n}) \mid n \geq 1\}$ has an infinite descending chain. $\square$

This proposition shows that for any given distribution μ, after finitely many transitions, the height of the resulting distributions will not change any more, which is important for the termination of the algorithm that we will soon present.

Theorem 2. Under the $\mathcal{L}_{\otimes}$, the algorithm terminates and is **EXPTIME**.

Proof. We first show the algorithm terminates. Proposition 5 says that for any chain of transitions $\mu \xrightarrow{a} \mu_1 \xrightarrow{b} \mu_2 \cdots$, the height of distributions in this chain will not change any more after finite number of steps. The proof of that proposition also says that the transition systems between distributions are finitary, which is important for the analysis below. For any $\mu, \nu \in \mathcal{F}(S)$ and $n \geq 0$, we know that

$$d_{n+1}(\mu,\nu) \;=\; \max(\bigvee_{a \in A} H_{d_n}(T_a(\mu), T_a(\nu)), \bot(\mu,\nu)). \quad (1)$$

If μ and ν are fixed, the choices on the right hand side are finite: $d_{n+1}(\mu,\nu)$ is determined by either $\bot(\mu,\nu)$ or $d_n(\mu_i,\nu_j)$ with $\mu \xrightarrow{a} \mu_i$ and $\nu \xrightarrow{a} \nu_j$, where there are only finitely many outgoing transitions from μ and ν. However, (1) holds for any $n \geq 0$. It must be the case that either

$$d_{n+1}(\mu,\nu) \;=\; \bot(\mu,\nu), \quad (2)$$

or

$$d_{n+1}(\mu,\nu) \;=\; d_n(\mu_i,\nu_j) \quad (3)$$

for some fixed transitions $\mu \xrightarrow{a} \mu_i$ and $\nu \xrightarrow{a} \nu_j$, but *infinitely* many different n's. In other words, the distance between μ and ν is either a fixed number $\bot(\mu,\nu)$ or determined by a pair of their successor distributions. In the latter case, the same remark can be made again. Eventually, the computation of **dist**(n,μ,ν) $(n \geq 0)$ boils down to the computations of $\bot(\mu_k,\nu_l)$, where μ_k and ν_l are elements in the chains generated by μ and ν, respectively. Hence, when the heights of distributions do not change, the values of $\bot(\mu_k,\nu_l)$ do not change either. Consequently, there exists some N such that **dist**(N,μ',ν') = **dist**$(N+1,\mu',\nu')$ for any μ' and ν' reachable from μ and ν, respectively, hence the algorithm will terminate.

We now consider the time complexity. Let $(S, A, \rightarrow)$ be the **NFTS** under consideration. Assume that the size of the state space is $n = |S|$ and the number of edges (viewing the

Algorithm 1 Compute the distance between two distributions

Input: A finite **NFTS**, and $\mu, \nu \in \mathcal{F}(S)$.
Output: $d_b(\mu,\nu)$.

```
procedure distance(μ, ν)
n ← 1;
for all μ_i ∈ generate(μ) do
   for all ν_j ∈ generate(ν) do
      n_ij ← 1;
      while dist(n_ij − 1, μ_i, ν_j) ≠ dist(n_ij, μ_i, ν_j) do
         n_ij ← n_ij + 1;
      end while
      if n_ij > n then
         n ← n_ij
      end if
   end for
end for
return dist(n, μ, ν);
end procedure

procedure generate(μ)
D ← {μ}
repeat
   for all a ∈ A do
      for all μ' ∈ D do
         D' ← D ∪ T_a(μ')
      end for
   end for
until D' = D
return D

procedure dist(n, μ, ν)
if n = 0 then
   return 1 − (ht(μ') ⇔ ht(ν'));
else
   for all a ∈ A do
      for all μ_i ∈ T_a(μ) do
         for all ν_j ∈ T_a(ν) do
            d_aij ← dist(n − 1, μ_i, ν_j);
         end for
      end for
   end for
   return ⋁_a((⋁_i ⋀_j d_aij) ∨ (⋁_j ⋀_i d_aij)) ∨ dist(0, μ, ν);
end if
end procedure
```

NFTS as a directed graph) labelled with possibility values is m. Suppose the density of the **NFTS** is k. That is, from any state s there are at most k outgoing transitions $s \xrightarrow{a} \mu_i$ with the same label a, for any $a \in A$. In the worst case, the support of μ is S. By Definition 4, from μ there are at most k^n outgoing transitions with the same label. Similarly for ν. Therefore, in the procedure **dist**(l, μ, ν), we will compute **dist**$(l-1, \mu_i, \nu_j)$ at most $|A|k^{2n}$ times. For any $\mu, \nu \in \mathcal{F}(S)$, let t_l be the time of computing **dist**(l, μ, ν). We see that $t_l = |A|k^{2n}t_{l-1}$. It follows that t_l is in $O((|A|k^{2n})^l)$. In the proof of Proposition 5, the sizes of the sets X and Y are bounded by n and m, respectively. Therefore, the procedure **dist**(l, μ', ν') will stabilize after $n+m$ iterations. Note that for each distribution μ' reachable from μ, its support is at most S and the possibility $\mu'(s)$ for any s in its support must be a possibility value that has already appeared and labelled some transition in the original **NFTS**. So the total number of those values is less than the total number m of edges of the graph of the original **NFTS**. It follows that the total number of distributions reachable from μ or ν is at most n^m. Therefore, the overall time consumed is thus in $O(n^m|A|^{n+m}k^{2n^2+2nm})$, which means that the algorithm takes exponential time. □

Remark 1. Let us consider the algorithm in three particular cases.

1) If an **NFTS** is deterministic, i.e. its density k is equal to 1, the above algorithm still takes exponential time. There are two time-consuming tasks: (1) to find some N such that $d_N(\mu, \nu)$ stabilizes; (2) to guarantee that d_N is stable not only for μ and ν, but also for any pair of distributions reachable from μ and ν. This obstacle reminds us of the problem of computing the bisimilarity metric for probabilistic systems [38]–[40], for which there does not seem to exist polynomial time algorithms. For example, van Breugel and Worrell [40] point out that the problem of computing the bisimilarity pseudometric on the state space of a probabilistic automaton is in **PPAD**, which stands for polynomial parity argument in a directed graph. It lies between the search problem versions of **P** and **NP**.
2) Given a finitary **NFTS** without loops, we know that it has a finite tree structure. Suppose the height of the tree is h. If we apply the above algorithm, it is easy to see that after at most $h+1$ rounds of iteration, the approximate computation of the distance between any pair of distributions will stabilize. That is, **dist**$(h, \mu, \nu) =$ **dist**$(h+1, \mu, \nu)$, and similarly for any pair of distributions reachable from μ and ν. Moreover, the procedure **distance** can be simplified: there is no need to call **generate** and it suffices to require the **while** loop to iterate $h+1$ times. The overall time thus consumed is $O(|A|^{h+1}k^{2n(h+1)})$.
3) If the **NFTS** in 2) is deterministic, then the running time of the algorithm becomes $O(|A|^{h+1})$. So in this very particular case the algorithm becomes polynomial.

V. Fuzzy Modal Logic

In this section, we introduce a fuzzy modal logic that can characterize distribution-based bisimilarity metric. Let A be a set of actions ranged over by $a, b, \cdots$, and $\top$ be a fixed proposition, which is taken from the classical Hennessy-Milner Logic [20]. The language $\mathcal{L}$ of formulae is the least set generated by the following **BNF** grammar:

$$\psi \quad ::= \quad \top \mid \psi_1 \wedge \psi_2 \mid \psi \to \bar{p} \mid \bar{p} \to \psi \mid \langle a \rangle \psi$$

where $p \in [0,1]$. We write $\varphi \leftrightarrow \bar{p}$ for $(\varphi \to \bar{p}) \wedge (\bar{p} \to \varphi)$.

Note that the formulae in $\mathcal{L}$ are defined inductively, so we can only generate formulae of finite depth such as $\top$, $\langle a \rangle \top$, $\langle a \rangle \langle b \rangle \top$, $\top \to \bar{p}$, $\bar{p} \to \langle a \rangle \top$, $(\langle a \rangle \top \to \bar{p}) \wedge (\bar{r} \to \langle b \rangle \top)$, etc. However, $\langle a \rangle \top \to \top$ is not a legitimate formula.

Let us fix an **NFTS** $\mathcal{M} = (S, A, \to)$. We define the semantic function $[\![\cdot]\!] : \mathcal{L} \to \mathcal{F}(S) \to [0,1]$ to calculate the degree for a distribution to satisfy a formula.

Definition 10. A formula $\psi \in \mathcal{L}$ evaluates in $\mu \in \mathcal{F}(S)$ as follows:

- $[\![\top]\!](\mu) \stackrel{def}{=} ht(\mu)$,
- $[\![\psi_1 \wedge \psi_2]\!](\mu) \stackrel{def}{=} \min([\![\psi_1]\!](\mu), [\![\psi_2]\!](\mu))$,
- $[\![\psi \to \bar{p}]\!](\mu) \stackrel{def}{=} [\![\psi]\!](\mu) \Rightarrow p$,
- $[\![\bar{p} \to \psi]\!](\mu) \stackrel{def}{=} p \Rightarrow [\![\psi]\!](\mu)$,
- $[\![\langle a \rangle \psi]\!](\mu) \stackrel{def}{=} \max_{\mu \xrightarrow{a} \mu'} [\![\psi]\!](\mu')$.

The roles of the operators in $\mathcal{L}_\otimes$ will be clearer when connecting them with the fuzzy modal logic $\mathcal{L}$.

1) In general, for any $\varphi, \psi \in \mathcal{L}$, $\varphi \vee \psi, \varphi \& \psi$, and $\varphi \to \psi$ need not be fuzzy modal logical formulae. This point is completely different from **BL** and **MTL**;
2) The operator $\min$ is used to interpret the min-conjunction $\wedge$ as **BL** and **MTL**-algebras do;
3) The operator $\max$ is used to interpret the formula $\langle a \rangle \psi$, which specifies the property for a distribution to perform action a and result in a possible distribution to satisfy ψ. Due to nondeterminism, from μ there may be several transitions labelled by the same action a, e.g. $\mu \xrightarrow{a} \mu_i$ with $i \in I$. We take the optimal case by letting $[\![\langle a \rangle \psi]\!](\mu)$ be the maximal $[\![\psi]\!](\mu_i)$ when $i \in I$;
4) The operator $\Rightarrow$ is used to interpret the formula $\psi \to \bar{p}$, where $p \in [0,1]$.

Definition 11. Let $\mathcal{S} = (S, A, \to)$ be an **NFTS**, for any distributions $\mu, \nu \in \mathcal{F}(S)$, the logical metric between μ and ν is defined by $d_l(\mu, \nu) = 1 - \inf_{\psi \in \mathcal{L}}([\![\psi]\!](\mu) \Leftrightarrow [\![\psi]\!](\nu))$.

By Proposition 1 (4), it is not difficult to verify that d_l is a $\oplus$-metric on $\mathcal{F}(S)$. In the following, we use d_l^L, d_l^G, d_l^P, and d_l^N to denote logical metrics based on $\mathcal{L}_\otimes^L$, $\mathcal{L}_\otimes^G$, $\mathcal{L}_\otimes^P$ and $\mathcal{L}_\otimes^N$, respectively.

The rest of this section is devoted to a logical characterization of d_b. It turns out that d_l coincides with d_b if we consider image-finite **NFTS**s. We split the proof of this coincidence result into two parts, to show that one metric is dominated by the other and vice versa.

Lemma 6. For image-finite **NFTS**s, we have $d_l \preceq d_b$.

Proof. Let $(S, A, \to)$ be an image-finite **NFTS**. For any distributions $\mu, \nu \in \mathcal{F}(S)$, we need to show the inequality $d_l(\mu, \nu) \leq d_b(\mu, \nu)$. It is sufficient to prove

$$[\![\psi]\!](\mu) \Leftrightarrow [\![\psi]\!](\nu) \;\geq\; 1 - d_b(\mu, \nu)$$

for all $\psi \in \mathcal{L}$.

We proceed by structural induction on formulae. We first analyze the structure of $\psi \in \mathcal{L}$.

- $\psi \equiv \mathsf{T}$. Then it is trivial to see that

  $$[\![\mathsf{T}]\!](\mu) \Leftrightarrow [\![\mathsf{T}]\!](\nu) = ht(\mu) \Leftrightarrow ht(\nu) \;\geq\; 1 - d_b(\mu, \nu)$$

  because $\bot(\mu, \nu) = 1 - (ht(\mu) \Leftrightarrow ht(\nu)) \leq d_b(\mu, \nu)$ and d_b is a bisimulation metric.
- $\psi \equiv \psi_1 \wedge \psi_2$. We infer that

  $$\begin{array}{rl} & [\![\psi_1 \wedge \psi_2]\!](\mu) \Leftrightarrow [\![\psi_1 \wedge \psi_2]\!](\nu) \\ = & \min([\![\psi_1]\!](\mu), [\![\psi_2]\!](\mu)) \Leftrightarrow \min([\![\psi_1]\!](\nu), [\![\psi_2]\!](\nu)) \\ \geq & \min([\![\psi_1]\!](\mu) \Leftrightarrow [\![\psi_1]\!](\nu), [\![\psi_2]\!](\mu) \Leftrightarrow [\![\psi_2]\!](\nu)) \\ \geq & 1 - d_b(\mu, \nu) \quad \text{(by induction on } \psi_1 \text{ and } \psi_2\text{).} \end{array}$$

  The second step holds because of Proposition 1 (5).
- $\psi \equiv \psi_1 \to \bar{p}$. We infer that

  $$\begin{array}{rl} & [\![\psi_1 \to \bar{p}]\!](\mu) \Leftrightarrow [\![\psi_1 \to \bar{p}]\!](\nu) \\ = & ([\![\psi_1]\!](\mu) \Rightarrow p) \Leftrightarrow ([\![\psi_1]\!](\nu) \Rightarrow p) \\ \geq & [\![\psi_1]\!](\mu) \Leftrightarrow [\![\psi_1]\!](\nu) \quad \text{(by Proposition 1 (6))} \\ \geq & 1 - d_b(\mu, \nu) \quad \text{(by induction on } \psi_1\text{).} \end{array}$$

- $\psi \equiv \bar{p} \to \psi_1$. Similar to the last case by using Proposition 1 (7).
- $\psi \equiv \langle a \rangle \psi'$. Let μ' be the distribution such that $\mu \xrightarrow{a} \mu'$ and $[\![\langle a \rangle \psi']\!](\mu) = [\![\psi']\!](\mu')$. Since d_b is a distribution-based bisimulation metric, by Definition 8, there exists some ν' such that $\nu \xrightarrow{a} \nu'$ and $d_b(\mu', \nu') \leq d_b(\mu, \nu)$. Without loss of generality, we assume $[\![\psi]\!](\mu) \geq [\![\psi]\!](\nu)$. In this case, $[\![\psi']\!](\mu') \geq [\![\psi']\!](\nu')$. We distinguish two possibilities:
  - $[\![\psi]\!](\mu) = [\![\psi]\!](\nu)$. Then

    $$\begin{array}{rl} & [\![\psi]\!](\mu) \Leftrightarrow [\![\psi]\!](\nu) \\ = & 1 \\ \geq & 1 - d_b(\mu, \nu) \end{array}$$

  - $[\![\psi]\!](\mu) > [\![\psi]\!](\nu)$. Then

    $$\begin{array}{rl} & [\![\psi]\!](\mu) \Leftrightarrow [\![\psi]\!](\nu) \\ = & [\![\psi]\!](\mu) \Rightarrow [\![\psi]\!](\nu) \quad \text{(by Proposition 1 (1))} \\ = & [\![\psi']\!](\mu') \Rightarrow \max_{\nu \xrightarrow{a} \nu''} [\![\psi']\!](\nu'') \\ \geq & [\![\psi']\!](\mu') \Rightarrow [\![\psi']\!](\nu') \\ \geq & 1 - d_b(\mu', \nu') \quad \text{(by induction on } \psi'\text{).} \\ \geq & 1 - d_b(\mu, \nu). \end{array}$$

□

Lemma 7. For image-finite **NFTS**s, we have $d_b \preceq d_l$.

Proof. Since d_b is the least distribution-based bisimulation metric, it suffices to prove that d_l is a distribution-based bisimulation metric. Let $(S, A, \to)$ be an image-finite **NFTS** and $\mu, \nu \in \mathcal{F}(S)$ be any two distributions. First, we prove the inequality $\bot \preceq d_l$. This holds because

$$\begin{array}{rcl} d_l(\mu, \nu) & = & 1 - \inf_{\psi \in \mathcal{L}}([\![\psi]\!](\mu) \Leftrightarrow [\![\psi]\!](\nu)) \\ & \geq & 1 - ([\![\mathsf{T}]\!](\mu) \Leftrightarrow [\![\mathsf{T}]\!](\nu)) \\ & = & 1 - (ht(\mu) \Leftrightarrow ht(\nu)) \\ & = & \bot(\mu, \nu). \end{array}$$

Now, assume that $\mu \xrightarrow{a} \mu'$ and $d_l(\mu, \nu) = r < 1$. We need to show that there exists some transition $\nu \xrightarrow{a} \nu'$ with $d_l(\mu', \nu') \leq d_l(\mu, \nu)$. Suppose for a contradiction that no a-transition from ν satisfies this condition. Then for each ν_i $(i \in I)$ with $\nu \xrightarrow{a} \nu_i$ we have $d_l(\mu', \nu_i) > r$. There must exist some formula $\psi_i \in \mathcal{L}$ such that $1 - ([\![\psi_i]\!](\mu') \Leftrightarrow [\![\psi_i]\!](\nu_i)) > r$, that is, $[\![\psi_i]\!](\mu') \Leftrightarrow [\![\psi_i]\!](\nu_i) < 1 - r$, for each $i \in I$. Note that by assumption the **NFTS** is image-finite, so the index set I is finite. Define $\psi'_i = \psi_i \leftrightarrow \overline{[\![\psi_i]\!](\mu')}$. Then for all $i \in I$,

$$[\![\psi'_i]\!](\mu') = 1 \text{ and } [\![\psi'_i]\!](\nu_i) \;=\; [\![\psi_i]\!](\mu') \Leftrightarrow [\![\psi_i]\!](\nu_i) \;<\; 1 - r.$$

Let $\psi = \langle a \rangle \bigwedge_{i \in I} \psi'_i$. Then we infer that

$$\begin{array}{rcl} [\![\psi]\!](\mu) & = & \max_{\mu \xrightarrow{a} \mu''} [\![\bigwedge_{i \in I} \psi'_i]\!](\mu'') \\ & \geq & [\![\bigwedge_{i \in I} \psi'_i]\!](\mu') \\ & = & \min_{i \in I} [\![\psi'_i]\!](\mu') \\ & = & 1. \end{array}$$

On the other hand, we have

$$\begin{array}{rcl} [\![\psi]\!](\nu) & = & \max_{\nu \xrightarrow{a} \nu_i} [\![\bigwedge_{j \in I} \psi'_j]\!](\nu_i) \\ & = & \max_{\nu \xrightarrow{a} \nu_i} \min_{j \in I} [\![\psi'_j]\!](\nu_i) \\ & \leq & \max_{\nu \xrightarrow{a} \nu_i} [\![\psi'_i]\!](\nu_i) \\ & < & 1 - r. \end{array}$$

It follows that

$$\begin{array}{rcl} d_l(\mu, \nu) & = & 1 - \inf_{\psi' \in \mathcal{L}}([\![\psi']\!](\mu) \Leftrightarrow [\![\psi']\!](\nu)) \\ & \geq & 1 - [\![\psi]\!](\mu) \Leftrightarrow [\![\psi]\!](\nu) \\ & = & 1 - [\![\psi]\!](\nu) \quad \text{(by Proposition 1 (2))} \\ & > & 1 - (1 - r) \\ & = & r \end{array}$$

which is in contradiction to $d_l(\mu, \nu) = r$. □

By combining the above two lemmas, we arrive at the following logical characterization theorem.

Theorem 3. For image-finite **NFTS**s, we have $d_l = d_b$.

For any $\mu, \nu \in \mathcal{F}(S)$ and any $\psi \in \mathcal{L}$, the proof of Lemma 6 implies $[\![\psi]\!](\mu) = [\![\psi]\!](\nu)$ when $\mu \sim \nu$. It follows that $\mu \sim \nu$ implies $d_l(\mu, \nu) = 0$. On the other hand, define the binary relation $\mathcal{R} \overset{def}{=} \{(\mu, \nu) \mid d_l(\mu, \nu) = 0\}$. Then $\mathcal{R}$ is obvious an equivalence relation. The proof of Lemma 7 implies that $\mathcal{R}$ is a distribution-based bisimulation. Hence $d_l(\mu, \nu) = 0$ implies $\mu \sim \nu$. Therefore, two distributions are bisimilar if and only of the logical distance between them is 0.

Corollary 4. Let μ and ν be two distributions over an image-finite **NFTS**. Then $\mu \sim \nu$ iff $d_l(\mu, \nu) = d_b(\mu, \nu) = 0$ iff $[\![\psi]\!](\mu) = [\![\psi]\!](\nu)$ for all $\psi \in \mathcal{L}$.

The above property can be used to judge whether two distributions are bisimilar.

Example 2. According to Corollary 4 and the computation results of Example 1, we can see that $\bar{s}$ and $\bar{t}$ are bisimilar, whereas $\bar{s}_1$ and $\bar{t}_2$ are not.

In addition, with Theorem 3 sometimes we can easily calculate distribution-based bisimilarity metric in terms of d_l. For instance, consider Example 1 again. The only relevant formula is $\langle b\rangle\top$, and it is direct to calculate that $[\![\langle b\rangle\top]\!](\bar{s}_1) = 0.8$ and $[\![\langle b\rangle\top]\!](\bar{t}_2) = 0.6$. Thus $[\![\langle b\rangle\top]\!](\bar{s}_1) \Leftrightarrow [\![\langle b\rangle\top]\!](\bar{t}_2) = 0.6$ under the Gödel implication. It follows from Theorem 3 that $d_b^G(\bar{s}_1, \bar{t}_2) = d_l^G(\bar{s}_1, \bar{t}_2) = 1 - 0.6 = 0.4$.

VI. Non-Expansiveness

In this section, we show the non-expansiveness of our distribution-based bisimilarity metric with respect to a parallel composition in the style of CSP [15]. It means that if the difference (with respect to bisimilarity metric) between μ_i and ν_i is ϵ_i, then the difference between $f(\mu_1, \cdots, \mu_n)$ and $f(\nu_1, \cdots, \nu_n)$ is no more than $\bigvee_{i=1}^n \epsilon_i$ [2], where f is a parallel composition operator with n arguments. This non-expansiveness makes compositional verification possible.

Given an **NFTS** $(S, A, \to)$ and two possibility distributions $\mu, \nu \in \mathcal{F}(S)$, the key to define a parallel composition is how to appropriately define the fuzzy set represented by $\mu\|\nu$. For example, let $\mu = \frac{0.3}{s_1} + \frac{0.5}{s_2}$ and $\nu = \frac{0.4}{s_3}$. Intuitively, in their parallel composition, (s_1, s_3) appears with possibility 0.3 and (s_2, s_3) appears with possibility of 0.4. That is, $\mu\|\nu$ is the fuzzy set $\frac{0.3}{(s_1,s_3)} + \frac{0.4}{(s_2,s_3)}$. Hence, $\mu\|\nu$ is given by

$$(\mu\|\nu)(s,t) = \mu(s) \wedge \nu(t)$$

for all $(s,t) \in S \times S$. Clearly, we have $\mu\|\nu \in \mathcal{F}(S \times S)$ and in general, $(\mu\|\nu)$ and $(\nu\|\mu)$ are different fuzzy sets on $\mathcal{F}(S \times S)$. For instance, $(\nu\|\mu) = \frac{0.3}{(s_3,s_1)} + \frac{0.4}{(s_3,s_2)}$. Note that $(\bar{s}\|\bar{t})(s', t') = 1$ if $(s', t') = (s,t)$ and 0 otherwise. The above fuzzy set $\mu\|\nu$ can be understood as $\frac{0.3}{\bar{s}_1\|\bar{s}_3} + \frac{0.4}{\bar{s}_2\|\bar{s}_3}$. For simplicity, we write $\frac{0.3}{s_1\|s_3} + \frac{0.4}{s_2\|s_3}$ instead of $\frac{0.3}{\bar{s}_1\|\bar{s}_3} + \frac{0.4}{\bar{s}_2\|\bar{s}_3}$.

The parallel composition operator we are going to define is asynchronous in the sense that the components can either synchronize or act independently. Given an **NFTS** $(S, A, \to)$, for any $\mu, \nu \in \mathcal{F}(S)$, the events intended to synchronize at μ and ν are listed in the set $A_\mu \cap A_\nu$ and the rest of the events can be performed independently, where A_μ stands for the set $\{a \in A : \exists \mu' \neq \emptyset, \mu \xrightarrow{a} \mu'\}$. Formally, the *parallel composition* $\mu\|\nu$ can perform action a in four different ways:

$$T_a(\mu\|\nu) = \begin{cases} \{\mu'\|\nu' : \mu' \in T_a(\mu), \nu' \in T_a(\nu)\} & \text{if } a \in A_\mu \cap A_\nu \\ \{\mu'\|\nu : \mu' \in T_a(\mu)\} & \text{if } a \in A_\mu \backslash A_\nu \\ \{\mu\|\nu' : \nu' \in T_a(\nu)\} & \text{if } a \in A_\nu \backslash A_\mu \\ \emptyset & \text{otherwise.} \end{cases}$$

Let us present an example of the parallel composition and compute the bisimilarity metric between distributions.

Example 3. Let $\mathcal{S} = (\{s_1, s_2, s_3, s_4\}, \{a\}, \to)$ be an **NFTS**, where $\to$ is defined by the transitions $\bar{s}_1 \xrightarrow{a} \{\frac{0.4}{s_3} + \frac{0.6}{s_4}\}$, $\bar{s}_2 \xrightarrow{a} \{\frac{0.6}{s_3} + \frac{0.3}{s_4}\}$, and $\bar{s}_3 \xrightarrow{a} \{\frac{0.4}{s_4}\}$. Then,

$$\bar{s}_1\|\bar{s}_2 \xrightarrow{a} \{\mu : \mu = \tfrac{0.4}{s_3\|s_3} + \tfrac{0.3}{s_3\|s_4} + \tfrac{0.6}{s_4\|s_3} + \tfrac{0.3}{s_4\|s_4}\}$$
$$\bar{s}_2\|\bar{s}_3 \xrightarrow{a} \{\nu : \nu = \tfrac{0.4}{s_3\|s_4} + \tfrac{0.3}{s_4\|s_4}\}$$

If we take the fuzzy logic $\mathcal{L}_\otimes^L$, then by a direct computation shown in Example 1, we can obtain that $d_b^L(\bar{s}_1, \bar{s}_2) = 0$ and $d_b^L(\bar{s}_2, \bar{s}_3) = 0.4$.

Now, for $n \geq 2$, we have that

$$\begin{aligned} & d_n(\bar{s}_1\|\bar{s}_2, \bar{s}_2\|\bar{s}_3) \\ = \ & \Delta(d_{n-1})(\bar{s}_1\|\bar{s}_2, \bar{s}_2\|\bar{s}_3) \\ = \ & \max(d_{n-1}(\mu,\nu), 1 - (ht(\bar{s}_1\|\bar{s}_2) \Leftrightarrow_L ht(\bar{s}_2\|\bar{s}_3))) \\ = \ & d_{n-1}(\mu,\nu) \\ = \ & \Delta(d_{n-2})(\mu,\nu) \\ = \ & \max(d_{n-2}(\tfrac{0.4}{s_4\|s_4}, \tfrac{0.4}{s_4\|s_4}), 1 - (ht(\mu) \Leftrightarrow_L ht(\nu))) \\ = \ & \max(0, 0.2) \\ = \ & 0.2. \end{aligned}$$

Hence $d_b^L(\bar{s}_1\|\bar{s}_2, \bar{s}_2\|\bar{s}_3) = 0.2$.

In a similar way, one can compute this distance under other fuzzy logics, which are shown in the following diagram.

	$(\bar{s}_1, \bar{s}_2)$	$(\bar{s}_2, \bar{s}_3)$	$(\bar{s}_1\|\bar{s}_2, \bar{s}_2\|\bar{s}_3)$
d_b^L	0	0.4	0.2
d_b^G	0	1	0.6
d_b^P	0	1	$\frac{1}{3}$
d_b^N	0	0.6	0.6

It is obvious that $d_b(\bar{s}_1\|\bar{s}_2, \bar{s}_2\|\bar{s}_3) \leq d_b(\bar{s}_1, \bar{s}_2) \vee d_b(\bar{s}_2, \bar{s}_3)$, i.e., the parallel composition is non-expansive. More generally, we have the following theorem.

Theorem 4. Let $(S, A, \to)$ be an **NFTS** and $\mu_i, \nu_i \in \mathcal{F}(S)$ with $d_b(\mu_i, \nu_i) = \epsilon_i$, where $i = 1, 2$. Then we have

$$d_b(\mu_1\|\mu_2, \nu_1\|\nu_2) \ \leq \ \epsilon_1 \vee \epsilon_2.$$

Proof. Let $\mathcal{F}(S)\|\mathcal{F}(S) = \{\mu\|\nu \mid \mu, \nu \in \mathcal{F}(S)\}$. We define the function $D : (\mathcal{F}(S)\|\mathcal{F}(S)) \times (\mathcal{F}(S)\|\mathcal{F}(S)) \to [0,1]$ by

$$D(\mu\|\mu', \nu\|\nu') \ \stackrel{def}{=} \ d_b(\mu,\nu) \vee d_b(\mu', \nu')$$

for any $\mu\|\mu', \ \nu\|\nu' \in \mathcal{F}(S)\|\mathcal{F}(S)$.

It is easy to check that D is a $\oplus$-metric on $\mathcal{F}(S)\|\mathcal{F}(S)$. We claim that D is a pre-fixed point of the function $\Delta : \mathcal{D}_{\mathcal{F}(S)\|\mathcal{F}(S)} \to \mathcal{D}_{\mathcal{F}(S)\|\mathcal{F}(S)}$. Let us verify it.

For any $\mu_1\|\mu_2, \nu_1\|\nu_2 \in \mathcal{F}(S)\|\mathcal{F}(S)$, we first check that $1 - (ht(\mu_1\|\mu_2) \Leftrightarrow ht(\nu_1\|\nu_2)) \ \leq \ D(\mu_1\|\mu_2, \nu_1\|\nu_2)$. It holds because

$$\begin{aligned} & D(\mu_1\|\mu_2, \nu_1\|\nu_2) \\ = \ & d_b(\mu_1, \nu_1) \vee d_b(\mu_2, \nu_2) \\ \geq \ & [1 - (ht(\mu_1) \Leftrightarrow ht(\nu_1))] \vee [1 - (ht(\mu_2) \Leftrightarrow ht(\nu_2))] \\ = \ & 1 - (ht(\mu_1) \Leftrightarrow ht(\nu_1)) \wedge (ht(\mu_2) \Leftrightarrow ht(\nu_2)) \\ \geq \ & 1 - (ht(\mu_1) \wedge ht(\mu_2) \Leftrightarrow ht(\nu_1) \wedge ht(\nu_2)) \\ = \ & 1 - (ht(\mu_1\|\mu_2) \Leftrightarrow ht(\nu_1\|\nu_2)). \end{aligned}$$

The second step holds because d_b is the bisimilarity metric, and the fourth step holds because of Proposition 1 (5). The last equality holds because of the following facts: there exist s_0, t_0 such that $ht(\mu_1) = \mu_1(s_0)$ and $ht(\mu_2) = \mu_2(t_0)$, then $ht(\mu_1\|\mu_2) = \sup_{(s,t)\in S\times S} \mu_1(s) \wedge \mu_2(t) = \mu_1(s_0) \wedge \mu_2(t_0) = ht(\mu_1) \wedge ht(\mu_2)$; similarly, $ht(\nu_1\|\nu_2) = ht(\nu_1) \wedge ht(\nu_2)$.

Next, suppose that $\mu_1\|\mu_2 \xrightarrow{a} \mu_1'\|\mu_2'$. We need to show that there exists some $\nu_1'\|\nu_2'$ such that $\nu_1\|\nu_2 \xrightarrow{a} \nu_1'\|\nu_2'$ and $D(\mu_1'\|\mu_2', \nu_1'\|\nu_2') \ \leq \ D(\mu_1\|\mu_2, \nu_1\|\nu_2)$. There are four cases: $a \in A_{\mu_1} \cap A_{\mu_2}$, $a \in A_{\mu_1} \backslash A_{\mu_2}$, $a \in A_{\mu_2} \backslash A_{\mu_1}$, and $a \notin A_{\mu_1} \cup A_{\mu_2}$. We only give the details for the first case; the other cases are simpler and can be proved in a similar way.

In the case of $a \in A_{\mu_1} \cap A_{\mu_2}$, we get by the definition of the parallel composition that there are $\mu_1 \xrightarrow{a} \mu_1'$ and $\mu_2 \xrightarrow{a} \mu_2'$. Since the least fixed point of Δ is a pre-fixed point as well. By Lemma 4 there are ν_1' and ν_2' such that $\nu_1 \xrightarrow{a} \nu_1'$ and $\nu_2 \xrightarrow{a} \nu_2'$ with $d_b(\mu_1', \nu_1') \leq d_b(\mu_1, \nu_1)$ and $d_b(\mu_2', \nu_2') \leq d_b(\mu_2, \nu_2)$. Clearly, we have $\nu_1 \| \nu_2 \xrightarrow{a} \nu_1' \| \nu_2'$ and it remains to verify that $D(\mu_1' \| \mu_2', \nu_1' \| \nu_2') \leq D(\mu_1 \| \mu_2, \nu_1 \| \nu_2)$. This follows because it is easy to see that

$$d_b(\mu_1', \nu_1') \vee d_b(\mu_2', \nu_2') \ \leq \ d_b(\mu_1, \nu_1) \vee d_b(\mu_2, \nu_2).$$

Based on the claim, it follows from the definition of d_b that $d_b \preceq D$, which implies

$$\begin{array}{rcl} d_b(\mu_1 \| \mu_2, \nu_1 \| \nu_2) & \leq & D(\mu_1 \| \mu_2, \nu_1 \| \nu_2) \\ & = & d_b(\mu_1, \nu_1) \vee d_b(\mu_2, \nu_2) \\ & = & \epsilon_1 \vee \epsilon_2. \end{array}$$

This completes the proof. □

As a consequence, we can show that the parallel composition preserves distribution-based bisimilarity.

Corollary 5. If $\mu_1 \sim \nu_1$ and $\mu_2 \sim \nu_2$, then $\mu_1 \| \mu_2 \sim \nu_1 \| \nu_2$.

Proof. It is straightforward by using Theorem 4 and Corollary 4. □

VII. An Illustrative example

We have seen from Corollary 4 that the problem of checking whether two distributions are bisimilar can be converted into a logical judgment of whether the two distributions have the same values on the same set of logical formulae, which can benefit from traditional logic theories and be assisted by some practical tools. In particular, it is very useful to judge the non-bisimilarity of two distributions because we only need to find a formula that witnesses the difference of the two distributions.

As an application, we consider an example related to medical diagnosis and treatment, as described by Cao et al. [2] and Qiu [41] (see also, [5], [42]). We assume that there is an unknown bacterial infection. Based on his experience, a physician believes that two drugs, say a_1 and a_2, may be useful for treating this disease. Three possible negative symptoms, e.g., $b_1,\ b_2,\ b_3$, must also be considered. A patient's condition can be in one of three rough types, e.g., "poor", "fair", and "excellent", which are denoted by the capital letters P, F, and E, respectively. A treatment (or a negative symptom) may lead to a state among multiple possible states with certain degree. For example, the transition $F \xrightarrow{a_1[0.6]} E$ means that the patient's condition has changed from "fair" to "excellent" with possibility 0.6 after using drug a_1, whereas $F \xrightarrow{b_1[0.3]} P$ means that the patient's condition has changed from "fair" to "poor" with possibility 0.3 if the patient has negative symptom b_1. The transition possibilities of these events are estimated by the physician. Different physical conditions of patients may lead to nondeterministic changes even if the patients are in the same state and are given the same treatment. For example, we may have the following two transition:

$$P \xrightarrow{a_1} \{\frac{0.1}{P} + \frac{0.9}{F}\} \qquad \text{and} \qquad P \xrightarrow{a_1} \{\frac{0.6}{F} + \frac{0.2}{E}\}.$$

So we obtain an **NFTS** $(S, A, \rightarrow)$, where $S = \{P, F, E\}$, $A = \{a_1, a_2, b_1, b_2, b_3\}$ and $\rightarrow \subseteq \mathcal{F}(S) \times A \times \mathcal{F}(S)$.

In order to compare with the work of Cao et al., we revisit Example 4 in [2]. According to the physician's estimation, the transition possibilities of these events among states are as follows.

$$\begin{array}{l} \overline{P} \xrightarrow{a_1} \{\frac{0.1}{P} + \frac{0.9}{F}\}, \quad \overline{P} \xrightarrow{a_2} \{\frac{0.6}{P} + \frac{0.4}{F}\} \\ \overline{F} \xrightarrow{a_1} \{\frac{0.5}{F} + \frac{0.4}{E}\}, \quad \overline{F} \xrightarrow{a_2} \{\frac{0.3}{F} + \frac{0.8}{E}\} \\ \overline{F} \xrightarrow{b_1} \{\frac{0.2}{P}\}, \quad \overline{F} \xrightarrow{b_2} \{\frac{0.3}{P}\}, \quad \overline{E} \xrightarrow{b_3} \{\frac{0.1}{P} + \frac{0.3}{F}\}. \end{array}$$

Suppose now that there are two patients: Alice and Bob. The physician describes the patients' states as $S_A = [0.8, 0.2, 0.1]$ and $S_B = [0.8, 0.3, 0]$, respectively. The fuzzy value vector, say S_B, means that Bob is in "poor" state with membership 0.8, and in "fair" state with membership 0.3. For simplicity, we assume that Alice and Bob have no negative symptoms when the drugs are used for the first time in therapy. It turns out that

$$\begin{array}{lcl} \overline{S_A} & \xrightarrow{a_1} & \{\frac{0.1}{P} + \frac{0.8}{F} + \frac{0.2}{E}\} \\ \overline{S_A} & \xrightarrow{a_2} & \{\frac{0.6}{P} + \frac{0.4}{F} + \frac{0.2}{E}\} \\ \overline{S_B} & \xrightarrow{a_1} & \{\frac{0.1}{P} + \frac{0.8}{F} + \frac{0.3}{E}\} \\ \overline{S_B} & \xrightarrow{a_2} & \{\frac{0.6}{P} + \frac{0.4}{F} + \frac{0.3}{E}\} \end{array}$$

We first show by Corollary 4 that the states of Alice and Bob are not bisimilar, which is not easy to judge directly by the definition of distribution-based bisimulation. In fact, we can find a formula $\langle a_1 \rangle \langle b_3 \rangle \top$ such that $[\![\langle a_1 \rangle \langle b_3 \rangle \top]\!](\overline{S_A}) = 0.2$ and $[\![\langle a_1 \rangle \langle b_3 \rangle \top]\!](\overline{S_B}) = 0.3$ and hence the states of Alice and Bob are not bisimilar.

Furthermore, we can obtain more quantitative information by measuring the distance between their states. Using $\mathcal{L}_{\otimes}^{N}$, it is not difficult to see that $d_b^N(\overline{S_A}, \overline{S_B}) = 0.3$ by a direct computation shown in Example 1. In the same way, we can also obtain that $d_b^N(\overline{S_A}, \overline{S_B}) = 0.5$ where $S_A = [0.5, 0.5, 0]$ and $S_B = [0, 0.5, 0.5]$. By Corollary 3, $\mathcal{S}(S_A, S_B) = 0.7$ and $\mathcal{S}(S_A, S_B) = 0.5$ in the above two cases, respectively, which measures the similarity between Alice's and Bob's conditions in the progress of treatment. A little surprise is that $d_b^N(\overline{S_A}, \overline{S_B})$ is equal to $d_f(S_A, S_B)$ in the above two cases, respectively, where d_f is the behavioural distance of Cao et al. [2]. However, Figure 1, Example 1, Example 2, Corollary 4 in this paper and Theorem 4 in [2] tell us that d_b^N is not always equal to d_f for any $s, t \in S$.

On the other hand, if we take a different fuzzy logic, then $d_b(\overline{S_A}, \overline{S_B})$ is in general different. For example, under $\mathcal{L}_{\otimes}^{P}$, we get that $d_b^P(\overline{S_A}, \overline{S_B}) = 0.4$ when $S_A = [0.5, 0.5, 0]$ and $S_B = [0, 0.5, 0.5]$, but $d_b^P(\overline{S_A}, \overline{S_B}) = \frac{1}{3}$ when $S_A = [0.8, 0.2, 0.1]$ and $S_B = [0.8, 0.3, 0]$, respectively.

VIII. Related Work

Simulations and bisimulations for fuzzy (or lattice-valued) systems have received much attention in the field. Errico and Loreti [9] proposed a notion of bisimulation and applied it to fuzzy reasoning. Kupferman and Lusting [11] studied a latticed simulation between two lattice-valued Kripke structures, which are applied to latticed games. P. Eleftheriou et al. [10] defined bisimulation for Heyting-valued modal languages. Pan et al. [12] investigated simulation for lattice-valued doubly labeled transition systems. Cao et al. [1] defined a bisimulation relation between two different **FTS**s by a correlational pair based on some relation. Ćirić et al. [3] introduced four types of bisimulations (forward, backward, forward-backward, and backward-forward) for fuzzy automata. Sun et al. [13] investigated forward and backward bisimulations for fuzzy automata. Qiu and Deng [4], and Xing et al. [5] addressed the supervisory control of fuzzy discrete event systems by using simulation equivalence and bisimulation equivalence, respectively. Fan [6] defined bisimulations for fuzzy Kripke structures. Damljanović et al. [8] also studied simulation and bisimulation for weighted automata in a similar manner, as also described in [3], [7]. Wu and Deng [14] considered bisimulations for **FTS**s.

All these approaches can be divided into two classes. In the first class, (bi)simulations are based on a crisp relation on the state space, and thus one state is either (bi)similar to another state or not. As [1], [5], [8]–[10], [13], [14], the present study belongs to this class. In the second class, (bi)simulations are based on a fuzzy relation (or a lattice-valued relation) on the state space, which shows the degree to which one state is (bi)similar to another. This approach was adapted in [3], [4], [6], [11], [12]. Recently, lattice-valued nondeterministic fuzzy automata have also been proposed [43].

This paper investigates distribution-based bisimulation, which is a crisp relation and to our knowledge, remains under-explored in fuzzy systems. We just compare our distribution-based bisimulation with the state-based bisimulation given by Cao et al [2].

An equivalence relation $R \subseteq S \times S$ is a state-based bisimulation if $s \; R \; t$ implies that whenever $s \xrightarrow{a} \mu$, there exists some transition $t \xrightarrow{a} \nu$ with $\mu([s']) = \nu([s'])$ for all R-equivalence classes $[s']$.

We can prove that if R is a state-based bisimulation, then the lifted relation $R^{\dagger}$ (see Definition 8 in [2]) is a distribution-based bisimulation. This means that for any $s, t \in S$, if s and t are related by a state-based bisimulation then $\bar{s}$ and $\bar{t}$ are related by a distribution-based bisimulation. However, the inverse does not hold, as can be seen from Figure 1 and Example 2. In this sense, we say that distribution-based bisimilarity is coarser than state-based bisimilarity.

As modal logics, Wu and Deng [14] gave a fuzzy style Hennessy-Milner logic, which characterizes bisimulations for **FTS**s soundly and completely. There the formula is interpreted on (bare) states and is two-valued, i.e., a state either satisfies a formula or not. Fan [6] characterized her fuzzy bisimulation in terms of Gödel modal logic $G(\Box\Diamond)$ as follows:

$$\varphi ::= p \mid \bar{c} \mid \varphi_1 \to \varphi_2 \mid \varphi_1 \wedge \varphi_2 \mid \Box\varphi \mid \Diamond\varphi.$$

Its semantic interpretation is given in terms of (bare) states and is quantitative, i.e., the semantic domain is the unit interval $[0, 1]$. This logic is different from the logic in the current work both in the syntax and in the semantics. In addition, the formula $\varphi \leftrightarrow \bar{p}$ in our logic plays a key role in proving Theorem 3. One can ask whether we can use $\psi_1 \to \psi_2$ instead of $\psi \to \bar{p}$ in our setting. The answer is negative. The reason is that Lemma 6 does not hold for $\psi \equiv \psi_1 \to \psi_2$. In addition, Fan [44] has pointed out that fuzzified Hennessy-Milner Theorem does not hold for modal logics based on Łukasiewicz t-norm and the product t-norm. However, this theorem holds in our framework under a class of fuzzy logics generated by a t-norm including Łukasiewicz t-norm and the product t-norm. Therefore, the results in this paper provide an answer to the question of Fan [44]: "to investigate alternative definitions of fuzzy bisimulation that can satisfy the Hennessy-Milner Theorem with respect to Łukasiewicz and product structures." In [10] Eleftheriou gave Heyting-valued modal logics $L^{\mathcal{H}}_{\Box\Diamond}$. The relationship between $L^{\mathcal{H}}_{\Box\Diamond}$ and $G(\Box\Diamond)$, and the corresponding bisimulations have been discussed in detail [6]. Another direction concerning modal logics is to investigate model checking of nondeterministic systems. For example, Li et al. [45], [46] studied quantitative computation tree logic for model checking based on possibility measures.

IX. Conclusion and Future Work

We have proposed a new notion of bisimulation, which is distribution-based. We have introduced an abstract metric based on triangular-conorm and constructed a distribution-based bisimilarity metric for measuring the behavioural distance between generalized states in **NFTS**s. The bisimilarity metric is a quantitative analogue of bisimilarity. The smaller the distance, the more alike the distributions are. In particular, two distributions are bisimilar if and only if they have distance 0. Moreover, we have given a fuzzy modal logic to characterize distributions-based bisimilarity metric soundly and completely. In addition, we have also shown that the CSP-style parallel composition with respect to bisimilarity metric is non-expansive. Unlike the exact notion of bisimulation, bisimilarity metric is more robust for fuzzy systems, which enables us to investigate the approximate equivalence of fuzzy systems.

There are several problems that are worth further study. First, it would be interesting to find an appropriate logic to characterize the behavioural distance d_f of Cao et al. [2]. In fact, we have a logic that can characterize the state-based bisimulation of Cao et al., but cannot characterize d_f. Second, it is unclear if our methodology can be generalized to the general framework of lattices. One subtlety lies in Lemma 7. In the current setting, $\sup\{x \in A \mid x < r\} < r$ holds when A is a finite subset of the unit interval $[0, 1]$. However, this property fails if we deal with a general lattice. Hence, it seems necessary to give a new proof of Theorem 3 or to change the modal logic. Last but not least, we would like to figure out whether it is possible and how to improve the algorithm for computing the bisimilarity metric so that they can be used in practical analysis of fuzzy systems.

X. Acknowledgment

The authors would like to thank the anonymous reviewers for their invaluable suggestions and comments.

REFERENCES

[1] Y. Cao, G. Chen, and E. E. Kerre, "Bisimulations for fuzzy transition systems," *IEEE Trans. Fuzzy Syst.,* vol. 19, no. 3, pp. 540-552, Jun. 2010.

[2] Y. Cao, S. X. Sun, H. Wang, and G. Chen, "A behavioral distance for fuzzy-transition systems," *IEEE Trans. Fuzzy Syst.,* vol. 21, no. 4, pp. 735-747, Aug. 2013.

[3] M. Ćirić, J. Ignjatović, N. Damljanović, and M. Bašić, "Bisimulations for fuzzy automata," *Fuzzy Sets Syst.,* vol. 186, no. 1, pp. 100-139, 2012.

[4] W. L. Deng and D. W. Qiu, "Supervisory control of fuzzy discrete event systems for simulation equivalence," *IEEE Trans. Fuzzy Syst.,* vol. 23, no. 1, pp. 178-192, Feb. 2015.

[5] H. Y. Xing, Q. S. Zhang, and K. S. Huang, "Analysis and control of fuzzy discrete event systems using bisimulation equivalence," *Theor. Comput. Sci.,* vol. 456, pp. 100-111, 2012.

[6] T. F. Fan, "Fuzzy bisimulation for Gödel logic," *IEEE Trans. Fuzzy Syst.,* vol. 23, no. 6, pp. 2387-2396, Dec. 2015.

[7] M. Ćirić, J. Ignjatović, I. Jančić, and N. Damljanović, "Computation of the greatest simulations and bisimulations between fuzzy automata," *Fuzzy Sets Syst.,* vol. 208, pp. 22-42, 2012.

[8] N. Damljanović, M. Ćirić, and J. Ignjatović, "Bisimulations for weighted automata over an additively idempotent semiring," *Theor. Comput. Sci.,* vol. 534, pp. 86-100, 2014.

[9] L. D'Errico and M. Loreti, "A process algebra approach to fuzzy reasoning," in *Proc. IFSA-EUSFLAT 2009,* pp. 1136-1141, 2009.

[10] P. Eleftheriou, C. Koutras, and C. Nomikos, "Notions of bisimulation for Heyting-valued modal languages," *J. Logic Comput.,* vol. 22, no. 2, pp. 213-235, 2012.

[11] O. Kupferman and Y. Lustig, "Latticed simulation relations and games," *Int. J. Found. Comput. S.,* vol. 21, no. 2, pp. 167-189, 2010.

[12] H. Y. Pan, Y. Z. Cao, M. Zhang, and Y. X. Chen, "Simulation for lattice-valued doubly labeled transition systems," *Int. J. Approx. Reason.,* vol. 55, pp. 797-811, 2014.

[13] D. D. Sun, Y. M. Li, and W. W. Yang, "Bisimulation relations for fuzzy finite automata," *Fuzzy Syst. Math.,* vol. 23, no. 4, pp. 92-100, 2009 (in Chinese).

[14] H. Y. Wu and Y. X. Deng, "Logical characterizations of simulation and bisimulation for fuzzy transition systems," *Fuzzy Sets Syst.,* vol. 301, pp. 19-36, 2016.

[15] C.A.R. Hoare, "Communicating Sequential Processes," Prentice-Hall, 1985.

[16] S. Crafa and F. Ranzato, "Logical characterizations of behavioral relations on transition systems of probability distributions," *ACM Trans. Comput. Logic.,* vol. 16, no. 1, pp. 1-24, 2014.

[17] P. R. D'Argenion, P. S. Terraf, and N. Wolovick, "Bisimulation for nondeterministic labelled Markov processes," *Math. Struct. Comp. Sci.,* vol. 22, pp. 43-68, 2012.

[18] Y. X. Deng, W. J. Du, and D. Gebler, "Modal characterizations of behavioral pseudometric," http://arxiv.org/abs/1509.03391.

[19] J. Desharnais, A. Edalat, and P. Panangaden, "Bisimulation for labelled Markov processes," *Inf. Comput.,* vol. 179, no. 2, pp. 163-193, 2002.

[20] M. Hennessy and R. Milner, "Algebraic laws for nondeterminism and concurrency," *J. ACM.,* vol. 32, no. 1, pp. 137-161, 1985.

[21] H. Hermanns, A. Parma et al., "Probabilistic logical characterization," *Inf. Comput.,* vol. 209, no. 2, pp. 154-172, 2011.

[22] K. G. Larsen and A. Skou, "Bisimulation through probabilistic testing," *Inf. Comput.,* vol. 94, no. 1, pp. 1-28, 1991.

[23] A. Parma and R. Segala, "Logical characterizations of bisimulations for discrete probabilistic systems," in *Proc. FOSSACS 2007*, LNCS 4423, pp. 287-301, 2007.

[24] R. Segala and N. A. Lynch, "Probabilistic simulations for probabilistic process," *Nord. J. Comput.,* vol. 2, no. 2, pp. 250-273, 1995.

[25] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.,* vol. 1, pp. 3-28, 1978.

[26] F. Esteva and L. Godo, "Monoidal t-norm based logic: towards a logic for left-continuous t-norms," *Fuzzy Sets Syst.,* vol. 124, pp. 271-288, 2001.

[27] E. P. Klement, R. Mesiar, and E. Pap, "Triangular Norms," Kluwer Academic press, Dordrecht, 2000.

[28] P. Hájek," Metamathematics of Fuzzy Logic," Kluwer Academic Press, Dordrecht, 1998.

[29] János C. Fodor, "Contrapositive symmetry of fuzzy implications," *Fuzzy Sets Syst.,* vol. 69, pp. 141-156, 1995.

[30] T. Flaminio and E. Marchioni, "T-norm-based logics with an independent involutive negation," *Fuzzy Sets Syst.,* vol. 157, pp. 3125-3144, 2006.

[31] F. Esteva, L. Godo, P. Hájek, and M. Navara, "Residuated fuzzy logics with an involutive negation," *Arch. Math. Logic.,* vol. 39, pp. 103-124, 2000.

[32] P. Cintula, E. P. Klement, R. Mesiar, and M. Navara, "Residuated logics based on strict t-norms with an involutive negation," *Math. Log. Quart.,* vol 52, pp. 269-282, 2006.

[33] S. S. Dai, D. W. Pei, and D. H. G, "Robustness analysis of full implication inference method," *Int. J. Approx. Reason.,* vol. 54, pp. 653-666, 2013.

[34] Y. Cao and Y. Ezawa, "Nondeterministic fuzzy automata," *Inf. Sci.,* vol. 191, no. 1, pp. 86-97, 2012.

[35] L. A. Zadeh, "Similarity relations and fuzzy orderings," *Inf. Sci.,* vol. 3, no. 2, pp. 177-200, 1971.

[36] W. Du, Y. Deng, and G. Gebler, "Behavioural Pseudometrics for Nondeterministic Probabilistic Systems," in *Proc. SETTA 2016*, LNCS 9984, pp. 67-84, 2016.

[37] B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order,* Cambridge, U. K.: Cambridge Univ. Press, 2002.

[38] F. van Breugela and J. Worrell, "A behavioural pseudometric for probabilistic transition systems," *Theor. Comput. Sci.,* vol. 331, pp. 115-142, 2005.

[39] F. van Breugela, B. Sharma and J. Worrell, "Approximating a behavioural pseudometric without discount for probabilistic systems," *Log. Meth. Comput. Sci.,* vol. 4, pp. 1-23, 2008.

[40] F. van Breugela and J. Worrell, "The complexity of computing a bisimilarity pseudometric on probabilistic automata," in *Panangaden Festschrift.*, LNCS 8464, pp. 191-213, 2014.

[41] D. W. Qiu, "Supervisory control of fuzzy discrete event systems: a formal approach," *IEEE Trans. Syst., Man, Cybern., B, Cybern.,* vol. 35, no. 1, pp. 72-88, Feb. 2005.

[42] F. Lin and H. Ying, "Modeling and control of fuzzy discrete event systems," *IEEE Trans. Syst., Man, Cybern., B, Cybern.,* vol. 32, no. 4, pp. 408-415, Aug. 2002.

[43] H. Y. Pan, Y. M. Li, Y. Z. Cao, and P. Li, "Nondeterministic fuzzy automata with membership values in complete residuated lattices," *Int. J. Approx. Reason.,* vol. 82, pp. 22-38, 2017.

[44] T. F. Fan and C. J. Liau, "Logical characterizations of regular equivalence in weighted social networks," *Artif. Intell.,* vol. 214, 66-88, 2014.

[45] Y. M. Li and L. J. Li, "Model checking of linear-time properties based on possibility measure," *IEEE Trans. Fuzzy Syst.,* vol. 21, no. 5, pp. 842-854, Oct. 2013.

[46] Y. M. Li and Z. Y. M, "Quantitative computation tree logic model checking based on generalized possibility measures," *IEEE Trans. Fuzzy Syst.,* vol. 23, no. 6, pp. 2034-2047, Dec. 2015.

Hengyang Wu received the B.S. degree from Xuzhou Normal University, in 1996, and M.Sc. and Ph.D. degrees from Shanghai Normal University, in 2004 and 2007, respectively, all in mathematics.

From 2008 to 2011, he was a Postdoctoral Researcher with East China Normal University. From 2011 to July 2016, he was an Associate Professor of computer science with Hangzhou Dianzi University. Currently, he is an Associate Research Fellow with East China Normal University. His current research interests include formal methods and domain theory.

Yuxin Deng received the B.Eng. and M.Sc. degrees from Shanghai Jiao Tong University in 1999 and 2002, respectively, and the Ph.D. in computer science from Ecole des Mines de Paris in 2005.

He is a Professor in East China Normal University. His research interests include concurrency theory, especially about process calculi, and formal semantics of programming languages, as well as formal verification. He authored the book titled *Semantics of Probabilistic Processes: An Operational Approach* (Springer, 2015).