

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 TSs (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, in Fig. 1, states s and t are not bisimilar. However, if we are mainly interested in the maximal possibilities of the occurrence 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 [1], [2] 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.

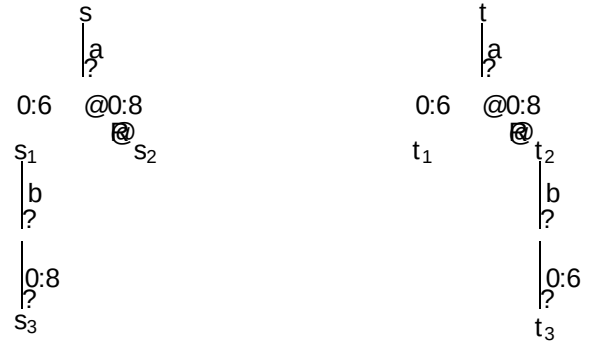


Fig. 1 s and t are distribution-bisimilar but not state-bisimilar

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 they 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 Δ -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 $\mathcal{D}(S)$, while the latter is defined on $\mathcal{S}$; (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, 61261130), Shanghai Municipal Natural Science Foundation (16ZR1409100), and NARS (12IS02001 PACE).

H. Wu is with MoE Engineering Center for Software/Hardware-Design Technology and Application, East China Normal University, Shanghai 200062, P.R. China. Email: wuhengy1974@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@ecnu.edu.cn.

triangular-conorms, while the later is only based on norm, finite set, say $s_1, s_2, \dots, s_n$, then a fuzzy set is written as which is a special case of ours. This generalization allows $u = \frac{r_1}{s_1} + \frac{r_2}{s_2} + \dots + \frac{r_n}{s_n}$; where $r_i \in (0, 1]$ and $r_i = (s_i)$ with according to the properties of different systems, to compute i . With a slight abuse of notations, we sometimes behavioural distances by choosing appropriate operators. write a possibility distribution to mean a fuzzy set

Modal logics and behavioural equivalences are closely related. Whenever a new equivalence is proposed, the question starts for the associated logic such that two systems (test) 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 Kripke structures in terms of Gödel modal logic; Wu and Deng [14] characterized bisimulations for FTSSs by using a fuzzy style Hennessy-Milner logic. In this paper, we present a fuzzy modal logic, which is directly interpreted on distribution-based bisimilarity metric for FTSSs 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 (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 FTSSs, which is distribution-based. Section IV embarks upon the development of distribution-based bisimilarity metric. We first define triangular-conorm based metric, called μ -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 of S is a function that assigns each element of S with a value (s) in the unit interval $[0, 1]$. The support of μ , written $\text{supp}(\mu)$, is the set $\{s \in S \mid \mu(s) > 0\}$. We denote by $F(S)$ the set of all fuzzy sets in S . For convenience, whenever $\text{supp}(\mu)$ is a

For any $p, q \in [0, 1]$, we write $p \sqcup q$ and $p \sqcap q$ 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 $I \neq \emptyset$ is a family of elements in $[0, 1]$. In particular, $\bigvee_{i \in \emptyset} p_i = 0$ and $\bigwedge_{i \in \emptyset} p_i = 1$. For any $c \in [0, 1]$ and $\mu \in F(S)$, the scalar multiplication of c and μ is defined by letting $(c \cdot \mu)(s) = c \wedge \mu(s)$, for each $s \in S$. Let $f_i \mid i \in I$ be a family of elements of $F(S)$, the union of f_i , denoted $\bigvee_{i \in I} f_i$, is given by $(\bigvee_{i \in I} f_i)(s) = \bigvee_{i \in I} f_i(s)$. The height of possibility distribution μ , written $\text{ht}(\mu)$, is defined as $\bigvee_{s \in S} \mu(s)$.

Throughout this paper, we assume an ordinary finite set 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 $(C; \cdot; 0, 1)$ be a bounded linearly ordered set. A function $\neg : C \rightarrow C$ is called an involution or strong negation function if it is order-reversing with $\neg(1) = 0$ and $\neg(\neg(x)) = x$ 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 $\neg(x) = 1 - x$ 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 residuum $\oslash$, as well as the standard negation. The ordering is as usual. This algebra will be denoted by L .

Notice that left-continuity is a necessary and sufficient condition for a t-norm $\otimes$ and its residuum $\oslash$, defined as $x \oslash y = \sup\{z \in [0, 1] \mid x \otimes z \leq y\}$, to verify the so-called residuation property: $x \otimes y \leq z$ iff $x \leq y \oslash z$. In this case, we say $(\otimes; \oslash)$ 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 ($x \otimes y = x \cdot y$), 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 L^G , L^P , and L^N , respectively. For convenience, we sometimes use the symbol

¹Strictly 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 $\mathbf{L}, \mathbf{L}^G, \mathbf{L}^P, \mathbf{L}^N$, 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 abstraction (see Section IV) and to characterize it (see Section V). Below we discuss the reasons of choosing by comparing it with some relevant algebras.

The standard semantics BL (resp. MTL) uses the unit interval $[0; 1]$ as the set of truth values and a continuous (resp. left-continuous) t-norm as an interpretation of the strong conjunction $\&$. The implication $\rightarrow$ is interpreted by the residuum of $\&$. It is well known that $([0; 1]; \min; \max; \rightarrow; 0; 1)$ is an MTL-algebra [26], and it is a BL-algebra provided that $\&$ is continuous [28]. In BL and MTL-algebras, the negation $x = x \rightarrow 0$ 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, IMTL-algebra [26], where δ stands for involution, is obtained by adding $\delta: x = x \rightarrow 0$ 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 $\mathbf{MTL}_n$ -algebra² is obtained by adding an independent involutive negation 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 $\mathbf{L}$ is like the $\mathbf{MTL}_n$ -algebra giving up the operator δ . In $\mathbf{L}$, we can obtain a t-conorm by letting $x \vee y = 1 - [(1 - x) \& (1 - y)]$, which is dual to

and $(\rightarrow; \rightarrow_n)$ is called a De Morgan triple

In this paper, the operator $\rightarrow$ is used for defining $\rightarrow_n$; the operator $\rightarrow_n$ is used for producing the adjoint pair $(\rightarrow; \rightarrow_n)$ and the De Morgan triple $(\rightarrow; \rightarrow_n)$. It is not used to interpret the strong conjunction $\&$ as BL and MTL-algebras do, since $\&$ in general is not a formula in our fuzzy modal logic; the induced operator is used for defining an abstract metric and for modeling the strong disjunction $\vee$ (see Definition 6). Note that it is not used to interpret the strong disjunctive connective between formulae, since $\vee$ 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 and is given by $x \leftrightarrow y = (x \rightarrow y) \& (y \rightarrow x)$ (see [3], [27]), which also deviates from BL and MTL-algebras, where it is defined as $x \leftrightarrow y = (x \rightarrow y) \vee (y \rightarrow x)$. For more differences between $\mathbf{L}$ and the above algebras, please see Section V.

The De Morgan triple $(\rightarrow; \rightarrow_n)$ and the adjoint pair $(\rightarrow; \rightarrow_n)$ guarantee the main results of this paper to hold.

The following properties will be useful.

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

- (1) $x \rightarrow y = 1$ if and only if $x \leq y$;
- (2) $1 \rightarrow x = x$,
- (3) $x \rightarrow (x \rightarrow y) = y$,
- (4) $(x \rightarrow y) \rightarrow (y \rightarrow z) = x \rightarrow z$,
- (5) $(x_1 \rightarrow y_1) \& (x_2 \rightarrow y_2) \rightarrow (x_1 \& x_2) \rightarrow (y_1 \& y_2)$,
- (6) $x_1 \rightarrow x_2 \rightarrow (x_1 \rightarrow y) \rightarrow (x_2 \rightarrow y)$,
- (7) $x_1 \rightarrow x_2 \rightarrow (y \rightarrow x_1) \rightarrow (y \rightarrow x_2)$,
- (8) $x \rightarrow (\bigvee_{i \geq 1} y_i) = \bigvee_{i \geq 1} (x \rightarrow y_i)$,
- (9) $x \rightarrow (\bigwedge_{i \geq 1} y_i) = \bigwedge_{i \geq 1} (x \rightarrow y_i)$.

Proof. The first few 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 x_1, x_2 . Then $t(x_1) \& x_2$ and $t(x_2) \& x_1$. It follows by adjoint condition that $t(x_1 \rightarrow x_2) \& t(x_2 \rightarrow x_1)$. Further, we have that $t(x_2 \rightarrow (x_1 \rightarrow y)) \& x_1 \rightarrow (x_1 \rightarrow y) \rightarrow y$ by (3). Again, by adjoint condition, we have that $(x_1 \rightarrow y) \rightarrow x_2 \rightarrow y$, and then $(x_1 \rightarrow y) \rightarrow (x_2 \rightarrow y)$. In a similar way, we can get $t(x_2 \rightarrow y) \rightarrow (x_1 \rightarrow y)$. So $t(x_1 \rightarrow y) \rightarrow (x_2 \rightarrow y)$. Thus, we obtain $(x_1 \rightarrow x_2) \rightarrow (x_1 \rightarrow y) \rightarrow (x_2 \rightarrow y)$.

Item (8) is trivial since $\rightarrow$ is monotonic. Item (9) can be obtained by using the duality between $\rightarrow$ and $\rightarrow_n$ and the left-continuity of $\&$. $\square$

III. DISTRIBUTION-BASED BISIMULATION

In this section, we introduce the notion of distributions and bisimulation for nondeterministic fuzzy transition systems.

Definition 3. [34] A nondeterministic fuzzy transition system (NFTS) is a triple $M = (S; A; \rightarrow)$, where S is a set of states, A is a set of actions, and $\rightarrow: S \times A \times F(S)$ 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 $s \xrightarrow{a}$ for $(s; a; \cdot) \in \rightarrow$. We write $s \not\xrightarrow{a}$ if there exists no δ such that $s \xrightarrow{a} \delta$. Let us define a transition relation between distributions in an NFTS.

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

- if $s \xrightarrow{a}$ then $s \xrightarrow{a}_d \delta$;
- if $s \not\xrightarrow{a}$ then $s \not\xrightarrow{a}_d \delta$;
- if $s \xrightarrow{a}_d \delta_1$ and $s \xrightarrow{a}_d \delta_2$ for each $s \in \text{supp}(\delta)$, then $s \xrightarrow{a}_d \delta$.

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

²We use $\rightarrow_n$ instead of $\rightarrow_n$ because this notion will be reserved for bisimilarity.

Note that if μ^a 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 μ^b , the distribution $\mu = \frac{0.6}{s_1} + \frac{0.8}{s_2}$ cannot perform action b , the distribution $\mu^b = \frac{0.6}{s_3}$ can make the transition!

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

- (i) $ht(\mu) = ht(\nu)$ and
- (ii) if μ^a , then there exists some ν^a such that μ^a and ν^a .

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

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 $F(S)$.

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 S on S , i.e., a function of type $S \times S \rightarrow [0, 1]$, that satisfies the following three conditions:

- (1) reflexive, i.e., $S(s; s) = 1$ for all $s \in S$,
- (2) symmetric, i.e., $S(s; t) = S(t; s)$ for all $s, t \in S$,
- (3) max-star transitive, i.e., $S(s; s^0) \geq S(s; t) \star S(t; s^0)$, or, more explicitly,

$$S(s; s^0) \geq \bigvee_{s^0 \in S} S(s; s^0) \star S(s^0; t)$$

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

Now, we take $\star$ to be a t-norm and define, for all $s, t \in S$, $d(s; t) = 1 - S(s; t)$. It turns out that, for all $s, s^0, t \in S$, we have $d(s; t) \leq d(s; s^0) \vee d(s^0; t)$. Moreover, $d(s; s) = 0$ and $d(s; t) = d(t; s)$. We call the distance function d a $\star$ -metric. Next, we generalize the $\star$ -metric from S to $F(S)$.

Definition 6. Under L , a function $d : F(S) \times F(S) \rightarrow [0, 1]$ is called a $\star$ -metric on $F(S)$ if for all $\mu, \nu \in F(S)$,

- (P1) $d(\mu; \mu) = 0$,
- (P2) $d(\mu; \nu) = d(\nu; \mu)$,
- (P3) $d(\mu; \nu) \leq d(\mu; \nu^0) \vee d(\nu^0; \nu)$.

We call $(F(S); d)$ a $\star$ -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 $\vee$ instead of $\wedge$ in Cao's definition, i.e., the max-disjunction is replaced by the strong disjunction. The following proposition indicates that metric always exists.

Proposition 2. Given a t-conorm $\star$ from L , the $\star$ -metric always exists.

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

$$\begin{aligned} d(\mu; \nu) &= 1 - \inf_{\mu \rightarrow \nu} (ht(\mu), ht(\nu)) \\ &= 1 - \inf_{\mu \rightarrow \nu} (ht(\mu), ht(\nu)) \\ &= 1 - \inf_{\mu \rightarrow \nu} (ht(\mu), ht(\nu)) \\ &= 1 - \inf_{\mu \rightarrow \nu} (ht(\mu), ht(\nu)) \end{aligned}$$

as expected. $\square$

We can understand $d(\mu; \nu)$ as how far distributions μ and ν are from the same height based on the metric. If μ is taken from L^L , then we have $d(\mu; \nu) = |ht(\mu) - ht(\nu)|$, which gives a pseudometric bounded by 1, often used in probabilistic systems. If μ is taken from L^G , then $d(\mu; \nu) = 0$ 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 $d(\mu; \nu)$ when μ is taken from L^N and L^P , respectively. A pseudo-ultrametric is a $\star$ -metric because $\star$ is the minimum t-conorm, but the inverse is not necessarily true, as witnessed by $\star$ when μ is taken from L^L , L^N , and L^P , respectively. Note that d is important for defining our bisimilarity metric and we use it as a special notation in the sequel.

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

It is easy to check that $\leq$ is indeed a partial order on $D_{F(S)}$. 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 $D_{F(S)}$ endowed with the above order forms a complete lattice.

Lemma 1. Under L , $(D_{F(S)}; \leq)$ is a complete lattice.

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

$$\begin{aligned} (\bigvee X)(\mu; \nu) &= \bigvee_{d \in X} d(\mu; \nu) \\ &\leq \bigvee_{d \in X} (d(\mu; \nu^0) \vee d(\nu^0; \nu)) \\ &= \bigvee_{d \in X} d(\mu; \nu^0) \vee \bigvee_{d \in X} d(\nu^0; \nu) \\ &= (\bigvee X)(\mu; \nu^0) \vee (\bigvee X)(\nu^0; \nu) \end{aligned}$$

for any $\mu, \nu \in F(S)$. Hence $\bigvee X$ is a $\star$ -metric. In addition, it is not hard to prove that $\bigvee X$ is the supremum of X . Hence $(D_{F(S)}; \leq)$ 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 metric on $F(S)$ to a $\star$ -metric on $P(F(S))$:

Definition 7. Let $(F(S); d)$ be a $\mathcal{L}$ -metric space. For any $\mathcal{L}$ -metric d on $F(S)$, define

$$d(\cdot; A) = \bigwedge_{2A} d(\cdot; \cdot):$$

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

$$H_d(A; B) = [\bigwedge_{2A} d(\cdot; B)] _ [\bigwedge_{2B} d(\cdot; A)]:$$

As expected H_d has the following property.

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

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

$$\begin{aligned} H_d(A; B) &= H_d(B; C) \quad \text{by Proposition 1 (8)} \\ &= [\bigwedge_{2A} d(\cdot; B)] _ [\bigwedge_{2B} d(\cdot; C)] \quad \text{(by Proposition 1 (8))} \\ &= [\bigwedge_{2A} \bigwedge_{2B} [d(\cdot; B) _ d(\cdot; C)]] \quad \text{(for some } 0 \in 2B) \\ &= [\bigwedge_{2A} [d(\cdot; 0) _ d(\cdot; C)]] \\ &= [\bigwedge_{2A} [d(\cdot; 0) _ (\bigwedge_{2C} d(\cdot; \cdot))]] \\ &= [\bigwedge_{2A} [\bigwedge_{2C} (d(\cdot; 0) _ d(\cdot; \cdot))]] \\ &= [\bigwedge_{2A} \bigwedge_{2C} d(\cdot; \cdot)] \\ &= [\bigwedge_{2A} d(\cdot; C)]: \end{aligned}$$

Similarly, we have $H_d(A; B) = H_d(B; C) _ \bigwedge_{2C} d(\cdot; A)$. Hence,

$$\begin{aligned} H_d(A; B) &= H_d(B; C) \\ &= [\bigwedge_{2A} d(\cdot; C)] _ [\bigwedge_{2C} d(\cdot; A)] \\ &= H_d(A; C): \end{aligned}$$

This completes the proof. $\square$

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

Lemma 3. Let $(F(S); d)$ be a $\mathcal{L}$ -metric space. If d_n for $n \in \mathbb{N}$ be an increasing chain, i.e. $d_0 \leq d_1 \leq d_2 \leq \dots$, and B be a finite subset of $F(S)$. Then for any $\mathcal{L}$ -metric d on $F(S)$, we have

$$\bigwedge_{n \in \mathbb{N}} d_n(\cdot; B) = \bigwedge_{n \in \mathbb{N}} d_n(\cdot; B):$$

Proof. It is sufficient to prove that

$$\inf_{2B} \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot) = \sup_{n \in \mathbb{N}} \inf_{2B} d_n(\cdot; \cdot):$$

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

$$\inf_{2B} d_n(\cdot; \cdot) \leq \inf_{2B} \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot):$$

This shows that $\inf_{2B} \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot)$ is an upper bound of the set $\{\inf_{2B} d_n(\cdot; \cdot) \mid n \in \mathbb{N}\}$.

Second, for any $\mathcal{L}$ -metric d , we consider $\sup_{n \in \mathbb{N}} d_n(\cdot; \cdot)$. For any $\mathcal{L}$ -metric d , there exists some n such that

$$d_n(\cdot; \cdot) > \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot):$$

Since B is finite and d_n for $n \in \mathbb{N}$ is an increasing chain, there exists n such that $d_n(\cdot; \cdot) > \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot)$ for any $\mathcal{L}$ -metric d . Hence,

$$\inf_{2B} d(\cdot; \cdot) > \inf_{2B} \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot) = \inf_{2B} \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot):$$

Thus, $\inf_{2B} \sup_{n \in \mathbb{N}} d_n(\cdot; \cdot)$ is the supremum of the set $\{\inf_{2B} d_n(\cdot; \cdot) \mid n \in \mathbb{N}\}$. We have completed the proof. $\square$

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

Corollary 1. Let $(F(S); d)$ be a $\mathcal{L}$ -metric space. If d_n for $n \in \mathbb{N}$ be an increasing chain. Then $H_d(A; B) = \bigwedge_{n \in \mathbb{N}} H_{d_n}(A; B)$.

Proof. Straightforward by using Lemma 3. $\square$

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 $(S; A; !)$ and $\mathcal{L}$, a $\mathcal{L}$ -metric d on $F(S)$ is called a distribution-based bisimulation metric if for all $\mathcal{L}$ -metric d on $F(S)$,

- (i) $d(\cdot; \cdot) \leq 1$ and
- (ii) $d(\cdot; \cdot) < 1$ implies that whenever $\mathcal{L}$ -metric d then there exists a transition $\mathcal{L}$ -metric d with $d(\cdot; \cdot) \leq d(\cdot; \cdot)$.

The first condition says that d should be at least 0. 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 $(S; A; !)$, the functional $\Phi : D_{F(S)} \rightarrow D_{F(S)}$ is defined as follows: for any $d \in D_{F(S)}$ and $\mathcal{L}$ -metric d on $F(S)$,

$$(\Phi d)(\cdot; \cdot) = \max_{a \in A} [\bigwedge_{2A} H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot)) _ ?(\cdot; \cdot)]:$$

Now, we check that Φ is well-defined.

Proposition 3. Given an image-finite NFTS $(S; A; !)$, then for any $d \in D_{F(S)}$, we have $(\Phi d) \in D_{F(S)}$.

Proof. It is easy to see that (Φd) satisfies P1 and P2. Now, we verify P3. For any $\mathcal{L}$ -metric d on $F(S)$, it is obvious that, for any $a \in A$, $(\Phi d)(\cdot; \cdot) \leq H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot))$ and symmetrically $(\Phi d)(\cdot; \cdot) \leq H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot))$. Hence, for any $a \in A$,

$$\begin{aligned} (\Phi d)(\cdot; \cdot) &\leq (\Phi d)(\cdot; \cdot) \\ &\leq H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot)) _ H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot)) \\ &\leq H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot)) \quad \text{(by Lemma 2)} \end{aligned}$$

Consequently,

$$(\Phi d)(\cdot; \cdot) \leq (\Phi d)(\cdot; \cdot) _ \bigwedge_{a \in A} H_d(T_a(\cdot; \cdot); T_a(\cdot; \cdot)):$$

On the other hand, we also have

$$(\Phi d)(\cdot; \cdot) \leq ?(\cdot; \cdot) \text{ and } (\Phi d)(\cdot; \cdot) \leq ?(\cdot; \cdot):$$

Hence,

$$(d)(;) \quad (d)(;) ? (;) ? (;) ? (;)$$

by Proposition 2. It follows that

$$\max_{a \in A} (d)(;) \quad (d)(;) \quad (d)(;) \text{ as desired. } \square$$

Moreover, d_b is monotonic with respect to the order, i.e., if $d_1 \leq d_2$, then we have $(d_1) \leq (d_2)$. 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 $(S; A; !)$, the functional $\lambda d. D_F(S) \circ d$ has a least fixed point given by

$$\min_{d \in D_F(S)} d$$

Let us present an explicit characterization of bisimulation metrics in terms of pre-fixed points of $\lambda d. D_F(S) \circ d$.

Lemma 4. Given an image-finite NFTS $(S; A; !)$, for any $d \in D_F(S)$, d is a pre-fixed point of $\lambda d. D_F(S) \circ d$ if and only if it is a distribution-based bisimulation metric.

Proof. Straightforward by the definitions of distribution-based bisimulation metric and $D_F(S)$. $\square$

Proposition 4 and Lemma 4 imply the following theorem.

Theorem 1. Given an image-finite NFTS $(S; A; !)$. The smallest distribution-based bisimulation metric exists, moreover it is just $\min_{d \in D_F(S)} d$. Let $d_b \stackrel{\text{def}}{=} \min_{d \in D_F(S)} d$ and call it distribution-based bisimilarity metric.

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

Lemma 5. Given an image-finite NFTS $(S; A; !)$, we have

$$(d_b)_{n \in \mathbb{N}} = (d_b^G)_{n \in \mathbb{N}}$$

where $(d_b)_{n \in \mathbb{N}}$ is an increasing chain in $D_F(S)$.

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

Corollary 2. Let $(S; A; !)$ be an image-finite NFTS. We define $d_b^0(?) \stackrel{\text{def}}{=} ?$ and $d_b^{n+1}(?) \stackrel{\text{def}}{=} (d_b^n(?))$. Then

$$d_b = \min_{n \in \mathbb{N}} d_b^n$$

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

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

Example 1. Consider Figure 1 again. Let $t_1 = \frac{0.6}{s_1} + \frac{0.8}{s_2}$, $t_2 = \frac{0.8}{s_3}$, $t_3 = \frac{0.6}{t_1} + \frac{0.8}{t_2}$ and $t_4 = \frac{0.6}{t_3}$. We write d_n for $d_b^n(?)$, clearly $d_i \leq d_{i+1}$ for any $i \in \mathbb{N}$. Now, we compute $d_b^G(s; t)$ and $d_b^G(s_1; t_2)$. For any $n \geq 3$,

$$\begin{aligned} d_n(s; t) &= (d_{n-1})(s; t) \\ &= \max(d_{n-1}(\frac{0.6}{s_1}; \frac{0.8}{s_2}); 1 - (ht(s), \mathcal{G} ht(t))) \\ &= d_{n-1}(\frac{0.6}{s_1}; \frac{0.8}{s_2}) \\ &= (d_{n-2})(\frac{0.6}{s_1}; \frac{0.8}{s_2}) \\ &= \max(d_{n-2}(\frac{0.6}{s_3}; \frac{0.8}{t_3}); 1 - (ht(\frac{0.6}{s_1}), \mathcal{G} ht(\frac{0.8}{s_2}))) \\ &= d_{n-2}(\frac{0.6}{s_3}; \frac{0.8}{t_3}) \\ &= (d_{n-3})(\frac{0.6}{s_3}; \frac{0.8}{t_3}) \\ &= \max(d_{n-3}(\frac{0.6}{t_4}; \frac{0.8}{t_4}); 1 - (ht(\frac{0.6}{s_3}), \mathcal{G} ht(\frac{0.8}{t_3}))) \\ &= 0; \end{aligned}$$

Hence $d_b^G(s; t) = 0$. In the same way, we can calculate that $d_b^G(s_1; t_2) = 0.4$.

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

Corollary 3. Given an image-finite NFTS $(S; A; !)$. For any $s, t \in S$, let $S(s; t) \stackrel{\text{def}}{=} 1 - d_b(s; t)$. Then S is a similarity relation on S .

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

We can see that $\mathcal{D}$ plays a key role in computing the distribution-based bisimilarity metric, which is obtained by an iteration of $D_F(S)$ starting from $?$. 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 behavioural distance. However, there is no essential difference because in [2] the partial order on $D_F(S)$ is reverse, i.e., $d_1 \leq d_2$ 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 $M = (S; A; !)$. Given any pair of possibility distributions $\mu, \nu \in F(S)$, 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 $d_b^n(\mu; \nu) = d_b^n(?)$ by Corollary 2, until an N appears such that $d_b^N(\mu; \nu) = d_b^{N+1}(\mu; \nu)$. Then $d_b(\mu; \nu) = d_b^N(\mu; \nu)$. The tricky point is that there may be several N satisfying $d_b^N(\mu; \nu) = d_b^{N+1}(\mu; \nu)$ and we cannot simply choose any of them to be d_b . We have to make sure that similar condition $d_b^N(\mu^0; \nu^0) = d_b^{N+1}(\mu^0; \nu^0)$ also holds for any μ^0 and ν^0 reachable from μ and ν , respectively.

Algorithm 1 gives the details. The main procedure is $\text{distance}(\mu; \nu)$, which first uses the procedure generate to generate all the generalized states reachable from μ and ν , then repeatedly call $\text{dist}(n; \mu^0; \nu^0)$ for computing $d_b^n(\mu^0; \nu^0)$. The procedure $\text{dist}(n; \mu^0; \nu^0)$ faithfully implements the inductive definition $d_b^n(?)$ for $n \geq 0$. When a sufficiently large number is obtained such that $d_b^N(\mu^0; \nu^0) = d_b^{N+1}(\mu^0; \nu^0)$ for any μ^0 and ν^0 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

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

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

Proof. Suppose for a contradiction that for all $n \geq 0$, $ht(\mu_{n+1}) < ht(\mu_n)$. So in particular, for $i = 0$, $ht(\mu_1) < ht(\mu_0)$. Then for $j = 1$, $ht(\mu_2) < ht(\mu_1)$. Continuing this process, we will get that $ht(\mu_n) < ht(\mu_{n-1})$ for any $n \geq 1$. Note that Definition 4 implies that $ht(\mu_n) = ht(\mu_{n-1})$ for any $n \geq 1$ with $a_n \in A$. Hence, we get that $ht(\mu_1) > ht(\mu_2) > \dots$, i.e., an infinite descending chain of non-zero values $ht(\mu_n)$. However, this is impossible. The reason is as follows. Let $X = \{s \in S \mid \mu_0(s) > 0\}$ and Y be the set defined by $\{s \in S \mid \mu_1(s) > 0\}$. Note that the two sets are finite because S is finite and the transitions are finite. Now, by Definition 4, $\mu_1 = \sum_{s \in S} \mu_0(s) \cdot a_s$, where the distribution μ_1 is obtained by 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, $ht(\mu_n)$ only takes values in these two sets, which is in contradiction with the fact that $ht(\mu_n)$ has an infinite descending chain. $\square$

This proposition shows that for any given distribution, 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 L₁ metric, the algorithm terminates and is EXPTIME.

Proof. We first show the algorithm terminates. Proposition 5 says that for any chain of transitions $\mu_0 \xrightarrow{a_1} \mu_1 \xrightarrow{a_2} \dots$, 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 $\mu \in F(S)$ and $n \geq 0$, we know that

$$d_{n+1}(\mu; \nu) = \max_{a \in A} (H_{d_n}(T_a(\mu); T_a(\nu)); ?(\mu; \nu)); \quad (1)$$

If μ and ν are fixed, the choices on the right hand side are finite: $d_{n+1}(\mu; \nu)$ is determined by either $?(\mu; \nu)$ or $d_n(\mu; \nu)$ with a_i and a_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

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

or

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

for some fixed transitions a_i and a_j , but in finitely many different n 's. In other words, the distance between μ and ν is either a fixed number $?(\mu; \nu)$ or determined by a pair of their successor distributions. In the latter case, the same remark can be made again. Eventually, the computation of $dist(n; \mu; \nu)$ ($n \geq 0$) boils down to the computations

Algorithm 1 Compute the distance between two distributions

Input: A finite NFTS, and $\mu, \nu \in F(S)$.

Output: $d_b(\mu; \nu)$.

```

procedure distance( $\mu, \nu$ )
   $n \leftarrow 1$ ;
  for all  $i \in \text{generate}(\mu)$  do
    for all  $j \in \text{generate}(\nu)$  do
       $n_{ij} \leftarrow 1$ ;
      while  $dist(n_{ij}, i; j) \neq dist(n_{ij}; i; j)$  do
         $n_{ij} \leftarrow n_{ij} + 1$ ;
      end while
      if  $n_{ij} > n$  then
         $n \leftarrow n_{ij}$ ;
      end if
    end for
  end for
  return  $dist(n; \mu; \nu)$ ;
end procedure

```

```

procedure generate( $\mu$ )
   $D \leftarrow \{ \mu \}$ ;
  repeat
    for all  $a \in A$  do
      for all  $\mu \in D$  do
         $D \leftarrow D \cup T_a(\mu)$ ;
      end for
    end for
  until  $D^0 = D$ ;
  return  $D$ ;

```

```

procedure dist( $n; \mu; \nu$ )
  if  $n = 0$  then
    return  $?(ht(\mu), ht(\nu))$ ;
  else
    for all  $a \in A$  do
      for all  $i \in T_a(\mu)$  do
        for all  $j \in T_a(\nu)$  do
           $d_{a,ij} \leftarrow dist(n-1; i; j)$ ;
        end for
      end for
    end for
    return  $\max_{a \in A} (\max_{i,j} d_{a,ij} - ?(\mu; \nu)) \cup dist(0; \mu; \nu)$ ;
  end if
end procedure

```

of $?(\mu_k; \mu_l)$, where μ_k and μ_l are elements in the chains generated by μ and ν , respectively. Hence, when the heights of distributions do not change, the values $?(\mu_k; \mu_l)$ do not change either. Consequently, there exists some n such that $dist(n; \mu; \nu) = dist(n+1; \mu; \nu)$ for any μ and ν reachable from μ_0 and ν_0 , respectively, hence the algorithm will terminate.

We now consider the time complexity. Let $(S; A; \mu_0)$ be the NFTS under consideration. Assume that the size of the state space is $|S| = n$ and the number of edges (viewing the

NFTS as a directed graph) labelled with possibility values is m . Suppose the density of the NFTS is k . That is, from any states there are at most k outgoing transitions, with the same label, 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 $\text{dist}(\mu; \nu)$, we will compute $\text{dist}(\mu_{i-1}; \nu_{i-1})$ at most $jAjk^{2n}$ times. For any $\mu \in F(S)$, let t_1 be the time of computing $\text{dist}(\mu; \nu)$. We see that $t_1 = jAjk^{2n}t_{1-1}$. It follows that t_1 is in $O((jAjk^{2n})^i)$. In the proof of Proposition 5, the sizes of the sets X and Y are bounded by n and m , respectively. Therefore, the procedure $\text{dist}(\mu; \nu)$ will stabilize after $n + m$ iterations. Note that for each distribution μ reachable from μ_0 , 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 of edges of the graph of the original NFTS. It follows that the total number of distributions reachable from μ_0 is at most n^m . Therefore, the overall time consumed is thus $O(n^m jAjk^{2n^2+2nm})$, 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 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 PSPACE, 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, $\text{dist}(\mu; \nu) = \text{dist}(\mu_{h+1}; \nu_{h+1})$, and similarly for any pair of distributions reachable from μ and ν . Moreover, the procedure 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 $O(jAjk^{h+1} k^{2n(h+1)})$.
- 3) If the NFTS in 2) is deterministic, then the running time of the algorithm becomes $O(jAjk^{h+1})$. 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 , and T be a fixed proposition, which is taken from the classical Hennessy-Milner Logic [20]. The language of formulae is the least set generated by the following BNF grammar:

$$\begin{aligned} & ::= T \mid j \mid 1 \wedge 2 \mid j \mid ! p \mid p \mid j \mid \text{hai} \\ & \text{where } p \in [0, 1]. \end{aligned}$$

We write $\$ p$ for $(! p) \wedge (p!)$. Note that the formulae in $\mathcal{L}$ are defined inductively, so we can only generate formulae of finite depth such as $\text{hai } T$, $T! p$, $p! \text{hai } T$, $(\text{hai } T! p) \wedge (r! \text{hai } T)$, etc. However, $\text{hai } T! T$ is not a legitimate formula. Let us examine an NFTS $M = (S; A; !)$. We define the semantic function $\llbracket \cdot \rrbracket : \mathcal{L} \rightarrow F(S) \rightarrow [0, 1]$ to calculate the degree for a distribution to satisfy a formula.

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

$$\begin{aligned} \llbracket T \rrbracket(\mu) &\stackrel{\text{def}}{=} \text{ht}(\mu), \\ \llbracket j \mid 1 \wedge 2 \rrbracket(\mu) &\stackrel{\text{def}}{=} \min(\llbracket j \mid 1 \rrbracket(\mu); \llbracket j \mid 2 \rrbracket(\mu)), \\ \llbracket j \mid ! p \rrbracket(\mu) &\stackrel{\text{def}}{=} \llbracket j \rrbracket(\mu) \wedge p, \\ \llbracket p! \rrbracket(\mu) &\stackrel{\text{def}}{=} p \wedge \llbracket j \rrbracket(\mu), \\ \llbracket \text{hai } j \rrbracket(\mu) &\stackrel{\text{def}}{=} \max_{i \in A} \llbracket j \rrbracket(\mu_i). \end{aligned}$$

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

- 1) In general, for any $\phi, \psi \in \mathcal{L}$, $\phi \wedge \psi$ and $\phi! \psi$ 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 $\text{hai } j$, which specifies the property for a distribution to perform action a and result in a possible distribution to satisfy j . 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 $\llbracket \text{hai } j \rrbracket(\mu)$ be the maximum $\llbracket j \rrbracket(\mu_i)$ when $i \in I$;
- 4) The operator $!$ is used to interpret the formula $p!$, where $p \in [0, 1]$.

Definition 11. Let $S = (S; A; !)$ be an NFTS, for any distributions $\mu, \nu \in F(S)$, the logical metric between μ and ν is defined by $d_L(\mu; \nu) = 1 - \inf_{\phi \in \mathcal{L}} |\llbracket \phi \rrbracket(\mu) - \llbracket \phi \rrbracket(\nu)|$:

By Proposition 1 (4), it is not difficult to verify that d_L is a $\mathcal{L}$ -metric on $F(S)$. In the following, we use d_L^T , d_L^G , d_L^P , and d_L^N to denote logical metrics based on $\mathcal{L}^T$, $\mathcal{L}^G$, $\mathcal{L}^P$ and $\mathcal{L}^N$, 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 NFTSs. 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 NFTSs, we have $d_L = d_b$.

Proof. Let $(S; A; !)$ be an image-finite NFTS. For any $\mu, \nu \in \mathcal{F}(S)$ and $\varphi \in 2^L$, we prove the inequality $d_l(\mu; \varphi) \leq d_b(\nu; \varphi)$. This holds because

$d_l(\mu; \varphi) = d_b(\mu; \varphi)$. It is sufficient to prove

$$d_l(\mu; \varphi) \leq d_b(\mu; \varphi)$$

for all $\mu \in \mathcal{F}(S)$.

We proceed by structural induction on formulae. We first analyze the structure of 2^L .

T. Then it is trivial to see that

$$d_l(\mu; \varphi), d_l(\nu; \varphi) = \text{ht}(\varphi), \text{ht}(\varphi) \leq 1 \leq d_b(\mu; \varphi)$$

because $d_b(\mu; \varphi) = 1 - \text{ht}(\varphi)$ and d_b is a bisimulation metric.

$\mu \wedge \nu$. We infer that

$$\begin{aligned} & d_l(\mu \wedge \nu; \varphi) = \min(d_l(\mu; \varphi), d_l(\nu; \varphi)) \\ & = \min(d_l(\mu; \varphi), d_l(\nu; \varphi)) \\ & = \min(d_l(\mu; \varphi), d_l(\nu; \varphi)) \\ & \leq d_b(\mu \wedge \nu; \varphi) \quad (\text{by induction on } \mu \text{ and } \nu). \end{aligned}$$

The second step holds because of Proposition 1 (5).

$\mu \rightarrow \nu$. We infer that

$$\begin{aligned} & d_l(\mu \rightarrow \nu; \varphi) = \min(d_l(\mu; \neg \varphi), d_l(\nu; \varphi)) \\ & = \min(d_l(\mu; \neg \varphi), d_l(\nu; \varphi)) \\ & = \min(d_l(\mu; \neg \varphi), d_l(\nu; \varphi)) \\ & \leq d_b(\mu \rightarrow \nu; \varphi) \quad (\text{by induction on } \mu). \end{aligned}$$

$\mu \rightarrow \nu$. Similar to the last case by using Proposition 1 (7).

$\mu \rightarrow \nu$. Let μ^0 be the distribution such that $\mu^0 \rightarrow \mu$ and $d_l(\mu^0; \varphi) = d_l(\mu; \varphi)$. Since d_b is a distribution-based bisimulation metric, by Definition 8, there exists some ν^0 such that $\nu^0 \rightarrow \nu$ and $d_b(\mu^0; \nu^0) = d_b(\mu; \nu)$. Without loss of generality, we assume $d_l(\mu^0; \varphi) = d_l(\mu; \varphi)$. In this case, $d_l(\mu^0; \varphi) = d_l(\mu; \varphi)$. We distinguish two possibilities:

– $d_l(\mu^0; \varphi) = d_l(\mu; \varphi)$. Then

$$\begin{aligned} & d_l(\mu; \varphi) = d_l(\mu^0; \varphi) \\ & = d_b(\mu^0; \nu^0) \\ & \leq d_b(\mu; \nu) \end{aligned}$$

– $d_l(\mu^0; \varphi) > d_l(\mu; \varphi)$. Then

$$\begin{aligned} & d_l(\mu; \varphi) = d_l(\mu^0; \varphi) \\ & = d_l(\mu^0; \varphi) \max_{\alpha \in 2^L} d_l(\mu^0; \alpha) \\ & = d_l(\mu^0; \varphi) \max_{\alpha \in 2^L} d_l(\mu^0; \alpha) \\ & \leq d_b(\mu^0; \nu^0) \quad (\text{by induction on } \varphi). \\ & \leq d_b(\mu; \nu); \end{aligned}$$

Lemma 7. For image-finite NFTSs, we have $d_b = d_l$.

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 $(S; A; !)$ be an image-finite NFTS

$$\begin{aligned} d_l(\mu; \varphi) &= 1 - \inf_{\alpha \in 2^L} (d_l(\mu; \alpha), d_l(\nu; \alpha)) \\ &= 1 - \inf_{\alpha \in 2^L} (d_l(\mu; \alpha), d_l(\nu; \alpha)) \\ &= 1 - \inf_{\alpha \in 2^L} (d_l(\mu; \alpha), d_l(\nu; \alpha)) \\ &= d_b(\mu; \nu); \end{aligned}$$

Now, assume that $\mu \rightarrow \nu$ and $d_l(\mu; \varphi) = r < 1$. We need to show that there exists some transition $\mu \rightarrow \nu$ with $d_l(\mu^0; \nu^0) = d_l(\mu; \nu)$. Suppose for a contradiction that no α -transition from μ satisfies this condition. Then for each $(i \in I)$ with $\mu^0 \rightarrow_i \nu^0$, we have $d_l(\mu^0; \nu^0) > r$. There must exist some formula $\varphi_i \in 2^L$ such that $d_l(\mu^0; \varphi_i) > r$, that is, $d_l(\mu^0; \varphi_i) > r$, for each $i \in I$. Note that by assumption the NFTS is image-finite, so the index set I is finite. Define $\mu^0 = \sum_{i \in I} \mu^0(i) \varphi_i$. Then for all $i \in I$,

$$d_l(\mu^0; \varphi_i) = 1 \text{ and } d_l(\nu^0; \varphi_i) = d_l(\nu; \varphi_i) < 1 - r.$$

Let $\mu = \sum_{i \in I} \mu^0(i) \varphi_i$. Then we infer that

$$\begin{aligned} d_l(\mu; \varphi) &= \max_{\alpha \in 2^L} d_l(\mu; \alpha) \\ &= \max_{\alpha \in 2^L} d_l(\mu; \alpha) \\ &= 1 - r; \end{aligned}$$

On the other hand, we have

$$\begin{aligned} d_l(\nu; \varphi) &= \max_{\alpha \in 2^L} d_l(\nu; \alpha) \\ &= \max_{\alpha \in 2^L} d_l(\nu; \alpha) \\ &= \max_{\alpha \in 2^L} d_l(\nu; \alpha) \\ &< 1 - r; \end{aligned}$$

It follows that

$$\begin{aligned} d_l(\mu; \nu) &= 1 - \inf_{\alpha \in 2^L} (d_l(\mu; \alpha), d_l(\nu; \alpha)) \\ &= 1 - \inf_{\alpha \in 2^L} (d_l(\mu; \alpha), d_l(\nu; \alpha)) \\ &= 1 - \inf_{\alpha \in 2^L} (d_l(\mu; \alpha), d_l(\nu; \alpha)) \\ &> 1 - (1 - r) \\ &= r \end{aligned}$$

which is in contradiction to $d_l(\mu; \nu) = r$. $\square$

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

Theorem 3. For image-finite NFTSs, we have $d_l = d_b$.

For any $\mu, \nu \in \mathcal{F}(S)$ and any $\varphi \in 2^L$, the proof of Lemma 6 implies $d_l(\mu; \varphi) = d_l(\nu; \varphi)$ when $\mu \rightarrow \nu$. It follows that $d_l(\mu; \varphi) = 0$. On the other hand, define the binary relation $R \stackrel{\text{def}}{=} \{(\mu, \nu) \mid d_l(\mu; \nu) = 0\}$. Then R is obvious an equivalence relation. The proof of Lemma 7 implies that R is a distribution-based bisimulation. Hence $d_l(\mu; \nu) = 0$ implies $\mu \rightarrow \nu$. Therefore, two distributions are bisimilar if and only if the logical distance between them is 0.

$\square$

Corollary 4. Let μ and ν be two distributions over an image-finite NFTS. Then $\mu \rightarrow \nu$ iff $d_l(\mu; \nu) = d_b(\mu; \nu) = 0$ iff $d_l(\mu; \varphi) = d_l(\nu; \varphi)$ for all $\varphi \in 2^L$.

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

Example 2. According to Corollary 4 and the computation of d_b^L , then by a direct computation results of Example 1, we can see that s_1 and t_1 are bisimilar, shown in Example 1, we can obtain that $d_b^L(s_1; s_2) = 0$ and $d_b^L(s_2; s_3) = 0.4$. whereas s_1 and t_2 are not.

In addition, with Theorem 3 sometimes we can easily calculate distribution-based bisimilarity metric in terms of formula. For instance, consider Example 1 again. The only relevant formula is $[[hi T]]$, and it is direct to calculate that $[[hi T]](s_1) = 0.8$ and $[[hi T]](t_2) = 0.6$. Thus $[[hi T]](s_1)$, $[[hi T]](t_2) = 0.6$ under the Gödel implication. It follows from Theorem 3 that $d_b^G(s_1; t_2) = d_b^G(s_1; t_2) = 1 - 0.6 = 0.4$.

VI. NON-EXPANSIVENESS

In this section, we show the non-expansiveness of our distribution-based bisimilarity metric with respect to parallel composition in the style of CSP [15]. It means that if the difference (with respect to bisimilarity metric) between α and β is i , then the difference between $\alpha \parallel \gamma$ and $\beta \parallel \gamma$ is no more than i [2], where f is a parallel composition operator with arguments. This non-expansiveness makes compositional verification possible.

Given an NFTS $(S; A; !)$ and two possibility distributions $\gamma_1, \gamma_2 \in F(S)$, the key to define a parallel composition is how to appropriately define the fuzzy set represented by $\gamma_1 \parallel \gamma_2$. For example, let $\gamma_1 = \frac{0.3}{s_1} + \frac{0.5}{s_2}$ and $\gamma_2 = \frac{0.4}{s_3}$. Intuitively, in their parallel composition $(s_1; s_3)$ appears with possibility 0.3 and $(s_2; s_3)$ appears with possibility 0.4. That is, k is the fuzzy set $\frac{0.3}{(s_1; s_3)} + \frac{0.4}{(s_2; s_3)}$. Hence, k is given by

$$(\gamma_1 \parallel \gamma_2)(s; t) = (\gamma_1(s) \wedge \gamma_2(t))$$

for all $(s; t) \in S \times S$. Clearly, we have $k \in F(S \times S)$ and in general $(\gamma_1 \parallel \gamma_2)$ and $(\gamma_2 \parallel \gamma_1)$ are different fuzzy sets on $F(S \times S)$. For instance $(\gamma_1 \parallel \gamma_2) = \frac{0.3}{(s_1; s_3)} + \frac{0.4}{(s_2; s_3)}$. Note that $(s; t)(s^0; t^0) = 1$ if $(s^0; t^0) = (s; t)$ and 0 otherwise. The above fuzzy set k can be understood as $\frac{0.3}{s_1 k s_3} + \frac{0.4}{s_2 k s_3}$. For simplicity, we write $\frac{0.3}{s_1 k s_3} + \frac{0.4}{s_2 k s_3}$ instead of $\frac{0.3}{s_1 k s_3} + \frac{0.4}{s_2 k s_3}$.

The parallel composition operator we are going to define is asynchronous in the sense that the components can either synchronize or act independently. Given NFTS $(S; A; !)$, for any $\gamma_1, \gamma_2 \in F(S)$, the events intended to synchronize at a and $\bar{a}$ are listed in the set $A \setminus A$ and the rest of the events can be performed independently, where a stands for the set $a \in A : \exists s^0 \in S, t^0 \in S, (s^0; t^0) \in \gamma_1 \parallel \gamma_2$. Formally, the parallel composition k can perform actions in four different ways:

$$T_a(\gamma_1 \parallel \gamma_2) = \begin{cases} \gamma_1 \parallel \gamma_2 & \text{if } a \in A \setminus A \\ \gamma_1 & \text{if } a \in A \text{ and } \bar{a} \notin A \\ \gamma_2 & \text{if } a \in A \text{ and } \bar{a} \notin A \\ 0 & \text{otherwise.} \end{cases}$$

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

Example 3. Let $S = (f s_1; s_2; s_3; s_4; g; !)$ be an NFTS, where $!$ is defined by the transitions $s_1 \xrightarrow{f} \frac{0.4}{s_3} + \frac{0.6}{s_4}$, $s_2 \xrightarrow{f} \frac{0.6}{s_3} + \frac{0.3}{s_4}$, and $s_3 \xrightarrow{f} \frac{0.4}{s_4}$. Then,

$$\begin{aligned} s_1 k s_2 & \xrightarrow{f} \frac{0.4}{s_3 k s_3} + \frac{0.3}{s_3 k s_4} + \frac{0.6}{s_4 k s_3} + \frac{0.3}{s_4 k s_4} \\ s_2 k s_3 & \xrightarrow{f} \frac{0.4}{s_3 k s_4} + \frac{0.3}{s_4 k s_4} \end{aligned}$$

Now, for $n = 2$, we have that

$$\begin{aligned} d_n(s_1 k s_2; s_2 k s_3) &= (d_n(s_1 k s_2; s_2 k s_3)) \\ &= \max(d_n(s_1 k s_2; s_2 k s_3), 1 - \text{ht}(s_1 k s_2), \text{ht}(s_2 k s_3)) \\ &= d_n(s_1 k s_2; s_2 k s_3) \\ &= (d_n(s_1 k s_2; s_2 k s_3)) \\ &= \max(d_n(s_1 k s_2; s_2 k s_3), 1 - \text{ht}(s_1 k s_2), \text{ht}(s_2 k s_3)) \\ &= \max(0; 0.2) \\ &= 0.2 \end{aligned}$$

Hence $d_b^L(s_1 k s_2; s_2 k s_3) = 0.2$.

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

	$(s_1; s_2)$	$(s_2; s_3)$	$(s_1 k s_2; s_2 k s_3)$
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 $d_b(s_1 k s_2; s_2 k s_3) = d_b(s_1; s_2) \wedge d_b(s_2; s_3)$, i.e., the parallel composition is non-expansive. More generally, we have the following theorem.

Theorem 4. Let $(S; A; !)$ be an NFTS and $\gamma_1, \gamma_2 \in F(S)$ with $d_b(\gamma_i; \gamma_i) = i$, where $i = 1, 2$. Then we have

$$d_b(\gamma_1 k \gamma_2; \gamma_1 k \gamma_2) = 1 - i$$

Proof. Let $F(S) \times F(S) = \{ \gamma_1 k \gamma_2 : \gamma_1, \gamma_2 \in F(S) \}$. We define the function $D : (F(S) \times F(S)) \rightarrow [0, 1]$ by

$$D(\gamma_1 k \gamma_2, \gamma_1 k \gamma_2) \stackrel{\text{def}}{=} d_b(\gamma_1; \gamma_1) \wedge d_b(\gamma_2; \gamma_2)$$

for any $\gamma_1, \gamma_2 \in F(S) \times F(S)$.

It is easy to check that D is a Δ -metric on $F(S) \times F(S)$. We claim that D is a pre-fixed point of the function $\Phi : F(S) \times F(S) \rightarrow [0, 1]$ defined by

For any $\gamma_1 k \gamma_2; \gamma_1 k \gamma_2 \in F(S) \times F(S)$, we first check that $D(\gamma_1 k \gamma_2; \gamma_1 k \gamma_2) = \Phi(\gamma_1 k \gamma_2; \gamma_1 k \gamma_2)$. It holds because

$$\begin{aligned} D(\gamma_1 k \gamma_2; \gamma_1 k \gamma_2) &= d_b(\gamma_1; \gamma_1) \wedge d_b(\gamma_2; \gamma_2) \\ &= [1 - \text{ht}(\gamma_1), \text{ht}(\gamma_1)] \wedge [1 - \text{ht}(\gamma_2), \text{ht}(\gamma_2)] \\ &= 1 - \text{ht}(\gamma_1) \wedge \text{ht}(\gamma_1) \wedge (1 - \text{ht}(\gamma_2) \wedge \text{ht}(\gamma_2)) \\ &= 1 - \text{ht}(\gamma_1) \wedge \text{ht}(\gamma_2), \text{ht}(\gamma_1) \wedge \text{ht}(\gamma_2) \\ &= 1 - \text{ht}(\gamma_1 k \gamma_2), \text{ht}(\gamma_1 k \gamma_2) \end{aligned}$$

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 $\text{ht}(\gamma_1) = \gamma_1(s_0)$ and $\text{ht}(\gamma_2) = \gamma_2(t_0)$, then $\text{ht}(\gamma_1 k \gamma_2) = \sup_{(s; t) \in S \times S} \gamma_1(s) \wedge \gamma_2(t) = \gamma_1(s_0) \wedge \gamma_2(t_0) = \text{ht}(\gamma_1) \wedge \text{ht}(\gamma_2)$; similarly, $\text{ht}(\gamma_1 k \gamma_2) = \text{ht}(\gamma_1) \wedge \text{ht}(\gamma_2)$.

Next, suppose that $\gamma_1 k \gamma_2 \xrightarrow{a} \gamma_1' k \gamma_2'$. We need to show that there exists some $\gamma_1'' k \gamma_2''$ such that $\gamma_1 k \gamma_2 \xrightarrow{a} \gamma_1'' k \gamma_2''$ and $D(\gamma_1 k \gamma_2; \gamma_1' k \gamma_2') = D(\gamma_1 k \gamma_2; \gamma_1'' k \gamma_2'')$. There are four

cases: $a_2 \in A_1 \setminus A_2$, $a_2 \in A_1 \cap A_2$, $a_2 \in A_2 \setminus A_1$, and the same state and are given the same treatment. For example, $a_2 \in A_1 \cap A_2$. We only give the details for the first case; the other cases may have the following two transition: other cases are simpler and can be proved in a similar way.

In the case of $a_2 \in A_1 \setminus A_2$, we get by the definition of the parallel composition that there are $\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}$ and $\frac{0}{2} \xrightarrow{a_2} \frac{0}{2}$. Since the least fixed point of d_b is a pre-fixed point as well. By Lemma 4 there are $\frac{0}{1}$ and $\frac{0}{2}$ such that $\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}$ and $\frac{0}{2} \xrightarrow{a_2} \frac{0}{2}$ with $d_b(\frac{0}{1}; \frac{0}{1}) = d_b(\frac{0}{1}; \frac{0}{1})$ and $d_b(\frac{0}{2}; \frac{0}{2}) = d_b(\frac{0}{2}; \frac{0}{2})$. Clearly, we have $\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}$ and it remains to verify that $D(\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}; \frac{0}{1} \xrightarrow{a_2} \frac{0}{1}) = D(\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}; \frac{0}{1} \xrightarrow{a_2} \frac{0}{1})$. This follows because it is easy to see that

$$d_b(\frac{0}{1}; \frac{0}{1}) = d_b(\frac{0}{2}; \frac{0}{2}) \quad d_b(\frac{0}{1}; \frac{0}{1}) = d_b(\frac{0}{2}; \frac{0}{2}):$$

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

$$\begin{aligned} d_b(\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}; \frac{0}{1} \xrightarrow{a_2} \frac{0}{1}) &= D(\frac{0}{1} \xrightarrow{a_2} \frac{0}{1}; \frac{0}{1} \xrightarrow{a_2} \frac{0}{1}) \\ &= d_b(\frac{0}{1}; \frac{0}{1}) = d_b(\frac{0}{2}; \frac{0}{2}) \\ &= \frac{0}{1} = \frac{0}{2}. \end{aligned}$$

This completes the proof. $\square$

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

Corollary 5. If $\mu_1 \sim \mu_1$ and $\mu_2 \sim \mu_2$, then $\mu_1 \times \mu_2 \sim \mu_1 \times \mu_2$.

Proof. It is straightforward by using Theorem 4 and Corollary 4. $\square$

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 be assisted by using 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, 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 S , 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 $F \xrightarrow{b_1[0.3]} S$ 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

$$P \xrightarrow{a_1} \frac{0.1}{P} + \frac{0.9}{F}g \quad \text{and} \quad P \xrightarrow{a_2} \frac{0.6}{F} + \frac{0.2}{E}g;$$

So we obtain an NFTS $(S; A; !)$, where $S = \{P; F; E\}$, $A = \{a_1; a_2; b_1; b_2; b_3; g\}$ and $!F(S) = A \setminus F(S)$.

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.

$$\begin{aligned} \overline{P} \xrightarrow{a_1} \frac{0.1}{P} + \frac{0.9}{F}g; \quad \overline{P} \xrightarrow{a_2} \frac{0.6}{P} + \frac{0.4}{F}g \\ \overline{F} \xrightarrow{a_1} \frac{0.5}{F} + \frac{0.4}{E}g; \quad \overline{F} \xrightarrow{a_2} \frac{0.3}{F} + \frac{0.8}{E}g \\ \overline{F} \xrightarrow{b_1} \frac{0.2}{P}g; \quad \overline{F} \xrightarrow{b_2} \frac{0.3}{P}g; \quad \overline{E} \xrightarrow{b_3} \frac{0.1}{P} + \frac{0.3}{F}g; \end{aligned}$$

Suppose now that there are two patients: Alice and Bob. The physician describes the patients' states $S_A = [0.8; 0.2; 0.1]$ 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

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

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 $\varphi_{a_1, b_3, i} T$ such that $\llbracket \varphi_{a_1, b_3, i} T \rrbracket(\overline{S_A}) = 0.2$ and $\llbracket \varphi_{a_1, b_3, i} T \rrbracket(\overline{S_B}) = 0.3$ 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 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, $S(S_A; S_B) = 0.7$ and $S(S_A; S_B) = 0.5$ in the above two cases, respectively, which measures the similarity between Alice's and Bob's conditions in the progress of treatment. A little surprising that $d_b^N(\overline{S_A}; \overline{S_B})$ is equal to $d_f(S_A; S_B)$ in the above two cases, 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 they are not always equal to d_f for any $s; t \in S$.

On the other hand, if we take a different fuzzy logic, then $d_b^N(\overline{S_A}; \overline{S_B})$ is in general different. For example, under $\mathcal{F}$, we get that $d_b^N(\overline{S_A}; \overline{S_B}) = 0.4$ when $S_A = [0.5; 0.5; 0]$ and $S_B = [0; 0.5; 0.5]$, but $d_b^N(\overline{S_A}; \overline{S_B}) = \frac{1}{3}$ 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 Lioreti [9] proposed a notion of bisimulation and applied it

fuzzy reasoning. Kupferman and Lustig [11] studied a detailed simulation between two lattice-valued Kripke structures, which are applied to latticed games. P. Eleftheriou et al. [10] studied bisimulation for Heyting-valued modal languages. Pan et al. [12] investigated simulation for lattice-valued doubly indexed transition systems. Cao et al. [1] defined a bisimulation relation between two different FTSSs 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. Qian 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. Valjoo [6] defined bisimulations for fuzzy Kripke structures. Damjanović et al. [8] also studied simulation and bisimulation for fuzzy automata in a similar manner, as also described in [3], [7]. Wu and Deng [14] considered bisimulations for FTSSs.

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

We can prove that if R is a state-based bisimulation, then the lifted relation R^Y (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 s and 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 bisimulation for FTSSs 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 $G(\cdot)$ as follows:

$$s \models \langle a \rangle \langle b \rangle c \text{ iff } \exists s' (s \xrightarrow{a} s' \wedge s' \models b) \text{ and } \forall s' (s \xrightarrow{a} s' \Rightarrow s' \models c)$$

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 $\langle a \rangle \langle b \rangle c$ in our logic plays a key role in proving for their invaluable suggestions and comments.

Theorem 3. One can ask whether we can use $\langle a \rangle \langle b \rangle c$ instead of $\langle a \rangle \langle b \rangle c$ in our setting. The answer is negative. The reason is that Lemma 6 does not hold for $\langle a \rangle \langle b \rangle c$. In addition, Fan [44] has pointed out that the 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 logics L^H . The relationship between L^H and $G(\cdot)$, 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 abstraction metric based on triangular-conorm and constructed a distribution-based bisimilarity metric for measuring the behavioural distance between generalized states in FTSSs. 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 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 it. 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 settings, $\sup_{j \in J} x_j \leq 2 \wedge \inf_{j \in J} x_j < r < g < r$ 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 is 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

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. Damjanović, and M. Basić, "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. Jancić, and N. Damjanović, "Computation of the greatest simulations and bisimulations between fuzzy automata," *Fuzzy Sets Syst.* vol. 208, pp. 22-42, 2012.
- [8] N. Damjanović, 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. Loret, "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," *Logic Comput.* vol. 22, no. 2, pp. 213-235, 2012.
- [11] O. Kupferman and Y. Lustig, "Lattice simulation relations and games," *Int. J. Found. Comput. Syst.* vol. 21, no. 2, pp. 167-189, 2010.
- [12] H. Y. Pan, Y. Z. Cao, M. Zhang, and Y. X. Chen, "Simulation 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. Terra, 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/1503.03911>.
- [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 simulations for discrete probabilistic systems," *Proc. FOSSACS 2007*, LNCS 4423, pp. 287-301, 2007.
- [24] R. Segala and N. A. Lynch, "Probabilistic simulations of probabilistic processes," *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: 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 fuzzy 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 ordering," *Inf. Sci.* vol. 3, no. 2, pp. 177-200, 1971.
- [36] W. Du, Y. Deng, and G. Gebler, "Behavioural Pseudometric 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. Liao, "Logical characterizations of regular equivalence in weighted social networks," *Artif. Intell.* vol. 214, pp. 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).

