Bridging LTL$_f$ Inference to GNN Inference for Learning LTL$_f$ Formulae (Supplement)

 # 5484

Proof of Theorem 1

We first recall the basic notations, definitions, and property.

Definition 1. *Let ϕ be an LTL$_f$ formula. The set of* sub-formulae $\mathsf{sub}(\phi)$ *of ϕ is defined recursively as follows:*

- *if $\phi = p$, then* $\mathsf{sub}(\phi) = \{p\}$*;*
- *if $\phi = \mathsf{o}_1\ \phi'$, then* $\mathsf{sub}(\phi) = \{\phi\} \cup \mathsf{sub}(\phi')$*;*
- *if $\phi = \phi'\ \mathsf{o}_2\ \phi''$, then* $\mathsf{sub}(\phi) = \{\phi\} \cup \mathsf{sub}(\phi') \cup \mathsf{sub}(\phi'')$*,*

where $p \in \mathbb{P} \cup \{\top, \bot\}$, $\mathsf{o}_1 \in \{\neg, \mathsf{X}\}$, $\mathsf{o}_1 \in \{\wedge, untilOP\}$, and ϕ', ϕ'' are LTL$_f$ formulae.

We given the proof of Property 1.

Property 1. *Let ϕ be an LTL$_f$ formula and π a trace. For any sub-trace π_i of π, it fulfills following property:*

- *if $\phi = p$, then $\pi_i \models \phi$ if and only if $\pi_i \models p$;*
- *if $\phi = \neg\phi_1$, then $\pi_i \models \phi$ if and only if $\pi_i \not\models \phi_1$;*
- *If $\phi = \mathsf{X}\phi_1$, then $\pi_i \models \phi$ if and only if $\pi_{i+1} \models \phi_1$;*
- *if $\phi = \phi_1 \wedge \phi_2$, then $\pi_i \models \phi$ if and only if $\pi_i \models \phi_1$ and $\pi_i \models \phi_2$;*
- *if $\phi = \phi_1 \mathsf{U} \phi_2$, then $\pi_i \models \phi$ if and only if it fulfills two conditions: (1) $\pi_i \models \phi_2$ or $\pi_i \models \phi_1$; (2) $\pi_i \models \phi_2$ or $\pi_{i+1} \models \phi$;*

where ϕ, ϕ_1, ϕ_2 are LTL$_f$ formulae, and $p \in \mathbb{P} \cup \{\top, \bot\}$.

Proof. When $\phi = p$, Property 1 obviously holds. When $\phi = \neg\phi_1$, $\pi_i \models \phi$, *i.e.*, $\pi_i \models \neg\phi_1$, if and only if $\pi_i \not\models \phi_1$ because the definition of satisfaction relation. Similarly, when $\phi = \phi_1 \wedge \phi_2$, Property 1 holds. When $\phi = \phi_1 \mathsf{U} \phi_2$, $\pi_i \models \phi$, *i.e.*, $\pi_i \models \phi_1 \mathsf{U} \phi_2$, if and only if $\exists i \leq k \leq n, \pi_k \models \phi_2$ and $\forall i \leq j < k, \pi_j \models \phi_1$, where n is the size of π, because the definition of satisfaction relation. The right condition includes the following two cases: (1) if $k = i$, $\pi_i \models \phi_2$; (2) if $i < k \leq n$, $\pi_i \models \phi_1$, $\forall i+1 \leq j < k, \pi_j \models \phi_1$, and $\pi_k \models \phi_2$. Because $\pi_{i+1} \models \phi_1 \mathsf{U} \phi_2$ if and only if $\exists i+1 \leq k \leq n, \pi_k \models \phi_2$ and $\forall i+1 \leq j < k, \pi_j \models \phi_1$, where n is the size of π, the condition (2) is equivalent to that if $i < k \leq n$, $\pi_i \models \phi_1$ and $\pi_{i+1} \models \phi_1 \mathsf{U} \phi_2$. Therefore, $\pi_i \models \phi$ if and only if at least one of the following two conditions holds. (1) If $k = i$, $\pi_i \models \phi_2$. (2) $\pi_i \models \phi_1$ and $\pi_{i+1} \models \phi_1 \mathsf{U} \phi_2$. That is, when $\phi = \phi_1 \mathsf{U} \phi_2$, Property 1 holds. In summary, Property 1 have been proved. □

Definition 2. *Let ϕ be an LTL$_f$ formula. The* semantics unfolding *of ϕ, denoted by* $\mathsf{unfold}(\phi)$*, and its* semantics elements *of ϕ, denoted by* $\mathsf{element}(\phi)$*, are defined as follows:*

- *if $\phi = p$, then* $\mathsf{unfold}(\phi) = p$ *and* $\mathsf{element}(\phi) = \{p\}$*;*
- *if $\phi = \neg\phi_i$, then* $\mathsf{unfold}(\phi) = \neg\phi_i$ *and* $\mathsf{element}(\phi) = \{\phi_i\}$*;*
- *if $\phi = \phi_i \wedge \phi_j$, then* $\mathsf{unfold}(\phi) = \phi_i \wedge \phi_j$ *and* $\mathsf{element}(\phi) = \{\phi_i, \phi_j\}$*;*
- *if $\phi = \mathsf{X}\phi_i$, then* $\mathsf{unfold}(\phi) = \mathsf{X}\phi_i$ *and* $\mathsf{element}(\phi) = \{\mathsf{X}\phi_i\}$*;*
- *if $\phi = \phi_i \mathsf{U} \phi_j$, then* $\mathsf{unfold}(\phi) = \phi_j \vee (\phi_i \wedge \mathsf{X}\phi)$ *and* $\mathsf{element}(\phi) = \{\phi_i, \phi_j, \mathsf{X}\phi\}$*,*

where $p \in \mathbb{P} \cup \{\top, \bot\}$ and ϕ_i, ϕ_j are LTL$_f$ formulae.

We given the proof of Property 2.

Property 2. *Let ϕ be an LTL$_f$ formula. For any $\phi_i \in \mathsf{sub}(\phi)$,* $\mathsf{element}(\phi_i) \subseteq \{\phi_j, \mathsf{X}\phi_j | \phi_j \in \mathsf{sub}(\phi)\}$ *holds.*

Proof. We prove that by considering all the shape of ϕ_i as follows, where $p \in \mathbb{P} \cup \{\top, \bot\}$ and ϕ_j, ϕ_k are LTL$_f$ formulae. If $\phi_i = p$, we have $\mathsf{element}(\phi_i) = \{p\}$. Because $p \in \mathsf{sub}(\phi_i)$ and $\phi_i \in \mathsf{sub}(\phi)$, $p \in \mathsf{sub}(\phi)$ holds. If $\phi_i = \neg\phi_j$, we have $\mathsf{element}(\phi_i) = \{\phi_j\}$. Because $\phi_j \in \mathsf{sub}(\phi_i)$ and $\phi_i \in \mathsf{sub}(\phi)$, $\phi_j \in \mathsf{sub}(\phi)$ holds. If $\phi_i = \phi_j \wedge \phi_k$, we have $\mathsf{element}(\phi_i) = \{\phi_j, \phi_k\}$. Because $\phi_j, \phi_k \in \mathsf{sub}(\phi_i)$ and $\phi_i \in \mathsf{sub}(\phi)$, $\phi_j, \phi_k \in \mathsf{sub}(\phi)$ holds. If $\phi_i = \mathsf{X}\phi_j$, we have $\mathsf{element}(\phi_i) = \{\mathsf{X}\phi_j\}$. Because $\mathsf{X}\phi_j \in \mathsf{sub}(\phi_i)$ and $\phi_i \in \mathsf{sub}(\phi)$, $\mathsf{X}\phi_j \in \mathsf{sub}(\phi)$ holds. If $\phi_i = \phi_j \mathsf{U} \phi_k$, we have $\mathsf{element}(\phi_i) = \{\phi_j, \phi_k, \mathsf{X}\phi_i\}$. Because $\phi_j, \phi_k, \mathsf{X}\phi_i \in \{\phi_l, \mathsf{X}\phi_l | \phi_l \in \mathsf{sub}(\phi_i)\}$ and $\phi_i \in \mathsf{sub}(\phi)$, $\phi_j, \phi_k, \mathsf{X}\phi_i \in \{\phi_l, \mathsf{X}\phi_l | \phi_l \in \mathsf{sub}(\phi)\}$ holds. Therefore, Property 2 have been proved. □

Definition 3. *Let ϕ be an LTL$_f$ formula. Its* LTL$_f$ graph *G_ϕ is a four-tuple $(V_\phi, E_\phi, \mathsf{W}_\phi, \mathsf{b}_\phi)$ defined as follows, where V_ϕ is a set of vertex, $E_\phi \subseteq V_\phi \times V_\phi$, $\mathsf{W}_\phi \colon E_\phi \to \mathbb{N}$, and $\mathsf{b}_\phi \colon V_\phi \to \mathbb{N}$. V_ϕ and E_ϕ are initialized as $\{v_\phi\}$ and $\emptyset$, respectively. For each sub-formula $\phi_i \in \mathsf{sub}(\phi)$, V_ϕ, E_ϕ, W_ϕ, and b_ϕ iteratively are constructed as follows:*

- *if* $\mathsf{unfold}(\phi_i) = p$*, then $V_\phi = V_\phi \cup \{v_p\}$ and $\mathsf{b}_\phi(v_p) = 0$;*
- *if* $\mathsf{unfold}(\phi_i) = \neg\phi_j$*, then $V_\phi = V_\phi \cup \{v_{\phi_j}\}$, $E_\phi = E_\phi \cup \{\langle v_{\phi_j}, v_{\phi_i} \rangle\}$, $\mathsf{W}_\phi(\langle v_{\phi_j}, v_{\phi_i} \rangle) = -1$, and $\mathsf{b}_\phi(v_{\phi_i}) = 1$;*

- *if* $\mathsf{unfold}(\phi_i) = \phi_j \wedge \phi_k$, *then* $V_\phi = V_\phi \cup \{v_{\phi_j}, v_{\phi_k}\}$, $E_\phi = E_\phi \cup \{\langle v_{\phi_j}, v_{\phi_i}\rangle, \langle v_{\phi_k}, v_{\phi_i}\rangle\}$, $\mathsf{W}_\phi(\langle v_{\phi_j}, v_{\phi_i}\rangle) = 1$, $\mathsf{W}_\phi(\langle v_{\phi_k}, v_{\phi_i}\rangle) = 1$, *and* $\mathsf{b}_\phi(v_{\phi_i}) = -1$;
- *if* $\mathsf{unfold}(\phi_i) = \mathsf{X}\phi_j$, *then* $V_\phi = V_\phi \cup \{v_{\mathsf{X}\phi_j}\}$, $E_\phi = E_\phi \cup \{\langle v_{\mathsf{X}\phi_j}, v_{\phi_i}\rangle\}$, $\mathsf{W}_\phi(\langle v_{\mathsf{X}\phi_j}, v_{\phi_i}\rangle) = 1$, *and* $\mathsf{b}_\phi(v_{\phi_i}) = \mathsf{b}_\phi(v_{\mathsf{X}\phi_j}) = 0$;
- *if* $\mathsf{unfold}(\phi_i) = \phi_k \vee (\phi_j \wedge \mathsf{X}\phi_i)$, *then* $V_\phi = V_\phi \cup \{v_{\phi_k}, v_{\phi_j}, v_{\mathsf{X}\phi_i}\}$, $E_\phi = E_\phi \cup \{\langle v_{\phi_k}, v_{\phi_i}\rangle, \langle v_{\phi_j}, v_{\phi_i}\rangle, \langle v_{\mathsf{X}\phi_i}, v_{\phi_i}\rangle\}$, $\mathsf{W}_\phi(\langle v_{\phi_k}, v_{\phi_i}\rangle) = 2$, $\mathsf{W}_\phi(\langle v_{\phi_j}, v_{\phi_i}\rangle) = 1$, $\mathsf{W}_\phi(\langle v_{\mathsf{X}\phi_i}, v_{\phi_i}\rangle) = 1$, $\mathsf{b}_\phi(v_{\phi_i}) = -1$, *and* $\mathsf{b}_\phi(v_{\mathsf{X}\phi_i}) = 0$,

where $p \in \mathbb{P} \cup \{\top, \bot\}$ *and* ϕ_j, ϕ_k *are LTL$_f$ formulae.*

Definition 4. *Let* ϕ *be an LTL$_f$ formula such that* $|\mathsf{sub}(\phi)| = L$, $(V_\phi, E_\phi, \mathsf{W}_\phi, \mathsf{b}_\phi)$ *an LTL$_f$ graph of* ϕ, *and* $\pi = s_0, s_1, \ldots, s_n$ *a trace. For each state* $s_i \in \pi$, *let* $\mathbf{x}_{s_i} \in \mathbb{R}^L$ *represent a vector of the state* $s_i \in \pi$. $\mathbf{x}_{s_i}$ *is defined as a vector* $[x_1, \ldots, x_L]$ *one-to-one corresponding to* $\mathsf{sub}(\phi)$ *such that for all* $1 \leq i \leq L$, x_i *corresponds to* $\phi_i \in \mathsf{sub}(\phi)$ *and that for all* $1 \leq j < k \leq L$, $\phi_k \notin \mathsf{sub}(\phi_j)$. *Let* $\mathbf{x}_{s_i}^{(t)}$ *be the value of* $\mathbf{x}_{s_i}$ *at time* t. $\mathbf{x}_{s_i}^{(0)}$ *is defined such that for all* $1 \leq j \leq L, (\mathbf{x}_{s_i}^{(0)})_j = 1$ *if* $\phi_j \in s_i$ *or* $(\mathbf{x}_{s_i}^{(0)})_j = 0$ *otherwise.* $\mathbf{x}_{s_i}^{(t)}$ *is defined recursively as follows:*

$$\mathbf{x}_{s_i}^{(t)} = \sigma(\mathbf{C}_\phi \mathbf{x}_{s_i}^{(t-1)} + \mathbf{A}_\phi \mathbf{x}_{s_{i+1}} + \mathbf{b}_\phi), \tag{1}$$

where $t \geq 1$, $\sigma(x) = \min(\max(0, x), 1)$, *and* $\mathbf{C}_\phi, \mathbf{A}_\phi \in \mathbb{R}^{L \times L}$ *and* $\mathbf{b}_\phi \in \mathbb{R}^L$ *are parameters defined as follows:*

$$(\mathbf{C}_\phi)_{ij} = \begin{cases} \mathsf{W}_\phi(\langle v_{\phi_j}, v_{\phi_i}\rangle), & \text{if } \langle v_{\phi_j}, v_{\phi_i}\rangle \in E_\phi \text{ and } \phi_j \in \mathsf{sub}(\phi) \\ 0, & \text{otherwise,} \end{cases}$$

$$(\mathbf{A}_\phi)_{ij} = \begin{cases} \mathsf{W}_\phi(\langle v_{\phi_j}, v_{\phi_i}\rangle), & \text{if } \langle v_{\phi_j}, v_{\phi_i}\rangle \in E_\phi \text{ and } \phi_j \in \{\mathsf{X}\phi_k | \phi_k \in \mathsf{sub}(\phi)\} \\ 0, & \text{otherwise,} \end{cases}$$

$$(\mathbf{b}_\phi)_i = \mathsf{b}_\phi(v_{\phi_i}), \quad \text{for all } v_{\phi_i} \in V_\phi \text{ and } \phi_i \in \mathsf{sub}(\phi).$$

By $\mathcal{S}_\phi$ *we denote the* state classifier $\mathcal{S}_\phi$ *which accepts two vectors* $\mathbf{x}_{s_i}^{(0)}, \mathbf{x}_{s_{i+1}}$ *and a number of iterations* $T \in \mathbb{N}$ *as input, and outputs* $\mathbf{x}_{s_i}^{(T)}$, *i.e.,* $\mathbf{x}_{s_i}^{(T)} = \mathcal{S}_\phi(\mathbf{x}_{s_i}^{(0)}, \mathbf{x}_{s_{i+1}}, T)$.

Definition 5. *Let* ϕ *be an LTL$_f$ formula such that* $|\mathsf{sub}(\phi)| = L$, *and* $\pi = s_0, s_1, \ldots, s_n$ *a trace. By* $\mathcal{T}_\phi$ *we denote the* trace classifier *which takes a vector* $\mathbf{x}_{s_i}^{(0)}$ *and a number of iterations* $T \in \mathbb{N}$ *as input and* $\mathbf{x}_{s_i}^{(T)}$ *as output; i.e.,* $\mathbf{x}_{s_i}^{(T)} = \mathcal{T}_\phi(\mathbf{x}_{s_i}^{(0)}, T)$, *where*

$$\mathcal{T}_\phi(\mathbf{x}_{s_i}^{(0)}, T) = \begin{cases} \mathcal{S}_\phi(\mathbf{x}_{s_i}^{(0)}, \mathcal{T}_\phi(\mathbf{x}_{s_{i+1}}^{(0)}, T), T), & 0 \leq i < n \\ \mathcal{S}_\phi(\mathbf{x}_{s_i}^{(0)}, \mathbf{0}, T), & i = n \end{cases} \tag{2}$$

Now, we given the proof of Theorem 1.

Theorem 1. *Let* ϕ *be an LTL$_f$ formula such that* $|\mathsf{sub}(\phi)| = L$. *For every trace* $\pi = s_0, s_1, \ldots, s_n$, $(\mathcal{T}_\phi(\mathbf{x}_{s_0}^{(0)}, L))_L = 1$ *if and only if* $\pi \models \phi$.

Proof. By induction on states, we prove that,

$$\text{for every } s_i \in \pi, \mathcal{S}_\phi(\mathbf{x}_{s_i}^{(0)}, \mathbf{x}_{s_{i+1}}, L) \text{ reaches a fixpoint and } (\mathcal{T}_\phi(\mathbf{x}_{s_i}^{(0)}, L))_L = 1 \text{ if and only if } \pi_i \models \phi. \tag{3}$$

The basic step is that if $s_i = s_n$ and $s_i = s_{n-1}$, the property (3) holds. We first prove that for every $\phi_i \in \mathsf{sub}(\phi)$ it holds that

$$\forall t \geq i, (\mathbf{x}_{s_n}^{(t)})_i = 1 \text{ if } \pi_n \models \phi_i; \text{ otherwise } (\mathbf{x}_{s_n}^{(t)})_i = 0. \tag{4}$$

By the equation (1), we have the expression of $(\mathbf{x}_{s_n}^{(t)})_i$:

$$\sigma(\sum_{j=1}^{L} (\mathbf{C}_\phi)_{ij} (\mathbf{x}_{s_n}^{(t-1)})_j + 0 + (\mathbf{b}_\phi)_i). \tag{5}$$

We prove the property (4) by induction on the number of sub-formulae of every ϕ_i.

The basic step is that ϕ_i has one sub-formula, *i.e.*, ϕ_i is an atomic proposition. We first prove that $(\mathbf{x}_{s_n}^{(1)})_i = 1$ if and only if $\pi_n \models \phi_i$. Since ϕ_i is an atomic proposition, we know that $(\mathbf{C}_\phi)_{ii} = 1$, $(\mathbf{C}_\phi)_{ij} = 0$ for all $j \neq i$, $(\mathbf{A}_\phi)_{ij} = 0$ for all j, and $(\mathbf{b}_\phi)_i = 0$. From the equation (5), we obtain that $(\mathbf{x}_{s_n}^{(1)})_i = \sigma((\mathbf{x}_{s_n}^{(0)})_i) = 1$ if and only if $(\mathbf{x}_{s_n}^{(0)})_i \geq 1$ that can only happen when $(\mathbf{x}_{s_n}^{(0)})_i = 1$. Then $(\mathbf{x}_{s_n}^{(1)})_i = 1$ if and only if $\pi_n \models \phi_i$. It is easy to prove that, for every $t > 1$, $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $\pi_n \models \phi_i$.

The inductive hypothesis is that ϕ_i has more than one sub-formula and that for every sub-formulae of ϕ_i the property (4) holds.

In the inductive step, we consider the following cases, corresponding to the cases for the shape of LTL$_f$ formula.

If $\phi_i = \neg\phi_j$, we know that $(\mathbf{C}_\phi)_{ij} = -1$, $(\mathbf{b}_\phi)_i = 1$ and others values in the i^{th} row of $\mathbf{C}_\phi$, $\mathbf{A}_\phi$, and $\mathbf{b}_\phi$ are 0. From the equation (5), we obtain that $(\mathbf{x}_{s_n}^{(t)})_i = \sigma(-(\mathbf{x}_{s_n}^{(t-1)})_j + 1)$. Because the inductive hypothesis, we have, for $t \geq j$ where $j < i$, $(\mathbf{x}_{s_n}^{(t)})_j = 1$ if $\pi_n \models \phi_j$; otherwise $(\mathbf{x}_{s_n}^{(t)})_j = 0$. When $t \geq i$, since $(\mathbf{x}_{s_n}^{(t)})_i = \sigma(-(\mathbf{x}_{s_n}^{(t-1)})_j + 1)$, we have that $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $1 - (\mathbf{x}_{s_n}^{(t-1)})_j \geq 1$ that can only happen when $(\mathbf{x}_{s_n}^{(t-1)})_j = 0$. Then $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $\pi_n \models \neg\phi_j$.

If $\phi_i = \phi_j \wedge \phi_k$, we know that $(\mathbf{C}_\phi)_{ij} = 1$, $(\mathbf{C}_\phi)_{ik} = 1$, $(\mathbf{b}_\phi)_i = -1$, $(\mathbf{C}_\phi)_{il} = 0$ for all $l \neq j, k$, and $(\mathbf{A}_\phi)_{il} = 0$ for all l. From the equation (5), we obtain that $(\mathbf{x}_{s_n}^{(t)})_i = \sigma((\mathbf{x}_{s_n}^{(t-1)})_j + (\mathbf{x}_{s_n}^{(t-1)})_k - 1)$. Because the inductive hypothesis, we have, for $t \geq j$ where $j < i$, $(\mathbf{x}_{s_n}^{(t)})_j = 1$ if $\pi_n \models \phi_j$; otherwise $(\mathbf{x}_{s_n}^{(t)})_j = 0$. Similarly, we have, for $t \geq k$ where $k < i$, $(\mathbf{x}_{s_n}^{(t)})_k = 1$ if $\pi_n \models \phi_k$; otherwise $(\mathbf{x}_{s_n}^{(t)})_k = 0$. When $t \geq i$, since $(\mathbf{x}_{s_n}^{(t)})_i = \sigma((\mathbf{x}_{s_n}^{(t-1)})_j + (\mathbf{x}_{s_n}^{(t-1)})_k - 1)$, we have that $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $(\mathbf{x}_{s_n}^{(t-1)})_j + (\mathbf{x}_{s_n}^{(t-1)})_k - 1 \geq 1$ that can only happen when $(\mathbf{x}_{s_n}^{(t-1)})_j = 1$ and $(\mathbf{x}_{s_n}^{(t-1)})_k = 1$. Then $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $\pi_n \models \phi_j \wedge \phi_k$.

If $\phi_i = \mathsf{X}\phi_j$, we know that $(\mathbf{A}_\phi)_{ij} = 1$, $(\mathbf{C}_\phi)_{ik} = 0$ for all k, $(\mathbf{A}_\phi)_{ik} = 0$ for all $k \neq j$, and $(\mathbf{b}_\phi)_i = 0$. From the equation (5), we obtain that $(\mathbf{x}_{s_n}^{(t)})_i = \sigma(0) = 0$. Because s_n is the last state, $\pi_n \models \mathsf{X}\phi_j$ always does not hold. It is easy to prove that, for every $t \geq i$, $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $\pi_n \models \mathsf{X}\phi_j$.

If $\phi_i = \phi_j \mathsf{U} \phi_k$, we know that $(\mathbf{C}_\phi)_{ik} = 2$, $(\mathbf{C}_\phi)_{ij} = 1$, $(\mathbf{A}_\phi)_{ii} = 1$, $(\mathbf{b}_\phi)_i = -1$, $(\mathbf{C}_\phi)_{il} = 0$ for all $l \neq j, k$, and $(\mathbf{A}_\phi)_{il} = 0$ for all $l \neq i$. From the equation (5), we obtain that $(\mathbf{x}_{s_n}^{(t)})_i = \sigma(2(\mathbf{x}_{s_n}^{(t-1)})_k + (\mathbf{x}_{s_n}^{(t-1)})_j + 0 - 1)$. Because the inductive hypothesis, we have, for $t \geq j$ where $j < i$, $(\mathbf{x}_{s_n}^{(t)})_j = 1$ if $\pi_n \models \phi_j$; otherwise $(\mathbf{x}_{s_n}^{(t)})_j = 0$. Similarly, we have, for $t \geq k$ where $k < i$, $(\mathbf{x}_{s_n}^{(t)})_k = 1$ if $\pi_n \models \phi_k$; otherwise $(\mathbf{x}_{s_n}^{(t)})_k = 0$. When $t \geq i$, since $(\mathbf{x}_{s_n}^{(t)})_i = \sigma(2(\mathbf{x}_{s_n}^{(t-1)})_k + (\mathbf{x}_{s_n}^{(t-1)})_j + 0 - 1)$, we have that $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $2(\mathbf{x}_{s_n}^{(t-1)})_k + (\mathbf{x}_{s_n}^{(t-1)})_j - 1 \geq 1$ that can only happen when $(\mathbf{x}_{s_n}^{(t-1)})_k = 1$. Then $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $\pi_n \models \phi_k$. Because s_n is the last state, $\pi_n \models \phi_j \wedge \mathsf{X}\phi_i$ always does not hold. Therefore, $(\mathbf{x}_{s_n}^{(t)})_i = 1$ if and only if $\pi_n \models \phi_k \vee (\phi_j \wedge \mathsf{X}\phi_i)$, *i.e.*, $\pi_n \models \phi_j \mathsf{U} \phi_k$.

Now, the property (4) has been proved. We have that $t \geq L, (\mathbf{x}_{s_n}^{(t)})_L = 1$ if $\pi_n \models \phi_L$, *i.e.*, $\pi_n \models \phi$; otherwise $(\mathbf{x}_{s_n}^{(t)})_L = 0$. Therefore, we have proved that $\mathcal{S}_\phi(\mathbf{x}_{s_n}^{(0)}, \mathbf{0}, L)$ reaches a fixpoint and $(\mathcal{T}_\phi(\mathbf{x}_{s_n}^{(0)}, L))_L = 1$ if and only if $\pi_n \models \phi$.

Then, we prove that for every $\phi_i \in \mathsf{sub}(\phi)$ it holds that

$$\forall t \geq i, (\mathbf{x}_{s_{n-1}}^{(t)})_i = 1 \text{ if } \pi_{n-1} \models \phi_i; \text{otherwise } (\mathbf{x}_{s_{n-1}}^{(t)})_i = 0. \tag{6}$$

By the equation (1), we have the expression of $(\mathbf{x}_{s_{n-1}}^{(t)})_i$:

$$\sigma(\sum_{j=1}^{L} (\mathbf{C}_\phi)_{ij} (\mathbf{x}_{s_{n-1}}^{(t-1)})_j + \sum_{j=1}^{L} (\mathbf{A}_\phi)_{ij} (\mathbf{x}_{s_n})_j + (\mathbf{b}_\phi)_i). \tag{7}$$

We prove the property (6) by induction on the number of sub-formulae of every ϕ_i.

The basic step is that ϕ_i has one sub-formula. The proof is similar to the proof of the basic step of the property (4).

The inductive hypothesis is that ϕ_i has more than one sub-formula and that for every sub-formulae of ϕ_i the property (6) holds.

In the inductive step, we consider the following cases, corresponding to the cases for the shape of LTL$_f$ formula. The proof of the cases $\phi_i = \neg\phi_j$ and $\phi_i = \phi_j \wedge \phi_k$ are respectively similar to the proof of the cases $\phi_i = \neg\phi_j$ and $\phi_i = \phi_j \wedge \phi_k$ of the inductive step of the property (4). We prove the cases $\phi_i = \mathsf{X}\phi_j$ and $\phi_i = \phi_j \mathsf{U} \phi_k$ as follows.

If $\phi_i = \mathsf{X}\phi_j$, we know that $(\mathbf{A}_\phi)_{ij} = 1$, $(\mathbf{C}_\phi)_{ik} = 0$ for all k, $(\mathbf{A}_\phi)_{ik} = 0$ for all $k \neq j$, and $(\mathbf{b}_\phi)_i = 0$. From the equation (7), we obtain that $(\mathbf{x}_{s_{n-1}}^{(t)})_i = \sigma((\mathbf{x}_{s_n})_j)$. Because the property (4) holds, before computing $\mathbf{x}_{s_{n-1}}^{(t)}$, $\mathbf{x}_{s_n} = \mathcal{S}_\phi(\mathbf{x}_{s_n}^{(0)}, \mathbf{0}, L)$ has reached a fixpoint. Therefore, we have, for $t \geq 0$, $(\mathbf{x}_{s_n})_j = 1$ if $\pi_n \models \phi_j$; otherwise $(\mathbf{x}_{s_n})_j = 0$. When $t \geq i$, since $(\mathbf{x}_{s_{n-1}}^{(t)})_i = \sigma((\mathbf{x}_{s_n})_j)$, we have that $(\mathbf{x}_{s_{n-1}}^{(t)})_i = 1$ if and only if $(\mathbf{x}_{s_n})_j \geq 1$ that can only happen when $(\mathbf{x}_{s_n})_j = 1$, *i.e.*, $\pi_n \models \phi_j$. Therefore, $\pi_{n-1} \models \mathsf{X}\phi_j$.

If $\phi_i = \phi_j \mathsf{U} \phi_k$, we know that $(\mathbf{C}_\phi)_{ik} = 2$, $(\mathbf{C}_\phi)_{ij} = 1$, $(\mathbf{A}_\phi)_{ii} = 1$, $(\mathbf{b}_\phi)_i = -1$, $(\mathbf{C}_\phi)_{il} = 0$ for all $l \neq j, k$, and $(\mathbf{A}_\phi)_{il} = 0$ for all $l \neq i$. From the equation (7), we obtain that $(\mathbf{x}_{s_{n-1}}^{(t)})_i = \sigma(2(\mathbf{x}_{s_{n-1}}^{(t-1)})_k + (\mathbf{x}_{s_{n-1}}^{(t-1)})_j + (\mathbf{x}_{s_n})_i - 1)$. Because the property (4) holds, before computing $\mathbf{x}_{s_{n-1}}^{(t)}$, $\mathbf{x}_{s_n} = \mathcal{S}_\phi(\mathbf{x}_{s_n}^{(0)}, \mathbf{0}, L)$ has reached a fixpoint. Therefore, we have, for $t \geq 0$, $(\mathbf{x}_{s_n})_i = 1$ if $\pi_n \models \phi_i$; otherwise $(\mathbf{x}_{s_n})_i = 0$. Because the inductive hypothesis, we have, for $t \geq j$ where $j < i$, $(\mathbf{x}_{s_{n-1}}^{(t)})_j = 1$ if $\pi_{n-1} \models \phi_j$; otherwise $(\mathbf{x}_{s_{n-1}}^{(t)})_j = 0$. Similarly, we have, for $t \geq k$ where $k < i$, $(\mathbf{x}_{s_{n-1}}^{(t)})_k = 1$ if $\pi_{n-1} \models \phi_k$; otherwise $(\mathbf{x}_{s_{n-1}}^{(t)})_k = 0$. When $t \geq i$, since $(\mathbf{x}_{s_{n-1}}^{(t)})_i = \sigma(2(\mathbf{x}_{s_{n-1}}^{(t-1)})_k + (\mathbf{x}_{s_{n-1}}^{(t-1)})_j + (\mathbf{x}_{s_n})_i - 1)$, we have that $(\mathbf{x}_{s_{n-1}}^{(t)})_i = 1$ if and only if $2(\mathbf{x}_{s_{n-1}}^{(t-1)})_k + (\mathbf{x}_{s_{n-1}}^{(t-1)})_j + (\mathbf{x}_{s_n})_i - 1 \geq 1$ that can only happen when $(\mathbf{x}_{s_{n-1}}^{(t-1)})_k = 1$ or $(\mathbf{x}_{s_{n-1}}^{(t-1)})_j = 1 \wedge (\mathbf{x}_{s_n})_i = 1$. If $(\mathbf{x}_{s_{n-1}}^{(t-1)})_k = 1$, then $(\mathbf{x}_{s_{n-1}}^{(t)})_i = 1$ if and only if $\pi_{n-1} \models \phi_k$. If $(\mathbf{x}_{s_{n-1}}^{(t-1)})_j = 1 \wedge (\mathbf{x}_{s_n})_i = 1$, then $(\mathbf{x}_{s_{n-1}}^{(t)})_i = 1$ if and only if $\pi_{n-1} \models \phi_j$ and $\pi_n \models \phi_i$, *i.e.*, $\pi_{n-1} \models \phi_j \wedge \mathsf{X}\phi_i$. Therefore, $(\mathbf{x}_{s_{n-1}}^{(t)})_i = 1$ if and only if $\pi_{n-1} \models \phi_k \vee (\phi_j \wedge \mathsf{X}\phi_i)$, *i.e.*, $\pi_{n-1} \models \phi_j \mathsf{U} \phi_k$.

Now, the property (6) has been proved. We have that $t \geq L, (\mathbf{x}_{s_{n-1}}^{(t)})_L = 1$ if $\pi_{n-1} \models \phi_L$, *i.e.*, $\pi_{n-1} \models \phi$; otherwise $(\mathbf{x}_{s_{n-1}}^{(t)})_L = 0$. Therefore, we have proved that $\mathcal{S}_\phi(\mathbf{x}_{s_{n-1}}^{(0)}, \mathbf{x}_{s_n}, L)$ reaches a fixpoint and $(\mathcal{T}_\phi(\mathbf{x}_{s_{n-1}}^{(0)}, L))_L = 1$ if and only if $\pi_{n-1} \models \phi$.

The basic step of the property (3) has been proved. The inductive hypothesis is that if $s_i = s_l$, where $l \in [1, n-1)$, the property (3) holds.

The inductive step is that if $s_i = s_{l-1}$, the property (3) holds. We prove that for every $\phi_i \in \mathsf{sub}(\phi)$ it holds that

$$\forall t \geq i, (\mathbf{x}_{s_{l-1}}^{(t)})_i = 1 \text{ if } \pi_{l-1} \models \phi_i; \text{otherwise } (\mathbf{x}_{s_{l-1}}^{(t)})_i = 0. \tag{8}$$

By the equation (1), we have the expression of $(\mathbf{x}_{s_{l-1}}^{(t)})_i$:

$$\sigma(\sum_{j=1}^{L} (\mathbf{C}_\phi)_{ij} (\mathbf{x}_{s_{l-1}}^{(t-1)})_j + \sum_{j=1}^{L} (\mathbf{A}_\phi)_{ij} (\mathbf{x}_{s_l})_j + (\mathbf{b}_\phi)_i). \tag{9}$$

We prove the property (8) by induction on the number of sub-formulae of every ϕ_i.

The basic step is that ϕ_i has one sub-formula. The proof is similar to the proof of the basic step of the property (4).

The inductive hypothesis is that ϕ_i has more than one sub-formula and that for every sub-formulae of ϕ_i the property (8) holds.

In the inductive step, we consider the following cases, corresponding to the cases for the shape of LTL$_f$ formula. The proof of the cases $\phi_i = \neg\phi_j$ and $\phi_i = \phi_j \wedge \phi_k$ are respectively similar to the proof of the cases $\phi_i = \neg\phi_j$ and $\phi_i = \phi_j \wedge \phi_k$ of the inductive step of the property (4). We prove the cases $\phi_i = \mathsf{X}\phi_j$ and $\phi_i = \phi_j \mathsf{U} \phi_k$ as follows.

If $\phi_i = \mathsf{X}\phi_j$, we know that $(\mathbf{A}_\phi)_{ij} = 1$, $(\mathbf{C}_\phi)_{ik} = 0$ for all k, $(\mathbf{A}_\phi)_{ik} = 0$ for all $k \neq j$, and $(\mathbf{b}_\phi)_i = 0$. From the equation (9), we obtain that $(\mathbf{x}_{s_{l-1}}^{(t)})_i = \sigma((\mathbf{x}_{s_l})_j)$. Because the inductive hypothesis, before computing $\mathbf{x}_{s_{l-1}}^{(t)}$, $\mathbf{x}_{s_l} = \mathcal{S}_\phi(\mathbf{x}_{s_l}^{(0)}, \mathbf{x}_{s_{l+1}}, L)$ has reached a fixpoint. Therefore, we have, for $t \geq 0$, $(\mathbf{x}_{s_l})_j = 1$ if $\pi_l \models \phi_j$; otherwise $(\mathbf{x}_{s_l})_j = 0$. When $t \geq i$, since $(\mathbf{x}_{s_{l-1}}^{(t)})_i = \sigma((\mathbf{x}_{s_l})_j)$, we have that $(\mathbf{x}_{s_{l-1}}^{(t)})_i = 1$ if and only if $(\mathbf{x}_{s_l})_j \geq 1$ that can only happen when $(\mathbf{x}_{s_l})_j = 1$, *i.e.*, $\pi_{l-1} \models \mathsf{X}\phi_j$.

If $\phi_i = \phi_j \mathsf{U} \phi_k$, we know that $(\mathbf{C}_\phi)_{ik} = 2$, $(\mathbf{C}_\phi)_{ij} = 1$, $(\mathbf{A}_\phi)_{ii} = 1$, $(\mathbf{b}_\phi)_i = -1$, $(\mathbf{C}_\phi)_{il} = 0$ for all $l \neq j, k$, and $(\mathbf{A}_\phi)_{il} = 0$ for all $l \neq i$. From the equation (9), we obtain that $(\mathbf{x}_{s_{l-1}}^{(t)})_i = \sigma(2(\mathbf{x}_{s_{l-1}}^{(t-1)})_k + (\mathbf{x}_{s_{l-1}}^{(t-1)})_j + (\mathbf{x}_{s_l})_i - 1)$. Because the inductive hypothesis, before computing $\mathbf{x}_{s_{l-1}}^{(t)}$, $\mathbf{x}_{s_l} = \mathcal{S}_\phi(\mathbf{x}_{s_l}^{(0)}, \mathbf{x}_{s_{l+1}}, L)$ has reached a fixpoint. Therefore, we have, for $t \geq 0$, $(\mathbf{x}_{s_l})_i = 1$ if $\pi_l \models \phi_i$; otherwise $(\mathbf{x}_{s_l})_i = 0$. Similarly, we have, for $t \geq j$ where $j < i$, $(\mathbf{x}_{s_{l-1}}^{(t)})_j = 1$ if $\pi_{l-1} \models \phi_j$; otherwise $(\mathbf{x}_{s_{l-1}}^{(t)})_j = 0$. And, we have, for $t \geq k$ where $k < i$, $(\mathbf{x}_{s_{l-1}}^{(t)})_k = 1$ if $\pi_{l-1} \models \phi_k$; otherwise $(\mathbf{x}_{s_{l-1}}^{(t)})_k = 0$. When $t \geq i$, since $(\mathbf{x}_{s_{l-1}}^{(t)})_i = \sigma(2(\mathbf{x}_{s_{l-1}}^{(t-1)})_k + (\mathbf{x}_{s_{l-1}}^{(t-1)})_j + (\mathbf{x}_{s_l})_i - 1)$, we have that $(\mathbf{x}_{s_{l-1}}^{(t)})_i = 1$ if and only if $2(\mathbf{x}_{s_{l-1}}^{(t-1)})_k + (\mathbf{x}_{s_{l-1}}^{(t-1)})_j + (\mathbf{x}_{s_l})_i - 1 \geq 1$ that can only happen when $(\mathbf{x}_{s_{l-1}}^{(t-1)})_k = 1$ or $(\mathbf{x}_{s_{l-1}}^{(t-1)})_j = 1 \wedge (\mathbf{x}_{s_l})_i = 1$. If $(\mathbf{x}_{s_{l-1}}^{(t-1)})_k = 1$, then $(\mathbf{x}_{s_{l-1}}^{(t)})_i = 1$ if and only if $\pi_{l-1} \models \phi_k$. If $(\mathbf{x}_{s_{l-1}}^{(t-1)})_j = 1 \wedge (\mathbf{x}_{s_l})_i = 1$, then $(\mathbf{x}_{s_{l-1}}^{(t)})_i = 1$ if and only if $\pi_{l-1} \models \phi_j$ and $\pi_l \models \phi_i$, *i.e.*, $\pi_{l-1} \models \phi_j \wedge \mathsf{X}\phi_i$. Therefore, $(\mathbf{x}_{s_{l-1}}^{(t)})_i = 1$ if and only if $\pi_{l-1} \models \phi_k \vee (\phi_j \wedge \mathsf{X}\phi_i)$, *i.e.*, $\pi_{l-1} \models \phi_j \mathsf{U} \phi_k$.

Now, the property (8) has been proved. In summary, the property (3) has been proved. If $s_i = s_0$ in the property (3), we have $\mathcal{S}_\phi(\mathbf{x}_{s_0}^{(0)}, \mathbf{x}_{s_1}, L)$ reaches a fixpoint and $(\mathcal{T}_\phi(\mathbf{x}_{s_0}^{(0)}, L))_L = 1$ if and only if $\pi_0 \models \phi$, *i.e.*, $\pi \models \phi$. Theorem 1 has been proved. □

Preliminary Results

In this section, we illustrate detail experimental results.

Performance and Robustness

In Figure 1 and Figure 2, accuracies are the classification accuracies of traces in the testing set by the best formula computed by `GLTLf` and `BayesLTL`.

When $\delta \geq 0.1$, `C.&M.` and `MaxSAT-DT` failed (timed out), so we did not mark them in the figures. Since noise data potentially cause the formula perfectly classifying traces to become very long, `C.&M.` designed for noise-free data fails to handle. Although `MaxSAT-DT` supports noise, their experiments (Gaglione et al. 2021) only considered about $\delta = 0.05$. Our results show that, as the noise increases further, `MaxSAT-DT` will fail. The results confirm that their approaches are not robust to noise.

Figure 1 and Figure 2 illustrate that, for all cases, our approach obviously dominates `BayesLTL`, which confirms that our approach has a better performance compared with `BayesLTL`.

We also consider the influence of different k_g on our approach. Overall, the accuracies corresponding to different k_g have the same changing trend. Figure 1 shows that larger networks achieve better accuracies in the same noise rate, even when solving short formulae. However, this advantage is lost as noise increases. Figure 2 shows that the performance our approach does not drop significantly when $\delta \leq 0.3$, which is slightly better than that of `BayesLTL`. However, when $\delta = 0.4$, the downward trend of performance of our approach is significantly weaker than that of `BayesLTL`. The results confirm that our approach is more robust to noise compared with `BayesLTL`.

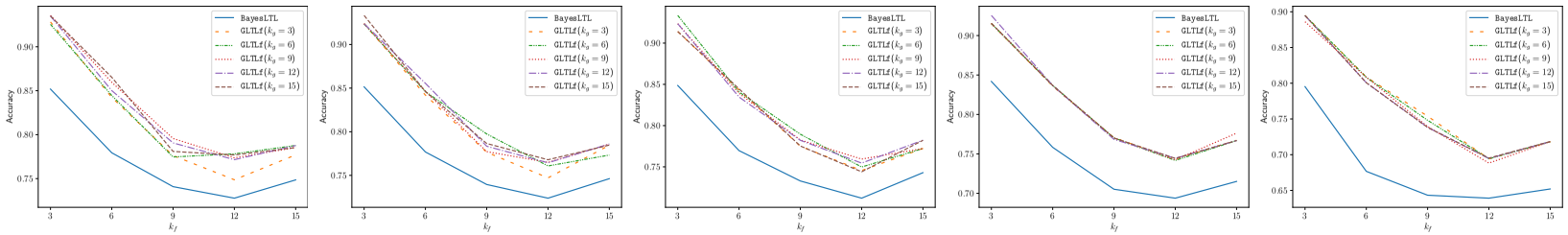
In summary, our approach is superior to the SOTA approaches. Specifically, our approach has the better performance for noise data. Moreover, our approach is stronger robust to k_g and noise data.

Performance of Interpreting

We used that the net accuracy minus the interpreting accuracy to measure the differences between net accuracies and interpreting accuracies (accuracy differences for short), where net accuracies are the classification accuracies of traces in the testing set by the learned GNN classifier and interpreting accuracies are the classification accuracies of traces in the testing set by the formula obtained by interpreting the parameters of the GNN classifier.

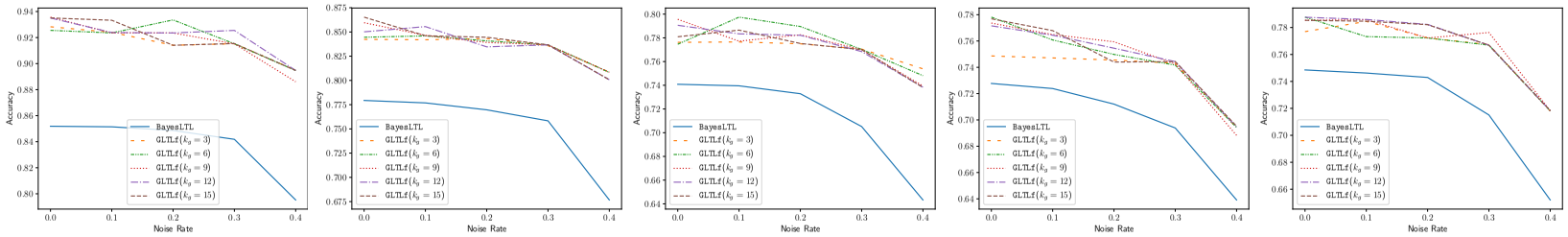
Overall, Figure 3 and Figure 4 shows that the accuracy differences corresponding to different k_g have the same changing trend. We observe that, for all noise rates, the accuracy differences first increases and then decreases with the increase of k_f in Figure 3. As the length of the formula increases, the difficulty of interpreting increases is obvious, which explains why the accuracy differences increases in the beginning. When the length of the formula reaches a certain scale, the decrease of the accuracy differences may be the reason that there is a short formula equivalent to the long formula. It corresponds to the increase of accuracies when $k_f = 15$ in Figure 1. Figure 4 shows that, for all k_f, the accuracy differences first decreases and then increases with the increase of the noise rate. The results show that as the noise increases, the error of interpretation helps to tolerate the noise. But this effect is limited. Studying the relationship between noise and the error of interpretation is a way to improve the noise tolerance of the model.

In summary, the results show that there is a gap between net accuracies and interpreting accuracies. It is interesting and challenging to guide networks to learn interpretable parameters and better interpret network parameters, which is our future work.



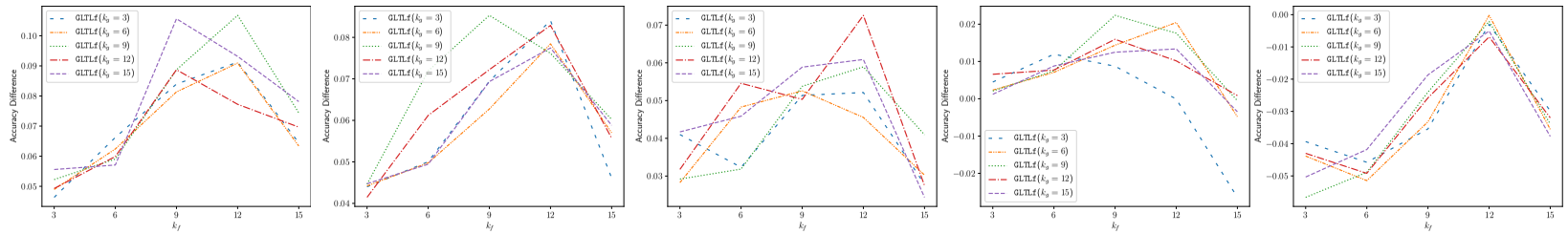
(a) $\delta = 0.0$ (b) $\delta = 0.1$ (c) $\delta = 0.2$ (d) $\delta = 0.3$ (e) $\delta = 0.4$

Figure 1: Accuracies among different δ.



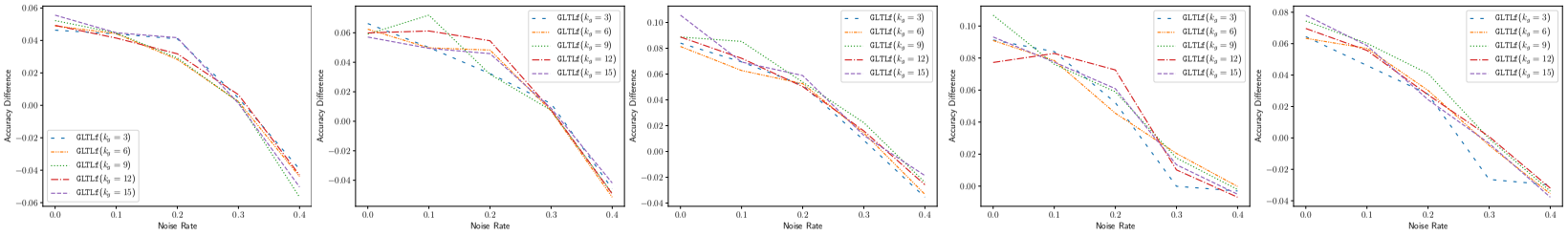
(a) $k_f = 3$ (b) $k_f = 6$ (c) $k_f = 9$ (d) $k_f = 12$ (e) $k_f = 15$

Figure 2: Accuracies among different k_f.



(a) $\delta = 0.0$ (b) $\delta = 0.1$ (c) $\delta = 0.2$ (d) $\delta = 0.3$ (e) $\delta = 0.4$

Figure 3: Difference between net accuracies and interpreting accuracies among different δ.



(a) $k_f = 3$ (b) $k_f = 6$ (c) $k_f = 9$ (d) $k_f = 12$ (e) $k_f = 15$

Figure 4: Difference between net accuracies and interpreting accuracies among different k_f.

References

Gaglione, J.; Neider, D.; Roy, R.; Topcu, U.; and Xu, Z. 2021. Learning Linear Temporal Properties from Noisy Data: A MaxSAT Approach. *CoRR*, abs/2104.15083.