# Reliable Truth Discovery for Dynamic and Dependent Sources

He Zhang, Shuang Wang, Long Chen, Xiaoping Li, *Senior Member, IEEE*, Qing Gao, and Quan Z. Sheng

***Abstract*—In the era of Big Data and generative artificial intelligence (AI), discovering the truth about various objects from different sources has become a pressing topic. Existing studies primarily focus on dependent sources with conflicting information, where sources may copy information from each other. However, real-world scenarios are often more complex, with dynamic dependence relationships among sources over time. This complexity makes it much more difficult to discover the truth. One of the key challenges centers on measuring the dynamic dependence among sources. To address this challenge, we have developed three models: $Depen\_Simple$, $Depen\_Complex$, and $Depen\_Dynamic$. These models are based on the Hidden Markov Model (HMM) and are designed to handle different types of dependencies, namely *simple source dependence*, *complex source dependence*, and *dynamic source dependence*. Based on the constructed models, we propose a generic framework for discovering the latent truth which are evaluated by three HMM-based methods. We conduct extensive experiments on three real-world datasets to evaluate the performance of the proposed methods, and the results demonstrate that all three methods achieve high accuracy over the state-of-the-art methods.**

***Index Terms*—Truth discovery, source dependence, hidden Markov model, big data, time series data.**

## I. Introduction

With advancements in Big Data and AI, integrating reliable information is essential for various applications like Web applications [1], [2], crowd sensing [3], [4], [5], social applications [6], [7], [8] and large language model (LLM), such as ChatGPT, which requires reliable sources for training data. However, data is often noisy and conflicting due to factors [9], [10], [11] like outdated information and data loss. Additionally, some sources may copy information from others, leading to potential inaccuracies and biases. For example, in social media [12], news aggregators [13], or even academic literature [14], dependent sources may propagate errors or outdated information by copying data. The widespread practice of copying among sources, especially on the web [15], further complicates truth discovery. For instance, weather data from different providers may vary due to dynamic changes like inter-source dependencies, leading to potential propagation of errors. In such cases, dependent sources typically have lower quality since they often lack the capability to independently explore the truth and simply copy data, even errors from others. Conversely, independent sources, which are less influenced by others, can provide more reliable information by offering unique perspectives and reducing the risk of shared inaccuracies, making them more reliable. Therefore, discovering dependencies among sources is essential for evaluating source reliability and ensuring the accuracy of truth discovery. This paper focuses on the truth discovery problem with dynamic and dependent sources.

In real-world applications [16], [17], sources report various claims for the same objects (the weather condition in which we are interested), and the claims (temperature) always change with time, making it hard to discover which claim is true. Therefore, it is necessary to discover the constantly changing truth in a dynamic world. Since a high-reliable source is probable to provide the truth value, the claim provided by a high-reliable source is more likely to be the truth. Therefore, how to measure the reliability of sources is critical. In addition, some sources may copy information from other sources and their reliability is alarming. How to distinguish the dependence among sources becomes an important task.

To uncover hidden source dependencies, we address specific challenges: (1) **Source Dependence**: While various truth discovery methods exist [18], [19], [20], [21], most overlook source interdependence or only estimate basic copying probabilities, such as the copying probability between two sources. In real scenarios, copying can occur simultaneously across multiple sources, where a source can copy information from several sources at the same time. The variety of source dependence makes it difficult to accurately measure source reliability. (2) **Partial Independence**: Dependent sources can also self-report, a state known as partial independence. When a source is in

Received 9 November 2024; revised 5 September 2025; accepted 5 November 2025. Date of publication 11 November 2025; date of current version 25 November 2025. This work was supported in part by the National Key Research and Development Program under Grant 2022YFF0902800, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220803, in part by the National Natural Science Foundation of China under Grant 62302095, in part by the Southeast University Interdisciplinary Research Program for Young Scholars, and in part by the Big Data Computing Center of Southeast University. Recommended for acceptance by F. Chiang. *(Corresponding author: Shuang Wang.)*

He Zhang, Shuang Wang, and Long Chen are with the School of Computer Science and Engineering, Key Laboratory of New Generation Artificial Intelligence Technology and its Interdisciplinary Applications, Ministry of Education, Southeast University, Nanjing 211189, China (e-mail: zhanghezhedu@163.com; shuangwang@seu.edu.cn; chen_long@seu.edu.cn).

Xiaoping Li is with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China (e-mail: xpli@seu.edu.cn).

Qing Gao is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: gaoqing@buaa.edu.cn).

Quan Z. Sheng is with the School of Computing, Macquarie University, Sydney, NSW 2109, Australia (e-mail: michael.sheng@mq.edu.au).

Digital Object Identifier 10.1109/TKDE.2025.3631376

partial independence, it is hard to distinguish whether the claims are copied or reported by the source. Their reliability needs to be re-measured, as they exhibit characteristics of both independence and plagiarism which makes it challenging to measure. (3) **Dynamic Source Dependence**: Few studies [22], [23] have explored the dynamics of source dependence, which is crucial for time-series data. Previous stuides oversimplify source modeling, neglecting the intricate and dynamic dependencies among sources. The evolving nature of dependencies over time and their impact on partial independence further complicates the problem.

Our main contributions in addressing the challenges are:

- We introduce a detailed measurement of source reliability encompassing accuracy, coverage, and dependence. We develop three types of Dependent Hidden Markov Models (DepenHMMs) to analyze dynamic source dependencies, categorized as simple ($Depen\_Simple$), complex ($Depen\_Complex$), and dynamic ($Depen\_Dynamic$).
- We investigate the simple dependence, partial dependence, and dynamic dependence among sources by Hidden Markov Models where the initial probabilities, transition probabilities, and the observation probabilities are reconstructed in terms of three different scenarios.
- Based on the constructed DepenHMM framework, we analyze the probability for each state and predict the latent truth. The experiments on the three real-word datasets show that the proposed approach can significantly improve accuracy compared with existing methods in truth discovery.

The remainder of this paper is organized as follows. The related research is reviewed and discussed in Section II. The mathematical model is defined for the dynamic truth discovery problem in Section III. In Section IV, we demonstrate three types of source dependence, assess the source reliability among sources, and compute the latent truth. The experimental results with real-world datasets are analyzed and reported in Section V. Finally, the conclusion is summarized in Section VI with several highlighting remarks.

## II. Related Work

Truth discovery is essential due to the rise in mixed-quality sources, especially in social sensing [7] and online data sharing [24], [25]. It's crucial to measure the reliability of sources for entity profiling [26], conflict resolution [27], and event forecasting [28]. Source dependence is a hot topic in truth discovery [1], [15], [29], [30], [31], [32], [33], [34], and is first analyzed in [15], noting its potential to provide additional information. In [30], the integration of large-scale data is processed with source dependence. The simple dependence can be labeled in semi-supervised approaches [33], [35]. Except simple copy relationships among sources, there are complex replication relationships among sources [31]. The simple copy typically refers to the act of directly copying a source's claims without proper citation or attribution, and the complex replication refers to a more subtle form of plagiarism, where copiers incorporate some original claims alongside the plagiarized data, but fails to clearly identify whether they are copied from others.

To identify the high complexity of dependence computation, researches focus on the selection of reliable sources [36] according to which Expectation-Maximization (EM) algorithm is widely used to model the dependence [7]. The replication relationship is captured based on the joint recall rate and false-true rate [1]. Some researchers [37] consider indirect independence. For example, the dependence is reflected in the latent evaluation consistency (the latent rating evaluation of different objects is consistent) shared by multiple sources in [37]. Although these studies [1], [15], [29], [30], [31], [32] have focused on source dependence, the methods for evaluating plagiarism ignore the dynamic property of dependence for sources. Compared to the above studies, we provide novel contributions by considering the dynamic property of sources.

HMMs have been widely used in various areas, including speech recognition, bioinformatics, and finance. In recent years, HMMs have also facilitated truth discovery by inferring truths from multi-source claims, even with lossy sampling [38]. They provide a structured way to model sequences with domain-informed probabilities. A dynamic truth inference method using object and temporal correlations is proposed in [39], and a physical constraint-aware HMM for variable truth inference is developed by Daniel et al. [40]. However, these approaches do not comprehensively integrate inter-source dependencies and other source quality, such as coverage, resulting in incomplete evaluations of source reliability. We present a novel study on source dependencies, integrating time sensitivity and partial independence, and combine source coverage by updating frequency over time, extending beyond prior work on simple copy relationships [23].

There have also been significant progress on integrating and conflict fusion for time-sensitive data [20], [23], [39], [41], [42], [43], [44] while few studies focus on the dynamic source dependence. For example, the spreading misinformation problem is studied on dynamic data in [44] and Gibbs Sampling-based algorithm has been proposed to discover the correct values of objects by the ground-truth values from historical data in [22]. Moreover, a robust framework is proposed to detect the dynamic periodicity for time series tasks in [41]. A novel online truth discovery framework [20] is studied with dynamical multi-source information and temporal patterns. There have been a significant amount of researches focused on evolution interpretation [45] and clustering of time series [42], [43] in dynamic data integration. However, these studies have overlooked the dependence among sources in dynamic time, which significantly impacts the reliability of the sources.

This paper investigates dynamic truth discovery, aiming to uncover the truth values of objects from evolving sources. The changing nature of object truths affects source reliability, influencing the accuracy and efficiency of truth discovery. We focus on measuring source dependence and recognizing dynamic source reliability to reveal the latent truths of objects for truth discovery problems.

## III. Problem Definition

In this paper, we focus on the truth discovery problem with dynamic and dependent sources. Formally, an object set is defined as $O = \{o_i | i = 1, 2, \ldots, Q\}$, where $Q$ is the number of objects. A source set which provides claims for the object

TABLE I
IMPORTANT NOTATION DEFINITION

| Symbol | Definition |
|---|---|
| $o_i$ | The $i^{th}$ object |
| $s_j$ | The $j^{th}$ source |
| $t^n$ | The $n^{th}$ time point |
| $Q$ | The amount of objects |
| $M$ | The amount of sources |
| $N$ | The amount of time points |
| $c_{i,j}^n$ | The claim provided by $s_j$ for object $o_i$ at $t^n$ time point |
| $C(s_j)$ | The coverage of source $s_j$ |
| $A(s_j)$ | The accuracy of source $s_j$ |
| $D(s_j)$ | The dependence of source $s_j$ |
| $I$ | The state that all sources are independent |
| $Cj_c$ | The state that source $s_j$ is dependent |
| $Cj_{\neg c}$ | The state that source $s_j$ is partial dependent |
| $Cj_{\neg c}^n$ | The state that source $s_j$ is partial dependent at $t^n$ time point |
| $\overline{UI_{s_j}}$ | The claim set that are only reported by the source $s_j$ |
| $\overline{UD_{s_j}}$ | The claim set reported by $s_j$ that has been reported by others |
| $c_j$ | The probability of source $s_j$ being a dependent source |
| $ck_j$ | The probability of the dependent $s_j$ keeping its dependence |
| $cc_j$ | The probability of copying for the dependent source $s_j$ |

is defined as $S = \{s_j | j = 1, 2, \ldots, M\}$. We assume that each object is associated with only one value at a time point while at different time points, the value can be changed. Let $T = \{t^n | n = 1, 2, \ldots, N\}$ be the time series set, where at each time point, the sources provide series claims for objects. Therefore, the claim set is defined as $C = \{c_{ij}^n | i = 1, 2, \ldots, Q; j = 1, 2, \ldots, M; n = 1, 2, \ldots, N\}$, where $c_{ij}^n$ denotes the claim provided by the $j^{th}$ source for the $i^{th}$ object at the $n^{th}$ time point. The objective for this problem is to discover the latent truth values (the most reliable value in the claim set) dynamically from various sources. Trustworthy sources reveal the truth [10], according to *accuracy*, *coverage*, and *dependence*. *Accuracy* is the consistency of correct information. *Coverage* is the comprehensiveness of details and *dependence* is the degree of independence from other sources. Reliable sources demonstrate high accuracy and extensive coverage, with minimal dependence. Table I presents important notations in this paper.

## IV. DEPENDENT HMM METHODS

In practical situations, it's advisable to trust high-reliability sources like. edu and. gov domains over less credible ones. Assessing source reliability, affected by *accuracy*, *coverage*, and *dependence*, is crucial. While accuracy and coverage can be statistically measured (details in Section IV-A), gauging source dependence is more complex, especially when it varies over time. To tackle this, three models are introduced: $Depen\_Simple$, $Depen\_Complex$, and $Depen\_Dynamic$. These models assess source dependence through hidden states to dynamically infer the most reliable truths. The DepenHMM framework, outlined in Algorithm 1, is designed to reveal latent truths using selected models. The chosen model dictates the hidden states, which are used to calculate transition and observation probabilities. For each model, the source dependence is computed by the hidden states extracted from the Viterbi algorithm [46]. Source reliability is assessed by (4), which helps in determining the final truth values of objects by (5).

To illustrate the DepenHMM algorithm, we assume that three sources provide the temperature for the same city in 4 time points. The first time point is part of the history data, and the objective is to compute the truth of time point 1,2 and 3. The claims of each time points are listed in Table II.

TABLE II
A WEATHER TRUTH DISCOVERY EXAMPLE

| Source | Time Point 0 | Time Point 1 | Time Point 2 | Time Point 3 |
|---|---|---|---|---|
| $s_1$ | 32°C | 31°C | 34°C | 37°C |
| $s_2$ | 33°C | 32°C | 35°C | 37°C |
| $s_3$ | 32°C | 32°C | 36°C | 38°C |

**Algorithm 1:** The DepenHMM Framework.

**Input:** $S$, $O$, $C$, $T$.
**Output:** Truths of $O$.
**for** *o in O* **do**
  **for** *s in S* **do**
    Compute source coverage and accuracy;
  Choose one MODEL from $Depen\_Simple$, $Depen\_Complex$, $Depen\_Dynamic$;
  Determine hidden states by the selected MODEL;
  Compute initial probabilities by the MODEL;
  Compute transition probabilities by the MODEL;
  Compute observation probabilities by the MODEL;
  Compute source dependence $D(s_j)$ by Equation (1) or (2) or (3) by the selected MODEL;
  **for** *s in S* **do**
    Compute source reliability by Equation (4);
  Discover object truth by Equation (5).

### A. Source Accuracy and Coverage Evaluation

Generally, high-reliability sources offer high accuracy and comprehensive coverage, quickly updating and accurately reflecting changes. To explain these metrics in detail, suppose there are $Q$ objects at the initialization stage, and source $s_j$ totally provides $m_j$ claims for $q_j$ objects. The number of correct claims made by source $s_j$ is denoted as $n_j$. We define the coverage and accuracy metrics as the following:

- *Coverage* ($C(s_j)$): The coverage of source $s_j$ is computed by the ratio of the number of objects $q_j$ covered by that source to the total number of objects ($Q$). Mathematically, it can be represented as: $C(s_j) = \frac{q_j}{Q}$.
- *Accuracy* ($A(s_j)$): The accuracy of source $s_j$ is defined as the ratio of the number of correct claims ($n_j$) to the total number of claims made by that source ($m_j$). Mathematically, it can be represented as: $A(s_j) = \frac{n_j}{m_j}$.

In the example, since we only have little history data, we set the same coverage and randomly accuracy of each source. Specifically, we set $C(s_1) = C(s_2) = C(s_3) = 0.5$ and $A(s_1) = A(s_2) = 0.8, A(s_3) = 0.85$.

### B. Source Dependence Evaluation

Source dependence is unique from accuracy and coverage, as it's difficult to detect when sources copy information. We categorize source dependence into three types: Dependent Sources that solely rely on others for claims. Partial Dependent Sources that occasionally make independent claims despite their dependence. Dynamic Dependent Sources, whose reliance on others can

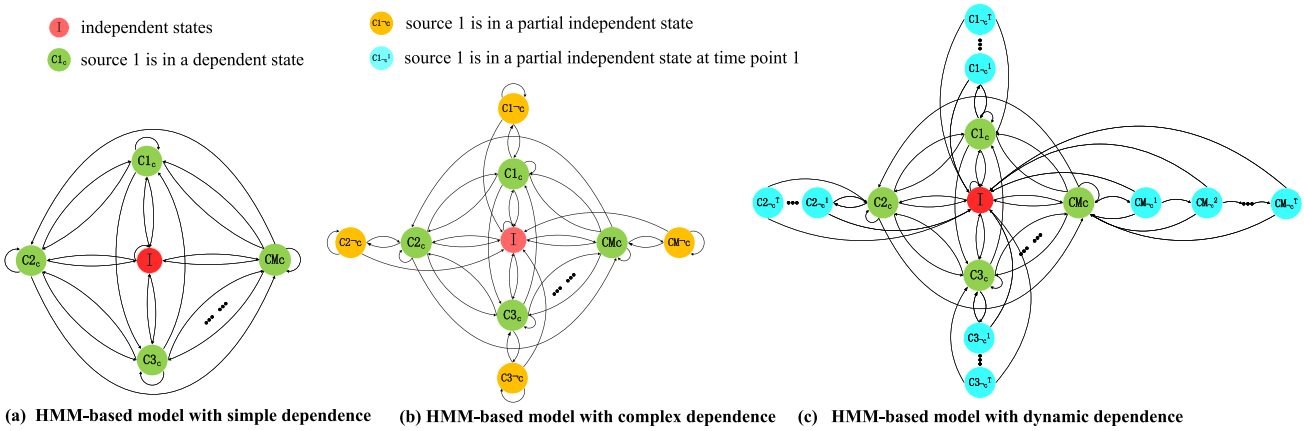


Fig. 1. State transition process: (a) An independent state can be converted to a dependent state (or remains independent) and a dependent state can be converted to another dependent state (or remains dependent), (b) A dependent state can be converted to a partial dependent state and a partial dependent state can be converted to a dependent state (or remains partial dependent), and (c) A partial dependent state can be converted to another partial dependent state at different time points and a partial dependent state can be converted to independent states after a series of time periods.

evolve over time, potentially shifting from copying to independent claim-making. We introduce three distinct models designed to address the dependence associated with these three categories of sources, as illustrated in Fig. 1.

*1) Simple Source Dependence:* To infer the dependence among sources, a simple hidden Markov model depicted in Fig. 1(a), termed $Depen\_Simple$, has been formulated. Within this model, each of the $M$ sources is represented by two states: *independent* and *dependent*. In the independent state, sources provide claims of objects in an independent manner. Conversely, in the dependent state, sources copy claims that have been reported by other sources. By these two states, the $Depen\_Simple$ model enables the inference of source dependence.

*Hidden states:* The hidden states are determined by the number of sources. Assume that there is only one dependent source at most which copies information from other sources at the current time. Therefore, there are $M+1$ hidden states: $I, C1_c, C2_c, C3_c, \ldots, CM_c$. State $I$ indicates that all sources are independent. State $Cj_c (j = 1, 2, 3, \ldots, M)$ implies that $s_j$ is dependent while other sources are independent. These hidden states can be converted to each other. According to Fig. 1(a), the state $I$ can be converted to $I, C1_c, C2_c, C3_c, \ldots, CM_c$. At the same time, the dependent source can also be converted to each other, i.e., $C1_c$ can be converted to $C1_c, C2_c, C3_c, \ldots, CM_c, I$. In the example, we have 4 hidden states: $I, C1_c, C2_c, C3_c$.

*Initial probability:* To analyze the initial probabilities in the $Depen\_Simple$ model, all sources may be in a state of complete independent or dependent state at the beginning. We assume the probability that all sources are independent is $\alpha$, which can be inferred by the Baum-Welch algorithm [47], and the probability that the remaining sources are dependent sources is equal. The initial probability of all states can be obtained by: $P(I) = \alpha$ and $P(C1_c) = P(\cdots) = P(CM_c) = \frac{1-\alpha}{M}$. In the example, we set $\alpha = 0.1$. Therefore, $P(I) = 0.1$ and $P(C1_c) = P(C2_c) = P(C3_c) = 0.3$.

*Transition probability:* To analyze the transition probability, $c_j$ is defined as the probability of $s_j$ being a dependent source, where $0 \leq c_j \leq 1$. All transition probabilities are described by above parameters. A matrix $T_{(M+1)(M+1)}$ is defined to store the transition. For the convenience of description, we use $T_{ij}$ to represent the transition from state $i$ to state $j$. The state transition probability is described as follows:

- $T_{II}$: For independent sources, the transition probability is computed by $P(T_{II}) = \prod_{j=1}^{M}(1 - c_j)$ since all the sources keep independent.
- $T_{ICj_c}$: The transition probability is computed by $P(T_{ICj_c}) = (1 - \prod_{k=1}^{M}(1 - c_k))\frac{c_j}{\sum_{k=1}^{M} c_k}$, when an independent source $s_j$ converts to a dependent state while other sources remain independent, with the constraint that the probabilities of all elements in the same row must sum to 1.
- $T_{Cj_cI}$: The transition probability is computed by $P(T_{Cj_cI}) = \prod_{k=1}^{M}(1 - c_k)$ when dependent source $s_j$ becomes independent, while all other sources stay independent.
- $T_{Cj_cCk_c}$: The transition probability is computed by $P(T_{Ck_cCj_c}) = c_k \prod_{m=1 \wedge m \neq k}^{M}(1 - c_m)$ which implies dependent source $s_k$ becomes independent, independent source $s_j$ becomes dependent and all other sources stay independent.
- $T_{Cj_cCj_c}$: The transition probability from dependent state $s_j$ to $s_j$ (itself) is calculated by $P(T_{Cj_cCj_c}) = 1 - P(T_{Cj_cI}) - P(T_{Cj_cC1_c}) - P(T_{Cj_cC2_c}) \cdots - P(T_{Cj_cCk_c})$ $(k \neq j) \cdots - P(T_{Cj_cCM_c}) = 1 - \prod_{k=1}^{M}(1 - c_k) - \sum_{m=1 \wedge m \neq j}^{M} c_m \prod_{n=1 \wedge n \neq m}^{M}(1 - c_n)$.

According to above calculation procedures, the state transition matrix of the $Depen\_Simple$ model is obtained. In the example, we set $c_1 = 0.4, c_2 = 0.9, c_3 = 0.3$ and the state transition matrix is as followed.

$$\begin{pmatrix} 0.042 & 0.2395 & 0.538875 & 0.179625 \\ 0.042 & 0.562 & 0.378 & 0.018 \\ 0.042 & 0.028 & 0.912 & 0.018 \\ 0.042 & 0.028 & 0.378 & 0.552 \end{pmatrix}$$

*Observation probability:* At each time, different sources have different updating claims. It is obviously inefficient to analyze all the updated claims separately. In order to improve efficiency,

the claims are divided into two categories for each source: only reported by the source ($s_j$), and previously reported by other sources, represented by $\overline{UI_{s_j}}$ and $\overline{UD_{s_j}}$ respectively. Usually, if $s_j$ has more $\overline{UI_{s_j}}$, it is more likely to indicate that the source is independent. But if $s_j$ has more $\overline{UD_{s_j}}$, the source may be dependent, or the source updates slowly. The reliability of a source is closely tied to its accuracy and coverage, which influence its dependence classification: (1) Sources that frequently echo others' false claims are considered dependent. (2) Sources that typically exhibit low coverage yet occasionally achieve high coverage of repeated claims are classified as dependent sources. (3) Sources that are generally inaccurate but occasionally demonstrate accuracy on repeated claims are identified as dependent sources.

In the $Depen\_Simple$ model, observation probabilities for different hidden states depend on error rates, coverage, and accuracy. For an object $O$ with $m$ errors, the model assumes equal probabilities for a source $s_j$ making a unique correct claim ($\overline{UI_{s_j}}$) or an error that others have also made ($\overline{UD_{s_j}}$) in the independent state $I$. Correct claims reflect timely and accurate updates of the object. In this case, the observation probability for independent sources in the $I$ state is described as follows:

$P(\overline{UI_{s_j}}, s_j|I, \overline{UI_{s_j}} = true) = P(\overline{UD_{s_j}}, s_j|I, \overline{UD_{s_j}} = true) = A(s_j)C(s_j)$.

Otherwise, if the claim is false, the source does not capture the update accurately:

$P(\overline{UI_{s_j}}, s_j|I, \overline{UI_{s_j}} = false) = P(\overline{UD_{s_j}}, s_j \mid I, \overline{UD_{s_j}} = false) = \frac{(1-A(s_j))}{m}$.

Combing these two conditions, when all sources are independent, the observation probability in $I$ state is described as follows:

$P(\overline{UI_{s_j}}, s_j|I) = P(\overline{UD_{s_j}}, s_j|I) = \frac{(1-A(s_j))}{m} + A(s_j)C(s_j)$.

For the $Depen\_Simple$ model, when source $s_j$ is dependent, it is impossible to give a claim which other sources has never been reported. The observation probability in $Cj_c$ state is described as: $P(\overline{UI_{s_j}}, s_j|Cj_c) = 0$, $P(\overline{UD_{s_j}}, s_j|Cj_c) = 1$.

When $s_j$ is dependent, according to one dependent source assumption of the $Depen\_Simple$ model, other sources are independent. The dependence of $s_j$ has no effect on other sources. Therefore, the observation probability in $s_k$ state is described as follows:

$P(\overline{UI_{s_k}}, s_k|Cj_c) = P(\overline{UD_{s_k}}, s_k|Cj_c) = \frac{(1-A(s_k))}{m} + A(s_k)C(s_k)$

Since all the observation probability of state $Cj_c$ should be added to 1, the scaling of Equations is required. In the example, we first estimate the claims category in time point 1-3. We can find that in time point 1, the claims of $s_2$ (32°C) and $s_3$ (32°C) have already been reported in time point 0 previously. Therefore, in time point 1, the observation list is ($\overline{UI_{s_1}}, \overline{UD_{s_2}}, \overline{UD_{s_3}}$), and in time point 2 and 3, the observation list is ($\overline{UI_{s_1}}, \overline{UI_{s_2}}, \overline{UI_{s_3}}$). We set $m = 2$. Therefore, $P(\overline{UI_{s_1}}, s_1|I) = P(\overline{UD_{s_1}}, s_1|I) = 0.5$, $P(\overline{UI_{s_1}}, s_1|C1_c) = 0$, $P(\overline{UD_{s_1}}, s_1|C1_c) = 1$, $P(\overline{UI_{s_1}}, s_1|C2_c) = 0.5$. The observation probability of $s_2$ and $s_3$ can be calculated in the same way.

*Source Dependence Calculation:* After setting the parameters of the Simple source dependence model, the hidden state of each time point can be extracted by the Viterbi algorithm. The final source dependence is calculated by (1). The more times a source is in a dependent state, the higher its dependence.

$$D(s_j) = \frac{|Cj_c|}{|Cj_c| + |I|} \tag{1}$$

In the example, there are three claims observed in each time point. We need to calculate the whole three observation probability. In time point 1, the whole observation probability of $I$ state is $P(\overline{UI_{s_1}}, \overline{UD_{s_2}}, \overline{UD_{s_3}}, s_1, s_2, s_3|I) = (0.5 + 0.5 + 0.5)/3 = 0.5$, state $C1_c$ is 0.33, state $C2_c$ is 0.67 and state $C3_c$ is 0.67. For time point 2 and 3, the whole observation probabilities of state $I, C1_c, C2_c, C3_c$ are 0.5,0.33,0.33 and 0.33 respectively. After been calculated by Viterbi algorithm, the hidden state state is $C2_c, C2_c, C2_c$ for time point 1-3. Finally, by Equation(1), $D(s_1) = D(s_3) = 0, D(s_2) = 1$.

*2) Complex Source Dependence:* For partial dependent sources, even if they are plagiarism sources, it is still possible for them to provide claims independently. $Depen\_Simple$ cannot capture the partial dependence of sources. A $Depen\_Complex$ model is constructed to capture it in Fig. 1(b).

*Hidden state and initial probability:* When a dependent source has the probability to report an independent claim, the state of source dependence is required to be split. A new type of hidden states is added: $Cj_{\neg c}$, which implies that source $s_j$ is dependent with other sources but report claims independently. According to Fig. 1(b), the convert among state $I$ and state $Cj_c$ is the same as simple dependence in the $Depen\_Simple$ model. It is obvious that the premise of source $s_j$ entering hidden state $Cj_c$ satisfies that source $s_j$ is dependent instead of independent. State $I$ is unable to convert to $Cj_{\neg c}$ directly while $Cj_c$ can convert to $Cj_{\neg c}$. When at $Cj_{\neg c}$ state, source $s_j$ can maintain its partly independence ($Cj_{\neg c}$ convert to $Cj_{\neg c}$), or copying ($Cj_{\neg c}$ converts to $Cj_c$), or abandon its dependence ($Cj_{\neg c}$ converts to $I$). Although the hidden state changes, the initial probability of hidden state is still computed as same as the $Depen\_Simple$ model. For the example in $Depen\_Complex$ model, there are 7 hidden states : $I, C1_c, C2_c, C3_c, C1_{\neg c}, C2_{\neg c}, C3_{\neg c}$.

*Transition probability:* To measure the partial independence, $ck_j$ is defined as the probability of a dependent source $s_j$, which keeps its dependence while $cc_j$ is represented as the copy probability for a dependent source $s_j$. The transition probability matrix is refined in terms of these parameters as follows.

- $T_{II}, T_{ICj_c}, T_{Cj_cCk_c}$: These transition probabilities are the same as the $Depen\_Simple$ model.
- $T_{ICj_{\neg c}}, T_{Cj_cCk_{\neg c}}, T_{Cj_{\neg c}Ck_c}$: These converts are non-existent, which implies that $P(T_{ICj_{\neg c}}) = P(T_{Cj_cCk_{\neg c}}) = P(T_{Cj_{\neg c}Ck_c}) = 0$.
- $T_{Cj_cCj_c}$: Dependent source $s_j$ maintains its dependence where $P(T_{Cj_cCj_c}) = c_j ck_j$.
- $T_{Cj_cCj_{\neg c}}$: Dependent source $s_j$ has partial independence where $P(T_{Cj_cCj_{\neg c}}) = c_j ck_j(1 - cc_j)$.
- $T_{Cj_cI}$: Dependent source $s_j$ converts to be independent, and other sources are independent at the same

time. $P(T_{Cj_cI}) = 1 - P(T_{Cj_cCj_{\neg c}}) - P(T_{Cj_cC1_c}) - P(T_{Cj_cC2_c}) \cdots - P(T_{Cj_cCM_c}) = 1 - c_j ck_j(2 - cc_j) - \sum_{m=1 \wedge m \neq j}^{M} c_m \prod_{n=1 \wedge n \neq m}^{M}(1 - c_n)$.

- $T_{Cj_{\neg c}Cj_c}$: The dependent source $s_j$ abandons its partial independence but still maintains its dependence. $P(T_{Cj_{\neg c}Cj_c}) = c_j ck_j cc_j$.
- $T_{Cj_{\neg c}Cj_{\neg c}}$: The dependent source $s_j$ maintains its partial independence. $P(T_{Cj_{\neg c}Cj_{\neg c}}) = c_j ck_j(1 - cc_j)$.
- $T_{Cj_{\neg c}I}$: The dependent source $s_j$ with partial independence abandons its dependence completely. $P(T_{Cj_{\neg c}I}) = 1 - P(T_{Cj_{\neg c}Cj_{\neg c}}) - P(T_{Cj_{\neg c}Cj_c}) = 1 - c_j ck_j$.

For the example in $Depen\_Complex$ model, we set $ck_1 = 0.5, ck_2 = 0.6, ck_3 = 0.6, cc_1 = 0.2, cc_2 = 0.05, cc_3 = 0.8$. Therefore, $T_{C1_cC1_c} = 0.2$, $T_{C1_cC1_{\neg c}} = 0.16$, $T_{C1_cI} = 1 - 0.2 - 0.378 - 0.018 - 0.16 = 0.244$, $T_{C1_{\neg c}C1_c} = 0.04$, $T_{C1_{\neg c}C1_{\neg c}} = 0.16$, $T_{C1_{\neg c}I} = 1 - 0.04 - 0.16 = 0.8$ and the transition probability of other source can also be computed in the same way.

*Observation probability:* In complex source dependence, a dependent source's accuracy is influenced by its copied sources, and its coverage is affected by both inherent characteristics and plagiarism. In the independence state $I$, observation probabilities mirror the $Depen\_Simple$ model. However, in the dependence state $Cj_c$, these probabilities differ, reflecting a slim chance for the dependent source $s_j$ to report independently. The presence of identical claim values increases the observation probability of $Cj_c$ (Proof in the Theorem 1), while incorrect values from independent sources, when copied, reduce accuracy. This necessitates a reassessment of the source's accuracy and coverage. Assuming that $s_j$ is a copier of $s_k$, the coverage and accuracy of the source $s_j$ are updated by: $C(s_j|s_j \, copy \, s_k) = C(s_j) - c_j C(s_k)$ and $A(s_j|s_j \, copy \, s_k) = A(s_j) - c_j(1 - A(s_k))$.

Similar to the observation probabilities of independent states, the observation probabilities of dependent states are calculated as follows. The partial independence is measured by $P_c(s_j)$. If we observe the claims which are only reported by source $s_j$, i.e., $\overline{UI_{s_j}}$, there is no doubt that the partial independence plays a role. The observation probability of $\overline{UI_{s_j}}$ in the source dependence state is computed by the dependent source probability $P_c(s_j)$: $P(\overline{UI_{s_j}}, s_j|Cj_c) = P_c(s_j)$ and $P_c(s_j) = \frac{1}{M-1}\sum_{k=1 \wedge k \neq j}^{M}((\frac{1 - A(s_j|s_j \, copy \, s_k)}{m}) + A(s_j|s_j \, copy \, s_k)C(s_j|s_j \, copy \, s_k))$.

When the claims of dependent source $s_j$ are observed and these claims have already been reported by others, i.e., $\overline{UD_{s_j}}$, there are two possible cases. On the one hand, the dependent source directly copies claims from others. On the other hand, the dependent source reports claims by itself, but the claims happen to be the same as others. By combining these two conditions, we compute the observation probability of $\overline{UD_{s_j}}$ for dependent states as $P(\overline{UD_{s_j}}, s_j|Cj_c) = cc_j + (1 - cc_j)P_c(s_j)$. In state $Cj_{\neg}c$, the observation probability is the same as $I$, since the dependent source in this state is able to report its own claims. For the example in $Depen\_Complex$ model, in time point 1, the observation probability $P(\overline{UI_{s_1}}, s_1|C1_c) = 0.35, P(\overline{UD_{s_2}}, s_2|C2_c) = 0.25, P(\overline{UD_{s_3}}, s_3|C3_c) = 0.8$, and in time point 2 and 3, $P(\overline{UI_{s_2}}, s_2|C2_c) = 0.21, P(\overline{UI_{s_3}}, s_3|C3_c) = 0.38$.

*Theorem 1:* If $cc_j > \frac{2}{m}$ and $cc_j > 2C(s_j)$, the addition of a claim to $\overline{UD_{s_j}}$ leads to an increase in the observation probability of state $Cj_c$.

*Proof:* if $\overline{UD_{s_j}} = true$, the observation probability in the $I$ state is the same as $Depen\_Simple$ model. The partial observation probability of $\overline{UD_{s_j}} = true$ in the $Cj_c$ state is computed in $P(\overline{UD_{s_j}}, s_j|Cj_c)$. Since $C(s_j) < \frac{cc_j}{2}$ and $A(s_j) \leq 1$:

$$P(\overline{UD_{s_j}}|I, \overline{UD_{s_j}} = true) = A(s_j)C(s_j) \leq C(s_j) < \frac{cc_j}{2}$$

$$< \frac{cc_j}{2} + \frac{(1 - cc_j)}{M - 1}\sum_{k=1 \wedge k \neq j}^{M} A(s_j|s_j \, copy \, s_k)C(s_j|s_j \, copy \, s_k)$$

$P(\overline{UD_{s_j}}|I, \overline{UD_{s_j}} = true) < P(\overline{UD_{s_j}}|Cj_c, \overline{UD_{s_j}} = true)$, and the probability of state $Cj_c$ increases when adding a correct claim in $\overline{UD_{s_j}}$.

If $\overline{UD_{s_j}} = false$, the observation probability of $\overline{UD_{s_j}} = false$ in $I$ state is calculated as same as $Depen\_Simple$ model. The observation probability $\overline{UD_{s_j}} = false$ in $Cj_c$ state is calculated by the part of the observation probability $P(\overline{UD_{s_j}}, s_j|Cj_c)$:

$$P(\overline{UD_{s_j}}|Cj_c, \overline{UD_{s_j}} = false)$$

$$= \frac{cc_j}{2} + \frac{(1 - cc_j)}{M - 1}\sum_{k=1 \wedge k \neq j}^{M} \frac{1 - A(s_j|s_j \, copy \, s_k)}{m}$$

Since $m > \frac{2}{cc_j}$ and $A(s_j) \leq 1$:

$$P(\overline{UD_{s_j}}|I, \overline{UD_{s_j}} = false) = \frac{1 - A(s_j)}{m} \leq \frac{1}{m} < \frac{cc_j}{2}$$

$$< \frac{cc_j}{2} + \frac{(1 - cc_j)}{M - 1}\sum_{k=1 \wedge k \neq j}^{M} \frac{1 - A(s_j|s_j \, copy \, s_k)}{m}$$

$P(\overline{UD_{s_j}}|I, \overline{UD_{s_j}} = flase) < P(\overline{UD_{s_j}}|Cj_c, \overline{UD_{s_j}} = false)$. The probability of state $Cj_c$ increases when adding a wrong claim in $\overline{UD_{s_j}}$.

Therefore, we can conclude $P(\overline{UD_{s_j}}|I, \overline{UD_{s_j}}) < P(\overline{UD_{s_j}}|Cj_c, \overline{UD_{s_j}})$. The addition of a claim to $\overline{UD_{s_j}}$ leads to an increase in the observation probability of the state $Cj_c$. □

*Source Dependence Calculation:* For complex source dependence model, the hidden state sequence can also be extracted by the Viterbi algorithm. The final source dependence for complex dependence model is calculated by (2). For the example in $Depen\_Complex$ model, we first calculate the whole observation probability. In time point 2, $P(\overline{UI_{s_1}}, \overline{UI_{s_2}}, \overline{UI_{s_3}}, s_1, s_2, s_3|I) = 0.5$, state $C1_c$ is 0.45, state $C2_c$ is 0.40, state $C3_c$ is 0.46 and state $C2_{\neg c}$ is 0.5. After computed by Viterbi algorithm, the whole the extracted hidden state sequence is $C2_C, C2_{\neg c}, C2_{\neg c}$. Therefore, the final source dependence for three sources are $D(s_1) = 0, D(s_2) =$

$0.67, D(s_3) = 0$.

$$D(s_j) = \frac{|Cj_c| + 0.5 \times |Cj_{\neg c}|}{|I| + |Cj_{\neg c}| + |Cj_c|} \quad (2)$$

*3) Dynamic Source Dependence:* For the $Depen\_Complex$ model, sources provide partial claims independently in the dependent state while it is ignored that plagiarism sources are slowly shifting from a dependent state to an independent state. To capture the dynamic characteristic, the $Depen\_Dynamic$ model is constructed in Fig. 1(c), which describes the state transition process.

*Hidden state and initial probability:* For the $Depen\_Dynamic$ model, we detail the temporal expansion of the partial independent state $Cj_{\neg}c$. When a dependent source $s_j$ first enters partial independence, it transitions to $Cj^1_{\neg}c$. Maintaining this state extends the chain with $Cj^2_{\neg}c$, eventually leading to the independent state $I$. The transitions among states are illustrated using $s_j$. Like in $Depen\_Complex$, state $I$ transitions to $Cj_c$ or itself. State $Cj_c$ can shift to another dependent state $Ck_c$, back to $I$, to its partial independent state $Cj^1_{\neg}c$, or remain the same. The transition of partial independence varies by time. For instance, from the $t^{th}$ partial independent state $Cj^t_{\neg}c$, if the source copies next time, it reverts to $Cj_c$. If it retains partial independence, it moves to $Cj^{t+1}_{\neg}c$, or transitions to $I$ otherwise. The initial probabilities follow the $Depen\_Simple$ model. For the example in dynamic source dependence model, the hidden states are $I, C1_c, C2_c, C3_c, C1^1_{\neg c}, C2^1_{\neg c}, C3^1_{\neg c}, C1^2_{\neg c}, C2^2_{\neg c}, C3^2_{\neg c}$.

*Transition and observation probability:* In real-world situations, a partially independent dependent source is less likely to copy claims and more likely to sustain this independence until full independence is achieved. A time decay function is used to measure the copying ability of such sources: $f(j,t) = cc_j e^{-(t-1)}$ Here, $t$ indicates the time slot in which source $j_i$ maintains partial independence.

The transition probability matrix for $Depen\_Dynamic$ model is calculated as follows:

- $T_{II}, T_{ICj_c}, T_{Cj_cCj_c}, T_{Cj_cCk_c}$: These transition probabilities are the same as the $Depen\_Simple$ model.
- $T_{ICj^t_{\neg c}}, T_{Cj_cCk^t_{\neg c}}, T_{Cj^t_{\neg c}Ck_c}$: These converts are non-existent with 0 probability.
- $T_{Cj_cCj^1_{\neg c}}$: The dependent source begins to report claims by itself. $P(T_{Cj_cCj^1_{\neg c}}) = c_j c k_j (1 - f(j,1))$.
- $T_{Cj_cI}$: Dependent source $s_j$ is gradually becoming independent, and all the other sources are independent. $P(T_{Cj_cI}) = 1 - P(T_{Cj_cCj^1_{\neg}c}) - P(T_{Cj_cC1_c}) - P(T_{Cj_cC2_c}) \cdots - P(T_{Cj_cCM_c}) = 1 - c_j c k_j (2 - f(j,1)) - \sum_{m=1 \wedge m \neq j}^{M} c_m \prod_{n=1 \wedge n \neq m}^{M} (1 - c_n)$.
- $T_{Cj^t_{\neg c}Cj_c}$: At the $t$ time point, the dependent source $s_j$ abandons its partial independence but still maintains its dependence. $P(T_{Cj^t_{\neg c}Cj_c}) = c_j c k_j f(j,t)$.
- $T_{Cj^t_{\neg c}Cj^{t+1}_{\neg c}}$: The dependent source $s_j$ maintains its partial independence at next time point. $P(T_{Cj^t_{\neg c}Cj^{t+1}_{\neg c}}) = c_j c k_j (1 - f(j,t))$.
- $T_{Cj^t_{\neg c}I}$: The dependent source $s_j$ with partial independence abandons its dependence completely. $P(T_{Cj^t_{\neg c}I}) = 1 - c_j c k_j$.

TABLE III
DATASET STATISTICS

| | Stock | Flight | Weather |
|---|---|---|---|
| Sources | 55 | 38 | 18 |
| Record Days | 20 | 31 | 7 |
| Objects | 1000 | 1200 | 1263 |
| Claims | 55*1000*20 | 38*1200*31 | 33970 |

The observation probability for dynamic sources is the same as that in the $Depen\_Complex$ model, since no matter what time point is, when dependent source $s_j$ at partial independent state, it still has the same observation probability as the independent state for $\overline{UI_{s_j}}$ and $\overline{UD_{s_j}}$. For the example in dynamic source dependence model, the transition probability $T_{C1^1_{\neg c}C1_c} = 0.04, T_{C1^1_{\neg c}C1^2_c} = 0.16$. Note that the normalization is required. Other parameters can also be calculated by this setting.

*Source Dependence Calculation:* In the dynamic source dependence model, we also use Viterbi algorithm to compute the hidden state sequence. The final source dependence is calculated by (3). For the example in dynamic source dependence model, we first calculated the whole observation probability. In time point 3, the observation sequence is $\overline{UI_{s_1}}, \overline{UI_{s_2}}, \overline{UI_{s_3}}$, the whole observation probability are 0.5, 0.45, 0.40, 0.46, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5 for state $I, C1_c, C2_c, C3_c, C1^1_{\neg c}, C2^1_{\neg c}, C3^1_{\neg c}, C1^2_{\neg c}, C2^2_{\neg c}, C3^2_{\neg c}$ respectively. The extracted hidden state sequence is $C2_c, C2^1_{\neg c}, C2^3_{\neg c}$. Therefore, the final source dependence in dynamic source dependence model are $D(s_1) = D(s_3) = 0, D(s_2) = 0.5$.

$$D(s_j) = \frac{|Cj_c| + \sum_{t=1}^{T} \frac{|Cj^t_{\neg c}|}{t+1}}{|I| + |Cj_c| + \sum_{t=1}^{T} |Cj^t_{\neg c}|}. \quad (3)$$

### C. Source Reliability Evaluation

After calculating source dependence, we evaluate the source reliability by Equation(4) with accuracy, coverage and dependency. The source reliability is defined by:

$$R(s_j) = a_1 A(s_j) + a_2 C(s_j) + (1 - a_1 - a_2)(1 - D(s_j)) \quad (4)$$

We use $a_1$ and $a_2$ to weight source accuracy, coverage, and dependence tested in experiment. Algorithm 1 is executed iteratively and the final truth is the claim that reported by the most reliable source.

$$Truth = claim_{max_{j \in M} R(s_j)} \quad (5)$$

In the example, source 3 has the highest reliability for all the three models. Therefore, in each model, the truths for time point 1-3 are 32 °C, 36 °C, and 38 °C respectively.

The time complexity of Algorithm 1 depends on the number of the observations and hidden states. In the proposed model, the Vitebi algorithm is used to search for the most possible hidden state sequence. According to Algorithm 1, there are three variants denoted as DepenHMM_Simple, DepenHMM_Complex, and DepenHMM_Dynamic algorithms. For DepenHMM_Simple algorithm, there are $M + 1$ states and the time complexity is $O(|C|M^2)$, where$|C|$ is the number of the observation set and M is the number of the source.

Similarly, the time complexities of DepenHMM_Complex and DepenHMM_Dynamic are $O(|C|M^2)$ and $O(|C|(NM)^2)$, respectively.

## V. Experimental Analysis

### A. Experimental Setup

*Datasets:* Our experiments are conducted according to three real-world datasets[1]: *Stock*, *Flight*, and *Weather* proposed in [48] and [31] for data fusion. The ground truth is also provided in the dataset. For Stock dataset, it contains trading data of 1,000 stock symbols from 55 sources on every workday. The ground truth comes from NASDAQ100 stocks, where 100 stocks are randomly selected by taking the majority values provided by five stock providers. For Flight dataset, it contains information of over 1,200 flights from 38 sources over 1-month period. The ground truth contains departure/arrival information on 100 randomly selected flights provided by corresponding airline websites. For Weather dataset, it contains weather data on 30 major USA cites from 18 websites every 45 minutes on a day (total 7 days and every day has 1,833 timestamps). The ground truth standard is provided in ref [31]. The detailed information is described as follows in Table III.

*Baseline methods:* We assess the efficacy and efficiency of the proposed three methods against the current state-of-the-art approaches. A concise overview of the baseline methods is as follows.

- TruthFinder [49]: It analyses the similarity among different claims and weights all the claims of the sources.
- Two-Estimates [9]: It calculates the source credibility by aggregating votes. If source $s$ provides claim $c$, $c$ is considered to oppose all claims except $c$.
- Three-Estimates [9]: It is an improved version of the Two-Estimates algorithm and adds the error factor of the claim in truth discovery.
- Depen [15]: It is the first Bayesian truth detection model considering source dependence.
- CSS [36]: It studies the key source selection problem, which determines a subset of key sources.
- EMMutiF [7]: It considers the characteristics and propagation patterns of multimodal content.
- HMMN [23]: It focuses on the source dependence and applies HMM to model the relevance between two sources.
- SRTD [44]: It estimates claim truthfulness for dynamic truth from the credibility analysis on the claims and the historical contributions of sources.
- CTD [11]: It incorporates denial constraints into the process of truth discovery. In Stock Dataset, we add two constrains (stock price should be lower than the highest price and higher than the lowest price) and we add one constraint for Flight Dataset (actual departure time is equal to or later than the expected time) and Book Dataset (each city has only one temperature at each time).
- DART [50]: It integrates domain expertise based on data richness in different domains into source reliability.
- DSMFA [51]: It proposes a hyper-parameter recommendation strategy for DART algorithm based on data augmentation on the input dataset.

*Implementation details:* To evaluate performance of truth discovery methods over continuous time, we partition the datasets into sub-datasets based on different time intervals. Each sub-dataset corresponds to the data generated on a specific day. We use five-day intervals for evaluation. Three models ($Depen\_Simple$, $Depen\_Complex$ and $Depen\_Dynamic$) serve as the basis for generating the DepenHMM_Simple, DepenHMM_Complex, and DepenHMM_Dynamic methods, by Algorithm 1 where performance is compared with other methods. All methods are implemented in Python 3.8.8 and conducted on a server equipped with an Intel Xeon(R) Gold 6148 CPU@2.40 GHz with 80 cores.

*Parameter Calibration:* To calibrate the parameters of $a_1$ and $a_2$ in (4), we tested $Depen\_Dynamic$ model on 15 valid combinations of in $\{0.2, 0.3, 0.4, 0.5, 0.6\}$ on the Stock Dataset. The accuracy result is shown in Fig. 5. According to the results, (0.3, 0.2, 0.5) is used to measure the accuracy, coverage, and dependence.

### B. Experimental Results

To assess the efficacy and efficiency of the three HMM-based methodologies, we employ a suite of metrics including Accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Recall, Precision, $F_1$-score, and execution time. For the Stock and Flight datasets, the time in $X$-axis represents a five-day period from Monday to Friday in a specific week. The Weather dataset has a finer granularity, with continuous five-minute intervals. The scalability test and ablation study are meticulously designed to comprehensively evaluate the algorithm's performance across varying scales and the contributions of individual components.

*Accuracy comparison:* Fig. 2 shows that DepenHMM_Dynamic outperforms DepenHMM_Complex in accuracy, which in turn consistently outperforms DepenHMM_Simple across datasets. The EMMutiF method lags, particularly on the Stock dataset, due to its limited capacity to capture source dependencies, and even on the Flight dataset, it shows significant accuracy fluctuations, suggesting difficulty in managing dynamic source relationships. The Depen and Two-Estimates methods also show lower accuracy on the Stock and Weather datasets for not accounting for source dynamics. HMMN's performance is average, as its freshness parameter, designed to measure update speed, is less relevant given that all sources update at the same rate in these datasets. The DART and DSMFA algorithms exhibit mediocre performance, whereas the CTD algorithm demonstrates relatively inferior performance, because the fused values derived in the quadratic programming step in CTD fail to align with the actual truth. Overall, the DepenHMM methods demonstrate higher accuracy and robustness compared to other methods, underscoring their effectiveness.

*MAE comparison:* Fig. 3 illustrates the Mean Absolute Error (MAE) between predicted and actual truths across three datasets.

[1]Datasets are available in http://lunadong.com/fusionDataSets.htm.

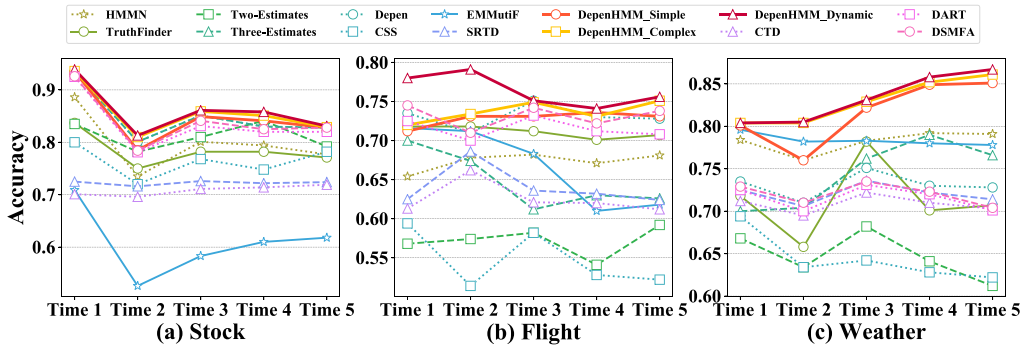


Fig. 2. Accuracy comparison results for three datasets.

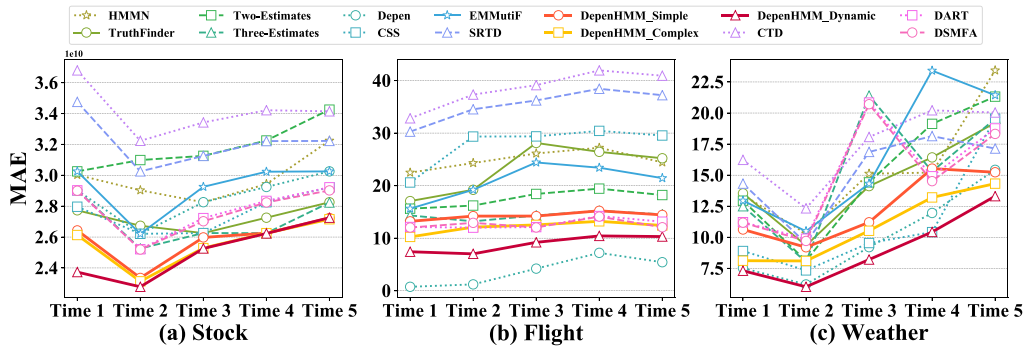


Fig. 3. MAE comparison results for three datasets.

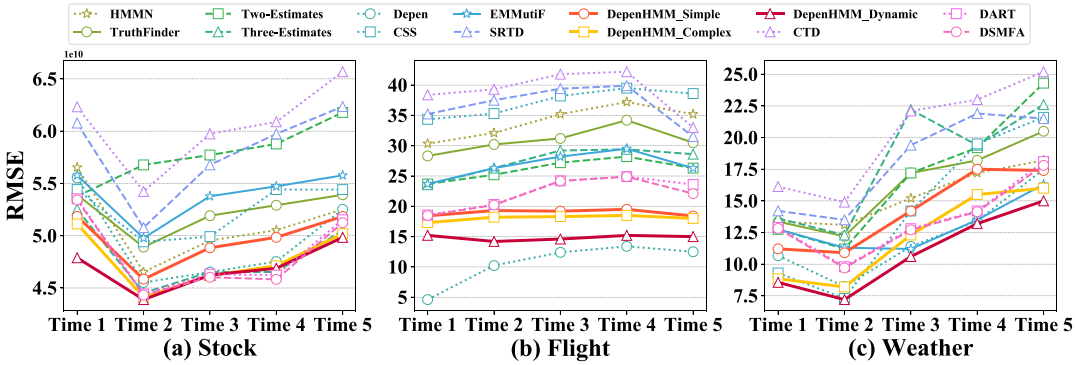


Fig. 4. RMSE comparison results for three datasets.

The DepenHMM_Dynamic method consistently records lower MAE than DepenHMM_Simple and DepenHMM_Complex, due to its decay function-based modeling of source dependencies, which adeptly manages complex source relationships and incomplete data for more accurate predictions. Comparing the DepenHMM methods to other benchmarks, they consistently show lower MAE on all datasets, highlighting their strength in tracking the dynamic truth over time. Due to the presence of data at the billion-scale in the Stock and Flight datasets, the magnitude of MAE and RMSE significantly exceed those observed in the Weather dataset. In the Stock dataset, the three methods respectively post average MAE of $2.58 \times 10^{10}$, $2.55 \times 10^{10}$, and $2.50 \times 10^{10}$. In the Flight dataset, the Depen method has the lowest MAE, closely followed by DepenHMM_Dynamic. In the Weather dataset, the average MAE are 12.3, 10.8, and 9.1, respectively, for the three methods, outperforming all but Depen and CSS. In summary, the DepenHMM methods consistently and accurately identify the dynamic truths of objects, as evidenced by their lower MAE across datasets.

*RMSE comparison:* Root Mean Squared Error (RMSE) measures the difference between estimated and actual truths, considering both error magnitude and direction, and is sensitive to large errors. RMSE is sensitive to scale, and results are presented in Fig. 4. The DepenHMM_Dynamic method outperforms the others in RMSE, indicating higher accuracy in truth prediction. In the Stock dataset, it achieves an RMSE of $4.69 \times 10^{10}$, better than the benchmark's $4.81 \times 10^{10}$. In the Flight dataset, DepenHMM_Dynamic's RMSE is 14.8, slightly

TABLE IV
COMPARISON ON RECALL, PRECISION AND $F_1$-SCORE

| Algorithms | Stock | | | Flight | | | Weather | | |
|---|---|---|---|---|---|---|---|---|---|
| | Recall | Precision | $F_1$-score | Recall | Precision | $F_1$-score | Recall | Precision | $F_1$-score |
| DepenHMM_Simple | 0.673 | 0.833 | 0.745 | 0.721 | 0.661 | 0.690 | 0.802 | 0.821 | 0.811 |
| DepenHMM_Complex | 0.675 | 0.835 | 0.747 | 0.722 | 0.671 | 0.696 | 0.792 | 0.804 | 0.798 |
| DepenHMM_Dynamic | **0.679** | **0.836** | **0.749** | **0.731** | **0.672** | **0.700** | **0.812** | **0.824** | **0.818** |
| TruthFinder | 0.663 | 0.826 | 0.736 | 0.691 | 0.652 | 0.671 | 0.781 | 0.794 | 0.787 |
| Two-Estimates | 0.571 | 0.776 | 0.658 | 0.634 | 0.663 | 0.648 | 0.792 | 0.782 | 0.787 |
| Three-Estimates | 0.673 | 0.821 | 0.740 | 0.693 | 0.652 | 0.672 | 0.794 | 0.791 | 0.792 |
| Depen | 0.624 | 0.826 | 0.711 | 0.681 | 0.621 | 0.650 | 0.761 | 0.753 | 0.757 |
| CSS | 0.663 | 0.826 | 0.736 | 0.702 | 0.602 | 0.648 | 0.735 | 0.782 | 0.758 |
| EMMutiF | 0.663 | 0.562 | 0.608 | 0.661 | 0.591 | 0.624 | 0.704 | 0.751 | 0.727 |
| HMMN | 0.645 | 0.823 | 0.723 | 0.712 | 0.653 | 0.681 | 0.805 | 0.819 | 0.812 |
| SRTD | 0.615 | 0.623 | 0.619 | 0.663 | 0.582 | 0.620 | 0.715 | 0.761 | 0.737 |
| CTD | 0.563 | 0.616 | 0.588 | 0.572 | 0.546 | 0.559 | 0.702 | 0.757 | 0.728 |
| DART | 0.661 | 0.803 | 0.725 | 0.684 | 0.645 | 0.664 | 0.776 | 0.754 | 0.765 |
| DSMFA | 0.661 | 0.813 | 0.729 | 0.691 | 0.653 | 0.671 | 0.781 | 0.763 | 0.772 |

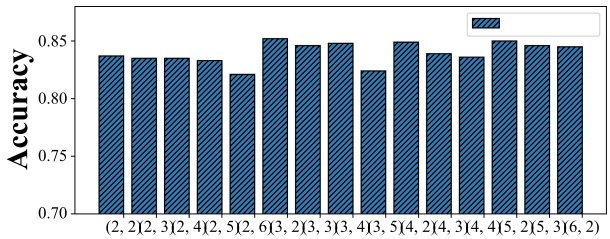


Fig. 5. Parameter calibration.

TABLE V
COMPARISON RESULTS WITH DYNAMIC DATA FOR THE STOCK, FLIGHT, AND WEATHER DATESETS ON EXECUTION TIME

| Method | Stock(/s) | Flight(/s) | Weather(/s) |
|---|---|---|---|
| DepenHMM_Simple | 5,429.6 | 438.5 | 331.4 |
| DepenHMM_Complex | 10,859.4 | 982.3 | 782.3 |
| DepenHMM_Dynamic | 12,060.3 | 1,083.4 | 792.4 |
| TruthFinder | 360.6 | 34.3 | 23.2 |
| Two-Estimates | 43,678.6 | 3,424.7 | 2,263.6 |
| Three-Estimates | 52,988.9 | 4,315.4 | 3,174.2 |
| Depen | 234,578.8 | 32,145.8 | 23,522.4 |
| CSS | 4,232.5 | 103.27 | 83.4 |
| EMMutiF | 406.0 | 35.22 | 21.3 |
| HMMN | 72,425.7 | 4,372.5 | 3,245.4 |
| SRTD | 246.3 | 22.5 | 16.4 |
| CTD | 1,882.9 | 164.2 | 145.5 |
| DART | 341.5 | 31.1 | 25.9 |
| DSMFA | 2,845.4 | 343.3 | 206.3 |

higher than Depen due to its penalization mechanism for dependent sources, enhancing stability over time. In the Weather dataset, DepenHMM_Dynamic records the lowest RMSE at 10.9. Overall, the DepenHMM methods provide an effective approach for truth discovery in dynamic settings, with lower RMSE values confirming their accuracy and reliability.x

*Recall, Precision, and $F_1$-score comparison:* To more comprehensively evaluate model performance, we employs a binarization strategy for gaining additional insights from regression models [52], [53]. For numerical datasets, instances with true values above the dataset mean are labeled as the positive class, while others are labeled as the negative class. For different datasets, samples with prediction errors within 1% of the true values are defined as positive samples; otherwise, they are defined as negative samples. Table IV shows that DepenHMM_Dynamic excels across all datasets and metrics, with improvements over the best-performing methods as follows: for the Stock dataset, it gains 0.9% in recall, 0.6% in precision, and 0.8% in $F_1$-score; for the Flight dataset, it gains 0.9% in recall, 0.1% in precision, and 0.5% in $F_1$-score; and for the Weather dataset, it gains 0.8% in recall, 1.0% in precision, and 0.9% in $F_1$-score. The DART algorithm exhibits average performance, mainly because the domain richness indicator does not show significant effectiveness in three datasets. The DSMFA algorithm shows some improvement over the DART algorithm, as it employs data augmentation to identify more optimal parameters for DART. However, due to the relatively small differences in richness among different entities in the dataset used in this experiment, its performance remains only moderately enhanced. These results underscore the effectiveness of the DepenHMM methods in dynamic truth discovery by analyzing source dependencies and dynamics. Overall, the DepenHMM methods outperform existing methods on all three datasets.

*Execution time comparison:* To assess the efficiency of the methods, we compare the execution time of all the methods for the three datasets, providing insights into their scalability and practicality for real-world applications. The results are presented in Table V. From the table we can see that the Depen method consumes the most execution time, followed by Three-Estimates and Two-Estimates. HMMN algorithm exhibits a time cost approximately 240 times greater than that of the DepenHMM_Dynamic algorithm in the Stock dataset because HMMN requires the estimation of dependencies among all pairwise sources. It becomes less efficient when dealing with many sources. The DART algorithm executes swiftly due to its lack of complex computations and rapid iterations. The CTD algorithm, though requiring constraint calculations, is efficient as it avoids iterations, while the DSMFA algorithm demands more time due to its data augmentation process. In terms of the DepenHMM methods, the execution time of DepenHMM_Complex is greater than that of DepenHMM_Simple but smaller than that of DepenHMM_Dynamic. Specifically, DepenHMM_Simple takes approximately half the execution time of DepenHMM_Dynamic, within the middle range compared to the other evaluated methods. As we can see, Depen_Simple has lower complexity, making it more suitable for large-scale

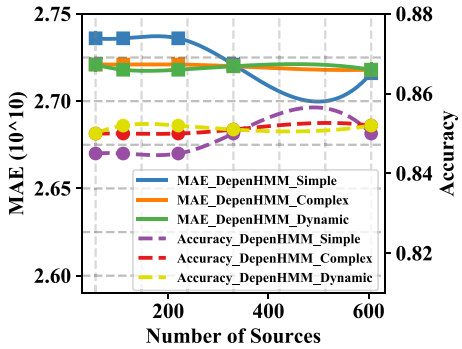


Fig. 6. Scalability test.

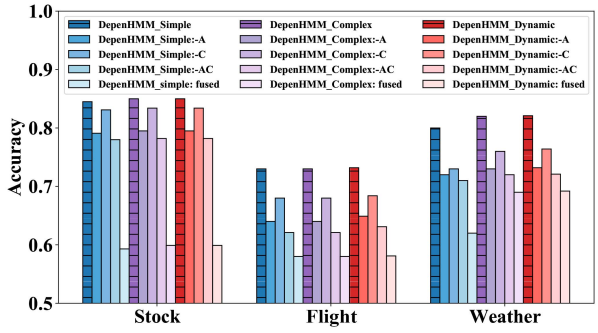


Fig. 7. Ablation test.

datasets. The superior performance of the Depen_Simple hidden Markov model compared to baseline models can be used for large-scale datasets in future. By partitioning a larger dataset into multiple source sets, each can contain at most one source of plagiarism.

Although the three DepenHMM methods may not be the fastest among the evaluated methods, their execution times remain within acceptable limits for the given datasets. Furthermore, it is worth noting that the execution time of the three DepenHMM methods scales linearly with the number of objects in the dataset, indicating their ability to handle large-scale data efficiently. Overall, the three proposed DepenHMM methods provide improved solutions for truth discovery problems in dynamic scenarios, with reasonable execution times that make them practical and scalable for a wide range of applications.

*Scalability Test:* To assess the effectiveness of the proposed method in handling large volumes of data, we conduct a scalability test by expanding the number of sources in the **Stock** dataset. We expand the Stock dataset (originally with 55 sources) by 1, 3, 5, and 10 times, resulting in 110, 220, 330, and 605 sources, respectively. MAE and Accuracy are utilized to measure the performance across these varying scales and the experimental results are presented in Fig. 6. As the number of sources increases, the performance of the three algorithms generally remains stable overall, particularly DepenHMM_Complex and DepenHMM_Dynamic. DepenHMM_Simple exhibits minor fluctuations within a limited range. Moreover, with the increase in the number of sources, the accuracy of DepenHMM_Simple improves and the Mean Absolute Error (MAE) decreases because due to the algorithm's design, at most only one source can be considered dependent. Therefore, the increase in the number of sources does not significantly affect the reliability assessment of the sources, thereby ensuring the stability.

*Ablation test:* In the ablation test, we aim to investigate the roles of accuracy, coverage, and the selection of the most reliable sources in truth discovery. For the three models, we conducted experiments by removing accuracy (-A), removing coverage (-C), removing both accuracy and coverage (-AC), and fusing the claims instead of selecting the most reliable source declarations (fused). We tested the accuracy on three datasets, and the results are shown in the Fig. 7. It can be observed that the accuracy of truth discovery decreases when either accuracy or coverage is removed, especially when both are removed simultaneously because the reliability of sources needs to be assessed from multiple perspectives, and relying solely on source dependence leads to singular evaluation. Additionally, the performance of fusing claims drops significantly because most of the fused claims are entirely new values, which may lead to unexpected and potentially inaccurate data that do not reflect any real-world situations.

## VI. Conclusion

The presence of conflicting information from multiple sources makes it critical to distinguish between what is true and what is not. This paper focuses on addressing the dynamic truth discovery problem with dependent sources where the dynamic characteristics of objects have received limited attention in the existing researches. To tackle this challenge, we construct three hidden Markov models and propose a novel approach to measure the dynamic dependence of sources and discover latent truths in dynamic scenarios. To cater to various application scenarios, we develop three distinct source dependence models. Each model is designed to measure source dependence in a more refined manner, capturing completed, partial and dynamic dependencies among sources. The experimental results show that all three developed methods perform high accuracy and robustness with acceptable execution time, making them suitable for real-world applications. In addition, sources may focus special domains on objects, which implies that they could provide more trustworthy claims on some objects than others. In the future, we will particularly study the biased trustworthiness for objects among different sources.

## References

[1] Y. U. Dong, D. R. Shen, Y. Kou, T. Z. Nie, and Y. U. Ge, "Web data integration oriented truth discovery algorithms," *J. Chin. Comput. Syst.*, vol. 37, no. 8, pp. 1633–1638, 2016.

[2] J. Pasternack and D. Roth, "Latent credibility analysis," in *Proc. 22nd Int. Conf. World Wide Web*, New York, NY, USA, 2013, pp. 1009–1020.

[3] H. Lu, X. Gao, and G. Chen, "Efficient crowdsourcing-aided positioning and ground-truth-aided truth discovery for mobile wireless sensor networks in urban fields," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1652–1664, Mar. 2022.

[4] X. Pang, Z. Wang, D. Liu, J. C. S. Lui, Q. Wang, and J. Ren, "Towards personalized privacy-preserving truth discovery over crowdsourced data streams," *IEEE/ACM Trans. Netw.*, vol. 30, no. 1, pp. 327–340, Feb. 2022.

[5] P. Sun et al., "Towards personalized privacy-preserving incentive for truth discovery in mobile crowdsensing systems," *IEEE Trans. Mobile Comput.*, vol. 21, no. 1, pp. 352–365, Jan. 2022.

[6] M. Zhao and J. Jiao, "Police: An effective truth discovery method in intelligent crowd sensing," in *Proc. 6th Int. Conf. Artif. Intell. Secur.*, Berlin, Heidelberg: Springer-Verlag, 2020, pp. 384–398.
[7] H. Shao et al., "Truth discovery with multi-modal data in social sensing," *IEEE Trans. Comput.*, vol. 70, no. 9, pp. 1325–1337, Sep. 2021.
[8] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web*, New York, NY, USA, 2011, pp. 675–684.
[9] A. Galland, S. Abiteboul, A. Marian, and P. Senellart, "Corroborating information from disagreeing views," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, New York, NY, USA, 2010, pp. 131–140.
[10] Y. Li et al., "A survey on truth discovery," *ACM SIGKDD Explorations Newslett.*, vol. 17, no. 2, pp. 1–16, 2016.
[11] C. Ye et al., "Constrained truth discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 205–218, Jan. 2022.
[12] D. Wang et al., "Using humans as sensors: An estimation-theoretic perspective," in *Proc. 13th Int. Symp. Inf. Process. Sensor Netw.*, 2014, pp. 35–46.
[13] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci.*, vol. 497, pp. 38–55, 2019.
[14] A. Tatsioni, N. G. Bonitsis, and J. P. A. Ioannidis, "Persistence of contradicted claims in the literature," *JAMA*, vol. 298, no. 21, pp. 2517–2526, Dec. 2007.
[15] X. L. Dong, L. Berti-Equille, and D. Srivastava, "Integrating conflicting data: The role of source dependence," in *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 550–561, 2009.
[16] D. Wang, L. Kaplan, and T. F. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Trans. Sensor Netw.*, vol. 10, no. 2, 2014, Art. no. 30.
[17] B. I. Aydin, Y. S. Yilmaz, Y. Li, Q. Li, J. Gao, and M. Demirbas, "Crowdsourcing for multiple-choice question answering," in *Proc. 28th AAAI Conf. Artif. Intell.*, AAAI Press, 2014, pp. 2946–2953.
[18] V. Beretta, S. Harispe, S. Ranwez, and I. Mougenot, "Truth selection for truth discovery models exploiting ordering relationship among values," *Knowl.-Based Syst.*, vol. 159, pp. 298–308, 2018.
[19] J. Feng, J. Chen, and J. Lu, "Novel approach for multi-valued truth discovery," in *Proc. 13th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Cham: Springer International Publishing, 2019, pp. 1015–1028.
[20] L. Yao et al., "Online truth discovery on time series data," in *Proc. 2018 SIAM Int. Conf. Data Mining*, Elsevier, 2018, pp. 162–170.
[21] Y. Wang, F. Ma, L. Su, and J. Gao, "Discovering truths from distributed data," in *Proc. 2017 IEEE Int. Conf. Data Mining*, 2017, pp. 505–514.
[22] A. Pal, V. Rastogi, A. Machanavajjhala, and P. Bohannon, "Information integration over time in unreliable and uncertain environments," in *Proc. 21st Int. Conf. World Wide Web*, New York, NY, USA, 2012, pp. 789–798.
[23] X. L. Dong, L. Berti-Equille, and D. Srivastava, "Truth discovery and copying detection in a dynamic world," in *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 562–573, 2009.
[24] X. S. Fang, "Truth discovery from conflicting multi-valued objects," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 711–715.
[25] X. S. Fang, Q. Z. Sheng, X. Wang, D. Chu, and A. H. Ngu, "SmartVote: A full-fledged graph-based model for multi-valued truth discovery," *World Wide Web*, vol. 22, no. 4, pp. 1855–1885, Jul. 2019.
[26] F. Li, M. L. Lee, and W. Hsu, "Entity profiling with varying source reliabilities," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, New York, NY, USA, 2014, pp. 1146–1155.
[27] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. 2014 ACM SIGMOD Int. Conf. Manage. Data*, New York, NY, USA, 2014, pp. 1187–1198.
[28] L. Zhao, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, New York, NY, USA, 2016, pp. 2085–2094.
[29] L. Jiang, X. Niu, J. Xu, D. Yang, and L. Xu, "Incentivizing the workers for truth discovery in crowdsourcing with copiers," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst.*, 2019, pp. 1286–1295.
[30] A. D. Sarma, X. L. Dong, and A. Halevy, "Data integration with dependent sources," in *Proc. 14th Int. Conf. Extending Database Technol.*, New York, NY, USA, 2011, pp. 401–412.
[31] X. L. Dong, L. Berti-Equille, Y. Hu, and D. Srivastava, "Global detection of complex copying relationships between sources," in *Proc. VLDB Endowment*, vol. 3, no. 1/2, pp. 1358–1369, Sep. 2010.
[32] H. Zhang et al., "Influence-aware truth discovery," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, New York, NY, USA, 2016, pp. 851–860.
[33] H. Cui, T. Abdelzaher, and L. Kaplan, "A semi-supervised active-learning truth estimator for social networks," in *Proc. World Wide Web Conf.*, New York, NY, USA, 2019, pp. 296–306.
[34] X. S. Fang, Q. Z. Sheng, X. Wang, and A. H. Ngu, "Value veracity estimation for multi-truth objects via a graph-based approach," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 777–778.
[35] X. Yin and W. Tan, "Semi-supervised truth discovery," in *Proc. 20th Int. Conf. World Wide Web*, New York, NY, USA, 2011, pp. 217–226.
[36] D. Wang, N. Vance, and C. Huang, "Who to select: Identifying critical sources in social sensing," *Knowl.-Based Syst.*, vol. 145, pp. 98–108, 2018.
[37] L. Ge, J. Gao, X. Li, and A. Zhang, "Multi-source deep learning for information trustworthiness estimation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, New York, NY, USA, 2013, pp. 766–774.
[38] D. Y. Zhang et al., "Towards scalable and dynamic social sensing using a distributed computing framework," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst.*, 2017, pp. 966–976.
[39] Y. Yang, Q. Bai, and Q. Liu, "A probabilistic model for truth discovery with object correlations," *Knowl.-Based Syst.*, vol. 165, pp. 360–373, 2019.
[40] D. Y. Zhang, D. Wang, and Y. Zhang, "Constraint-aware dynamic truth discovery in Big Data social media sensing," in *Proc. 2017 IEEE Int. Conf. Big Data*, 2017, pp. 57–66.
[41] Q. Wen, K. He, L. Sun, Y. Zhang, M. Ke, and H. Xu, "RobustPeriod: Robust time-frequency mining for multiple periodicity detection," in *Proc. 2021 Int. Conf. Manage. Data*, New York, NY, USA, 2021, pp. 2328–2337.
[42] D. Tiano, A. Bonifati, and R. Ng, "FeatTS: Feature-based time series clustering," in *Proc. 2021 Int. Conf. Manage. Data*, New York, NY, USA, 2021, pp. 2784–2788.
[43] S. Y. Shah et al., "AutoAI-TS: AutoAI for time series forecasting," in *Proc. 2021 Int. Conf. Manage. Data*, New York, NY, USA, 2021, pp. 2584–2596.
[44] D. Zhang, D. Wang, N. Vance, Y. Zhang, and S. Mike, "On scalable and robust truth discovery in Big Data social media sensing applications," *IEEE Trans. Big Data*, vol. 5, no. 2, pp. 195–208, Jun. 2019.
[45] Y. Chen and S. Huang, "TSExplain: Surfacing evolving explanations for time series," in *Proc. 2021 Int. Conf. Manage. Data*, New York, NY, USA, 2021, pp. 2686–2690.
[46] G. Forney, "The viterbi algorithm," in *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
[47] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, vol. 41, no. 1, pp. 164–171, 1970.
[48] X. Li, X. L. Dong, K. Lyons, W. Meng, and D. Srivastava, "Truth finding on the deep web: Is the problem solved?," in *Proc. VLDB Endowment*, vol. 6, no. 2, pp. 97–108, 2012.
[49] X. Yin, J. Han, and P. S. Yu, "Truth discovery with multiple conflicting information providers on the web," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 6, pp. 796–808, Jun. 2008.
[50] X. Lin and L. Chen, "Domain-aware multi-truth discovery from conflicting sources," in *Proc. VLDB Endowment*, vol. 11, no. 5, pp. 635–647, Oct. 2018.
[51] S. Chen, X. Ding, Z. Liang, Y. Tang, and H. Wang, "Hyper-parameter recommendation for truth discovery," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, Singapore: Springer Nature, 2025, pp. 277–292.
[52] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. New York, NY, USA: Springer, 2009.
[53] S. S. Lokhande and N. Dawande, "A survey on document image binarization techniques," in *Proc. 2015 Int. Conf. Comput. Commun. Control Automat.*, 2015, pp. 742–746.

**He Zhang** received the BSc degree from Chien-Shiung Wu College, Southeast University, in 2022, and the MSc degree from the School of Computer Science and Engineering, Southeast University, in 2025. Her research interests focus on truth discovery and cloud computing.

**Shuang Wang** received the BSc degree from the College of Sciences, Nanjing Agricultural University, in 2015, and the PhD degree from the School of Computer Science and Engineering, Southeast University, Nanjing, China, in 2020. She was a visiting PhD student with the School of Computing, Macquarie University, Sydney, Australia, from 2019 to 2020, and a postdoctoral research fellow from 2020 to 2021. She is currently an associate professor with Southeast University. Her research has been published in international journals and conferences such as *IEEE Transactions on Computers*, *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Services Computing*, *IEEE Transactions on Network and Service Management* and ICSOC. Her main research interests focus on service computing, Big Data analytics, and truth discovery.

**Long Chen** received the BSc and PhD degrees in computer science and engineering from Southeast University, Nanjing, China, in 2009 and 2018, respectively. He is currently a associate professor with the Department of Computer Science and Engineering, Southeast University, Nanjing. He has published more than 20 papers in international journals and conferences, such as *IEEE Transactions on Services Computing*, *IEEE Transactions on Cloud Computing*, and *IEEE Transactions on Automation Science and Engineering*. His main interests include task scheduling in cloud computing, service-oriented computing, evolutionary computation, data normalization in smart building.

**Xiaoping Li** (Senior Member, IEEE) received the BSc and MSc degrees in applied computer science from the Harbin University of Science and Technology, in 1993 and 1999 respectively, and the PhD degree in applied computer science from the Harbin Institute of Technology, in 2002. He is a full professor with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China. Prior to that, he was a full professor with the School of Computer Science and Engineering, Southeast University, Nanjing, China. He is the author or co-author more than 100 academic papers, some of which have been published in international journals such as the *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Services Computing*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Automation Science and Engineering*, *IEEE Transactions on Cloud Computing*, *IEEE Transactions on Systems, Man and Cybernetics: Systems*, etc. His research interests include scheduling in cloud computing, scheduling in cloud manufacturing, service computing, Big Data and machine learning.

**Qing Gao** received the PhD degree in mechatronics engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2014. From 2014 to 2016, he was with the School of Engineering and Information Technology, University of New South Wales, Canberra, Australian Defence Force Academy, as a postdoctoral research associate. Since 2018, he has joined the School of Automation Science and Electrical Engineering, Beihang University as a full professor. His research interests include intelligent control and quantum control. He is the recipient of the Alexander von Humboldt Fellowship of Germany and the 21st Guan Zhao-Zhi Award.

**Quan Z. Sheng** received the PhD degree in computer science from the University of New South Wales (UNSW) and did his postdoc as a research scientist with CSIRO ICT Centre. He is a distinguished professor and head of the School of Computing, Macquarie University, Sydney, Australia. His research interests include service-oriented computing, distributed computing, Internet computing, and Internet of Things (IoT). He is the recipient of AMiner Most Influential Scholar Award in IoT, in 2019, ARC (Australian Research Council) Future Fellowship, in 2014, Chris Wallace Award for Outstanding Research Contribution, in 2012, and the Microsoft Research Fellowship, in 2003. He is ranked by Microsoft Academic as one of the Most Impactful Authors in Services Computing (ranked the 4th all-time). He is the vice chair of the Executive Committee of the IEEE Technical Community on Services Computing (IEEE TCSVC), the associate director (Smart Technologies) of Macquarie University Smart Green Cities Research Centre, and a member of the ACS (Australian Computer Society) Technical Advisory Board on IoT.