Tsinghua Science and Technology

Volume 22 | Issue 6 Article 5

2017

Truth Discovery with Memory Network

Luyang Li
Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001, China.

Bing Qin
Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001, China.

Wenjing Ren
Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001, China.

Ting Liu
Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001, China.

Recommended Citation
Luyang Li, Bing Qin, Wenjing Ren et al. Truth Discovery with Memory Network. Tsinghua Science and Technology 2017, 22(6): 609-618.

TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 05/15 pp609-618
Volume 22, Number 6, December 2017

Truth Discovery with Memory Network

Luyang Li, Bing Qin*, Wenjing Ren, and Ting Liu

Abstract: Truth discovery aims to resolve conflicts among multiple sources and find the truth. Conventional methods for truth discovery mainly investigate the mutual effect between the reliability of sources and the credibility of statements. These methods use real numbers, which have a lower representation capability than vectors to represent the reliability. In addition, neural networks have not been used for truth discovery. In this work, we propose memory-network-based models to address truth discovery. Our proposed models use feedforward and feedback memory networks to learn the representation of the credibility of statements. Specifically, our models adopt a memory mechanism to learn the reliability of sources for truth prediction. The proposed models use categorical and continuous data during model learning by automatically assigning different weights to the loss function on the basis of their own effects. Experimental results show that our proposed models outperform state-of-the-art methods for truth discovery.

Key words: truth discovery; memory networks; source reliability

1 Introduction

In the present age of information abundance, numerous conflicts exist among statements from multiple sources. For example, different booking sites provide different boarding times for the same flight on the same date. Conflicting information seriously affects people's daily lives. It's more common in the social media[1]. Truth discovery aims to identify the most credible statement among the conflicts[2–4]. The development of research on truth discovery has also benefited natural language processing tasks such as knowledge management[5], question answering[6], and information retrieval[7].

Previous methods have mainly utilized a voting mechanism to predict the truth[2–4, 8–14]. In voting, the reliability of sources is used as the weight for computing the credibility of statements. Reliability is treated as a real number and as other intermediate variables. The first challenge of the voting mechanism, however, is that real numbers have weak representation capability. Existing methods apply multiplication to combine the reliability of sources in computing the credibility of statements. The second challenge of the voting mechanism is that the combination method should be more complex to better reflect reality. The third challenge is that multiple types of data, such as categorical and continuous data exist in the real world. Both types of data are helpful in evaluating the credibility of statements. Except for the Conflict Resolution on Heterogeneous (CRH) data framework method, few approaches have used multiple data types to predict the truth[15]. However, CRH treats two types of data equally, that intuitively does not conform to reality.

In this work, we propose models that are based on Neural Networks (NNs), which have clear advantages in numerous natural language processing problems, to address the truth discovery task[16–23]. Our method resolves the three abovementioned challenges by using a memory-network-based model to learn the reliability of sources and predict the credibility of statements. The reliability of sources can be treated as the

• Luyang Li, Bing Qin, Wenjing Ren, and Ting Liu are with Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001, China. E-mail: {lyli, qinb, wjren, tliu}@ir.hit.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2016-12-31; revised: 2017-03-29; accepted: 2017-04-21

background knowledge of the statements. We vectorize the statement and the reliability of sources, and use a memory component to store and update the reliability of sources. Our proposed memory-network-based models can incorporate the representation of source reliability into the evaluation of the credibility of information. The components of the memory-network-based model iteratively update the representation of source reliability and of information credibility. Thus, memory network is appropriate for using in truth discovery. In the model, we apply an NN model to learn the representation of the information and the reliability of sources because the feedforward and feedback NN models are two classical NN models. We apply both of them to combine with memory mechanism (memory network model) and develop the Feedforward Memory Network (FFMN) and Feedback Memory Network (FBMN). During the computation of the credibility of statements, multiple types of data are utilized with different weights. The weight of each type of data is automatically computed on the basis of the effect in model learning. We validate the effectiveness of the proposed method by performing experiments with benchmark datasets. Experimental results show that our proposed method outperforms the state-of-the-art method.

We make the following contributions:

- In a novel approach, we utilize vectors, which have better representation capability than real numbers, to represent the statement and the reliability of sources.
- We propose memory-network-based models to resolve the truth discovery problem. The reliability of sources is treated as latent background knowledge and can be stored in the memory unit to help compute the credibility of statements.
- To better learn the reliability of sources, we utilize categorical data and continuous data. These two types of data are automatically aligned on the basis of different weights through model optimization.

2 Methodology

We first formulate the problem. Then we introduce the framework and the central functions of the proposed model.

2.1 Problem formulation

In the real world, the *statement* is an information unit used to describe an item or an event. We investigate the truth discovery problem on the basis of the statements. The *object* and the *property* of the statement constitute an *entry*[15]. The *observation* is the *value* of the *entry*. According to the same entry, observations from different sources may contain conflicting observations. Our goal is to find the correct statement. In other words, we aim to find the correct observation for a given entry.

EXAMPLE 1. In the statement “Flight A will take off at 01:19 PM”, “Flight A” is an *object*. “The takeoff time of flight A” is an *entry*. The *observation* of the *entry* is “01:19 PM”.

The *object* may have more than one *property* in a statement, and we separate the statement into multiple entries. Specifically, we take an entry as the basic unit in the prediction model, and the credibility of the observations drawn from the same statement will be evaluated separately.

EXAMPLE 2. “Flight A will take off at 01:19 PM at gate 42B” has two entries, “the takeoff time of A” and “the gate of A”.

Observations can be classified into two data types: categorical and continuous data. Categorical data is class-style data. Continuous data are numbers, or can be converted into real numbers. In the prediction model, we want to predict the correct category for categorical data and predict the closest value to the true value for continuous data.

EXAMPLE 3. “Gate 42B” is categorical datum, and “01:19 PM” can be treated as continuous datum when converted into minutes.

Suppose there are N entries, each of which has K observations offered by K sources $\{s_1, s_2, ..., s_K\}$. Given an entry e_i, the observations are the value set $V_i = \{v_{i1}, v_{i2}, ..., v_{iK}\}$.

2.2 Basic framework

The lack of gold standard data is a challenge in truth discovery. The CRH method[15] uses an unsupervised framework to resolve this problem. Li et al.[15] suggested that reliable sources offer trustworthy observations that must be close to the truth. We cautiously think that observations that are more trustworthy should be closer to the truth. Given an entry e_i, D is the function for computing the distance between the truth t_i and the observation value v_{ik}. The truth-finding task is treated as an optimization problem. The objective of optimization is to minimize the following loss function, where r_{ik} stands for the credibility of the observation provided by source S_k. We compute the credibility of the observation by using the memory-network-based model, which is introduced in Section 3.

$$f_{\text{loss}} = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \cdot D(t_i, v_{ik}) \quad (1)$$

Categorical data and continuous data have different distance functions. When entry e_i belongs to the categorical dataset U_{cate}, the distance function d_{cate} is as follows.

$$d_{\text{cate}}(t_i, v_{ik}) = \begin{cases} 1, & v_{ik} \neq t_i; \\ 0, & v_{ik} = t_i \end{cases} \quad (2)$$

If the observations of entry e_i belong to the continuous dataset U_{con}, the distance function d_{con} is as follows. The denominator of the function is the mean square error of value set $V_i = \{v_{i1}, v_{i2}, ..., v_{ik}, ..., v_{iK}\}$ according to entry e_i. $\tilde{v_i}$ stands for the mean value of the value set V_i.

$$d_{\text{con}}(t_i, v_{ik}) = \frac{|t_i - v_{ik}|}{\sqrt{(v_{i1} - \tilde{v_i})^2 + \cdots + (v_{iK} - \tilde{v_i})^2}} \quad (3)$$

The truth t_i, which can minimize the overall weighted absolute distance, is the weighted median[15]. Given the observation set $\{v_{i1}, v_{i2}, ..., v_{ik}, ..., v_{iK}\}$ with the credibility set $\{r_{i1}, r_{i2}, ..., r_{ik}, ..., r_{iK}\}$, the weighted median of the set is v_{im} which satisfies the following condition:

$$\sum_{k: v_{ik} < v_{im}} r_{ik} < \frac{1}{2} \sum_{k=1}^{K} r_{ik},$$
$$\sum_{k: v_{ik} > v_{im}} r_{ik} \leqslant \frac{1}{2} \sum_{k=1}^{K} r_{ik} \quad (4)$$

The loss function in our model is as follows and is the sum of the loss functions of categorical data and continuous data. Specifically, the penalty values α and β are automatically learned through model learning.

$$f_{\text{loss}} = \alpha \sum_{i \in U_{\text{cate}}} \sum_{k=1}^{K} r_{ik} \cdot d_{\text{cate}}(t_i, v_{ik}) + \beta \sum_{j \in U_{\text{con}}} \sum_{k=1}^{K} r_{jk} \cdot d_{\text{con}}(t_j, v_{jk}) \quad (5)$$

3 Memory-Network-Based Model for Truth Discovery

3.1 Embedding learning

In the truth discovery task, the observation with high frequency for an object empirically has a high probability to be right. Embedding algorithms can learn the closeness between objects and observations by considering objects and observations as the context for each other. Multiple observations of the same object are related. Thus, we posit that embedding algorithms can learn the latent law of the structure and interaction among data.

We learn the word embedding of data to vectorize every information from a dataset. There are three kinds of data: objects, properties, and values. We use Word2Vec[24] to obtain the vector of each datum. Specifically, we take a source and its observation of an entry (i.e., object, property, and value) as a context term in learning word embedding. We think that the relationship between data can be learned on the basis of the context-based word-embedding-learning algorithm.

3.2 Introduction to the memory network

In this section, we introduce the memory network. The memory network is a framework with long-term memory as an inference component[25–27]. It consists of memory M with four components I, G, O, and R. The memory M can be read and written to store long-term information that is useful in prediction. The I component learns the representation of inputs. The G component generates new memories on the basis of new input. The O component produces an output on the basis of the new input and current memories. The R component converts the output into the response format.

In our problem, source reliability is long-term information to be used in truth prediction. Source reliability should be updated during model learning by inputting each sample and combined with input to generate a new output. We use memory $M = \{\boldsymbol{m}_1, \boldsymbol{m}_2, ..., \boldsymbol{m}_K\}$ to store the reliability of K sources and update memories on the basis of the input observations and current derivative of back propagation. The inputs are the vectorized observations, and the response of the model is the credibility of observations. We learn the representation of the input data with memories using feedforward NN and feedback NN models. We designate these models as FFMN and FBMN.

3.3 FFMN

In the truth discovery task, the reliability of sources and the credibility of information influence each other. In traditional methods, the reliability of sources is represented by a real number and is treated as a weight in computing the credibility of information. We vectorize the reliability of sources and treat it as a weight matrix in evaluating the credibility of information in the FFMN.

The feedforward NN is similar to a directed acyclic graph without feedback through the network. The FFNN is feedforward NN with a memory network mechanism. The architecture of FFMN is shown in Fig. 1. The input $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_j, ..., \boldsymbol{x}_K\}$ of each iteration in the I component is the set of vectors of observations according to a same entry. The observation vector is the concatenation of the vectors of object, property, and value data. M stores K memory vectors, which represent the reliability of sources. The memories and inputs are computed through element-wise multiplication. In the FFMN, the memory vectors serve as a weight matrix and are combined with input information through matrix multiplication. Thus, similar to other parameters in the FFMN, the memory vector $\boldsymbol{m}_j$ is updated through error backpropagation each time. Furthermore, $\sigma : R \rightarrow [0, 1]$ is a softmax function. The response of the FFMN is $\{r_1, r_2, ..., r_j, ..., r_K\}$ and is computed similar as in the following formula:

$$r_j = \sigma(\boldsymbol{m}_j \odot \boldsymbol{x}_j) \qquad (6)$$

3.4 FBMN

No unique answer exists for the roles of source reliability in the evaluation of information credibility. We attempt to take a more complex approach to combine source reliability with information credibility. We apply FBMN to treat source reliability as external knowledge and incorporate it in the composition of the hidden-layer representation.

Feedback NN is a kind of network whose neuron feedback is output to other neurons as input after a time step. Long Short-Term Memory (LSTM) is a typical feedback NN. LSTM uses the hidden layer H to store "short-term memory" and uses the cell unit C to store "long-term memory". It adopts input gate $\boldsymbol{x}_t$, forget gate $\boldsymbol{f}_t$ to control the updating of "long-term memory" $\boldsymbol{c}_t$, and adopts output gate $\boldsymbol{o}_t$ to compute the hidden vector at a current time step.

The FBMN is shown in Fig. 2. We add a memory component M to LSTM to store the reliability of sources. The memory component M has a different role and operation than cell C in LSTM. The input of the model is a series of values from different sources according to a same entry. The time step is the order of input values. The response of the model is the result of the softmax operation of the hidden vector of the last time step. Memories in M are updated on the basis of the response through the back-propagation of the

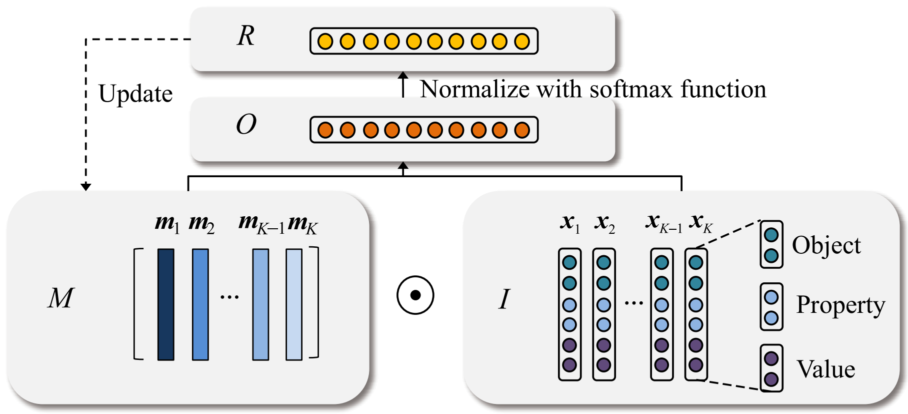

Fig. 1 Architecture of FFMN.

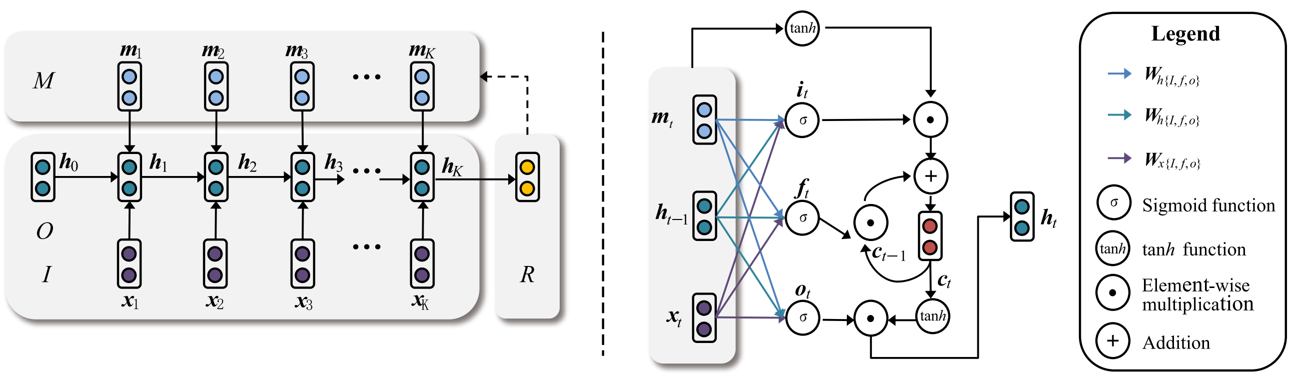

Fig. 2 Architecture (left) of FBMN and the detailed schematic (right) for hidden vector generation.

derivative.

The hidden vector $\boldsymbol{h}_t$ is computed on the basis of memory $\boldsymbol{m}_t$, the hidden vector on last time step $\boldsymbol{h}_{t-1}$, and the current input $\boldsymbol{x}_t$. Memory $\boldsymbol{m}_t$ stores the reliability of the t-th source. The σ is a sigmoid function and $\odot$ is element-wise multiplication. $\boldsymbol{c}_t$ stores the long-term memory at t time. $\boldsymbol{W}_{ix}$, $\boldsymbol{W}_{ih}$, $\boldsymbol{W}_{ic}$, and $\boldsymbol{W}_{im}$ are the weight matrices of the input vecter $\boldsymbol{x}_t$, the hidden vector $\boldsymbol{h}_{t-1}$, the cell vecroe $\boldsymbol{c}_{t-1}$, and the memory vector $\boldsymbol{m}_t$ through computing the input gate $\boldsymbol{i}_t$. The meaning of $\boldsymbol{W}$ in Eqs. (8)–(10) is similar with the above weight matrix. $\boldsymbol{b}$ is the bais. The detailed operating mechanism is shown in the right hand side of Fig. 2.

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{ix}\boldsymbol{x}_t + \boldsymbol{W}_{ih}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{ic}\boldsymbol{c}_{t-1} + \boldsymbol{W}_{im}\boldsymbol{m}_t + \boldsymbol{b}_i) \tag{7}$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{fx}\boldsymbol{x}_t + \boldsymbol{W}_{fh}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{fc}\boldsymbol{c}_{t-1} + \boldsymbol{W}_{fm}\boldsymbol{m}_t + \boldsymbol{b}_f) \tag{8}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tan h(\boldsymbol{W}_{cx}\boldsymbol{x}_t + \boldsymbol{W}_{ch}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{cm}\boldsymbol{m}_t + \boldsymbol{b}_c) \tag{9}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{ox}\boldsymbol{x}_t + \boldsymbol{W}_{oh}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{oc}\boldsymbol{c}_{t-1} + \boldsymbol{W}_{om}\boldsymbol{m}_t + \boldsymbol{b}_o) \tag{10}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \tan h(\boldsymbol{c}_t) \tag{11}$$

4 Experiment and Analysis

We present the results of the experiments that we conducted to validate the effectiveness of our memory-network-based models. We first introduce the dataset and evaluation criteria. We then present and analyze the experimental results. We also discuss the influence of different parameters on the results.

4.1 Datasets

We use two public datasets (http://lunadong.com/fusion-DataSets.htm.)[28] to demonstrate the effectiveness of the proposed method. The statistics of the utilized datasets are listed in Table 1. We perform data pre-processing to eliminate redundancy, which causes multiple different values according to a same entry in the ground-truth set. The entries contained in ground truths are part of the whole entries in the dataset. The ground truths are used only in the evaluation.

Stock Dataset. Li et al.[28] collected stock data from 55 sources on every work data in July 2011. The dataset contains 1000 stock symbols with 16 properties. The ground truths contain NASDAQ100 stocks and other 100 randomly selected stocks. These stock data are acquired by taking the majority of the values provided by five sources: Nasdaq.com, Yahoo Finance, Google Finance, Bloomberg, and MSN Finance. To verify the feasibility of utilizing categorical data and continuous data, Li et al.[15] considered the properties, Volume, Shares Outstanding, and Market Cap, as continuous data and other properties as categorical data[15]. We follow their example in our experiments.

Table 1 Statistics of the datasets used in the validation experiments.

	Number of observations	Number of entries	Number of ground truths
Stock Dataset	12 056 684	335 975	29 207
Flight Dataset	2 703 448	204 414	16 276

Flight Dataset. The flight dataset was collected from 38 sources over a one-month period (December 2011). It consists of 1200 flights with six properties. We treat departure gate and arrival gate as categorical data and treat other properties as continuous data. We treat time data as real numbers by converting them into minutes. The ground truths are about 100 randomly selected flights.

4.2 Evaluation criteria

We use two evaluation criteria, error rate and Mean Normalized Absolute Distance (MNAD)[15], to evaluate our methods. These criteria are used to evaluate the two types of data separately. Lower values for these two criteria indicate better results.

Error rate: Error rate is the percentage of the wrong prediction of categorical data. If the output of the method is different from the ground truth, it is a wrong prediction. This evaluation criterion reflects the prediction capability of a method using categorical data.

MNAD: MNAD is used to evaluate the closeness between the prediction output and the ground truths of continuous data. Given that the values of different entries share different scales, the absolute distance between the prediction output $\boldsymbol{o}_i$ and ground truth $\boldsymbol{t}_i$ is normalized by the mean square error of the entry by a given entry $\boldsymbol{e}_i$.

$$\text{MNAD} = \frac{1}{|M|}\sum_{i=1}^{|\boldsymbol{M}|}\frac{|\boldsymbol{t}_i - \boldsymbol{o}_i|}{\sqrt{(v_{i1} - \tilde{v_i})^2 + \cdots + (v_{iK} - \tilde{v_i})^2}} \tag{12}$$

4.3 Baseline methods

We compare the following baseline methods with our methods.

Mean: The mean method, which is used on continuous data, averages all values of the same entry as the prediction.

Median:: The median method, which is used on continuous data, finds the median value of all values of the same entry as the prediction.

Gaussian Truth Model[29]: The Gaussian Truth Model (GTM) is a Bayesian probabilistic method that works only with continuous data. Note that this method only uses continuous data to learn the model and make a prediction. Insufficient data may cause GTM to perform poorly than other methods.

Voting: This method takes the value with the highest occurrence as the predicted value.

Investment[30]: In this approach, a source uniformly "invests" its reliability in the observations it provides and collects credits back from the credibility of the observations.

PooledInvestment[30]: Unlike the Investment method, PooledInvestment linearly scales the credibility of observations.

2-Estimates[4]: This method was proposed on the basis of the assumption that "one and only one true value exists for each entry". If a source provides an observation for an entry, 2-Estimates assumes that this source votes against different observations for this entry.

3-Estimates[4]: 3-Estimates improves 2-Estimates by considering the difficulty of obtaining the truth for each entry. The estimation of truth will affect the source weight.

TruthFinder[2]: TruthFinder adopts Bayesian analysis, wherein the confidence of each observation is calculated as the product of its provided reliability degrees. The similarity function is used to adjust the vote of a value by considering the influences between facts.

AccuSim[3]: AccuSim also applies Bayesian analysis and adopts the usage of the similarity function. Meanwhile, it considers the complement vote which is adopted by 2-Estimates and 3-Estimates.

CRH[15]: The method is the current state-of-the-art method for datasets with categorical and continuous data. This method performs truth discovery through iteratively computing the reliability of sources.

Bi-LSTM: Bidirectional LSTM is a variant of LSTM and has been validated as an effective or even the state-of-the-art method for NLP tasks.

4.4 Truth discovery experiment

We compare the effectiveness of the FFMN and FBMN models with that of the baseline methods. The experimental results shown in Table 2 verify the effectiveness of our models.

From the results, we can see that our memory-network-based models outperform the previous methods. On Stock Dataset, FFMN has the best prediction capacity for categorical data and the lowest error rate. The LSTM-based models, Bi-LSTM and FBMN, perform better with continuous data and have the lowest MNAD. On Flight Dataset, FFMN performs best with categorical and continuous data. CRH has a similar framework as our methods. The results of the two methods verify the effectiveness of using NN-based models to resolve the truth discovery problem.

Comparing LSTM, Bi-LSTM, and FBMN shows that feedback NN with the memory mechanism can provide improved performance. On the stock data, FBMN yields the best result for categorical and continuous data. This excellent performance verifies the effectiveness of the memory mechanism.

Comparing the performances of the NN-based methods reveals that the memory-network-based models obtain the best results. FBMN has good MNAD results on continuous data. Despite its simple architecture, FFMN outperforms models that are more complex.

4.5 Experiment on different data scales

In reality, the scales of the available data may not be as large as the above benchmark datasets. We want to

Table 2 Classifier performance for cross-domain test data.

Method	Stock Dataset		Flight Dataset	
	Error rate	MNAD	Error rate	MNAD
Previous non NN results				
Mean	NA	7.1858	NA	8.2894
Median	NA	3.9334	NA	7.8471
GTM	NA	3.4253	NA	7.6703
Voting	0.0817	NA	0.0859	NA
Investment	0.0983	2.8081	0.0919	6.4153
PooledInvestment	0.0990	2.7940	0.0925	5.8562
2-Estimates	0.0726	2.8509	0.0885	7.4347
3-Estimates	0.0818	2.7749	0.0881	7.1983
TruthFinder	0.1194	2.7140	0.0950	8.1351
AccuSim	0.0726	2.8503	0.0881	7.3204
CRH	0.0700	2.6445	0.0823	4.8613
NN results				
LSTM	0.0884	2.4742	0.0013	1.8111
Bi-LSTM	0.0737	**1.4211**	0.0170	1.7657
Proposed method results				
FFMN	**0.0207**	1.5105	**0.0008**	**1.2600**
FBMN	0.0644	**1.4211**	0.0038	1.7711

test the learning capability of the model for different data scales. We sample data from benchmark datasets from 10% to 100% for model learning. Given that our model is an unsupervised method, the model learns on the basis of observations and tests its performance on ground truths. Ground truths are a small portion of observations with labels that belong to whole observations. We need to ensure the fairness of the comparison. Thus, the sampling datasets all contain the ground truths. We use the FBMN model and perform the experiment using the flight dataset. The results are shown in Table 3.

We can see that our model is not highly affected by the data scales. The sampling dataset has less noise than the original dataset because it also contains the ground truths. When using 50% data for model learning, we obtain the best result with 0.0021 error rate and 1.7657 MNAD.

4.6 Parameter setting

We conducted a series of experiments to analyze the effect of parameters on the results of truth discovery. We analyzed the following parameters, embeddings learned by different algorithms, different embedding lengths, and different learning rates.

Embedding. In the truth discovery task, the observation with a high frequency for an object empirically has a high probability to be correct. Embedding algorithms can learn the closeness between objects and observations by regarding objects and observations as the context for each another. Multiple observations of the same object have relationships. Embedding algorithms can possibly learn the latent law of the structure and interaction among data.

We learn the word embedding of data to vectorize every piece of information from the dataset. Three types of data exist: objects, properties, and values. We use Word2Vec[24] to obtain the vector of each data. Specifically, we take a source and its observation of an entry (i.e., object, property, and value) context terms in learning word embedding. We believe that the relevance between data can be learned on the basis of the context-based, word-embedding learning algorithm. We adopt the LINE[31] word embedding learning algorithm to compare with Word2Vec[24] in the truth discovery experiment. LINE can learn embedding from network-structure information which preserves first-order proximity and second-order proximity. We use embedding in the FFMN model and run experiments on flight data. The results in Table 4 show that embedding learned by Word2Vec is more suitable for our problem. We also can see that LINE_second order performs better than LINE_first order.

Embedding Length. We try multiple numbers of dimensionality from 50 to 300. We find that dimensionality negligibly affects the results of the models. Thus, embedding length is set to 50 in the final experiment.

Learning Rate. We try learning rates of 0.0003 to 0.3. We find that learning rate has limited effects on the results of the models. Thus, we set learning rate as 0.3. The results are shown in Fig. 3.

Table 3 Results of models learning with different data scales.

Scale of sampling (%)	Error rate	MNAD
10	0.0125	1.8111
20	0.0021	1.7711
30	0.0125	1.8111
40	0.0021	1.7711
50	0.0021	1.7657
60	0.0021	1.7711
70	0.0035	1.8111
80	0.0030	1.8111
90	0.0030	1.7711
100	0.0038	1.7711

Table 4 Effects of different embeddings on truth discovery.

	Error rate	MNAD
Word2Vec	**0.0008**	**1.2600**
LINE_first order	0.0017	1.7824
LINE_second order	**0.0008**	1.9463

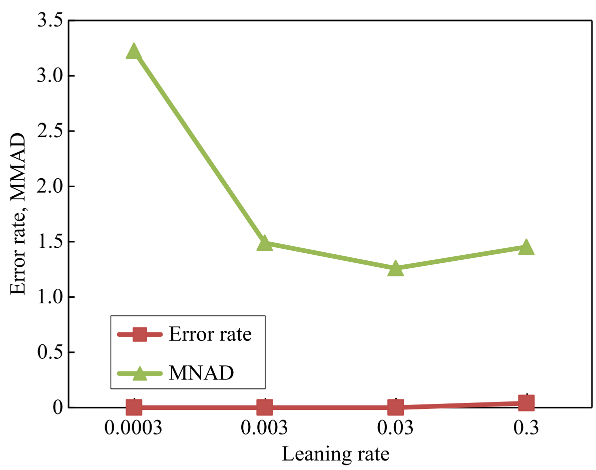

Fig. 3 Results under different values of learning rates.

5 Related Work

Truth discovery involves finding the most credible statement. Most methods for truth discovery utilize voting and similarity mechanisms. The more popular a statement is, the more likely it is to be true. Similar statements have similar credibility. The statements from sources with similar reliability have similar credibility. The existing methods account for some important features, such as the reliability of sources, number of sources which post the same statements, difficulty of the statement, uncertainty in information extraction[30], and similarity between statements and copying relationship. Li et al.[28] categorized the previous methods for truth discovery and found that the most basic method uses the voting strategy. Web-link-based methods share a similar strategy for computing the credibility of statements on the basis of links between statements and sources. Corresponding methods are HUB[32], AvgLog[14], Investment[14], and PoolInvestment[14]. Specifically, the Investment method works on the principle that the source uniformly "invests" its reliability in its statements. The credibilities of statements are computed on the basis of the assessed reliability of sources. The PoolInvestment method adds linear scaling to each entry through computing the credibility of statements. IR-based methods are inspired by the similarity computing approach in information retrieval. Given a value of the object, credibility is computed on the basis of the supporting and opposing sources. Corresponding methods are Cosine[4], 2-Estimates[4], and 3-Estimates[4]. Specifically, the 3-Estimates method accounts for the likelihood of correctness by voting on the value.

Bayesian-based truth discovery methods apply Bayesian analysis to predict the probability of a statement as truthful based on observed information. The corresponding methods include TruthFinder[2], AccuPr[3], AccuSim[3], AccuFormat[3], LCA[33], and CRH[15]. Specially, the TruthFinder method considers similarity between statements, and the AccuPr method considers that different statements on the same entry should be disjointed. The LCA method is a probabilistic model that analyzes latent credibility factors by using them as parameters to find the maximum *a posteriori*. To estimate the reliability of sources and predict truth, the CRH method[15] uses heterogeneous datasets that consist of categorical and continuous data. Copying affected methods, such as AccuCopy, discount votes from copied observations in computing credibility[3].

6 Conclusion

Truth discovery is a fundamental research problem in natural language processing and data mining. Previous approaches to truth discovery have mostly treated the reliability of sources as a real number and have not yet used NNs. Our proposed model vectorizes statements and the reliability of sources and uses memory-network-based models for truth discovery. Specifically, the proposed model adopts the memory mechanism to learn the reliability of sources and incorporate it to represent the credibility of statements. We utilize two types of data and account for their different contributions to truth discovery by automatically assigning weights in the loss function. Experiments with two benchmark datasets show that our methods considerably outperform the state-of-the-art method. The FFMN has the best performance.

Acknowledgment

This work was supported by the National High-Tech Development (863) Program of China (No. 2015AA015407) and the National Natural Science Foundation of China (Nos. 61632011 and 61370164).

References

[1] W. Han X. Zhu, Z. Zhu, W. Chen, W. Zheng, and J. Lu, A comparative analysis on Weibo and Twitter, *Tsinghua Sci. Technol.*, vol. 21, no. 1, pp. 1–16, 2016.

[2] X. Yin, J. Han, and P. S. Yu, Truth discovery with multiple conflicting information providers on the web, *IEEE Transactions on Knowledge & Data Engineering*, vol. 20, no. 6, pp. 796–808, 2008.

[3] X. L. Dong, L. Berti-Equille, and D. Srivastava, Integrating conflicting data: The role of source dependence, *Proceedings of the Vldb Endowment*, vol. 2, no. 1, pp. 550–561, 2009.

[4] A. Galland, S. Abiteboul, A. Marian, and P. Senellart, Corroborating information from disagreeing views, in *ACM International Conference on Web Search & Data Mining*, 2010, pp. 131–140.

[5] R. S. Poston and C. Speier, Effective use of knowledge management systems: A process model of content ratings and credibility indicators, *Mis Quarterly*, vol. 29, no. 2, pp. 221–244, 2005.

[6] P. Banerjee and H. Han, Answer credibility: A language modeling approach to answer validation, in *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings*, 2009, pp. 157–160.

[7] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, Web credibility: Features exploration and credibility prediction, in *European Conference on Advances in Information Retrieval*, 2013, pp. 557–568.

[8] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng, Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks, in *EMNLP'08 Proceedings of the Conference on Empriical Methods in Natural Language Processing*, 2008.

[9] X. L. Dong, B. Saha, and D. Srivastava, Less is more: Selecting sources wisely for integration, *Proceedings of the VLDB Endowment*, vol. 6, no. 2, pp. 37–48, 2012.

[10] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, On truth discovery in social sensing: A maximum likelihood estimation approach, in *ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2012, pp. 233–244.

[11] B. Zhao, B. I. P. Rubinstein, J. Gemmell, and J. Han, A bayesian approach to discovering truth from conflicting sources for data integration, *Proceedings of the Vldb Endowment*, vol. 5, no. 6, pp. 550–561, 2012.

[12] Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han, A confidence-aware approach for truth discovery on long-tail data, *Proceedings of the Vldb Endowment*, vol. 8, no. 4, pp. 425–436, 2014.

[13] Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han, A survey on truth discovery, *ACM Sigkdd Explorations Newsletter*, vol. 17, no. 2, pp. 1–16, 2016.

[14] J. Pasternack and R. Dan, Knowing what to believe (when you already know something), in *COLING'10 Proceedings of the 23rd, International Conference on Computational Linguistics*, Beijing, China, 2010, pp. 877–885.

[15] 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 *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, 2014, pp. 1187–1198.

[16] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch, *The Journal of Machine Learning Research*, vol. 12, pp. 2493–2537, 2011.

[17] J. Turian, L. Ratinov, and Y. Bengio, Word representations: A simple and general method for semi-supervised learning, in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 384–394.

[18] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, Parsing with compositional vector grammars, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013, pp. 455–465.

[19] A. Mnih and G. E. Hinton, A scalable hierarchical distributed language model, in *Advances in Neural Information Processing Systems*, 2009, pp. 1081–1088.

[20] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, A neural probabilistic language model, *The Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.

[21] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013, p. 1642.

[22] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, Learning sentiment-specific word embedding for twitter sentiment classification, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014, pp. 1555–1565.

[23] P. Wang, B. Xu, J. Xu, G. Tian, C. L. Liu, and H. Hao, Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification, *Neurocomputing*, vol. 174, pp. 806–814, 2016.

[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv: 1301.3781, 2013.

[25] J. Weston, S. Chopra, and A. Bordes, Memory networks, arXiv preprint arXiv: 1410.3916, 2014.

[26] S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus, End-to-end memory networks, in *Advances in Neural Information Processing System 28 (NIPS 2015)*, 2015.

[27] D. Tang, B. Qin, and T. Liu, Aspect level sentiment classification with deep memory network, arXiv preprint arXiv: 1605.08900, 2016.

[28] X. Li, X. L. Dong, K. Lyons, W. Meng, and D. Srivastava, Truth finding on the deep web: Is the problem solved? *Proceedings of the Vldb Endowment*, vol. 6, no. 2, pp. 97–108, 2012.

[29] B. Zhao and J. Han, A probabilistic model for estimating real-valued truth from conflicting sources, in *Proc.of Intl. Workshop on Quality in Databases*, 2012.

[30] J. Pasternack and D. Roth, Making better informed trust decisions with generalized fact-finding, in *International Joint Conference on Artificial Intelligence*, 2011.

[31] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, Line: Large-scale information network embedding, in *International Conference on World Wide Web*, 2015.

[32] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[33] J. Pasternack and R. Dan, Latent credibility analysis, in *International Conference on World Wide Web*, 2013, pp. 1009–1020.

Luyang Li received PhD degree from Harbin Institute of Technology in 2017. She is currently a lecturer at Department of Computer Science and Technology, Beijing Foreign Studies University. Her current research interests are truth discovery, fake review detection, and deep learning.

Wenjing Ren received the master degree from Harbin Institute of Technology in 2017. She is currenthy a programmer in Netease company. Her current research interests are truth discovery, fake review detection, and deep learning.

Bing Qin received the PhD degree in computer science from Harbin Institute of Technology in 2005. She is currently a professor at Department of Computer Science and Technology, Harbin Institute of Technology. Her current research interests are information extraction, sentiment classification, textual entailment, and machine learning.

Ting Liu received the PhD degree in computer science from Harbin Institute of Technology in 1998. He is currently a professor at Department of Computer Science and Technology, Harbin Institute of Technology. His current research interests are social computing, information extraction, question & answer, and machine learning.