Understanding Electricity-Theft Behavior via Multi-Source Data

Wenjie Hu
Zhejiang University

Yang Yang*
Zhejiang University

Jianbo Wang
State Grid Power Supply Co. Ltd.

Xuanwen Huang
Zhejiang University

Ziqiang Cheng
Zhejiang University

ABSTRACT

Electricity theft, the behavior that involves users conducting illegal operations on electrical meters to avoid individual electricity bills, is a common phenomenon in the developing countries. Considering its harmfulness to both power grids and the public, several mechanized methods have been developed to automatically recognize electricity-theft behaviors. However, these methods, which mainly assess users' electricity usage records, can be insufficient due to the diversity of theft tactics and the irregularity of user behaviors.

In this paper, we propose to recognize electricity-theft behavior via multi-source data. In addition to users' electricity usage records, we analyze user behaviors by means of regional factors (non-technical loss) and climatic factors (temperature) in the corresponding transformer area. By conducting analytical experiments, we unearth several interesting patterns: for instance, electricity thieves are likely to consume much more electrical power than normal users, especially under extremely high or low temperatures. Motivated by these empirical observations, we further design a novel hierarchical framework for identifying electricity thieves. Experimental results based on a real-world dataset demonstrate that our proposed model can achieve the best performance in electricity-theft detection (e.g., at least +3.0% in terms of F0.5) compared with several baselines. Last but not least, our work has been applied by the State Grid of China and used to successfully catch electricity thieves in Hangzhou with a precision of *15%* (an improvement from 0% attained by several other models the company employed) during monthly on-site investigation.

CCS CONCEPTS

• **Applied computing → Law, social and behavioral sciences**.

KEYWORDS

User modeling, electricity-theft detection, hierarchical recurrent neural network, power grids

*Corresponding author: Yang Yang, yangya@zju.edu.cn
[1]The source code is published at https://github.com/zjunet/HEBR.
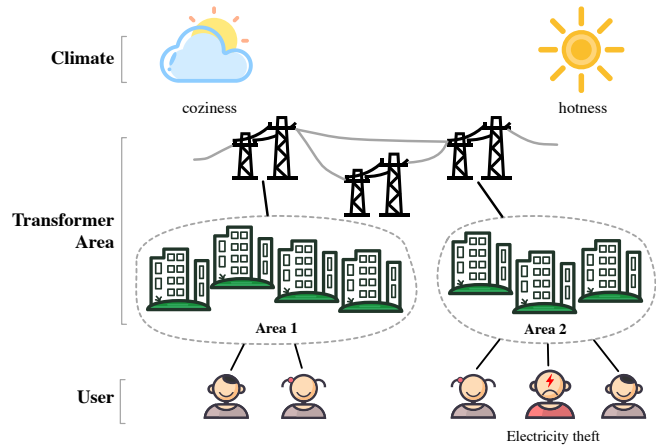
Figure 1: An example of electricity theft. The power is supplied to different transformer areas (area 1, area 2). The climate in area 2 is hotter than that in area 1; thus, the power consumption is higher. In order to avoid high electricity bills, several households in area 2 try to pilfer the electricity.

Taiwan, China. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3366423.3380291

1 INTRODUCTION

Electrical power is an important national energy resource [4]. One of the barriers to the stable provision of electrical power is *electricity theft*. Formally speaking, electricity theft refers to the illegal operations by which users unauthorizedly tamper the electricity meter or wires to reduce or avoid consumption costs. Electricity theft not only results in unbearable economic losses to the power suppliers, but also endangers the safety of electricity users, the electrical systems, and even the public at large. As reported in [31][1], electricity theft and other so-called "non-technical loss" result in a staggering $96 billion in losses globally per year; moreover, a shocking statistic is that in 2012, GDP in India was reported to drop 1.5% as a result of electricity theft, while Uttar Pradesh, the most populous state in India, lost 36% of its total electric power to theft of this kind [30].

Great efforts have been made to detect and prevent electricity theft. The most intuitive way of doing this is to utilize hardware-driven methods: in short, to find out how the thieves are pilfering the power, then design and upgrade the meter structures accordingly [11, 15, 19]. For instance, Guo et al. [19] surveys and summarizes the most commonly used electricity pilfering methods, which include changing the structure or wiring mode of the meter; then some countermeasures associated with the electrical meters are

[1]Source report is conducted by Northeast Group LLC.

proposed, including installing a centralized or fully closed meter box for residents. However, there are three major drawbacks of these hardware-driven methods: 1) they require expert domain knowledge to specify the techniques that the electricity thieves use; 2) it is difficult to design a general meter structure, as different regions may often have different pilfering tactics; 3) these methods lose their effectiveness once thieves change their tactics.

To address these problems, data-driven methodologies have been applied to the task of electricity-theft detection. Before reviewing the existing works on this topic, it is necessary to explain one associated technical term : *non-technical loss* (NTL) [5]. In practice, losses in a utility distribution grid are classified as technical and non-technical. Technical loss is mainly caused by unwanted effects (e.g., heating of resistive components, radiation, etc.), and is unavoidable[2]. By contrast, non-technical loss is defined as the energy that is distributed but not billed; in other words, this type of loss is caused by issues in the meter-to-cash processes. Although disparate issues may contribute to non-technical losses, a large proportion of the reasons that cause NTL are related to the electrical power pilfering and frauds [14]. It is therefore straightforward to detect electricity theft from abnormal NTL records [1, 35]. NTL provides transformer-area information. To capture individual patterns, many existing work have utilized electricity usage records or NTL as input [12, 20, 21, 25, 29, 39], and applied various machine learning techniques (e.g., SVM, CNNs, RNNs) to identify electricity theft. However, most of these methods have failed to obtain good performance, due to the diversity and irregularity of electricity usage behavior, which is almost impossible to fully understand either using NTL or electricity usage records alone.

Motivated by the abovementioned concerns, in this paper, we propose to recognize electricity-theft behavior by bridging three distinct levels of information: micro, meso, and macro. At the micro- and the meso-level, we seek to capture users' abnormal behavior from electricity usage records and NTL respectively. At the macro-level, moreover, we creatively study how climatic conditions influence electricity-theft behavior. We then effectively integrate these three levels of information into a uniform framework. Our work achieves significant progress in the real-world application of electricity-theft detection by deploying the proposed model in State Grid of China[3] and improving the theft detection performance

To be more specific, we employ a multi-source dataset, which contains two years' worth of daily electrical power consumption records from 311K users in Zhejiang province, China during June 2017 to April 2019, along with the corresponding daily NTL records and climatic condition data in all transformer areas covered by those users. In addition, we have 4,501 (1.45%) electricity-theft labels representing 4,626 cases of electricity theft (note that a single user may pilfer electrical power several times within the relevant timespan), all of which were confirmed during several large-scale on-site investigations conducted by State Grid staff over the two years in question. Our empirical observations find that: 1) at the macro-level, climatic condition (or more specifically, *temperature*, which is our main focus on in this paper) affects users' electricity consumption to some extent, while users belonging to different groups may show

different levels of correlations between electricity usage and temperature; 2) as previous works have pointed out [1, 35], NTL is an important factor in detecting electricity thieves at the meso-level; our detailed analysis shows that abnormal NTL patterns in the transformer areas are a strong signal on electricity theft; 3) at the micro-level, since electrical power consumption fluctuates with time, we can clearly see the temporal relationships within the individual electricity usage, such that the unusual patterns during a specific period may indicate abnormal user behavior. These findings are hierarchically illustrated in Figure 1, which indicates that users consume more electric power as the temperature gradually increase (due to e.g. the usage of air conditioners). In order to avoid high electricity bills, some users may employ some tactics to pilfer electricity, which leads to some abnormal fluctuations in their electricity-usage sequences as recorded by the smart meters. At the same time, the NTL of the corresponding transformer area can reveal anomalies related to abnormal electricity usage.

Despite the interesting insights provided by our empirical observation, the question of how to integrate multi-source information into a uniform framework remains a challenging one. To capture the temporal and spatial correlations of multi-source sequences, one straightforward method involves first concatenating them at each temporal point, then adopting a single latent representation to capture the overall patterns, such as *multiscale recurrent neural networks* (MPNN) [18]. However, concatenation from different sources widens the feature dimensions, which precludes capturing the significant influences of macro- or meso-level information on micro-level information (Section 3). Therefore, we propose a hierarchical framework, named Hierarchical Electricity-theft Behavior Recognition (HEBR), to extract features and fuse them step by step (Section 4). Experimental results of the electricity-theft detection task on a real-world dataset demonstrate the effectiveness of the proposed HEBR method (Section 5).

Most excitingly, HEBR has been employed for real-time electricity-theft detection in the State Grid of China. During the monthly on-site investigation in August of 2019, we successfully caught electricity thieves in Hangzhou, China, and punished them instantly by checking the suspected samples predicted by HEBR, representing an improvement in precision from **0%** to **15%** (Section 5.4). The success of the proposed model in this real-world application further illustrate its validity. Accordingly, the contributions of this paper can be summarized as follows:

- We analyze electricity-theft behaviors from three distinct levels of information based on multi-source data.
- Based on observational studies, we propose the Hierarchical Electricity-theft Behavior Recognition (HEBR) model, which identifies electricity theft by fusing different levels of information, and validate its effectiveness using a real-world dataset.
- We apply HEBR to catch electricity thieves on-site and achieve significant performance improvements.

2 PROBLEM DEFINITION

Let $\mathbf{U}$ be a set of users, while $\mathbf{A}$ is the set of corresponding transformer areas; that is, each user $u \in \mathbf{U}$ belongs to a specific transformer area $a \in \mathbf{A}$ based on the regional location. An area a usually

---

contains hundreds of users. Each user u has electricity usage records with T observations within a certain timespan, which we refer to as $\mathcal{X}_u^e = \{\mathbf{x}_1^e, ..., \mathbf{x}_T^e\}$. Each area $a \in \mathbf{A}$ has the NTL records, denoted as $\mathcal{X}_a^l = \{\mathbf{x}_1^l, ..., \mathbf{x}_T^l\}$, and the observation sequence of climatic conditions, written as $\mathcal{X}_a^c = \{\mathbf{x}_1^c, ..., \mathbf{x}_T^c\}$. The denotation $\mathbf{x}_t^* \in \mathbb{R}^{d_*}$ represents different quotas with various dimensions (see details in Section 3.1). In light of the above, we can define the problem addressed in this paper as follows:

*Definition 2.1. **Electricity-theft detection.*** Given a specific user u who belongs to the transformer area a, the goal is to estimate $\mathbf{P}(\mathcal{Y}|\mathcal{X}_u^e, \mathcal{X}_a^l, \mathcal{X}_a^c)$, which is the probability that u pilfers electrical power ($\mathcal{Y} = 1$) or not ($\mathcal{Y} = 0$).

In the following sections, we will conduct observational studies on the multi-source sequences to obtain potential insights. Based on these intuitions, we then propose a novel hierarchical framework for recognizing electricity-theft behavior.

3 EMPIRICAL OBSERVATIONS

Although existing data-driven electricity-pilfering-detection methods seek to capture the characteristics of users' electrical power consumption, it is rather difficult in real-world applications to efficiently catch electricity thieves if we only observe the user's electrical power consumption records; this is due to the diversity, complexity and irregularity of electricity-theft behavior. Therefore, we compile a multi-source dataset, with additional NTL and temperature records for each transformer area, in order to analyze the traces of users' electricity usage. The associated observational findings and insights are presented in this section.
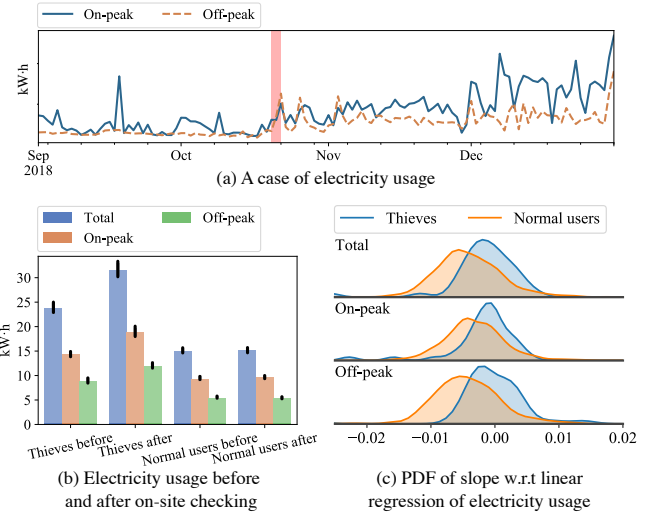
3.1 Dataset Description

Our dataset comprises three parts: the two sets of electricity-related records are provided by State Grid Zhejiang Power Supply Co. Ltd.[4], while the temperature records are collected from the official weather website. The overall statistics are summarized in Table 1.

Electricity Usage Records. This dataset covers the daily electrical power consumption records of **310,786** users in total, ranging from June 2017 to April 2019. For each user, we have the total, on-peak and off-peak electricity usage (kW·h) records for each day within the relevant timespan.

Non-Technical Loss Records. This dataset contains the daily meso-level electrical records from a total of **3,908** transformer areas, covering all of these 311K users, and has the same time range as the usage dataset. More specifically, for each area, the daily amount of electrical power (kW·h) lost due to non-technical loss (NTL) is recorded.

Temperature Records. We obtained the temperature records for all prefecture-level cities in Zhejiang during the same timespan as above from the Weather Radar[5]. For each city, these records contain the maximum and minimum temperatures (℃) for each day.

Labels of Electricity Thieves. Among all users, **4,501 (1.45%)** were confirmed by State Grid staff to be *electricity thieves* during their on-site investigations; it should be noted that there are a

(a) A case of electricity usage

(b) Electricity usage before and after on-site checking

(c) PDF of slope w.r.t linear regression of electricity usage

Figure 2: A study of micro-level factors and electricity theft. (a) presents a case of electricity usage with the red bar indicating the time at which the thief was caught. (b) presents the statistics of all users pertaining to electricity usage before and after on-site checking, while the trends of the usage are shown in (c). We can therefore see that the behaviors of electricity thieves are different from those of normal users.

total of **4,626** electricity-pilfering cases, since a single user may be caught committing theft several times during the two years. We then regard all remaining users (98.55%) as normal, i.e., as having engaged in no pilfering behavior over the entire timespan. While it is possible that a few users who adopt subtle ways to pilfer electricity, which are not caught; these cases bring in the noise but are very rare. We take these confirmed cases as ground-truth and collect the timestamps when electricity thieves were caught for detailed analysis and experiments.

3.2 Micro-level: User Observation

We first examine the micro-level electricity usage records to see whether abnormal patterns or characteristics in the user behavior exist that indicate electricity theft.

To better understand pilfering behaviors, we first have to recognize the time periods during which the thieves were pilfering electricity. According to the timestamps of on-site investigations, the records for each electricity theft can be divided into two parts: before and after being caught. For convenience we refer to the specific timestamp as a *checkpoint*. We further assume that a thief was stealing the power during the last 30 days before the checkpoint, then returning back to a normal state in the following 30 days after

Table 1: Overall statistics of the datasets.

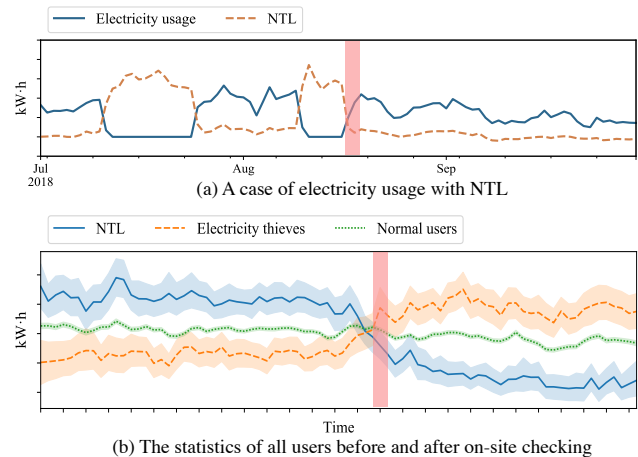| Metric	Statistics
#(users)	310,786
#(transformer areas)	3,908
#(electricity thieves)	4,501
#(pilfering-cases)	4,626
#(prefecture-level cities)	11

the checkpoint. This is based on the domain knowledge that we have chosen a month as the timespan, since an electricity thief often steals power for a long time (probably for a period longer than one month), and once caught and punished, he/she would instantly cease engaging in pilfering behavior and behave normally. To give a concrete example, Figure 2a presents a case of an electricity thief who was caught in the middle of October 2018 (red bar). We can see an abnormal pattern in the middle of September, showing that he had a sudden peak of electricity usage, while he had almost no off-peak electrical power usage during Sep. and Oct. Moreover, in this case, there is no significant decreasing trend of electricity usage before he was caught; in practice, however, users would probably use less electricity in Sep. and Oct. due to the temperature drop. These clues may revel the thief's abnormal state or behavior.

In light of the above assumption and case study, we provide an overview of the electricity consumption of both electricity thieves and normal users in Figure 2b, 2c. We draw the distributions of daily power usage at different times (before and after being caught) in Figure 2b; for each group of settings, we show the average value along with the standard error bar. We can clearly see that for electricity thieves, there is a significant increment of electricity usage after being caught compared with before (the two histograms on the left-hand side). This is reasonable, since pilfering behavior would reduce the electricity that they actually consume. An opposite trend is exhibited by the normal users in that the usage before and after the checkpoint seems unchanged; in other words, the characteristics of the electrical power consumption behaviors of normal users are relatively stable.

Another obvious observation from Figure 2b is that electricity thieves exhibit a much higher level of electricity usage compared with normal users, although pilfering behavior would cut down the recorded value of electrical power consumption. One reasonable explanation for this phenomenon comes from the idea that people typically engage in risky behavior only when they expect a higher benefit. As for this scenario, users whose electricity consumption is high are expected to have a far stronger motivation to steal power, as they would reduce their costs significantly through engaging in pilfering behavior. By contrast, if a user uses a very small amount of electricity, there is no need for him/her to undertake such illegal operations since this behavior would also be economically foolish, as once being caught, the user would be fined vast amounts.

To further illustrate the differences in the stability of electricity usage between electricity thieves and normal users, we conduct a linear regression on electricity usage during the three months before the theft was identified. For better visualization and explanation, we set a restriction on the checkpoint: in short we only sample the thieves who were caught during October. Under these settings, user are expected to reduce electricity usage stably at the end of Aug. and Sep. compared with Jul. due to the drop in temperature, such that the linear regression may be able to capture this decreasing trend. We use a probability density function (PDF) [32] to compare the distributions of slopes in the linear regression. As Figure 2c shows, the gap of PDF curves between electricity thieves and normal users represent the different trends in electricity usage. More specifically, and in line with our commonsense expectations, the usage of the majority of normal users (80.30%) exhibits a negative slope; by contrast, for electricity thieves, the peak of the PDF

(a) A case of electricity usage with NTL

(b) The statistics of all users before and after on-site checking

Figure 3: The study of meso-level factors to electricity theft. (a) presents an electricity-usage case with NTL and the red bar indicates the time of being caught as thief. (b) presents the statistics of all users and transformer areas. Both of them illustrate that the abnormal increment of NTL can be caused by pilfering electricity.

lies around 0, while nearly half of them even have a positive slope, which is entirely opposite to the common cases.

Combined with the analysis mentioned above, we can conclude that electricity thieves have a higher level of electrical power consumption with less stability.

3.3 Meso-level: Area Observation

Although we can observe several differences between electricity thieves and normal users in terms of electricity usage, due to the fact that user behaviors are very complicated and may be affected by many factors, focusing only on those micro-level differences prevents us from efficiently identifying electricity thieves. Here, another intuitive element of knowledge is that the transformer area records the overall electricity consumption of different users, i.e., non-technical loss (NTL). Hence, in order to reveal the meso-level characteristics of electricity thieves, we conduct an analysis combining regional information with individual electricity usage.

We begin with a real-world case study of an electricity theft in Figure 3a. The user was caught as a thief in the middle of August (red bar). A clear and interesting observation is that the trend of correlations between the user's electricity usage and NTL before he was caught is completely opposite to that of the period after on-site checking was conducted. More specifically, before the user was caught, the less electricity usage his records showed (probably a large part of the power used was stolen), the higher the NTL of the transformer area would be; after on-site checking, however, the NTL of the area became relatively stable.

We can confirm this finding on the whole dataset by observing the correlations between the averaged NTL of all transformer areas and the individual electricity usage of all users during the period before and after on-site checking (Figure 3b). In more detail, we move the sliding window to retrieve the electricity usage of each thief: 1) 50 days before, and 2) 30 days after the checkpoint. We then sample the usage records of all normal users in the same transformer area and during the same period as the thief. Again, we can see
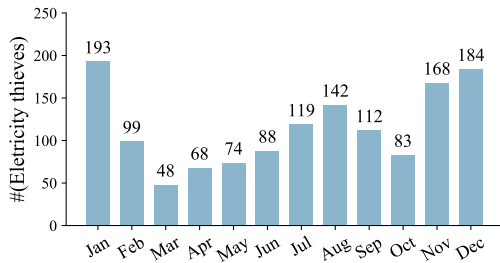
Figure 4: Number of electricity thieves in each month.

that the electricity usage of normal users (green line) are rather stable during the observed timespan; for electricity thieves (orange line), however, their averaged power consumption significantly increased once they were caught, while the NTL of their transformer areas dropped accordingly. It gives us the clue that the NTL in the transformer area may be a signal for indicating whether there are electricity thieves in this area, and additionally, that capturing such correlations could be helpful in electricity-theft detection.

3.4 Macro-level: Climate Observation

Previous work [23] has demonstrated the influence of climatic conditions on user behavior (taxi ordering). It would be interesting to determine whether some relationship between climate and electricity usage exists in our scenario, or whether the macro-level factors affect users' electrical power usage in a non-trivial way. Accordingly, we present the statistics of the seasonal effects on electricity-pilfering behavior (Figure 4), which reveal that most pilfering cases were caught in winter (Dec., Jan.) and summer (Jul., Aug.). This leads to the straightforward conclusion that the climatic conditions would influence the user behavior of electricity usage, especially for the electricity thieves. Inspired by the practical experiences, temperature is the only climatic variable we consider in the present research, since the most obvious difference between winter and summer is the temperature factor, and it is believed that temperature will influence users' daily behavior in an intuitive way. We use the averaged value of maximum and minimum in a day to represent the daily temperature; this setting is used throughout the entire paper unless otherwise indicated.

We first illustrate that a correlation does indeed exist between temperature and users' electricity usage, as shown in Figure 5a. We indicate the averaged daily total electricity consumption of all users over one year with an error bar (blue line), while the orange line indicates the temperature each day. Total electricity usage fluctuates with the temperature, as some sharp peaks and valleys are coincident; the most obvious correlations between these two factors is that extremely high or low temperatures are associated with increased electricity consumption. This would appear to be a commonsense observation since extreme temperatures are of course associated with the increased usage of high-powered appliances such as air conditioning or heaters; here we verify this assumption by means of a brief visualization.

We next examine the relationships between temperature and electricity usage among different groups of users. In Figure 5b, we aggregate the daily total electricity usage based on the temperature of that day, then draw a boxplot with respect to our two user groups

(a) The correlations between temperature and total electricity usage

(b) Electricity usage distribution in different temperatures

(c) A case of electricity thieves

(d) A case of normal users

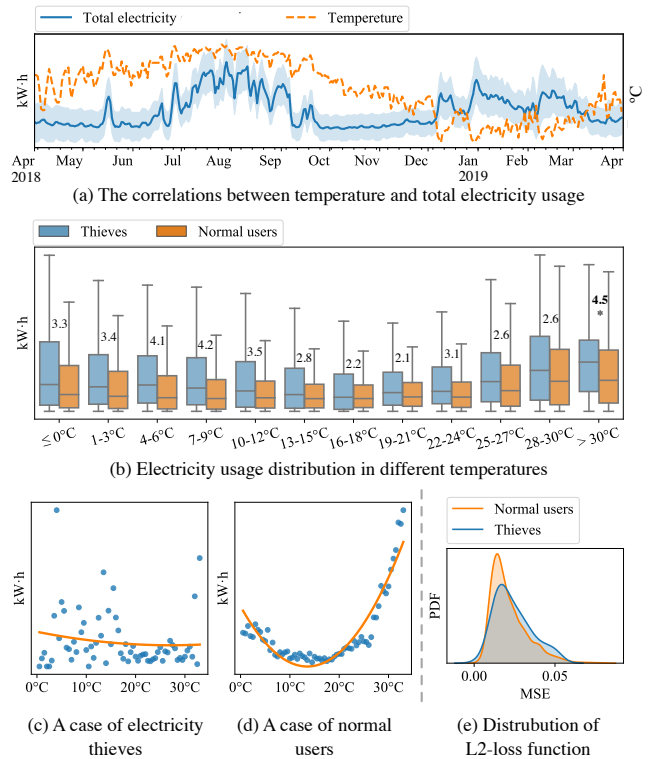(e) Distrubution of L2-loss function

Figure 5: The study of macro-level factors to electricity theft. (a) presents the strong correlations from climatic conditions to user behaviors. (b)-(e) illustrate the electricity-usage irregularity of thieves under different temperature conditions. The floating numbers in (b) indicate the Wasserstein distance of distribution between the thieves and normal users.

of interest, i.e. electricity thieves and normal users. In line with what we have observed in Section 3.2, we again find that electricity thieves have a much higher level of electrical power consumption regardless of the temperature. Furthermore, additional observation reveals the specific influence of temperature on electricity pilfering, as the gap of daily total electricity usage between thieves and normal users is significantly larger under extremely high or low temperatures compared with average temperatures: if we regard a temperature lower than 0℃ or higher than 30℃ as extreme conditions (as is accepted by the public), the average gap of daily total electricity usage between thieves and normal users under extreme conditions is 4.51kW·h, while that for non-extreme conditions (between 0 and 30℃) is 3.03kW·h (a decrease of 32.8%); moreover, if we restrict the temperature span to between 10℃ and 25℃, the average gap becomes 2.70kW·h. We further compute the Wasserstein distance (d_w) [9] of electricity usage distributions between thieves and normal users under different temperatures, and find that distances for extreme temperatures (d_w=4.5 (>30℃)) are much larger than other cases (e.g., d_w=2.1 (19 − 21℃)). This verifies our conclusion from a statistical perspective that electricity thieves are likely to consume much more electrical power than normal users, especially under extremely high or low temperature conditions.

Finally, we take a deeper look into these correlations by means of several case studies. We choose a typical electricity thief along with

a normal user in the same transformer area as examples, then show the pairs of daily total electricity and temperature in the form of 2-D figures. For better visualization and illustration, we also draw the second-order regression curve in both Figure 5c and 5d. Again, the trends of the normal user's curve is consistent with our experience suggesting that very high or low temperatures are associated with elevated electricity consumption, and the points are well fitted to a specific quadratic curve (Figure 5d); by contrast, the scatterplot for the case of the electricity thief seems disordered (Figure 5c). It is better to explore such correlations by merging users together within different groups; however, since the scale of each user's electricity consumption may vary substantially, the scatterplot of the data for all normal users may not well fit to a quadratic curve. An alternative approach would to individually fit the points for each user, then show the gap in the distributions of the fitting loss between normal users and electricity thieves (Figure 5e). The two PDFs simply reveal the fact that scatters of normal users can be fitted more easily to a quadratic curve than thieves.

Summarizing from the abovementioned observations, we can conclude that the analysis based on the combination of electricity usage, NTL and temperature can yield extra information related to electricity-theft behavior, and can also give us strong insights and motivations to consider the correlations among these three different levels of factors when detecting electricity theft.

4 OUR APPROACH

In this section, we integrate the insights gained from empirical observations (Section 3) into a hierarchical framework, named Hierarchical Electricity-theft Behavior Recognition (HEBR), to capture the behavioral patterns from multi-source observation sequences.

Overview. Motivated by Section 3, we model the user behaviors based on three distinct levels of information, as follows:

- *Macro-level.* We define $\mathcal{X}_a^c, a \in \mathbf{A}$, to represent the observation sequence of temperature, which reflects the climatic conditions that may influence users' electricity consumption patterns.
- *Meso-level.* We define $\mathcal{X}_a^l, a \in \mathbf{A}$, to represent the observation sequence of non-technical loss (NTL), which indicates the real-time status of the corresponding transformer area.
- *Micro-level.* We define $\mathcal{X}_u^e, u \in \mathbf{U}$, to represent the observation sequence of users' electricity usage, which presents the trace of individual electrical power consumption behaviors.

The empirical observations demonstrate that the macro- and meso-level information influence the micro-level behaviors to some extent. In order to integrate the multi-source information so as to capture the abnormal behavioral patterns of electricity thieves, intuitively, we develop a hierarchical framework to extract features from their respective different sources, and fuse them step by step.

4.1 Model Description

When modeling the multi-source sequences, a straightforward baseline can be established by first concatenating them at each temporal point, then using a single latent representation to capture the overall patterns, such as *MRNN* [18]. However, concatenation from different sources widens the feature dimensions, which may preclude capturing the significant correlations between different levels
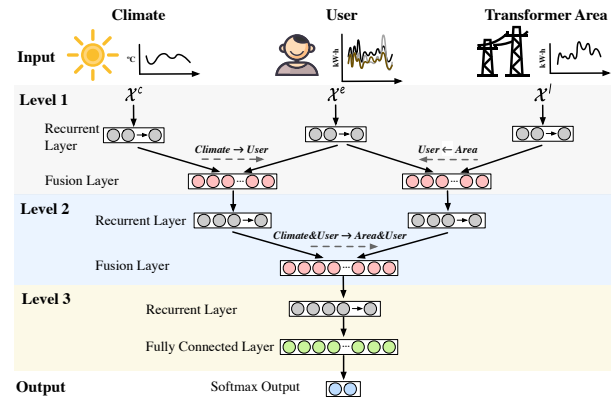
Figure 6: Architecture of HEBR. Given a multi-source observation sequence (climate, area and user), HEBR constructs a three-level framework: at each level, different sequences are inputted into different recurrent layers respectively, the latent representations of which are fused in pairs. Dashed lines indicate the direction of the fusion. The last layer outputs the probability of electricity theft.

of information. More specific, in our scenario, different users in the same transformer area may have the same observation sequence of NTL or temperature; therefore, the straightforward concatenation could result in a confusion of user behaviors from the regional and climatic levels. Accordingly, we opt to better extract the independent features of each source respectively, and then conduct the pairwise information fusion.

In Figure 6, we present the architecture of the proposed HEBR framework. In addition to input and output, HEBR contains three levels of feature extraction and hierarchical fusion, which aim to capture different levels of influence between data sources. Each level is described in more detail below.

- **Level 1:** Captures the temporal patterns in the observation sequence independently (e.g. the patterns in temperature ($\mathbf{h}^c$), NTL ($\mathbf{h}^l$) and user's electricity usage ($\mathbf{h}^e$)), and fuses them in pairs ($\mathbf{h}^c \rightarrow \mathbf{h}^e$, $\mathbf{h}^l \rightarrow \mathbf{h}^e$). It aims to model the influence from macro- or meso-level factors on user behaviors respectively.
- **Level 2:** Captures the temporal patterns after preliminary fusion at Level 1 (e.g. user-climate ($\mathbf{h}^{ec}$) and user-area ($\mathbf{h}^{el}$)) respectively, and then fuses the patterns of $\mathbf{h}^{ec} \rightarrow \mathbf{h}^{el}$. It aims to uniform the influence from macro and meso level on user behaviors.
- **Level 3:** Captures the overall temporal patterns in the multi-source sequence ($\mathbf{h}^{ecl}$). The hierarchically fused information is integrated to capture the behavioral patterns, which can be applied to estimate the probability of electricity theft.

Note that we do not fuse the representations $\mathbf{h}^c$ and $\mathbf{h}^l$ in the first level. This is because our observational studies (Section 3) suggest that these two distinct levels of information are uncorrelated, although they are closely related to user's electricity usage. Here, we aim to capture how these two factors influence user electricity consumption behavior.

Based on the hierarchical construction, HEBR can conduct feature extraction and fusion in each level and gradually integrate the information between multiple sources. As for the operations in each
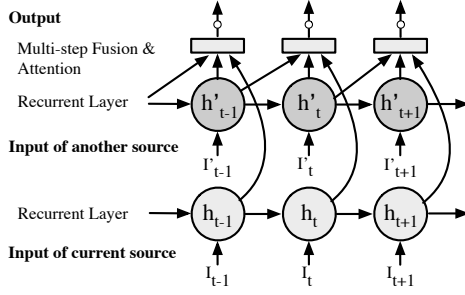
Figure 7: The architecture of recurrent and fusion layers on multi-source sequences. The sequential inputs of another source $\mathbf{I}'$ are integrated into the current ones $\mathbf{I}$ by means of the above architecture, which is a uniform component in Figure 6.

level, we define a uniform formulation: given the sequential input $\mathbf{I}_t^{(k)}$ and another source $\mathbf{I}_t'^{(k)}$ in the k-th level, the feature extraction and hierarchical fusion can be formulated as follows:

$$
\begin{aligned}
\mathbf{h}_t^{(k)} &= \mathcal{F}_{\text{recurrent}}\left(W_{\mathbf{I}\to\mathbf{h}} \cdot \mathbf{I}_t^{(k)}, \mathbf{h}_{t-1}^{(k)}\right), \\
\mathbf{h}_t'^{(k)} &= \mathcal{F}_{\text{recurrent}}\left(W_{\mathbf{I}'\to\mathbf{h}'} \cdot \mathbf{I}_t'^{(k)}, \mathbf{h}_{t-1}'^{(k)}\right), \\
\mathbf{I}_t^{(k+1)} &= \boldsymbol{\alpha}_t \cdot \mathcal{F}_{\text{act}}\left(\mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t^{(k)}, W_{\mathbf{h}'\to\mathbf{h}} \cdot \mathbf{h}_t'^{(k)}\right)\right)
\end{aligned}
\tag{1}
$$

$\mathbf{h}_t^{(k)}$ denotes the latent representation of $\mathbf{I}_t^{(k)}$ in the k-th level at temporal point t, which is updated by function $\mathcal{F}_{\text{recurrent}}$ based on its previous memory $\mathbf{h}_{t-1}^{(k)}$ and current input $\mathbf{I}_t^{(k)}$, while the apostrophe ($'$) denotes the same meaning of the sequences from another source. W_* are the trainable weighted matrices, and the function $\mathcal{F}_{\text{fuse}}$ aims to fuse the information from $\mathbf{I}_t'^{(k)}$ into the current sequence $\mathbf{I}_t^{(k)}$, then outputs the intermediate representations by means of an activation function ($\mathcal{F}_{\text{act}}$). Moreover, $\boldsymbol{\alpha}_t$ denotes the attention coefficient from $\mathbf{h}_t'$ to $\mathbf{h}_t$, which tries to automatically discover the attention weights from $\mathbf{I}'$ to $\mathbf{I}$ based on "end-to-end" learning. We will introduce the details of model inference and learning for detecting electricity thieves in the next section.

4.2 Model Inference and Learning

Herein, we adopt neural networks to implement HEBR, the parameters of which are learned by minimizing some specific loss. As for capturing the temporal patterns, we can apply several existing methods to implement the recurrent layer $\mathcal{F}_{\text{recurrent}}$ (see details in Table 6 of Section 5.3). As for the fusion function $\mathcal{F}_{\text{fuse}}$, we propose a new hierarchical fusion mechanism, containing multi-step fusion and attention operations, to effectively bridge the information from different sources. More details are introduced in the next paragraph.

Hierarchical fusion mechanism. In order to capture the correlations between different levels of information, we propose a multi-step fusion mechanism. The intuition here is that the influence between two distinct levels of sequences may be time-delayed. For instance, in addition to being affected by today's temperature, a user's electricity usage is also probably related to yesterday's weather; one concrete example is that if the previous day was very hot, people will tend to turn on the air conditioning for a longer time, even if today is colder. Therefore, we should try to fuse more

information in the temporal interval rather than just at the current temporal point. More specifically, as shown in Figure 7, the current latent representation $\mathbf{h}_t \in \mathbb{R}^{d_{\mathbf{h}}}$ and that from another data source $\mathbf{h}_t' \in \mathbb{R}^{d_{\mathbf{h}'}}$ are fused via the following formulation:

$$
\mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t, \mathbf{h}_t'\right) = \left(\mathbf{h}_t \odot W_{\mathbf{h}'\to\mathbf{h}}\mathbf{h}_{t-1}'\right) \oplus \left(\mathbf{h}_t \odot W_{\mathbf{h}'\to\mathbf{h}}\mathbf{h}_t'\right)
\tag{2}
$$

where $\oplus$ denotes the concatenation operator and $\odot$ is the pooling operator. For the t-th time step, $\left(\mathbf{h}_t \odot W_{\mathbf{h}'\to\mathbf{h}}\mathbf{h}_{t-1}'\right)$ and $\left(\mathbf{h}_t \odot W_{\mathbf{h}'\to\mathbf{h}}\mathbf{h}_t'\right)$ capture how $\mathbf{h}_{t-1}'$ and $\mathbf{h}_t'$ respectively influence $\mathbf{h}_t$.

However, it is impossible that the fused information at each time step will be equally important to behavioral patterns. For example, someone is probably an electricity theft if he consumes little electrical power in summer or winter, but this may not be true in autumn or spring, since users typically consume less electricity during these months due to seasonal effects. Liu et al. [27] suggests that models should try to measure such significant information at different temporal points. Hence, we design an attention mechanism to model the varying significance of the fused information at different time steps: given the current fusion level k, the input of the next level $\mathbf{I}^{(k+1)} \in \mathbb{R}^{T \times 2d_{\mathbf{h}^{(k)}}}$ is computed by a linear combination of the intermediate representations, weighted by a score vector $\boldsymbol{\alpha}^{(k+1)} \in \mathbb{R}^T$:

$$
\begin{aligned}
\mathbf{I}^{(k+1)} &= \sum_{t=0}^{T} \boldsymbol{\alpha}_t^{(k+1)} \cdot \tanh\left(\mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t^{(k)}, \mathbf{h}_t'^{(k)}\right)\right) \\
\boldsymbol{\alpha}^{(k+1)} &= \text{softmax}\left(W_{\mathbf{h}\to\boldsymbol{\alpha}} \cdot \sum_{t=0}^{T} \mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t^{(k)}, \mathbf{h}_t'^{(k)}\right)\right)
\end{aligned}
\tag{3}
$$

where $\sum$ denotes the concatenation, and $W_{\mathbf{h}\to\boldsymbol{\alpha}}$ is a trainable weighted matrix shared by all temporal points. The activation function tanh is used to activate the intermediate representations.

Model formulation. So far, we can materialize Eq 1 for HEBR by the abovementioned definitions, the complete mathematical formulations for which are as follows:

$$
\mathbf{I}^e = \mathcal{X}_u^e, \quad \mathbf{I}^l = \mathcal{X}_a^l, \quad \mathbf{I}^c = \mathcal{X}_a^c, \qquad u \in \mathbf{U}, a \in \mathbf{A}
\tag{4a}
$$

$$
\begin{bmatrix} \mathbf{h}_t^e \\ \mathbf{h}_t^l \\ \mathbf{h}_t^c \end{bmatrix} = \begin{bmatrix} \mathcal{F}_{\text{recurrent}}\left(\mathbf{I}_t^e, \mathbf{h}_{t-1}^e\right) \\ \mathcal{F}_{\text{recurrent}}\left(\mathbf{I}_t^l, \mathbf{h}_{t-1}^l\right) \\ \mathcal{F}_{\text{recurrent}}\left(\mathbf{I}_t^c, \mathbf{h}_{t-1}^c\right) \end{bmatrix}, \begin{bmatrix} \mathbf{I}_t^{el} \\ \mathbf{I}_t^{ec} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_t^{el} \\ \boldsymbol{\alpha}_t^{ec} \end{bmatrix} \cdot \tanh \begin{bmatrix} \mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t^e, \mathbf{h}_t^l\right) \\ \mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t^e, \mathbf{h}_t^c\right) \end{bmatrix}
\tag{4b}
$$

$$
\begin{bmatrix} \mathbf{h}_t^{ec} \\ \mathbf{h}_t^{el} \end{bmatrix} = \begin{bmatrix} \mathcal{F}_{\text{recurrent}}\left(\mathbf{I}_t^{ec}, \mathbf{h}_{t-1}^{ec}\right) \\ \mathcal{F}_{\text{recurrent}}\left(\mathbf{I}_t^{el}, \mathbf{h}_{t-1}^{el}\right) \end{bmatrix}, \mathbf{I}_t^{elc} = \boldsymbol{\alpha}_t^{elc} \cdot \tanh\left(\mathcal{F}_{\text{fuse}}\left(\mathbf{h}_t^{ec}, \mathbf{h}_t^{el}\right)\right)
\tag{4c}
$$

$$
\mathbf{h}_t^{elc} = \mathcal{F}_{\text{recurrent}}\left(\mathbf{I}_t^{elc}, \mathbf{h}_{t-1}^{elc}\right)
\tag{4d}
$$

$$
\mathbf{H}_u^{elc} = \prod_{t=0}^{T} \mathbf{h}_t^{elc}
\tag{4e}
$$

where multi-source sequences $(\mathcal{X}_u^e, \mathcal{X}_u^l, \mathcal{X}_u^c)$ initialize the input $\mathbf{I}^*$, after which three levels of feature extraction and hierarchical fusion (Eq 4a, Eq 4b, Eq 4c) are conducted. $\prod$ denotes the pooling operator that aggregates the fused representation $\mathbf{h}^{elc}$ in each temporal point, and the final output is the behavior embedding $\mathbf{H}_u^{elc}$ of each user u. In our experiment, we use mean pooling as the pooling operator. As for estimating the probability of each user being an electricity thief, we define a mapping function Ψ that maps

the feature embedding into a binary vector, and turn it into the probability ranging from 0 to 1 by means of the softmax function, as follows:

$$\mathbf{P}\left(\mathcal{Y} \mid \mathcal{X}_u^e, \mathcal{X}_a^l, \mathcal{X}_a^c\right) = \text{softmax}\left(\Psi\left(\mathbf{H}_u^{\text{elc}}\right)\right) \tag{5}$$

We can implement Ψ by means of fully connected networks or some well-known classifiers.

The whole hierarchical framework can also be implemented by HBRNN [13], which is proposed for recognizing skeleton-based actions by combining the multi-source time series data. The main difference compared with our approach is that HBRNN implements fusion functions by directly concatenating two representations at the same temporal point; by contrast, we design a multi-step fusion mechanism for HEBR. We will validate the effectiveness of such architecture by comparing both methods in our experiments (Section 5.2) and additional ablation studies (Section 5.3).

Learning. We use the Adam optimizer [24] for the parameter learning, where the objective function is defined as the binary cross-entropy:

$$\mathcal{L} = -\sum_{u \in \mathbf{U}} \hat{\mathcal{Y}}_u \log \mathbf{P}\left(\mathcal{Y} \mid \mathcal{X}_u^e, \mathcal{X}_a^l, \mathcal{X}_a^c\right) + \tag{6}$$
$$(1 - \hat{\mathcal{Y}}_u) \log \left(1 - \mathbf{P}\left(\mathcal{Y} \mid \mathcal{X}_u^e, \mathcal{X}_a^l, \mathcal{X}_a^c\right)\right)$$

where $\hat{\mathcal{Y}}_u \in \{0, 1\}$ is the ground truth of a user being an electricity thief, as confirmed by on-site investigations, and $\mathbf{P}\left(\mathcal{Y} \mid \mathcal{X}_u^e, \mathcal{X}_a^l, \mathcal{X}_a^c\right)$ is computed by Eq 5.

5 EXPERIMENTS

In this section, we conduct experiments on a real-world dataset (introduced in 3.1) to answer the following three questions:

- **Q1:** How does HEBR perform on electricity-theft detection tasks, compared with state-of-the-art baselines?
- **Q2:** How does the multi-source information contribute to the detection task?
- **Q3:** Can the proposed hierarchical fusion mechanism effectively bridge the information from different sources?

5.1 Experimental Setup

Baselines. We validate the effectiveness of HEBR compared with several different types of baselines. The first type is classification methods based on handcrafted features, which have been commonly used in existing work on electricity theft detection [11, 19, 21, 35, 36]. We list the handcrafted features we consider in Table 2 and employ the following classifiers: *logistic regression (LR)* [17], *support vector machine (SVM)* [38], *random forest (RF)* [16] and *extreme gradient boosting (XGB)* [6].

The second type of baseline is time series classification methods, including:

- *Nearest Neighbor*: This method determines whether a user u pilfers electrical power with reference to other users close to u. In particular, we consider the following different metrics to calculate the distance between two users' time series in our experiment: *Euclidean Distance (NN-ED)*, *Dynamic Time*

Table 2: List of handcrafted features related to electricity theft.

Feature	Description
Power usage	Mean, variance and slope of total, on-peak and off-peak electricity usage $\mathcal{X}_u^e$
NTL	Mean, variance and slope of non-technical loss $\mathcal{X}_a^l$
Temperature	Mean, variance and slope of maximum and minimum temperature $\mathcal{X}_a^c$
Usage vs. NTL	$\sqrt[2]{(\mathcal{X}_u^e - \mathcal{X}_a^l)^2}$, Euclidean distance between total usage and NTL $DTW(\mathcal{X}_u^e - \mathcal{X}_a^l)$, DTW distance between total usage and NTL
Usage vs. Temperature	$\sqrt[2]{(\mathcal{X}_u^e - \mathcal{X}_a^c)^2}$, Euclidean distance between total usage and temperatures $DTW(\mathcal{X}_u^e - \mathcal{X}_a^c)$, DTW distance between total usage and temperatures

Warping (NN-DTW) [3] and *Complexity Invariant Distance (NN-CID)* [2].

- *Fast Shapelets (FS)* [34]: This approach extracts shapelets, the representative segments of time series, as features for classification.
- *Time Series Forest (TSF)* [10]: This is a tree-ensemble method for time series classification.

As for the third type of baseline, we consider the following competitive deep learning methodologies:

- *MRNN* [18]: A multiscale recurrent neural network that takes the concatenated multi-source sequences ($\mathcal{X}_u^e \oplus \mathcal{X}_a^l \oplus \mathcal{X}_a^c$) as input.
- *HBRNN* [13]: This is a hierarchical recurrent neural network on multi-source sequences, which is proposed to recognize skeleton-based actions. For the fusion layer, it concatenates the latent representation at the same time points ($\mathcal{F}_{\text{fuse}}(\mathbf{h}_t, \mathbf{h}'_t) = \mathbf{h}_t \oplus \mathbf{h}'_t$).
- *WDCNN* [39]: A *wide&deep convolutional neural network* for detecting electricity theft that focuses on capturing periodic patterns of users' electricity usage.
- *HEBR*: The proposed method. We empirically set the dimension of $\mathbf{h}^e$, $\mathbf{h}^l$ and $\mathbf{h}^c$ in the first layer as 32, 8 and 8 respectively, and further set the learning rate as 0.01 with the reduction via a factor of 10 at every 20 iterations. We implement $\mathcal{F}_{\text{recurrent}}$ by LSTM, and will study how different implementations influence the performance later in Section 5.3.

Comparison metrics. We use precision, recall and two F-measures (F1, F0.5) as metrics. The F-measure is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test, with the following mathematical form:

$$F_\beta = (1 + \beta^2) \frac{precision \times recall}{\beta^2 \times precision + recall}$$

We prefer to use F0.5 as the metric for electricity-theft detection, as precision is more important than recall to real application.

Implementation details. To meet the demands of the application scenario, the input of all methods is the historical observation sequence spanning six months. The output is the probability of electricity theft, which can be validated in the next month. The first 80% of samples ordered by time are used for training, and we

Table 3: Comparison of classification performance (%). The bold indicates the best performance of all the methods.

Methods	Metrics	Precision	Recall	F1	F0.5	Variation
Handcrafted features	LR	9.52	7.12	8.14	8.92	±0.08
	SVM	11.11	5.21	7.09	9.06	±0.14
	RF	17.07	11.93	14.04	15.72	±0.22
	XGB	20.00	9.35	12.74	16.29	±0.23
Time Series Classification	NN-ED	0.00	0.00	0.00	0.00	±0.00
	NN-DTW	10.82	27.78	15.57	12.32	±0.12
	NN-CID	12.86	22.96	16.48	14.10	±0.24
	FS	2.26	20.11	4.06	2.75	±0.85
	TSF	18.67	24.12	21.05	19.55	±0.53
Neural Networks	MRNN	1.28	**100.0**	2.54	1.59	±0.00
	HBRNN	19.07	36.98	25.16	21.12	±0.85
	WDCNN	18.46	24.69	21.12	19.44	±1.31
Ours	HEBR	**22.54**	34.19	**27.17**	**24.19**	±0.85

test different methods on the remaining samples. We also use 10% of samples from the training set as validation set, for avoiding the overfitting. For baselines that require a classifier, we use XGB [6] with a batch size of 2000. We adopt a larger weight (e.g., the ratio of negative/positive) for positive samples to address class imbalance. All the experiments are ran on a single Nvidia GTX 1080Ti GPU.

5.2 Performance Comparison

We first compare the experimental results of HEBR with other baselines to answer **Q1**. As shown in Table 3, all handcraft-feature methods perform poorly, as these methods can only capture a limited number of patterns. Relatively speaking, the ensemble methods (i.e., *XGB* and *RF*) performs better (an average +7% of F0.5). By automatically capturing temporal features, time series classification methods achieve further performance improvements, especially for recall. However, these methods cannot effectively handle multi-source time series and therefore suffer in terms of precision. A similar phenomenon can be observed the neural network results. In particular, when simply concatenating all multi-source data and inputting it into a recurrent neural network (*MRNN*), we can see that it identifies all samples as instances of electricity theft. This suggests that improperly handling multi-source data will bring in more noise which hurts performance. *WDCNN* tries to capture the abnormal non-periodic behaviors of users, resulting in a performance improvement. However, the performance of this method is unstable (around 1.31 variation). As expected, moreover, models with hierarchical structure like HBRNN and HEBR are better able to handle multi-source data and consequently outperform other methods. Moreover, HEBR outperforms HBRNN by +3% in terms of F0.5. Through careful investigation, we find that with the help of the multi-step fusion and the attention operator, HEBR can not only better bridge the multi-source information, but have superior interpretability, the details of which are presented in later chapters.

5.3 Model Effectiveness

Effectiveness of multi-source information (Q2). We study whether or not the multi-source information can be effective in electricity theft detection. To do this, we remove the input sequences of temperature and NTL respectively from HEBR. It is notable that

Table 4: Effect of multi-source information (%).

Removed component	Precision	Recall	F1	F0.5
Temperature	20.65	28.58	23.98	21.86
NTL	19.01	22.33	20.54	19.59
Both	10.46	16.79	12.89	11.31
HEBR	**22.54**	**34.19**	**27.17**	**24.19**

Table 5: Effect of fusion and attention operators (%).

Removed component	Precision	Recall	F1	F0.5
Multi-step fusion	19.18	32.05	23.99	20.85
Attention	20.91	30.27	24.73	22.29
Both	18.66	29.47	22.85	20.14
HEBR	**22.54**	**34.19**	**27.17**	**24.19**

Table 6: Effect of temporal modeling (%).

Implementation	Precision	Recall	F1	F0.5
Average pooling	15.96	32.11	21.32	17.75
Linear RNN	19.53	30.24	23.73	21.02
GRU	21.04	**36.79**	26.77	23.01
LSTM	**22.54**	34.19	**27.17**	**24.19**

after each sequence is removed, the number of HEBR's levels decreases; once we remove both sequences, HEBR is transformed into a single recurrent neural network with the users' electricity usage records as input.

From Table 4, we can see that multi-source information is significant: the performance drops substantially when both NTL and temperature are simultaneously removed (-12.88% of F0.5). Notably, temperature is slightly less sensitive than NTL to the performance (+1.97% of F0.5).

Effectiveness of hierarchical fusion mechanism (Q3). The proposed hierarchical fusion mechanism includes the multi-step fusion operator and the attention operator. How do these elements contribute to bridging the information from different data sources? To answer question, we remove each operator in turn and assess the subsequent impact on performance. Note that after removing the fusion operator, we adopt a simple concatenation for the fusion layer: $\mathcal{F}_{\text{fuse}} = \mathbf{h}_t \oplus \mathbf{h}'_t$.

Results are presented in Table 5. From the table, we can see that the performance clearly drops when either the multi-step fusion operator or the attention operator is removed (-2.62% of F0.5 on average). Moreover, removing them both influences the performance more significantly, suggesting that these two operators work well together to bridge the information from different sources. We will later qualitatively demonstrate the effectiveness of our proposed hierarchical fusion mechanism through a specific application case.

Implementations of recurrent layers. Finally, we study how different implementations of the recurrent layers in our proposed model influence the performance. To do this, we use several common methods, such as average pooling, linear RNN [28], GRU [7] and LSTM[22]. As shown in Table 6, the gated networks (GRU, LSTM) outperform the linear methods (pooling, RNN), which illustrates that the appropriate temporal modeling is important for improving the performance.

5.4 Application and Case Study

In the past, the State Grid staff would employ several data-driven models to detect electricity theft. However, these models are inefficient in practice. For example, in 2018, *none* of the electricity thieves in Zhejiang were caught by these models. However, in order to catch thieves, large-scale on-site investigations are very costly and time-consuming.

Accordingly, in order to improve the accuracy of on-site investigation and validate the effectiveness of our model, we employed HEBR for monthly on-site investigation, in cooperation with State Grid Hangzhou Power Supply Co. Ltd.. More specifically, HEBR detected 20 high-risk users at the beginning of August 2019 and suggested that the State Grid staff should investigate and collect evidence. It turned out that our approach successfully caught three electricity thieves out of 20 with 15% precision in practice, which represented a significant improvement in on-site investigation precision (improved from 0%). Moreover, another six users among the remaining 17 were identified that staff in Hangzhou strongly suspected of being electricity thieves, although no clear evidence was found during investigation. It is likely that these six users pilfered electrical power in June or July, then restored electrical meter to its previous condition before on-site checking.

We next present a specific case identified by HEBR to demonstrate its effectiveness in real-world applications. In Figure 8, for a specific user, we present temperature, NTL, the user's electrical usage record, and three score vectors (α^{ec}, α^{el}, α^{elc}) in our model's fusion layers from top to bottom. We can see that the high-scoring positions (the brighter areas of the heat map) are where the electricity usages are low, along with hot weather and high NTL. This discovery is consistent with previous empirical observation (Section 3): when a user consumes little electricity in hot weather, while the NTL increases abnormally, he or she is likely to be pilfering electricity. Moreover, when examining the fusion layers from top to bottom, the brighter areas in the heat map (Figure 8) become increasingly clear, suggesting that HEBR captures more accurate patterns of multi-source information. We can see that users do not pilfer electrical power all the time, and that the attention score returns to a low point after on-site checking (red bar). From the case study and the performance comparison in Table 3, we can see that the attention operator can not only improve the electricity-theft detection performance, but also provides superior interpretability.

6 RELATED WORK

Additional related works on electricity theft. Electricity theft, or pilfering, is known as a common phenomenon in developing countries. Substantial effort has been expended to prevent or detect the behavior of electricity theft. As mentioned in Section 1, there are two main avenues of works related to electricity theft detection or prevention. The first of these is hardware-driven methods, and several representative works will be discussed herein. Fennell [15] proposed a special, pilfer-proofing, system incorporating a plug-in terminal block set and a meter box cover; Depuru et al. [11] first conclude that electricity theft makes up a significant proportion of non-technical loss (NTL). Secondly, rather than attempting to improve the hardware of electricity meters, data-driven methodologies have also been proposed that focus on the analysis of electrical
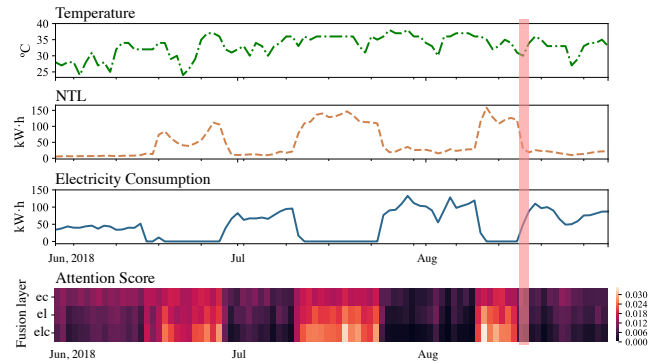
Figure 8: A case study of the attention operator shown together with different fusion layers ($\alpha^{ec,el,elc}$). The upper rows present different multi-source observation sequences, while the red bar indicates the time at which the electricity theft was detected. The lower row presents the score vector of the attention operator by heat map.

power consumption records. In addition to the works referenced in Section 1, Zheng et al. [39] propose a framework based on *wide & deep CNNs*, which aims at accurately identifying the periodicity and non-periodicity of electricity usage by utilizing 2-D electrical power consumption data. Moreover, Costa et al. [8] apply the use of knowledge-discovery in the database process based on artificial neural networks to conduct electricity-theft detection. However, the problem of how to utilize multi-source data (including electricity consumption records and other related information) to conduct electricity-theft detection remains unstudied.

Time series modeling. A basic but rather important characteristic of power consumption data is that these records are time series data. Time series modeling has been widely studied over the past decades. One traditional avenue here is to extract efficient features from the original data and develop a well-trained classifier, such as TSF [10], *shapelets* [26, 37] etc.; another avenue has focused on deep learning, such as RNNs and their variants (LSTM [22], GRU [18], etc.). In addition, since the dimensions of time series features may be large, and it is likely that different levels of correlations exist between features, hierarchical LSTM-based models are proposed to learn these hierarchical relationships (e.g., HBRNN [13]).

User behavior modeling. Another related domain of this work is that of the user behavior analysis. One typical case is that of the web search: for example, Radinsky et al. [33] develop a temporal modeling framework to predict user behavior using smoothing and trends. However, although these works may provide various insights relevant to human behavior modeling, fully understanding complicated user behaviors is quite difficult; thus, it may be necessary to tailor the specific analysis to the scenario in question.

7 CONCLUSION AND FUTURE WORK

In this paper, we study the problem of electricity-theft detection and analyze the influence of the macro-level (climate) and meso-level (non-technical loss) factors on users' electricity usage behavior. We proposed a hierarchical framework, HEBR, that encodes the correlations between different levels of information step by step. When evaluated on real-world datasets, the proposed method not only achieved significantly better results than other baselines, but

also helped to catch electricity thieves in practice. In future work, we plan to explore the following aspects: 1) investigating more factors from different sources to improve electricity-theft detection accuracy; 2) extending HEBR to other similar scenarios.

ACKNOWLEDGMENTS

REFERENCES

[1] Tanveer Ahmad. 2017. Non-technical loss analysis and prevention using smart meters. *Renewable and Sustainable Energy Reviews* 72 (2017), 573–589.

[2] Gustavo EAPA Batista, Xiaoyue Wang, and Eamonn J Keogh. 2011. A complexity-invariant distance measure for time series. *SDM* (2011), 699–710.

[3] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. *SIGKDD* (1994), 359–370.

[4] Hung-Po Chao and Stephen Peck. 1996. A market mechanism for electric power transmission. *Journal of regulatory economics* 10, 1 (1996), 25–59.

[5] Abhishek Chauhan and Saurabh Rajvanshi. 2013. Non-technical losses in power system: A review. In *International Conference on Power, Energy and Control (ICPEC)*. IEEE, 558–561.

[6] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *SIGKDD* (2016), 785–794.

[7] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2015. Gated Feedback Recurrent Neural Networks. *ICML* (2015), 2067–2075.

[8] Breno C Costa, Bruno LA Alberto, André M Portela, W Maduro, and Esdras O Eler. 2013. Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process. *International Journal of Artificial Intelligence & Applications* 4, 6 (2013), 17.

[9] Marco Cuturi and Arnaud Doucet. 2014. Fast Computation of Wasserstein Barycenters. *ICML* (2014), 685–693.

[10] Houtao Deng, George Runger, Eugene Tuv, and Martyanov Vladimir. 2013. A time series forest for classification and feature extraction. *Information Sciences* 239 (2013), 142–153.

[11] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktuni. 2011. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy* 39, 2 (2011), 1007–1015.

[12] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktuni. 2011. Support vector machine based data classification for detection of electricity theft. In *Power Systems Conference and Exposition (PES)*. IEEE, 1–8.

[13] Yong Du, Wei Wang, and Liang Wang. 2015. Hierarchical recurrent neural network for skeleton based action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1110–1118.

[14] Engerati. 2018. Non-technical losses today: the business challenges. https://www.engerati.com/transmission-and-distribution/article/theft-detection/non-technical-losses-today-business-challenges

[15] Robert B Fennell. 1983. Pilfer proofing system and method for electric utility meter box. US Patent 4,404,521.

[16] Robin Genuer, Jeanmichel Poggi, Christine Tuleaumalot, and Nathalie Villavialaneix. 2015. Random Forests for Big Data. *arXiv: Machine Learning* (2015).

[17] Steven L Gortmaker, David W Hosmer, and Stanley Lemeshow. 1994. Applied Logistic Regression. *Contemporary Sociology* 23, 1 (1994), 159.

[18] Alex Graves, Abdelrahman Mohamed, and Geoffrey E Hinton. [n.d.]. Speech recognition with deep recurrent neural networks. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* ([n.d.]).

[19] Li-Cai Guo, Zhi-Wei Peng, and Qiang Fan. 2010. A survey of electric energy metering and countermeasures to electric power stealing [J]. *High Voltage Apparatus* 46, 5 (2010), 86–88.

[20] Songlin HAN, Dezu SHANG, J Gao, et al. 2002. Talking about Products of Guard against Pilfering Electricity and Pilfering Electricity on the Contrary. *Electrical Measurement & Instrumentation* 39, 9 (2002), 10–15.

[21] Hu Hanmei and Wang Yuanlong. 2004. The analysis and discussion of electricity-stealing and precaution method. *Electrical Measurement & Instrumentation* 3 (2004).

[22] Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[23] Camille Kamga, M Anil Yazici, and Abhishek Singhal. 2013. Hailing in the rain: Temporal and weather-related variations in taxi ridership and taxi demand-supply equilibrium. In *Transportation Research Board 92nd Annual Meeting*.

[24] Diederik P Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. *ICLR* (2015).

[25] Blazakis Konstantinos and Stavrakakis Georgios. 2019. Efficient Power Theft Detection for Residential Consumers Using Mean Shift Data Mining Knowledge Discovery Process. *International Journal of Artificial Intelligence and Applications (IJAIA)* 10, 1 (2019).

[26] Jason Lines, Luke M Davis, Jon Hills, and Anthony Bagnall. 2012. A shapelet transform for time series classification. In *SIGKDD*. ACM, 289–297.

[27] Zongtao Liu, Yang Yang, Wei Huang, Zhongyi Tang, Ning Li, and Fei Wu. 2019. How Do Your Neighbors Disclose Your Information: Social-Aware Time Series Imputation. In *The World Wide Web Conference (WWW)*.

[28] Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Cernocký, and Sanjeev Khudanpur. [n.d.]. Recurrent neural network based language model. *Conference of the International Speech Communication Association (INTERSPEECH)* ([n.d.]).

[29] Jawad Nagi, Keem Siah Yap, Sieh Kiong Tiong, Syed Khaleel Ahmed, and Malik Mohamad. 2009. Nontechnical loss detection for metered customers in power utility using support vector machines. *Transactions on Power Delivery* 25, 2 (2009), 1162–1171.

[30] Zhiyu Chen: China Power News Network. 2018. The severe problem of electricity stealing in India. http://www.cpnn.com.cn/sd/gj/201805/t20180507_1068093.html

[31] PR Newswire. 2017. 96 Billion Is Lost Every Year To Electricity Theft: Utilities increasingly investing in solutions to combat theft and non-technical losses. https://www.prnewswire.com/news-releases/96-billion-is-lost-every-year-to-electricity-theft-300453411.html

[32] Emanuel Parzen. 1962. ON ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE. *Annals of Mathematical Statistics* 33, 3 (1962), 1065–1076.

[33] Kira Radinsky, Krysta Svore, Susan Dumais, Jaime Teevan, Alex Bocharov, and Eric Horvitz. 2012. Modeling and predicting behavioral dynamics on the web. In *The World Wide Web Conference (WWW)*. ACM, 599–608.

[34] Thanawin Rakthanmanon and Eamonn Keogh. 2013. Fast shapelets: A scalable algorithm for discovering time series shapelets. *ICDM* (2013), 668–676.

[35] Joaquim L Viegas, Paulo R Esteves, R Melício, VMF Mendes, and Susana M Vieira. 2017. Solutions for detection of non-technical losses in the electricity grid: a review. *Renewable and Sustainable Energy Reviews* 80 (2017), 1256–1268.

[36] Wu Xiaomei. 2003. A brief view of preventing the electricity power stealing. *Electrical Measurement & Instrumentation* 2 (2003), 1.

[37] Lexiang Ye and Eamonn Keogh. 2011. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *DMKD* 22, 1 (2011), 149–182.

[38] Ji Zheng and Baoliang Lu. 2011. A support vector machine classifier with automatic confidence and its application to gender classification. *Neurocomputing* 74, 11 (2011), 1926–1935.

[39] Zibin Zheng, Yatao Yang, Xiangdong Niu, Hong-Ning Dai, and Yuren Zhou. 2017. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Transactions on Industrial Informatics* 14, 4 (2017), 1606–1615.