ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096948

BAYESIAN NETWORK MODELING AND PREDICTION OF TRANSITIONS WITHIN THE HOMELESSNESS SYSTEM

Khandker Sadia Rahman[1], *Daphney-Stavroula Zois*[2], *Charalampos Chelmis*[1]

[1]Department of Computer Science, University at Albany, State University of New York, Albany, NY

[2]Department of Electrical and Computer Engineering, University at Albany, State University of New York, Albany, NY

ABSTRACT

Administrative data collected by homeless service providers offer a unique opportunity to understand how homeless individuals navigate the homeless system towards securing stable housing. However, the literature on predictive models in the context of homeless service provision has neglected the sequential nature of services that an individual receives over time. Our work addresses this gap by learning, from administrative data, a Bayesian network, which in turn can be used to accurately predict whether an individual will exit the system, or alternatively, the service she would be assigned to the next time she experiences homelessness. Experimental evaluation shows that the proposed approach outperforms prior art not only at predicting exit, but also the less frequent services (and thus more challenging to predict).

Index Terms— Complex systems, Bayesian network, human behavior, probabilistic modeling, trajectory prediction

1. INTRODUCTION

Homelessness is defined by the U.S. Department of Housing and Urban Development (HUD) as a situation where an individual experiences lack of fixed, regular, and adequate nighttime residence [1]. According to the U.S. Department of Housing and Urban Development (HUD), more than $326,000$ people experience homelessness on a given night [2]. Unfortunately, the number of homeless individuals typically exceeds available resources used to assist them, necessitating identifying the most vulnerable or in–need individuals, and matching them to appropriate housing resources. With this goal in mind, [3] explored the feasibility of an automated recommendation system designed to match individuals to homelessness services, when they first experience homelessness. On the other hand, numerous methods have been proposed to predict the potential of individuals to experience repeated episodes of homelessness [4, 5, 6] and/or their risk of becoming chronically homeless [7, 8]. Most such works are formulated as a binary–classification task, oversimplifying the complexities of the homeless service provision system.

In this work, we begin by modeling the homeless service provision task as a multi–class classification problem. Building on the literature of Bayesian learning, we propose a methodology that uses administrative data collected by homeless service providers, to learn a Bayesian network that can be used to accurately predict whether an individual will exit the system, or alternatively, predict the service she would be assigned to the next time she experiences homelessness. We demonstrate the effectiveness of our approach using a one–of–a–kind longitudinal dataset that spans 6 years. Our results show the ability of our approach to predict well not only exit or high frequency services (which are most probable, but least actionable), but also less frequent (and therefore more challenging to predict).

The rest of the paper is organized as follows. Section 2 delineates the problem statement. Section 3 introduces the proposed approach. Section 4 describes the data, metrics, and baselines used for evaluation, whereas Section 5 discusses the experimental results. Section 6 concludes with contributions, future work, and limitations of this study.

2. PROBLEM STATEMENT

Homeless service providers offer services that are organized by *type* (e.g., emergency shelters, transitional housing) [9]. We denote the set of types as $\mathcal{P}$, and the set of individuals as $\mathcal{C}$, each assigned to such services multiple times (i.e., reentering the homeless system more than once). Specifically, each $c \in \mathcal{C}$ is associated with a temporally ordered sequence of services, of which we focus on the K most recent services, $\mathcal{T}_c = \{p_{t_1}, p_{t_2}, ..., p_{t_K}\}$, where c is assigned to $p_{t_i} \in \mathcal{P}$ at timestamp t_i. Given $\mathcal{T}_c$, our goal is to first predict whether, at timestamp t_{K+1}, individual c is likely to exit ($E = 1$) or reenter ($E = 0$) the homeless system. When the individual is predicted to reenter, the goal is to additionally predict the service $p_{t_{K+1}}$ she is to be assigned to at t_{K+1}.

This material is based upon work supported by the National Science Foundation under Grants ECCS–1737443 & CNS–1942330.

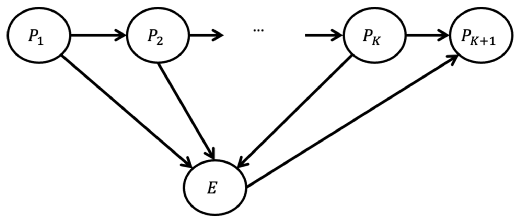

Fig. 1. Bayesian network model of K–length sequences.

3. PROPOSED APPROACH

The timeline of services provided to each individual, including the start and end date of specific services, and the transition between services, offers a unique opportunity to study how individuals navigate through the homeless system over time. To gain insights into the dynamics of this complex system, we propose PREVISE, a method that PRedicts Exit and serVice assIgnment using BayeSian nEtwork.

PREVISE begins by modeling the sequences of individuals in $\mathcal{C}$ using the Bayesian network $\mathcal{B} = (\mathcal{G}, \theta)$ shown in Figure 1. Specifically, $\mathcal{G}$ is a directed acyclic graph (DAG) comprising a finite set of random variables, $\mathcal{X} = \{P_1, P_2, ..., P_K, P_{K+1}, E\}$, where $P_t \in \mathcal{P}$ denotes the service an individual is assigned to at timestamp t, and E denotes exit (or reentry). Edges represent the direct influence of one variable to another, and θ is the set of parameters to be learned to fully represent the prior probability $Pr(P_1)$ and the conditional probability distributions $Pr(P_t|P_{t-1})$, for $t = 2, ..., K$, $Pr(P_{K+1}|P_K, E)$, and $Pr(E|P_1, P_2, ..., P_K)$.

The rationale for this particular structure of $\mathcal{B}$ stems from mathematical epidemiology [10], where the goal is to obtain the *simplest* possible model that can *accurately* replicate the empirical data. In this context, we make the following modeling assumptions:

1. To account for sequences of varying length, we model the possibility of individuals exiting (E) the homeless system at any point (P_t).
2. An individual is assigned to a service when she first enters, or upon reentering the system (i.e., $E = 0$). We thus model a direct influence of exit (E) on P_{k+1}.
3. To reduce the complexity of $\mathcal{B}$, we assume that P_{t+1} is only influenced by P_t, where $t = 1, 2, ..., K$ (i.e., Markov property).

Given $\mathcal{B}$ and the set of sequences $\mathcal{T} = \{\mathcal{T}_c, \forall c \in \mathcal{C}\}$, PREVISE learns the parameters $\hat{\theta}$ that satisfy:

$$\max_{\theta} L(\theta : \mathcal{T}), \tag{1}$$

where, likelihood $L(\theta : \mathcal{T}) = \prod_i \prod_c Pr(\mathcal{X}_i[c]|Pa_{\mathcal{X}_i[c]} : \theta)$, determines how the probability of the sequences changes as a function of θ, $\mathcal{X}_i \in \mathcal{X}$, and $Pa_{\mathcal{X}_i}$, the set of parent nodes of $\mathcal{X}_i$ in $\mathcal{B}$.

Once parameters $\hat{\theta}$ are learned, PREVISE is tasked with making a prediction about an individual given her history thus far. Specifically, presented with history $\mathcal{X}_e = \{P_1 = p_{t_1}, P_2 = p_{t_2}, ..., P_K = p_{t_K}\}$ obtained from sequence $\mathcal{T}_c = \{p_{t_1}, p_{t_2}, ..., p_{t_K}\}$ of individual c, PREVISE uses Maximum A Posteriori (MAP) to predict $\hat{E}$ and $\hat{p}_{t_{K+1}}$ as:

$$\begin{aligned} \hat{E} &= \text{argmax}_E Pr(E|X_e) \\ \hat{p}_{t_{K+1}} &= \text{argmax}_{P_{K+1}} Pr(P_{K+1}|X_e, E = 0) \end{aligned} \tag{2}$$

The individual is predicted to exit if $\hat{E} = 1$, or be assigned to service type $\hat{p}_{t_{K+1}}$ otherwise.

4. EXPERIMENTAL EVALUATION

4.1. Data

An anonymized set of $18,817$ records of all services provided by homeless service providers, in the Capital Region of the state of New York (NY), to a total of $6,011$ individuals over the time period of 2012 and 2018 was used for evaluation purposes in this study. The dataset was provided by the CARES of NY. The history of each individual in the dataset is sampled using a sliding window of length K. The best value of K was experimentally determined to be 4 as discussed in Section 5.

We illustrate the sampling process in Figure 2. Specifically, each sequence is sampled backwards (i.e., starting from the last entry before $p_{t_{K+1}}$) with the window of length K sliding by one–step until it reaches p_{t_1} (e.g., individual X in Figure 2 is sampled into X_1, X_2, X_3, X_4). Intuitively, when an individual has exited the system (i.e., $E = 1$), a value may not be available at p_{t_i}. We encode this scenario by introducing label -1. Similarly, we set $p_{t_{K+1}} = -1$ if $\mathcal{T}_c[K+1] \notin \mathcal{P}$. In addition, we set $E = 1$ if $\mathcal{T}_c[K+1] = -1$, and 0 otherwise. Finally, for trajectories shorter than K, the missing values are encoded with -1 to indicate unavailability of data for that individual (e.g., individual Z_1 in Figure 2).

We use $\mathcal{D}$ to denote our dataset after sampling, and perform a random $80 - 20$ split to obtain $\mathcal{D}_{train}$ and $\mathcal{D}_{test}$, the training and testing sets, accordingly. We use $\mathcal{D}_{train}$ to learn the set of parameters θ during training, and $\mathcal{D}_{test}$ to predict $\hat{E}$ and $\hat{p}_{t_{K+1}}$ during testing.

4.2. Evaluation Metrics

We evaluate the prediction capability of the proposed model for both when the individual exits the homeless system, and when she reenters another service in the next step using the following metrics: (i) **Accuracy** (i.e., how many times predicted services (and exit) are correct), (ii) **Precision** (how often each predicted service type (and exit) is correct), and (iii) **Recall** (how many times each service type (and exit) is identified correctly).

Table 1. Comparison between PREVISE and the baselines with respect to Accuracy (Acc), Mean Recall (mR), and Mean Precision (mP). For reference, K refers to window length.

Prediction	K	PREVISE			**RF**			**LOG–REG**			**UNI–RNDM**			**PA-RNDM**		
		Acc	mR	mP	Acc	mR	mP	Acc	mR	mP	Acc	mR	mP	Acc	mR	mP
$\hat{E}$	2	73.2	68.7	76.3	**73.3**	**68.8**	**76.4**	67.0	63.4	66.6	51.1	51.1	51.0	49.8	49.9	49.9
	3	**78.1**	**77.6**	**78.2**	71.7	71.6	72.0	69.8	69.9	70.0	50.4	50.4	50.4	50.9	51.0	50.9
	4	**79.5**	**79.6**	**79.3**	68.3	66.4	66.0	63.1	62.1	62.2	49.4	49.5	49.5	50.4	50.3	50.4
	5	**81.8**	**81.9**	**81.3**	63.4	59.1	63.9	53.5	50.2	49.1	49.9	50.1	50.0	50.1	50.1	50.2
	6	**82.0**	**82.1**	**81.1**	62.1	57.4	60.7	47.1	43.2	41.7	49.3	49.3	49.4	50.0	50.1	50.0
$\hat{p}_{t_{K+1}}$	2	68.0	**40.4**	**46.8**	**69.0**	37.1	41.7	68.0	40.6	49.2	11.1	10.3	11.1	35.5	11.0	10.9
	3	**74.1**	**46.2**	**48.0**	67.4	32.8	39.6	66.7	28.3	37.4	11.3	12.1	11.6	37.9	11.4	11.2
	4	**75.5**	**48.9**	**42.7**	59.3	22.6	26.2	54.2	19.4	19.6	10.6	11.7	12.3	41.3	12.6	12.7
	5	**77.5**	**50.6**	**45.6**	56.0	16.6	12.2	50.8	17.1	13.3	11.3	9.9	15.6	44.1	15.8	16.0
	6	**77.5**	**51.5**	**45.9**	56.0	17.3	11.7	49.6	15.8	12.6	13.5	10.8	18.5	45.7	18.5	18.4

Individual	Trajectory within the homeless system					
X	1	11	3	4	12	6
Y	11	4	11	1		
Z	11	1				

(a)

Individual	p_K			p_{K+1}	E
	1	2	3	4	
X_1	4	12	6	−1	1
X_2	3	4	12	6	0
X_3	11	3	4	12	0
X_4	1	11	3	4	0
Y_1	1	11	1	−1	1
Y_2	11	4	11	1	0
Z_1	−1	11	1	−1	1

(b)

Fig. 2. Data pre–processing illustrative example. (a) Historical data of service assignments (each row correspond to an individual). Services are numerically encoded as in [9]. (b) Each sequence in (a) is sampled with a sliding window of length $K = 3$ to obtain $\mathcal{D}$. The next service, $p_{t_{K+1}}$, received after p_K when $E = 0$ is also shown, and -1 is used to encode the fact that $E = 1$ (i.e., the individual exited the system, and was therefore not subsequently assigned to a service).

4.3. Baselines

We compare PREVISE with the baselines described below. Given the sequence $\mathcal{T}_c$ of client c, both $\hat{p}_{t_{K+1}}$ and $\hat{E}$ are:

- **LOG–REG**: predicted using logistic regression. The sequence is one–hot encoded such that each P_i is converted to a vector $\mathbf{P}_i[p] = \{1 \text{ if } P_i = p \text{ or } 0 \text{ otherwise}\}$, where $p \in \mathcal{P}$. Two logistic regression models are used to predict service type and exit, accordingly. Similar to PREVISE, the individual exits if predicted exit is 1, or is assigned to the next predicted service otherwise.

- **RF**: predicted using random forest. Similar to **LOG–REG**, two random forest classifiers are trained with one–hot encoded sequences obtained from $\mathcal{D}_{train}$ to predict service type and exit accordingly. Both classifiers comprise 100 trees each, with the maximum depth reached when the leaves of each tree cannot be further split. **RF**, like **LOG–REG** and PREVISE, predicts exit if predicted exit is 1, or assigns the next predicted service otherwise.

- **RNDM**: chosen randomly. We consider two variants, namely selecting a service uniformly at random (**UNI–RNDM**), and with a probability proportional to its frequency in the training data (**PA–RNDM**).

5. RESULTS AND ANALYSIS

Table 1 shows the accuracy, mean recall, and mean precision at different values of K (the length of the historical data used for prediction). As K increases, the accuracy of PREVISE in predicting both the next service assignment, $\hat{p}_{t_{K+1}}$, and exit, $\hat{E}$, increases. This indicates that longer sequences improve prediction accuracy by providing more historical data. Nevertheless, although the mean recall and mean precision for $\hat{E}$ and $\hat{p}_{t_{K+1}}$ increases as K increases, however both plateaus after $K = 4$. Overall, we find $K = 4$ to be the best choice with respect to accuracy, mean recall, and mean

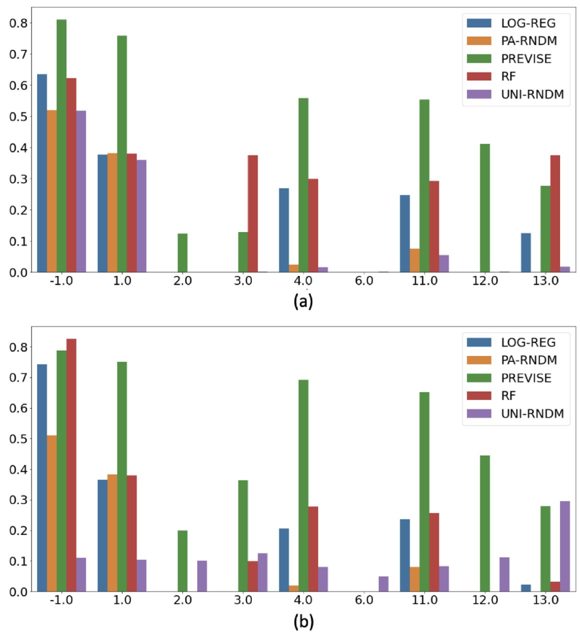

Fig. 3. (a) Precision and (b) Recall plot for service prediction at $K = 4$.

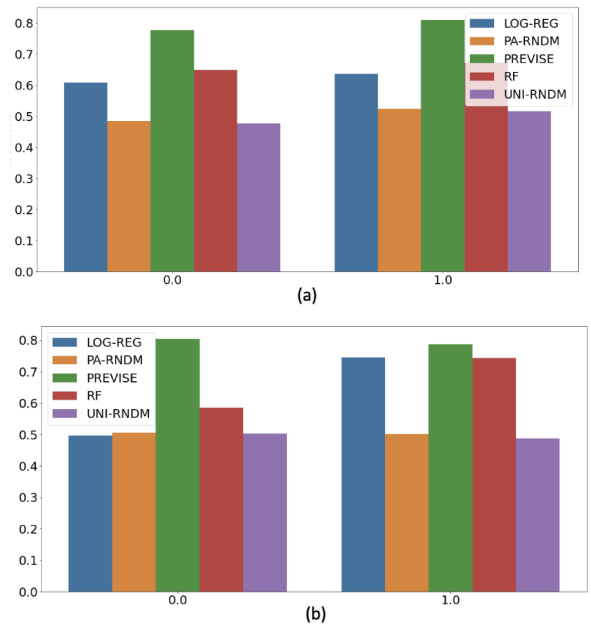

Fig. 4. (a) Precision and (b) Recall plot for exit prediction at $K = 4$.

precision. We also observe that the performance for each service at $K = 4$ is relatively better than the other values of K. Moreover, since an increase in window length K decreases the number of samples in $\mathcal{D}_{train}$ obtained from sequences in $\mathcal{T}$, the performance (accuracy, mean recall, and mean precision) of **LOG–REG** and **RF** drops with higher values of K. Furthermore, the change in $\mathcal{D}_{\sqcup\nabla\dashv\rangle\backslash}$ due to changing K causes the performance of random baselines to fluctuate. This indicates that PREVISE performs significantly better with lesser number of samples. Finally, compared with the baselines, PREVISE achieves a considerable improvement in mean recall as compared to the baselines. Specifically for $K = 4$, it achieves an improvement of 116% (19.8%), 151% (28.1%), 315% (60.6%), and 287% (58%) improvement in service (and exit) prediction compared to **RF**, **LOG–REG** and **UNI–RNDM**, and **PA-RNDM** respectively.

Finally, to better study how PREVISE (and the baselines) behaves for each individual service types and exit, we plot precision and recall for $K = 4$ (Figures 3 and 4). Evidently, PREVISE outperforms the baselines in the task of next service prediction (Figure 3) and exit prediction (Figure 4), while predicting highest number of service types. This result demonstrates the utility of the proposed model; beyond predicting exit and high frequency services (which would be accurate yet meaningless), PREVISE can additionally predict low-frequency service assignments with better precision.

6. CONCLUSION

In this study, we proposed a Bayesian network representation of individuals' navigation of the homeless system. To the best of our knowledge, this work is the first to bring attention to the sequential nature of services that an individual receives over time. Our experiments showed that inferring the next service assignment, or alternatively exit out of the system becomes possible, once the parameters of the Bayesian network are learned. In future work, we plan to study alternative Bayesian network structures to further improve prediction accuracy. We also plan to explore structure learning methods (e.g., [11]) to learn the structure of the network directly from the data. At the same time, our model does not make use of the rich set of features for each individual recorded in our dataset and the time delay between services, which we plan to include in more elaborate Bayesian models.

Finally, we wish to underscore the following crucial considerations before applying our approach to the real–world. The data is bounded to the Capital region of the New York State, and captures only the receipt of services, but not their availability. For instance, capacity constraints (e.g., number of available beds in an emergency shelter) are not currently recorded in the data. However, such factors may have an impact on the next service assignment process. Furthermore, unobserved confounders between service assignments and reentry are also not recorded, but may have caused biased estimates of service assignments, which in turn impact the learned model.

7. REFERENCES

[1] Andrew Alfred Sullivan, "What does it mean to be homeless? how definitions affect homelessness policy," *Urban Affairs Review*, p. 10780874221095185, 2022.

[2] Meghan Henry, Tanya de Sousa, Caroline Roddey, Swati Gayen, Thomas Joe Bednar, and Abt Associates, "The 2021 Annual Homeless Assessment Report (AHAR) to Congress. Part 1: Point-in-time estimates of homelessness," *The US Department of Housing and Urban Development*, 2021.

[3] Charalampos Chelmis, Wenting Qi, Wonhyung Lee, and Stephanie Duncan, "Smart homelessness service provision with machine learning," *Procedia Computer Science*, vol. 185, pp. 9–18, 2021.

[4] Yuan Gao, Sanmay Das, and Patrick Fowler, "Homelessness service provision: a data science perspective," in *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[5] Boyeong Hong, Awais Malik, Jack Lundquist, Ira Bellach, and Constantine E Kontokosta, "Applications of machine learning methods to predict readmission and length-of-stay for homeless families: The case of win shelters in new york city," *Journal of Technology in Human Services*, vol. 36, no. 1, pp. 89–104, 2018.

[6] Amanda Kube, Sanmay Das, and Patrick J Fowler, "Allocating interventions based on predicted outcomes: A case study on homelessness services," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 622–629.

[7] Blake VanBerlo, Matthew AS Ross, Jonathan Rivard, and Ryan Booker, "Interpretable machine learning approaches to prediction of chronic homelessness," *Engineering Applications of Artificial Intelligence*, vol. 102, pp. 104243, 2021.

[8] Geoffrey Messier, Caleb John, and Ayush Malik, "Predicting chronic homelessness: The importance of comparing algorithms using client histories," *Journal of Technology in Human Services*, pp. 1–12, 2021.

[9] United States Department of Housing and Urban Development, "HMIS Data Standards Manual," *Retrieved May 26, 2021, from* `https://www.hudexchange.info/resource/3824/hmis-data-dictionary/`, 2020.

[10] Odo Diekmann and Johan Andre Peter Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, vol. 5, John Wiley & Sons, 2000.

[11] Daphne Koller and Nir Friedman, *Probabilistic graphical models: principles and techniques*, MIT press, 2009.