Supplementary Material

Anonymous submission

Related Studies

Medication recommenders are divided into two categories (*spec.*, instance-based and longitudinal-based) according to whether or not a patient's *past health records* are used in recommending medications at her current visit. In this paper, we are interested in recommending *more-accurate* medications at her current visit. So, we review only longitudinal-based methods that are known to be more accurate than instance-based ones, distinguishing between chronic and acute disease (Shang et al. 2019b; Yang et al. 2021b). Then, we discuss the differences between them and VITA.

GAMENet. Longitudinal-based methods use a patient's past health records in addition to her current visit information (Choi et al. 2016; Shang et al. 2019a,b; Wang et al. 2021a,b; Yang et al. 2021a,b; Bhoi et al. 2021; Li et al. 2022a,b; Ma et al. 2022; Wu et al. 2022a,b). GAMENet (Shang et al. 2019b), one of the most popular methods, consists of an *encoder* that obtains the *patient representation* indicative of the *patient's current health status*, and a *predictor* that finds medications based on the patient representation. Specifically, in the encoder, GAMENet chronologically enumerates *all* past and current visit information of the patient and then uses them as the input of a RNN-based model to obtain her patient representation. Then, based on the patient representation obtained by the encoder, the predictor of GAMENet first obtains the following two medication representations: (Ra) current health-aware medication representation capturing the features of the medications that are necessary at her current visit when considering her current health status; (Rb) current health-relevant past medication representation capturing the probabilities that all medications prescribed at her past visits will be recommended at her current visit when considering her current health status. Finally, GAMENet aggregates these two medication representations and the patient representation to recommend accurate medications to her at the current visit.

Other longitudinal-based methods based on GAMENet. Since then, most longitudinal-based methods (Wang et al. 2021a,b; Yang et al. 2021b; Li et al. 2022a,b; Wu et al. 2022a) have been developed by improving usually one of the two components (*i.e.*, encoder and predictor) of GAMENet to improve accuracy. For instance, SafeDrug (Yang et al. 2021b) and MRSC (Wang et al. 2021b) employ the *encoder of GAMENet as it is*, and improve the predictor of GAMENet. Specifically, in the predictor, SafeDrug enriches the medication representation by utilizing additional external information about the molecular structures of medications to recommend safer medications to patients. MRSC, in the predictor, employs only the medications prescribed in the *three* past visits most-similar to the current visit instead of those in *all* of the patient's past visits to obtain the (Rb) current health-relevant past medication representation.

COGNet. On the other hand, COGNet (Wu et al. 2022b) improves both the encoder and predictor of GAMENet. Specifically, in the encoder, COGNet encodes *only* a patient's current visit information to obtain her current health status, while reflecting the relation between diagnoses (resp. procedures) given at the current visit via a transformer (Vaswani et al. 2017)-based model. Then, in obtaining the (Rb) current health-relevant past medication representation in the predictor, COGNet additionally takes into account the similarity between (i) medications that should be prescribed at her current visit by considering only her current visit information, and (ii) those prescribed in all of her past visits.

Discussions. All the aforementioned methods, except for COGNet, use the encoder of GAMENet as it is. Therefore, they use *all* past visit information, including those that are irrelevant to the patient's current visit, in encoding her current health status. Moreover, these methods aggregate the current and past visit information, considering simply the order of visits in all cases via a RNN-based model. Meanwhile, COGNet does not use *any* past visit information in encoding the patient's current health status, even including those that are relevant to her current one. Also, it does not consider the relevance score between the current and past visit information in obtaining her current health status, since it uses only her current visit information in the encoder. However, the past visit information that is relevant (resp. irrelevant) to the current one is more helpful (resp. rather harmful) in obtaining her current health status, which is the basis for recommending medications. Therefore, VITA aims to improve the encoder of existing methods to obtain a more-accurate her current health status (*i.e.*, patient representation), improving the accuracy of medication recommendation eventually.

Key Notations

Table 1 summarizes the key notations used in this paper.

Table 1: Key notations used in this paper

Notation	Description
$\mathbf{d}^t, \mathbf{p}^t, \mathbf{m}^t$	Multi-hot vectors of diagnoses, procedures, and medications given at a patient's t-th visit
$\mathbf{v}^t, \mathbf{v}^T$	Dense representations of the past t-th visit and current visit T
s^t	Probability of the past t-th visit being selected
$\mathcal{V}^{rel}$	Set of past visits relevant to the current visit
α^t	Relevance score between the current visit and the past t-th visit (w.r.t. dense visit representation)
$\mathbf{q}^T$	Patient representation at the current visit T
$r^t_{m,i}$	Medication-level relevance, which indicates how much each medication i prescribed at each of the patient's past visits t is related to the k-th medication that is necessary at her current visit T
r^t_v	Visit-level relevance between the current visit and the past t-th visit
$\mathbf{p}^T_k$	Current health-aware medication representation, which captures features of the k-th medication that are necessary at the patient's current visit T when considering her current health status
$\bar{\mathbf{p}}^T_k$	Current health-relevant past medication representation, which captures the probabilities that all medications prescribed at the patient's past visits will be recommended as the k-th medication at her current visit T when considering her current health status
$\hat{\mathbf{p}}^T_k$	Probability vector, which indicates the probabilities that all medications will be recommended as the k-th medication at the patient's current visit T

Table 2: Statistics of the MIMIC-III and -IV datasets

Item	MIMIC-III	MIMIC-IV
# patients	5,442	11,723
# visits	14,124	35,850
# diagnoses	1,958	1,979
# procedures	1,430	1,883
# medications	131	152
Avg. / max # of visits	2.59 / 29	3.05 / 41
Avg. / max # of diagnoses per visit	13.69 / 39	10.53 / 39
Avg. / max # of procedures per visit	4.58 / 32	2.93 / 29
Avg. / max # of medications per visit	19.25 / 53	14.61 / 56

Additional Evaluation

Statistics on Datasets

Table 2 provides some statistics on the MIMIC-III and -IV datasets (Johnson et al. 2016, 2023) used in this paper.

Implementation Details

Infrastructures and Implementations. All the experiments were performed on a server with NVIDIA RTX 3070 GPU (8GB memory), 96GB RAM, and Intel Core i9-10900K Processor. Our model is implemented using Pytorch 1.13.0 (Paszke et al. 2019).

Table 3: The effectiveness of VITA's two core ideas (relevant-visit selection and target-aware attention) on the MIMIC-IV dataset. The best result in each column (*i.e.*, each measure) is in bold.

Measures	Jaccard	PRAUC	F1
VITA	**0.5218 ± 0.0007**	**0.7148 ± 0.0024**	**0.6685 ± 0.0004**
VITA-RS	0.5197 ± 0.0002	0.7088 ± 0.0009	0.6590 ± 0.0004
VITA-RS$_{Top\text{-}1}$	0.4803 ± 0.0007	0.6943 ± 0.0005	0.6352 ± 0.0010
VITA-RS$_{sharp}$	0.5199 ± 0.0000	0.7090 ± 0.0014	0.6639 ± 0.0003
VITA-TA$_{avg.}$	0.5130 ± 0.0026	0.7069 ± 0.0016	0.6596 ± 0.0006
VITA-TA$_{RNN}$	0.5167 ± 0.0009	0.7012 ± 0.0012	0.6633 ± 0.0001
VITA-TA$_{attn.}$	0.5196 ± 0.0001	0.7094 ± 0.0004	0.6622 ± 0.0002

Hyperparameters. For a fair comparison, we set the dimensionality of the representations to 64 for all methods including VITA as in (Choi et al. 2016; Zhang et al. 2017; Shang et al. 2019b; Wang et al. 2021b; Yang et al. 2021b; Wu et al. 2022b); we investigated the range of all hyperparameters for all models used in this paper by referring to their original papers (Zhang et al. 2017; Shang et al. 2019b; Wang et al. 2021b; Yang et al. 2021b; Wu et al. 2022b). Then, we used the best hyperparameters of competitors and VITA obtained by extensive grid search on the validation set in the following ranges: learning rate $\in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$; batch size $\in \{1, 2, 4, 8, 16, 32\}$; temperature $\tau_a \in \{0.5, 1, 2, 5, 10, 20\}$ for the target-aware attention of VITA; temperature $\tau_g \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 5, 10\}$ for the Gumbel-softmax of the relevant-visit selection of VITA. For VITA, we set its hyperparameters as follows: learning rate=0.001; batch size=16; temperature τ_a=20; temperature τ_g=0.5 for the MIMIC-III and 0.6 for the MIMIC-IV dataset.

Additional Results for RQs

Results on the MIMIC-IV Dataset. Due to space limitations, except for RQ1, we excluded the results on the MIMIC-IV dataset in RQ2, RQ3, and RQ4 of Section 4.2. We present the results on the MIMIC-IV dataset for RQ2 in Table 3, for RQ3 in Figure 1, and for RQ4 in Figure 2.

As shown in the results, *all* results for each RQ exhibit a trend *very similar* to those observed in the MIMIC-III dataset. These results once again demonstrate the superiority of VITA and the effectiveness of VITA's two core ideas (relevant-visit selection and target-aware attention).

Results of VITA-RS for Different Values of τ_a. In RQ2, to obtain VITA-RS$_{sharp}$, we adjusted the temperature τ_a value of the target-aware attention module in VITA-RS from 1 to 0.2 in decrement of 0.2; the accuracy for each temperature τ_a value is shown in Figure 4.

Then, we finally chose VITA-RS with the best-performing τ_a value for most measures as VITA-RS$_{sharp}$; that is, for the MIMIC-III dataset with a τ_a value of 0.6, and for the MIMIC-IV dataset with a τ_a value of 0.2.

Results from PRAUC and F1 on the MIMIC-III dataset in RQ3. Figure 1 includes the results of four longitudinal-based methods (GAMENet, MRSC, SafeDrug, and COGNet) and their variants equipped with two core

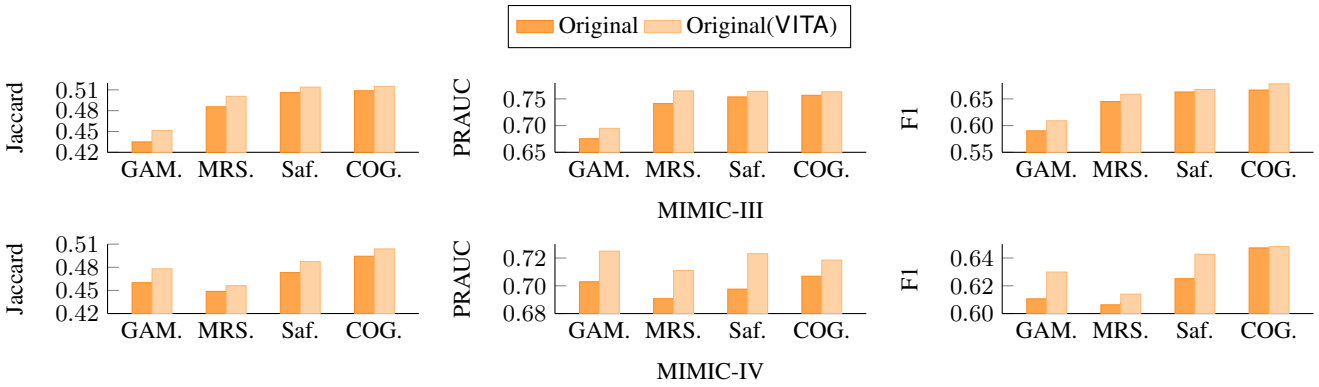

Figure 1: Accuracies of four longitudinal-based methods and their variants equipped with two core ideas of VITA on the MIMIC-III and -IV datasets. The results of each method are represented by its first three letters (*i.e.*, GAM.: GAMENet, MRS.: MRSC, Saf.: SafeDrug, and COG.: COGNet).

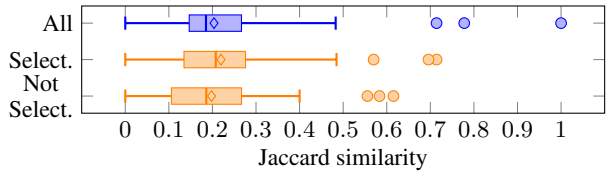

Figure 2: The statistics of the Jaccard similarity between patients' current visit and 'all' past visits, (a) (resp. (b)) the past visits 'selected' (resp. 'not selected') by the relevant-visit selection module on the MIMIC-IV dataset.

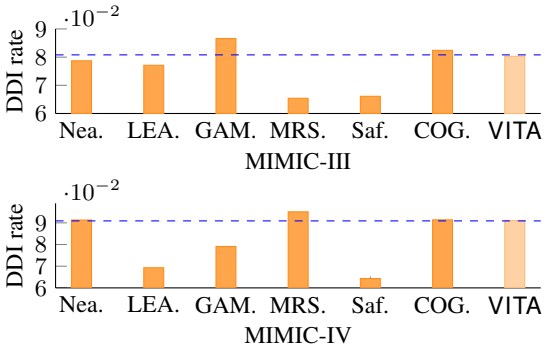

Figure 3: DDI rates of six competitors and VITA on the MIMIC-III and -IV datasets. The results of each method are represented by its first three letters (*i.e.*, Nea.: Nearest, LEA.: LEAP, GAM.: GAMENet, MRS.: MRSC, Saf.: SafeDrug, and COG.: COGNet).

ideas of VITA in terms of PRAUC and F1 on the MIMIC-III dataset for RQ3.

Additional Results in terms of DDI Rate

The DDI rate, which measures the rate of adverse effects between the medications recommended, is defined as follows:

$$\text{DDI rate} = \frac{\sum_{x}^{n}\sum_{t}^{l_x}\sum_{i=1}^{|\hat{\mathcal{M}}^t|}\sum_{j=i+1}^{|\hat{\mathcal{M}}^t|} 1\{(\hat{m}_i^t, \hat{m}_j^t) \in \mathcal{E}_{DDI}\}}{\sum_{x}^{n}\sum_{t}^{l_x}\sum_{i=1}^{|\hat{\mathcal{M}}^t|}\sum_{j=i+1}^{|\hat{\mathcal{M}}^t|} 1}, \tag{1}$$

where n and l_x denote the number of patients in the test set and the number of visits for the x-th patient in the test set, respectively; $\hat{m}_i^t \in \hat{\mathcal{M}}^t$ denotes the i-th recommended medication at t-th visit; $1\{\cdot\}$ is a function that returns 1 when the expression in $\{\cdot\}$ is true, otherwise 0. In other words, the lower the DDI rate is, the lower the possibility of adverse effects between recommended medications is.

In RQ1, we confirmed that VITA recommends more accurate medications than all of its competitors. Here, we measure the DDI rate for VITA and all of its competitors to make sure that they recommend safe medications. The results of all methods, including VITA, are illustrated in Figure 3; the names of the competitors are represented by their first three letters (*i.e.*, Nea.: Nearest, LEA.: LEAP, GAM.: GAMENet, MRS.: MRSC, Saf.: SafeDrug, and COG.: COGNet).

In Figure 3, VITA shows a lower DDI rate than GAMENet and COGNet, but a higher DDI rate than Nearest, LEAP, MRSC, and SafeDrug on the MIMIC-III dataset. On the MIMIC-IV dataset, VITA shows a lower DDI rate than Nearest, MRSC, and COGNet, but a higher DDI rate than LEAP, GAMENet, and SafeDrug. Here, we should note that there is an inevitable tradeoff between correctly identifying all medications in the medication sets in the test set and minimizing the DDI rate (*i.e.*, recommending medications that minimize the occurrence of adverse effects). This is because the average DDI rate between all medications in the medication sets in the test set (*i.e.*, between medications actually allowed to be prescribed) is 0.0808 (resp. 0.0909) on the MIMIC-III dataset (resp. the MIMIC-IV dataset), which is denoted by the blue dashed line in Figure 3. Therefore, to recommend the ground truth answers to patients, the DDI rate becomes inevitably close to 0.0808 (resp. 0.0909). Even though VITA shows a higher DDI rate than Nearest, LEAP, MRSC, and SafeDrug on the MIMIC-III dataset (resp. LEAP, GAMENet, and SafeDrug on the

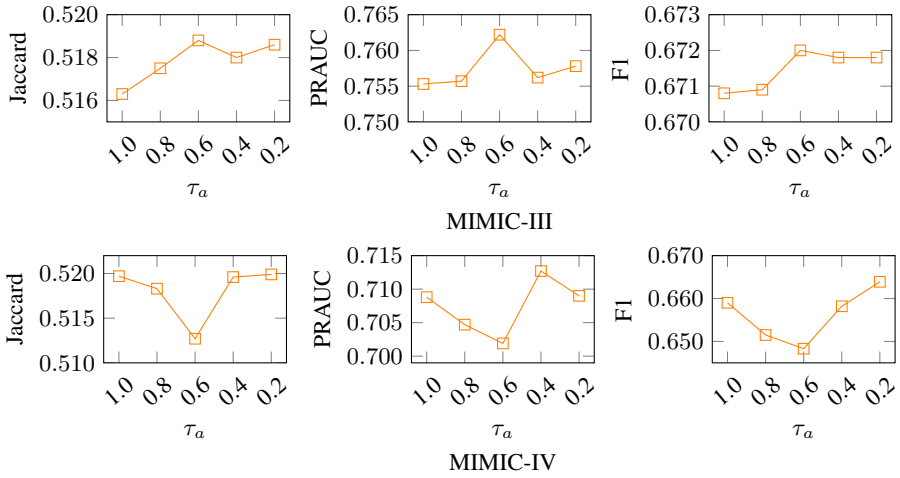

Figure 4: The effect of τ_a on the accuracies of VITA-RS on the MIMIC-III and -IV datasets.

MIMIC-IV dataset), the main goal of VITA is to recommend accurate medications. Thus, it provides more accurate recommendations than *all* of its competitors, including Nearest, LEAP, MRSC, and SafeDrug on the MIMIC-III dataset (resp. LEAP, GAMENet, and SafeDrug on the MIMIC-IV dataset), and it shows a lower DDI rate (0.0803) than the average DDI rate (0.0808) (resp. a DDI rate (0.0910) comparable to the average DDI rate (0.0909)) between all medications in the medication sets in the test set on the MIMIC-III dataset (resp. the MIMIC-IV dataset). Accordingly, VITA recommends not only the most accurate (as shown in RQ1) but also safe medications to patients.

References

Bhoi, S.; Lee, M.; Hsu, W.; Fang, A. H. S.; and Tan, N. C. 2021. Personalizing Medication Recommendation with a Graph-Based Approach. *ACM Transactions on Information Systems*, 40: 1–23.

Choi, E.; Bahadori, M. T.; Sun, J.; Kulas, J.; Schuetz, A.; and Stewart, W. F. 2016. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 3504–3512.

Johnson, A.; Bulgarelli, L.; Pollard, T.; Horng, S.; Celi, L. A.; and Mark, R. 2023. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data*, 10: 1.

Johnson, A. E.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3: 1–9.

Li, R.; Wang, J.; Lin, H.; Lin, Y.; Lu, H.; and Yang, Z. 2022a. Patient Condition Change Network for Safe Medication Recommendation. In *Proceedings of the International Conference on Bioinformatics and Biomedicine, (BIBM)*, 1046–1051.

Li, X.; Zhang, Y.; Wang, J.; Lu, M.; and Lin, H. 2022b. Knowledge-Enhanced Dual Graph Neural Network for Robust Medicine Recommendation. In *Proceedings of the International Conference on Bioinformatics and Biomedicine, (BIBM)*, 477–482.

Ma, X.; Wang, Y.; Chu, X.; Ma, L.; Tang, W.; Zhao, J.; Yuan, Y.; and Wang, G. 2022. Patient Health Representation Learning via Correlational Sparse Prior of Medical Features. *IEEE Transactions on Knowledge and Data Engineering*, 1–14.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Shang, J.; Ma, T.; Xiao, C.; and Sun, J. 2019a. Pre-training of Graph Augmented Transformers for Medication Recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI)*, 5953–5959.

Shang, J.; Xiao, C.; Ma, T.; Li, H.; and Sun, J. 2019b. GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination. In *proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1126–1133.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 5998–6008.

Wang, Y.; Chen, W.; Pi, D.; Yue, L.; Wang, S.; and Xu, M. 2021a. Self-Supervised Adversarial Distribution Regularization for Medication Recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI)*, 3134–3140.

Wang, Y.; Chen, W.; Pi, D.; Yue, L.; Xu, M.; and Li, X. 2021b. Multi-hop Reading on Memory Neural Network with Selective Coverage for Medication Recommendation. In *Proceedings of the ACM International Conference on Information & Knowledge Management (ACM CIKM)*, 2020–2029.

Wu, J.; Qian, B.; Li, Y.; Gao, Z.; Ju, M.; Yang, Y.; Zheng, Y.; Gong, T.; Li, C.; and Zhang, X. 2022a. Leveraging Multiple Types of Domain Knowledge for Safe and Effective Drug Recommendation. In *Proceedings of the ACM International Conference on Information & Knowledge Management (ACM CIKM)*, 2169–2178.

Wu, R.; Qiu, Z.; Jiang, J.; Qi, G.; and Wu, X. 2022b. Conditional Generation Net for Medication Recommendation. In *Proceedings of the international ACM conference on World Wide Web (ACM WWW)*.

Yang, C.; Xiao, C.; Glass, L.; and Sun, J. 2021a. Change Matters: Medication Change Prediction with Recurrent Residual Networks. In *Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI)*, 3728–3734.

Yang, C.; Xiao, C.; Ma, F.; Glass, L.; and Sun, J. 2021b. SafeDrug: Dual Molecular Graph Encoders for Recommending Effective and Safe Drug Combinations. In *Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI)*, 3735–3741.

Zhang, Y.; Chen, R.; Tang, J.; Stewart, W. F.; and Sun, J. 2017. LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity. In *Proceedings of the ACM International Conference on Knowledge Discovery & Data Mining (ACM KDD)*, 1315–1324.