CRAFT: COUNTERFACTUAL UTILITY-GUIDED DYNAMIC REASONING FOR TRADITIONAL CHINESE MEDICINE PRESCRIPTION RECOMMENDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Traditional Chinese Medicine (TCM) prescription recommendation is commonly formulated as static multi-label prediction, overlooking how therapeutic demands evolve during prescription construction. To address this limitation, we propose CRAFT, a Counterfactual utility-guided dynamic Reasoning framework with Adaptive Fusion for TCM prescription recommendation. CRAFT first performs hierarchical graph reasoning over symptom–symptom, symptom–herb, and herb–herb relations to capture structured clinical dependencies. It then reformulates prescription generation as an order-free dynamic decision process, where each step evaluates candidate herbs according to their neural compatibility and Counterfactual Marginal Utility (CMU). CMU performs candidate-level what-if evaluation by estimating the additional coverage of residual therapeutic demand if a candidate herb were added to the current partial prescription. CRAFT adaptively integrates this utility into neural decision scores and progressively constructs the prescription until a learned STOP action is selected. Experiments on two benchmark TCM prescription datasets demonstrate consistent improvements over competitive baselines, while further analyses show that CMU provides informative candidate-level signals throughout the dynamic reasoning process. The source code is publicly available at `https://anonymous.4open.science/r/CRAFT-DAD6/`.

1 INTRODUCTION

Traditional Chinese Medicine (TCM) prescription recommendation aims to generate appropriate herb combinations from patients' observed symptoms. Unlike conventional drug recommendation, a TCM prescription is typically composed of multiple herbs that jointly address different therapeutic demands (Zhang & Tan, 2025). Therefore, effective prescription recommendation requires not only identifying symptom–herb associations, but also modeling how the therapeutic contribution of each herb changes as the prescription is progressively constructed.

Recent advances in artificial intelligence have promoted computational TCM prescription recommendation. Early studies mainly formulated this task as multi-label classification, directly predicting a set of herbs from symptom representations. Subsequent studies introduced topic models and sequence-based approaches to capture latent symptom–prescription patterns and dependencies among herbs. More recently, graph-based methods have incorporated symptom–symptom, symptom–herb, and herb–herb relations to exploit structured clinical dependencies and improve prescription representations.

Despite their effectiveness, most existing approaches still formulate prescription recommendation as a largely *static* prediction problem. As illustrated in Fig. 1(a), conventional methods typically map patient symptoms to herb scores and directly return a fixed top-K set in a one-shot manner. This paradigm has three major limitations. First, *one-shot prediction* ignores the progressive nature of prescription construction, where previously selected herbs should affect subsequent decisions. Second, a *fixed top-K* strategy applies the same prescription size across patients despite varying symptom patterns and therapeutic complexity. Third, existing methods generally lack *utility aware-*

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

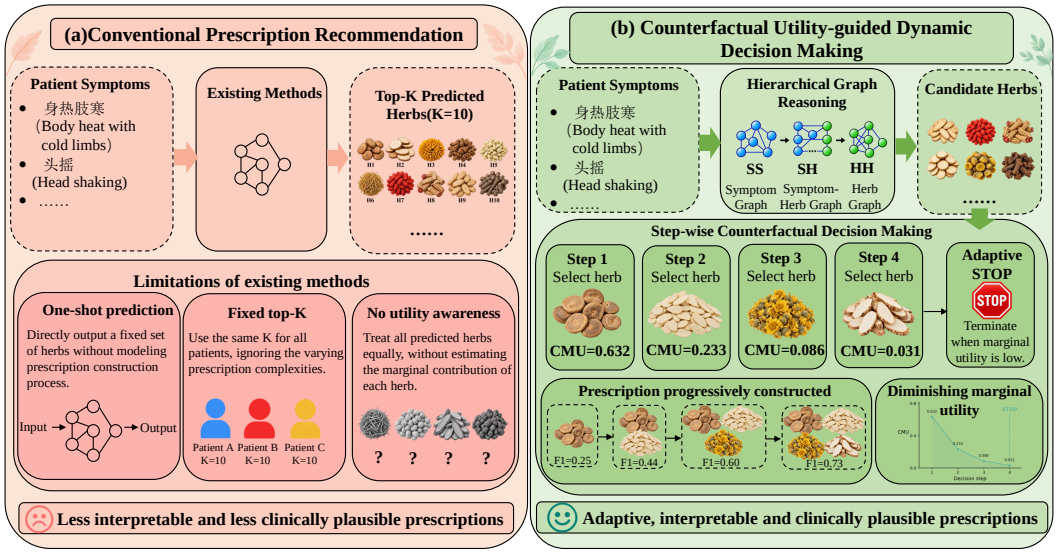


Figure 1: Comparison of conventional TCM prescription recommendation and CRAFT. Existing methods perform one-shot top-K prediction, whereas CRAFT progressively constructs prescriptions by dynamically evaluating herb marginal utility and adaptively terminating with STOP.

ness: predicted herbs are mainly ranked by static relevance scores without explicitly estimating the additional therapeutic contribution of adding each herb to the current prescription.

Prescription construction can exhibit a diminishing-return pattern: as herbs are progressively selected, therapeutic demands may become increasingly covered, reducing the additional contribution of subsequent candidates (Zhang et al., 2026; Cao & Zhang, 2022). Therefore, the relevance of a candidate herb should depend not only on the observed symptoms, but also on the current partial prescription and the remaining therapeutic demand (Yang et al., 2024). This raises two challenges. First, conventional sequence generation requires a predefined herb order, whereas prescriptions are inherently set-structured (Madaan et al., 2022; Yang et al., 2019). Second, neural decision scores do not explicitly quantify the additional therapeutic benefit of a candidate given the current prescription (Kleinberg et al., 2022; Yang et al., 2023). An effective model should therefore construct prescriptions progressively while estimating the *marginal utility* of candidate herbs. To address these limitations, we propose CRAFT, a Counterfactual utility-guided dynamic Reasoning framework with Adaptive Fusion for TCM prescription recommendation. As illustrated in Fig. 1(b), CRAFT reformulates prescription generation as an order-free dynamic decision process. At each step, candidate herbs are evaluated according to neural compatibility and Counterfactual Marginal Utility (CMU), which estimates their additional coverage of residual therapeutic demand. A learned STOP action enables patient-specific termination, while adaptive fusion integrates local dynamic evidence with holistic prescription-level compatibility. Our main contributions are summarized as follows:

1. We reformulate TCM prescription recommendation from static symptom-to-herb prediction into an order-free dynamic decision process, where herbs are progressively selected according to the evolving prescription state and construction is adaptively terminated through a learned STOP action.

2. We introduce Counterfactual Marginal Utility (CMU), a candidate-level what-if utility that estimates the additional coverage of residual therapeutic demand if a candidate herb were added to the current partial prescription. An adaptive utility fusion mechanism dynamically integrates CMU with neural compatibility scores without requiring additional utility supervision.

3. We evaluate CRAFT on two benchmark TCM prescription datasets. Extensive experiments demonstrate consistent improvements over competitive baselines, while utility analyses and counterfactual interventions further characterize the behavior of the proposed dynamic reasoning process.

2 Related Work

2.1 Sequential and Dynamic Recommendation

Sequential recommendation models capture dependencies among successive decisions by conditioning predictions on previous interactions or outputs (Hidasi et al., 2015; Kang & McAuley, 2018; Boka et al., 2024). More recently, reinforcement-learning-based recommendation formulates recommendation as a dynamic decision process, where actions are selected according to evolving states and long-term utility (Zou et al., 2019; Chen et al., 2019; Afsar et al., 2022; Chen et al., 2023). Adaptive fusion methods further combine multiple predictive signals according to the current state rather than fixed scoring rules (Zhou et al., 2018; Cheng et al., 2020; Cao et al., 2025). These studies highlight the importance of state-dependent decision making, but mainly focus on user–item interactions and behavioral rewards. In contrast, CRAFT considers the evolving state of a therapeutic herb set, where the value of a candidate depends on the herbs already selected and the therapeutic demand that remains uncovered.

2.2 TCM Prescription Recommendation

TCM prescription recommendation has been formulated as multi-label prediction (Li & Yang, 2019; Jin et al., 2020b), topic modeling (Yao et al., 2018), and sequence generation (Wang et al., 2019; Rong et al., 2022). Recent graph-based approaches further model structured symptom–herb dependencies using hierarchical, hypergraph, contrastive, and diffusion-based representations (Hu et al., 2025; Lu et al., 2025; Zhou et al., 2026). Clinical reasoning has also been incorporated by PresRecST (Dong et al., 2024), which models successive diagnostic and treatment stages, and SDPR (Yue et al., 2025), which captures symptom–syndrome–therapeutic method–herb interactions. TCM-KLLaMA (Zhuang et al., 2025) further combines TCM knowledge graphs with large language models. Sequential methods condition predictions on previously generated herbs (Zhao et al., 2023; Zheng et al., 2025), but do not explicitly model how candidate utility changes with the evolving prescription. In contrast, CRAFT progressively updates the prescription state and re-evaluates each candidate according to its marginal contribution to the remaining therapeutic demand.

3 Method

Given observed symptoms, CRAFT formulates TCM prescription recommendation as an order-free dynamic decision process rather than one-shot multi-label prediction. As illustrated in Fig. 2, CRAFT consists of three stages: (i) hierarchical graph reasoning to encode structured symptom–herb knowledge; (ii) counterfactual prescription reasoning to estimate residual therapeutic demand and candidate-level marginal utility; and (iii) dynamic prescription construction with a learned STOP action. A complementary holistic branch provides global prescription evidence, which is adaptively fused with the dynamic reasoning scores.

3.1 Problem Definition

Let $\mathcal{S} = \{s_1, \ldots, s_{N_s}\}$ and $\mathcal{H} = \{h_1, \ldots, h_{N_h}\}$ denote the symptom and herb sets. For each patient, the observed symptoms and target prescription are represented by multi-hot vectors $\mathbf{x} \in \{0,1\}^{N_s}$ and $\mathbf{y} \in \{0,1\}^{N_h}$, respectively, with ground-truth herb set $\mathcal{Y} = \{h \in \mathcal{H} \mid y_h = 1\}$. Instead of predicting $\mathcal{Y}$ in one shot, CRAFT maintains a partial prescription $\mathcal{P}_t \subseteq \mathcal{H}$ at step t and selects either an unselected herb or STOP. Since selected herbs are accumulated as a set, no predefined generation order is imposed.

3.2 Stage I: Hierarchical Graph Reasoning

As shown in Stage I of Fig. 2, CRAFT models structured clinical dependencies through $\mathcal{G} = \{\mathcal{G}_{SS}, \mathcal{G}_{SH}, \mathcal{G}_{HH}\}$, representing symptom co-occurrence, symptom–herb association, and herb co-occurrence, respectively. All graphs are constructed exclusively from the training set.

Let $\mathbf{S}^0 \in \mathbb{R}^{N_s \times d}$ and $\mathbf{H}^0 \in \mathbb{R}^{N_h \times d}$ denote learnable symptom and herb embeddings. Symptom information is first propagated over the SS graph and combined with the original embeddings through

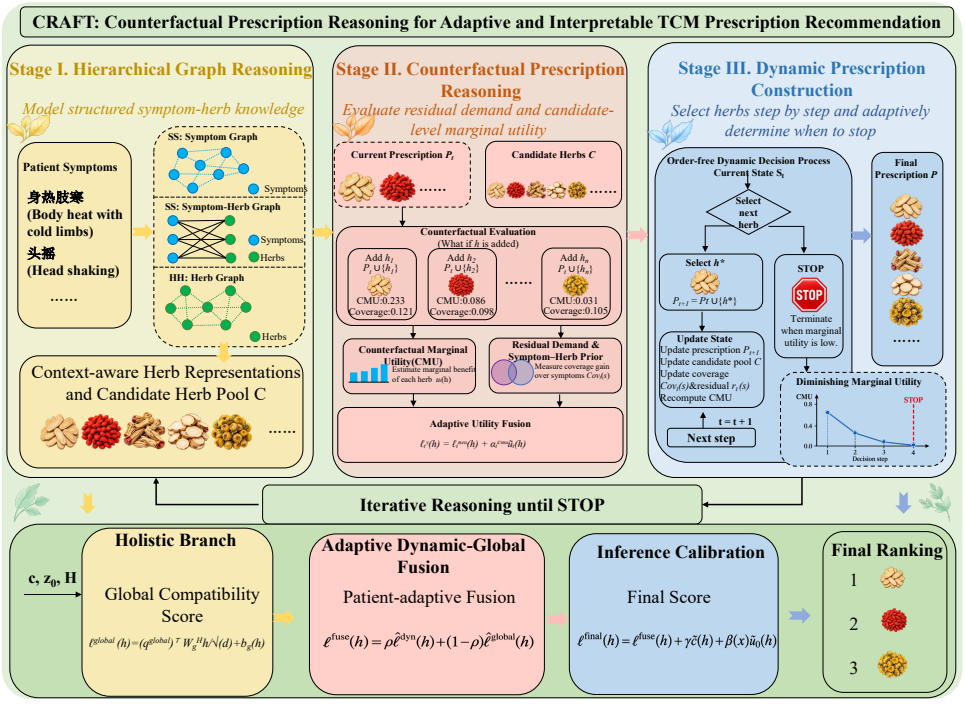


Figure 2: Overview of CRAFT. Given observed symptoms, CRAFT performs Hierarchical Graph Reasoning over symptom–symptom (SS), symptom–herb (SH), and herb–herb (HH) relations, followed by Counterfactual Prescription Reasoning with Counterfactual Marginal Utility (CMU) and Dynamic Prescription Construction. A holistic branch provides complementary global evidence, and prescription construction terminates through a learned STOP action.

gated residual fusion:

$$\mathbf{S}^{SS} = \mathrm{GAT}_{SS}(\mathbf{S}^0, \mathcal{G}_{SS}), \qquad \widetilde{\mathbf{S}} = \mathbf{G}^S \odot \mathbf{S}^{SS} + (1 - \mathbf{G}^S) \odot \mathbf{S}^0, \tag{1}$$

where $\mathbf{G}^S = \sigma(\mathbf{W}_g^S[\mathbf{S}^0 \| \mathbf{S}^{SS}])$. The refined symptoms are propagated to herbs through symptom-to-herb attention, followed by HH graph reasoning:

$$(\mathbf{S}^{SH}, \mathbf{H}^{SH}) = \mathrm{HGAT}_{SH}(\widetilde{\mathbf{S}}, \mathbf{H}^0, \mathcal{G}_{SH}), \qquad \mathbf{H}^{HH} = \mathrm{GAT}_{HH}(\mathbf{H}^{SH}, \mathcal{G}_{HH}). \tag{2}$$

The final herb representation is $\mathbf{H} = f_h([\mathbf{H}^0 \| \mathbf{H}^{SH} \| \mathbf{H}^{HH}])$. For each patient, active symptom representations are mean-pooled to obtain $\mathbf{s}^{SS}$ and $\mathbf{s}^{SH}$. We then construct the clinical context and initial prescription state as $\mathbf{c} = f_c([\mathbf{s}^{SS} \| \mathbf{s}^{SH}])$ and $\mathbf{z}_0 = f_z([\mathbf{c} \| \mathbf{s}^{SS}])$. Together, $\mathbf{H}$, $\mathbf{c}$, and $\mathbf{z}_0$ initialize the subsequent dynamic reasoning process.

3.3 Stage II: Counterfactual Prescription Reasoning

At each step, CRAFT evaluates candidate herbs relative to the current partial prescription $\mathcal{P}_t$. Since explicit herb-level therapeutic annotations are unavailable, symptom–herb associations estimated only from the training set serve as a data-driven proxy for therapeutic coverage. Accordingly, residual therapeutic demand represents observed symptom evidence not yet covered by the current prescription.

Symptom–Herb Prior and Residual Demand. To reduce noise from sparse symptom–herb frequencies, we construct an Empirical-Bayes prior

$$C_{sh} = \frac{N(s,h) + \kappa p(h)}{N(s) + \kappa}, \tag{3}$$

where $N(s,h)$ and $N(s)$ denote co-occurrence and symptom frequencies, $p(h)$ is the global herb prior, and κ controls shrinkage. Given $\mathcal{P}_t$, the accumulated coverage and residual demand of symptom s are

$$\mathrm{Cov}_t(s) = 1 - \exp\left(-\sum_{h\in\mathcal{P}_t} C_{sh}\right), \qquad r_t(s) = x_s[1 - \mathrm{Cov}_t(s)]. \tag{4}$$

The residual evidence is summarized as

$$\mathbf{q}_t^r = f_r\left(\frac{\sum_s r_t(s)\mathbf{S}^{SH}_{s,:}}{\max(\sum_s r_t(s), \epsilon)}\right).$$

Counterfactual Marginal Utility. For candidate herb h, CRAFT evaluates the hypothetical prescription $\mathcal{P}_t \cup \{h\}$. Here, "counterfactual" denotes a candidate-level what-if evaluation rather than causal counterfactual inference. The symptom-level marginal coverage gain is $\Delta_t(s,h) = r_t(s)(1-\exp(-C_{sh}))$. Counterfactual Marginal Utility (CMU) is then

$$u_t(h) = \frac{\sum_{s=1}^{N_s} \Delta_t(s,h)}{\max(\sum_s x_s, 1)}. \tag{5}$$

As prescription construction proceeds, covered symptom evidence yields smaller $r_t(s)$, naturally reducing the utility of redundant candidates. This explicitly models diminishing marginal returns. To align CMU with neural score scales, we standardize it within each patient as $\widetilde{u}_t(h) = [u_t(h) - \mu_t]/\max(\sigma_t, \epsilon)$.

Adaptive Utility Fusion. CRAFT combines the patient context, evolving prescription state, and residual evidence through

$$\mathbf{g}_t^r = \sigma(\mathbf{W}_g^R[\mathbf{c}\|\mathbf{z}_t\|\mathbf{q}_t^r])$$

and

$$\mathbf{q}_t = \mathbf{W}_z\mathbf{z}_t + \mathbf{W}_c\mathbf{c} + \mathbf{g}_t^r \odot \mathbf{W}_r\mathbf{q}_t^r.$$

The resulting neural compatibility is $\ell_t^{\mathrm{neu}}(h) = \mathbf{q}_t^\top \mathbf{W}_h\mathbf{h}/\sqrt{d} + b(\mathbf{h})$. Rather than assigning CMU a fixed weight, CRAFT predicts $\alpha_t^{\mathrm{cmu}} = \mathrm{Softplus}(f_\alpha([\mathbf{c}\|\mathbf{z}_t\|\mathbf{q}_t^r]))$. The final dynamic score is

$$\ell_t^{\mathrm{dyn}}(h) = \ell_t^{\mathrm{neu}}(h) + \alpha_t^{\mathrm{cmu}}\widetilde{u}_t(h). \tag{6}$$

Thus, the influence of explicit utility evidence varies across patients and decision steps without requiring additional CMU supervision.

3.4 Stage III: Dynamic Prescription Construction

At step t, the action space is $\mathcal{A}_t = (\mathcal{H}\setminus\mathcal{P}_t)\cup\{\mathrm{STOP}\}$. When herb h_t is selected, $\mathcal{P}_{t+1} = \mathcal{P}_t\cup\{h_t\}$, and the prescription state is updated as

$$\mathbf{z}_{t+1} = \mathrm{LN}(\mathbf{z}_t + \mathrm{GRU}(\mathbf{h}_t, \mathbf{z}_t)). \tag{7}$$

Termination is scored by $\ell_t^{\mathrm{stop}} = f_{\mathrm{stop}}([\mathbf{c}\|(\mathbf{z}_t + \mathbf{q}_t^r)])$. The STOP action competes directly with candidate herbs in the same decision space. After each selected herb, the prescription state, coverage, residual demand, and CMU are recomputed; selecting STOP terminates the rollout.

3.5 Dynamic–Holistic Prescription Modeling

Step-wise reasoning emphasizes local marginal contributions, whereas a complete prescription also requires global patient–herb compatibility. CRAFT therefore introduces a complementary holistic branch with $\mathbf{q}^{\mathrm{global}} = f_g([\mathbf{c}\|\mathbf{z}_0])$ and

$$\ell^{\mathrm{global}}(h) = \frac{(\mathbf{q}^{\mathrm{global}})^\top \mathbf{W}_g^H\mathbf{h}}{\sqrt{d}} + b_g(\mathbf{h}).$$

Because candidate relevance may emerge at different stages, dynamic trajectory evidence is aggregated by $\ell^{\mathrm{dyn}}(h) = \max_{t<T}\ell_t^{\mathrm{dyn}}(h)$. The dynamic and global scores are independently standardized to $\widehat{\ell}^{\mathrm{dyn}}$ and $\widehat{\ell}^{\mathrm{global}}$, and combined using the patient-specific coefficient $\rho = \sigma(f_\rho([\mathbf{c}\|\mathbf{z}_0]))$:

$$\ell^{\mathrm{fuse}}(h) = \rho\,\widehat{\ell}^{\mathrm{dyn}}(h) + (1-\rho)\,\widehat{\ell}^{\mathrm{global}}(h). \tag{8}$$

This fusion preserves both step-wise marginal evidence and holistic prescription compatibility.

3.6 TRAINING

CRAFT is trained end-to-end using an order-free sequential decision objective and a prescription-level multi-label objective.

Set-valued Sequential Decision Loss. At step t, the remaining ground-truth herbs are $\mathcal{R}_t = \mathcal{Y} \setminus \mathcal{P}_t$. Instead of imposing a unique generation order, the valid action set is $\mathcal{V}_t = \mathcal{R}_t$ if $\mathcal{R}_t \neq \emptyset$, and $\mathcal{V}_t = \{\text{STOP}\}$ otherwise. The sequential objective maximizes the probability mass assigned to all valid actions:

$$\mathcal{L}_{\text{seq}} = -\frac{1}{T}\sum_{t=0}^{T-1} \log \frac{\sum_{a\in\mathcal{V}_t} \exp(\ell_t(a))}{\sum_{a\in\mathcal{A}_t} \exp(\ell_t(a))}. \tag{9}$$

During training, the highest-scoring remaining ground-truth herb is used only to update the recurrent state. Since all herbs in $\mathcal{R}_t$ are simultaneously valid next actions, the supervision remains permutation-invariant and introduces no predefined herb order.

Prescription-level Loss. The fused scores additionally receive complete-prescription supervision via frequency-weighted binary cross-entropy:

$$\mathcal{L}_{\text{pres}} = -\frac{1}{N_h}\sum_{h=1}^{N_h} \left[w_h y_h \log \sigma(\ell^{\text{fuse}}(h)) + (1-y_h)\log(1-\sigma(\ell^{\text{fuse}}(h))) \right], \tag{10}$$

where w_h is a frequency-aware positive-class weight. The overall objective is

$$\mathcal{L} = \lambda_{\text{seq}}\mathcal{L}_{\text{seq}} + \lambda_{\text{pres}}\mathcal{L}_{\text{pres}}. \tag{11}$$

CMU introduces no auxiliary supervision; its contribution is learned end-to-end through the prescription objectives.

3.7 INFERENCE

At inference, CRAFT performs dynamic rollout until STOP or the maximum number of steps is reached. We further exploit two non-parametric trajectory signals: the normalized initial CMU $\widetilde{u}_0(h)$ and a consensus score $\widetilde{c}(h)$ measuring how consistently a candidate appears among top-ranked herbs across active steps. The final ranking score is

$$\ell^{\text{final}}(h) = \ell^{\text{fuse}}(h) + \gamma\widetilde{c}(h) + \beta(\mathbf{x})\widetilde{u}_0(h), \tag{12}$$

where γ controls consensus calibration and $\beta(\mathbf{x})$ adaptively adjusts CMU calibration according to prediction uncertainty. Specifically, with $d(\mathbf{x}) = \ell^{(1)} - \frac{1}{K}\sum_{k=1}^{K}\ell^{(k)}$, we use

$$\beta(\mathbf{x}) = \beta_0 \ \text{clip}\left(1 - \xi\frac{d(\mathbf{x})-\mu_d}{\max(\sigma_d,\epsilon)}, b_{\min}, b_{\max}\right).$$

Thus, uncertain predictions receive stronger support from the explicit therapeutic prior. This calibration introduces no additional trainable parameters. The dynamic rollout itself produces an adaptive-length prescription. For comparison with prior work, we additionally rank the complete herb vocabulary by $\ell^{\text{final}}(h)$ and report fixed-cutoff metrics at $K \in \{5, 10, 20\}$; these cutoffs are used only for benchmark evaluation and do not constrain the internal prescription length.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Dataset. We evaluate CRAFT on two TCM prescription datasets. TCM_PD (Yang et al., 2022) contains 33,765 prescriptions with 390 symptoms and 805 herbs, split into 20,709 training and 13,056 test samples. TCM_SMGCN (Yao et al., 2018) contains 26,360 prescriptions with 360 symptoms and 753 herbs, including 22,917 training and 3,443 test samples. Prescriptions are represented as multi-hot symptom and herb vectors. To avoid information leakage, all SS, SH, and HH graphs and CMU coverage statistics are constructed solely from the training set.

Baselines. We compare CRAFT with six representative TCM prescription recommendation methods: SMGCN (Jin et al., 2020a), KDHR (Yang et al., 2022), SMRGAT (Yang & Ding, 2023), SCEIKG (Liu et al., 2023), PresRecST (Dong et al., 2024), and TCMRGCL (Hu et al., 2025). These baselines cover graph convolution, knowledge-enhanced recommendation, graph attention, condition-evolution modeling, progressive clinical reasoning, and hierarchical contrastive learning. All methods are evaluated under the same data split and evaluation protocol for fair comparison.

Evaluation Metrics. Following prior TCM prescription recommendation studies, we report Precision@K, Recall@K, and F1-score@K for $K \in \{5, 10, 20\}$. Precision measures the proportion of recommended herbs that appear in the ground-truth prescription, while recall measures the proportion of ground-truth herbs successfully retrieved. F1-score is their harmonic mean. All metrics are computed over the full candidate herb set.

4.2 Overall Performance

Table 1: Performance on TCM_PD and TCM_SMGCN compared with the baseline models.

Dataset	Model	Precision			Recall			F1-score		
		P@5	P@10	P@20	R@5	R@10	R@20	F1@5	F1@10	F1@20
TCM_PD	SMGCN	0.1933	0.1568	0.1153	0.1265	0.2031	0.3027	0.1529	0.1770	0.1670
	KDHR	0.2138	0.1660	0.1251	0.1510	0.2284	0.3414	0.1770	0.1922	0.1832
	SMRGAT	0.2461	0.1873	0.1343	0.1821	0.2777	0.3970	0.2093	0.2237	0.2007
	SCEIKG	0.2482	0.1942	0.1446	0.1722	0.2691	0.3973	0.2033	0.2256	0.2119
	PresRecST	0.2238	0.1749	0.1290	0.1512	0.2338	0.3465	0.1805	0.2001	0.1879
	TCMRGCL	0.2543	0.1984	0.1477	0.1852	0.2841	0.4192	0.2143	0.2337	0.2184
	CRAFT	**0.2820**	**0.2216**	**0.1607**	**0.2129**	**0.3347**	**0.4853**	**0.2427**	**0.2667**	**0.2414**
	Improve	10.89%	11.69%	8.80%	14.96%	17.81%	15.77%	13.25%	14.12%	10.53%
TCM_SMGCN	SMGCN	0.2878	0.2287	0.1664	0.2070	0.3229	0.4631	0.2409	0.2677	0.2449
	KDHR	0.2698	0.2103	0.1537	0.1932	0.2967	0.4299	0.2251	0.2461	0.2264
	SMRGAT	0.2710	0.2157	0.1574	0.1892	0.3009	0.4334	0.2228	0.2512	0.2309
	SCEIKG	0.2735	0.2189	0.1621	0.1921	0.3066	0.4482	0.2257	0.2555	0.2381
	PresRecST	0.2807	0.2248	0.1656	0.1988	0.3173	0.4618	0.2327	0.2632	0.2438
	TCMRGCL	0.2884	0.2310	0.1689	**0.2082**	0.3266	0.4707	0.2418	0.2706	0.2486
	CRAFT	**0.2944**	**0.2325**	**0.1702**	0.2073	**0.3275**	**0.4793**	**0.2433**	**0.2719**	**0.2512**
	Improve	2.08%	0.65%	0.77%	-0.43%	0.28%	1.83%	0.62%	0.48%	1.05%

As shown in Table 1, CRAFT achieves strong and consistent performance on both datasets. On TCM_PD, CRAFT outperforms the strongest baseline, TCMRGCL, across all nine evaluation metrics. In particular, F1@5, F1@10, and F1@20 increase from 0.2143, 0.2337, and 0.2184 to 0.2427, 0.2667, and 0.2414, corresponding to relative improvements of 13.25%, 14.12%, and 10.53%, respectively. The gains are particularly pronounced in recall, with R@5, R@10, and R@20 improved by 14.96%, 17.81%, and 15.77%, respectively. On TCM_SMGCN, where the baseline performance is substantially stronger, CRAFT still achieves the best results on eight of the nine metrics. Specifically, it improves F1@5, F1@10, and F1@20 from 0.2418, 0.2706, and 0.2486 to 0.2433, 0.2719, and 0.2512, respectively, while achieving the second-best R@5 with only a marginal difference from the best baseline. These results demonstrate that CRAFT consistently improves the precision–recall trade-off across different recommendation sizes and generalizes effectively across datasets with different performance characteristics.

4.3 Ablation Study

Ablation Study. Table 2 shows that removing individual components generally degrades performance on both datasets, particularly at smaller cutoffs. Hierarchical graph reasoning and CMU consistently improve top-ranked recommendations, while dynamic decision reasoning is especially important on TCM_SMGCN, where its removal reduces F1@5 from 0.2433 to 0.2098 and F1@10 from 0.2719 to 0.2594. Adaptive dynamic–global fusion and inference calibration provide further gains. Although removing therapeutic coverage on TCM_SMGCN or dynamic reasoning on TCM_PD slightly improves F1@20, both reduce performance at smaller K, suggesting that CRAFT's dynamic components primarily improve the prioritization of highly relevant herbs.

Table 2: Ablation study with different components removed on TCM_PD and TCM_SMGCN.

Dataset	Model	Precision			Recall			F1-score		
		P@5	P@10	P@20	R@5	R@10	R@20	F1@5	F1@10	F1@20
TCM_PD	Full Model	**0.2820**	**0.2216**	0.1607	**0.2129**	**0.3347**	0.4853	**0.2427**	**0.2667**	0.2414
	w/o Hierarchical Graph Reasoning	0.2815	0.2206	0.1604	0.2126	0.3332	0.4846	0.2422	0.2655	0.2410
	w/o Dynamic Decision Reasoning	0.2789	0.2216	**0.1616**	0.2106	0.3347	**0.4881**	0.2400	0.2666	**0.2428**
	w/o Therapeutic Coverage	0.2807	0.2209	0.1600	0.2120	0.3336	0.4832	0.2415	0.2658	0.2404
	w/o Counterfactual Marginal Utility	0.2814	0.2207	0.1606	0.2125	0.3333	0.4851	0.2422	0.2656	0.2413
	w/o Adaptive Dynamic–Global Fusion	0.2802	0.2199	0.1601	0.2116	0.3321	0.4836	0.2411	0.2646	0.2406
	w/o Inference Calibration	0.2787	0.2197	0.1597	0.2104	0.3319	0.4823	0.2398	0.2644	0.2399
TCM_SMGCN	Full Model	**0.2944**	**0.2325**	0.1702	**0.2073**	**0.3275**	0.4793	**0.2433**	**0.2719**	0.2512
	w/o Hierarchical Graph Reasoning	0.2915	0.2324	0.1702	0.2053	0.3274	0.4793	0.2410	0.2719	0.2512
	w/o Dynamic Decision Reasoning	0.2538	0.2218	0.1696	0.1788	0.3124	0.4779	0.2098	0.2594	0.2504
	w/o Therapeutic Coverage	0.2929	0.2319	**0.1704**	0.2063	0.3266	**0.4802**	0.2421	0.2712	**0.2516**
	w/o Counterfactual Marginal Utility	0.2886	0.2306	0.1684	0.2033	0.3248	0.4743	0.2386	0.2697	0.2485
	w/o Adaptive Dynamic–Global Fusion	0.2892	0.2313	0.1695	0.2037	0.3258	0.4775	0.2390	0.2706	0.2502
	w/o Inference Calibration	0.2910	0.2310	0.1689	0.2049	0.3254	0.4757	0.2405	0.2702	0.2493

Table 3: Case study of step-wise counterfactual prescription reasoning in CRAFT. Red herbs overlap with the ground-truth prescription. CMU denotes counterfactual marginal utility.

Symptom	**Step**	**Decision (CMU)**	**P / R / F1**
身热肢寒 (Body heat with cold limbs)	1	马钱子 (Maqianzi) (.632)	1.000 / .143 / .250
	2	冬瓜子 (Dongguazi) (.233)	1.000 / .286 / .444
	3	野菊花 (Yejuhua) (.086)	1.000 / .429 / .600
	4	山药 (Shanyao) (.031)	1.000 / .571 / .727
	5	**STOP**	1.000 / .571 / .727
Prescription Coverage: 4/7		**CMU:** .632 → .233 → .086 → .031	
头摇 (Head shaking)	1	杠板归 (Gangbangui) (.632)	1.000 / .167 / .286
	2	苦参 (Kushen) (.145)	1.000 / .333 / .500
	3	甘草 (Gancao) (.088)	1.000 / .500 / .667
	4	**STOP**	1.000 / .500 / .667
Prescription Coverage: 3/6		**CMU:** .632 → .145 → .088	

4.4 Case Study

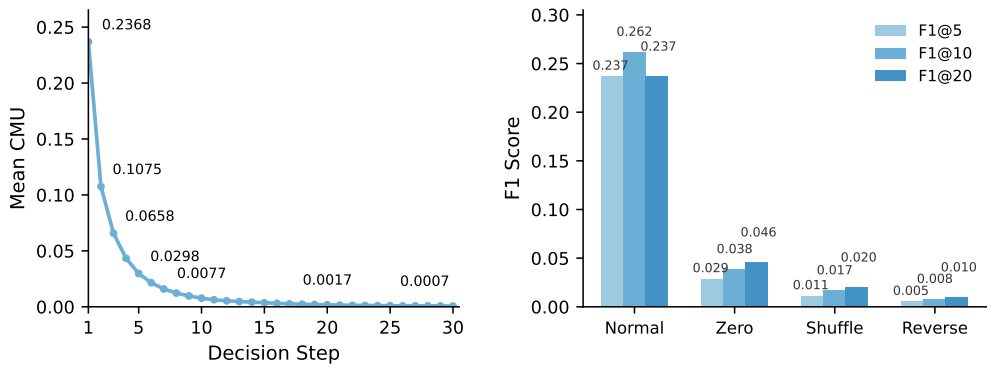
Table 3 illustrates the step-wise prescription reasoning of CRAFT. In both cases, CRAFT progressively identifies ground-truth herbs, improving F1 from 0.250 to 0.727 and from 0.286 to 0.667, respectively, while CMU decreases as the prescription becomes more complete. CRAFT then triggers STOP when further additions provide limited utility. These examples demonstrate that CRAFT constructs prescriptions through candidate-level marginal utility evaluation rather than one-shot top-K prediction, providing an interpretable trajectory with adaptive termination.

4.5 Counterfactual Utility Analysis

Table 4: Validity analysis of counterfactual marginal utility (CMU) over 86,324 valid decision steps. Gold denotes remaining ground-truth herbs at each decision step.

Metric	**Value**
Pairwise Utility AUC	**0.8943**
Top-1 Gold Hit	0.1026
Top-5 Gold Hit	0.2742
Mean Utility (Gold)	**0.0425**
Mean Utility (Non-Gold)	0.0042
Gold / Non-Gold Utility	**10.12**×

Utility Validity. We further investigate whether CMU itself provides a meaningful signal for distinguishing therapeutically relevant candidates. Table 4 summarizes the statistics over 86,324 valid decision steps. CMU achieves a pairwise utility AUC of 0.8943, indicating that a remaining ground-



(a) Utility evolution (b) Counterfactual intervention

Figure 3: Evolution of mean counterfactual marginal utility (CMU) across prescription decision steps. CMU progressively decreases as therapeutic demands are satisfied during prescription construction.

truth herb receives a higher marginal utility than a non-ground-truth candidate in approximately 89.4% of pairwise comparisons. Moreover, the average utility of ground-truth herbs is 0.0425, approximately 10.12 times that of non-ground-truth herbs (0.0042). These results demonstrate that the proposed counterfactual utility captures candidate-level therapeutic relevance before being fused with the neural decision score.

Dynamic Utility Evolution. We examine how counterfactual utility evolves throughout the prescription construction process. As illustrated in Fig. 3(a), the mean CMU decreases rapidly from 0.2368 at the first decision step to 0.0298 at the fifth step and 0.0077 at the tenth step, followed by a gradual convergence toward zero. This diminishing-return pattern suggests that early decisions address the dominant therapeutic demands, whereas subsequent herbs provide progressively smaller complementary gains as the residual demands are satisfied. The result provides dataset-level evidence for the dynamic therapeutic reasoning process modeled by CRAFT.

Counterfactual Intervention. To examine whether CRAFT genuinely exploits counterfactual utility rather than merely benefiting from an additional numerical signal, we intervene on CMU at inference time while keeping the trained model unchanged. As shown in Fig. 3 (b), the original CMU yields substantially better performance than all intervention variants. Zeroing CMU causes a pronounced degradation, while randomly shuffling its candidate-wise correspondence further reduces performance. Reversing CMU produces the largest deterioration. The consistent trend of *Normal* > *Zero* > *Shuffled* > *Reversed* indicates that CRAFT relies not only on the presence of the utility signal, but more importantly on its correct candidate-specific counterfactual structure.

5 Conclusion

In this paper, we proposed CRAFT, a counterfactual utility-guided dynamic reasoning framework for TCM prescription recommendation. CRAFT formulates prescription construction as an order-free dynamic decision process, combining hierarchical graph reasoning with Counterfactual Marginal Utility (CMU) to evaluate the marginal contribution of candidate herbs as the prescription evolves. Experiments on two benchmark datasets demonstrate consistent improvements over competitive baselines across different recommendation cutoffs. Further analyses show that CMU provides informative candidate-level signals and captures diminishing returns during prescription construction. These results highlight the effectiveness of explicitly modeling evolving prescription states and candidate-specific marginal utility for TCM prescription recommendation.

REFERENCES

M Mehdi Afsar, Trafford Crump, and Behrouz Far. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7):1–38, 2022.

Tesfaye Fenta Boka, Zhendong Niu, and Rama Bastola Neupane. A survey of sequential recommendation systems: Techniques, evaluation, and future directions. *Information Systems*, 125:102427, 2024.

Jie Cao and Yin Zhang. Otseq2set: An optimal transport enhanced sequence-to-set model for extreme multi-label text classification. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 5588–5597, 2022.

Yang Cao, Changhao Zhang, Xiaoshuang Chen, Kaiqiao Zhan, and Ben Wang. xmtf: A formula-free model for reinforcement-learning-based multi-task fusion in recommender systems. In *Proceedings of the ACM on Web Conference 2025*, pp. 3840–3849, 2025.

Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H Chi. Top-k off-policy correction for a reinforce recommender system. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pp. 456–464, 2019.

Xiaocong Chen, Lina Yao, Julian McAuley, Guanglin Zhou, and Xianzhi Wang. Deep reinforcement learning in recommender systems: A survey and new perspectives. *Knowledge-based systems*, 264:110335, 2023.

Weiyu Cheng, Yanyan Shen, and Linpeng Huang. Adaptive factorization network: Learning adaptive-order feature interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3609–3616, 2020.

Xin Dong, Chenxi Zhao, Xinpeng Song, Lei Zhang, Yu Liu, Jun Wu, Yiran Xu, Ning Xu, Jialing Liu, Haibin Yu, et al. Presrecst: a novel herbal prescription recommendation algorithm for real-world patients with integration of syndrome differentiation and treatment planning. *Journal of the American Medical Informatics Association*, 31(6):1268–1279, 2024.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*, 2015.

Hailong Hu, Yaqian Li, Zeyu Zheng, Wenjun Hu, Riyang Lin, and Yanlei Kang. A traditional chinese medicine prescription recommendation model based on contrastive pre-training and hierarchical structure network. *Expert Systems with Applications*, 268:126318, 2025.

Yuanyuan Jin, Wei Zhang, Xiangnan He, Xinyu Wang, and Xiaoling Wang. Syndrome-aware herb recommendation with multi-graph convolution network. In *2020 IEEE 36th international conference on data engineering (ICDE)*, pp. 145–156. IEEE, 2020a.

Yuanyuan Jin, Wei Zhang, Xiangnan He, Xinyu Wang, and Xiaoling Wang. Syndrome-aware herb recommendation with multi-graph convolution network. In *2020 IEEE 36th international conference on data engineering (ICDE)*, pp. 145–156. IEEE, 2020b.

Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, pp. 197–206. IEEE, 2018.

Jon Kleinberg, Emily Ryu, and Éva Tardos. Ordered submodularity and its applications to diversifying recommendations. *arXiv preprint arXiv:2203.00233*, 2022.

Wei Li and Zheng Yang. Exploration on generating traditional chinese medicine prescriptions from symptoms with an end-to-end approach. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 486–498. Springer, 2019.

Jingjin Liu, Hankz Hankui Zhuo, Kebing Jin, Jiamin Yuan, Zhimin Yang, and Zhengan Yao. Sequential condition evolved interaction knowledge graph for traditional chinese medicine recommendation. *ICLR*, 2023.

Meixi Lu, Long Tan, and Yeran Wang. Phcl: Prescription recommendation via hypergraph-enhanced contrastive learning. In *2025 IEEE International Conference on Big Data (BigData)*, pp. 1104–1112. IEEE, 2025.

Aman Madaan, Dheeraj Rajagopal, Niket Tandon, Yiming Yang, and Antoine Bosselut. Conditional set generation using seq2seq models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 4874–4896, 2022.

Chuitian Rong, Xueyan Li, Xuemei Sun, and Huabo Sun. Chinese medicine prescription recommendation using generative adversarial network. *IEEE access*, 10:12219–12228, 2022.

Zeyuan Wang, Josiah Poon, and Simon Poon. Tcm translator: A sequence generation approach for prescribing herbal medicines. In *2019 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pp. 2474–2480. IEEE, 2019.

Kuo Yang, Zecong Yu, Xin Su, Fengjin Zhang, Xiong He, Ning Wang, Qiguang Zheng, Feidie Yu, Tiancai Wen, and Xuezhong Zhou. Prescdrl: deep reinforcement learning for herbal prescription planning in treatment of chronic diseases. *Chinese medicine*, 19(1):144, 2024.

Liangwei Yang, Shengjie Wang, Yunzhe Tao, Jiankai Sun, Xiaolong Liu, Philip S Yu, and Taiqing Wang. Dgrec: Graph neural network for recommendation with diversified embedding generation. In *Proceedings of the sixteenth ACM international conference on web search and data mining*, pp. 661–669, 2023.

Pengcheng Yang, Fuli Luo, Shuming Ma, Junyang Lin, and Xu Sun. A deep reinforced sequence-to-set model for multi-label classification. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 5252–5258, 2019.

Xiaoyan Yang and Changsong Ding. Smrgat: A traditional chinese herb recommendation model based on a multi-graph residual attention network and semantic knowledge fusion. *Journal of ethnopharmacology*, 315:116693, 2023.

Yun Yang, Yulong Rao, Minghao Yu, and Yan Kang. Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation. *Neural Networks*, 146:1–10, 2022.

Liang Yao, Yin Zhang, Baogang Wei, Wenjin Zhang, and Zhe Jin. A topic modeling approach for traditional chinese medicine prescriptions. *IEEE Transactions on Knowledge and Data Engineering*, 30(6):1007–1021, 2018.

Wenjing Yue, Wendi Ji, Xinyu Wang, Xin Ma, Pengfei Wang, and Xiaoling Wang. Sdpr: Prescription recommendation with syndrome differentiation in traditional chinese medicine. *IEEE journal of biomedical and health informatics*, 29(5):3736–3749, 2025.

Chaobo Zhang and Long Tan. A hierarchical structure-enhanced personalized recommendation model for traditional chinese medicine formulas based on kg diffusion guidance. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pp. 4066–4075, 2025.

Mingchuan Zhang, Longfei Chai, Junlong Zhu, Wenxuan Xu, Junqiang Yan, Lin Wang, Liye Xia, and Qingtao Wu. A prescription recommendation method based on knowledge graph in traditional chinese medicine. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 25(2):1–24, 2026.

Zijuan Zhao, Xueting Ren, Kai Song, Yan Qiang, Juanjuan Zhao, Junlong Zhang, and Peng Han. Pregenerator: Tcm prescription recommendation model based on retrieval and generation method. *IEEE access*, 11:103679–103692, 2023.

Xinhan Zheng, Xueting Wang, Ruotai Li, Huyu Wu, Haopeng Jin, Yehan Yang, and Guodong Shan. Fmash: Advancing traditional chinese medicine formula recommendation with efficient fusion of multiscale associations of symptoms and herbs. *arXiv preprint arXiv:2503.05167*, 2025.

Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1059–1068, 2018.

Juanzi Zhou, Yin Zhang, Fang Hu, Xiong Zhang, and Pin-Han Ho. Cl-mhad: Contrastive learning-based multi-hypergraph aggregation and diffusion model for prescription recommendation. *Artificial Intelligence in Medicine*, pp. 103388, 2026.

Yi Zhuang, Lingkai Yu, Nan Jiang, and Yujia Ge. Tcm-kllama: Intelligent generation model for traditional chinese medicine prescriptions based on knowledge graph and large language model. *Computers in Biology and Medicine*, 189:109887, 2025.

Lixin Zou, Long Xia, Zhuoye Ding, Jiaxing Song, Weidong Liu, and Dawei Yin. Reinforcement learning to optimize long-term user engagement in recommender systems. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2810–2818, 2019.

A Appendix

A.1 Additional Experiments

Table 5: Performance of CRAFT under different prescription and symptom complexities.

Group	N	**F1@5**	**F1@10**	**F1@20**
Prescription Length				
Short (≤ 5)	5,918	0.1798	0.1656	0.1291
Medium (6–10)	5,108	**0.2754**	0.3096	0.2816
Long (> 10)	2,030	0.2434	**0.3256**	**0.3565**
Number of Symptoms				
Few (≤ 2)	11,609	0.2340	0.2564	0.2300
Moderate (3–5)	1,346	0.2562	**0.3002**	0.2836
Many (> 5)	101	**0.2603**	0.2998	**0.3115**

Performance under Different Complexities. We additionally analyze CRAFT across samples with different prescription lengths and symptom complexities. As shown in Table 5, CRAFT maintains competitive performance across different complexity groups. Notably, F1@10 reaches 0.3256 for prescriptions containing more than 10 herbs, while cases with three or more symptoms achieve approximately 0.30 F1@10. These observations suggest that CRAFT remains effective on relatively complex prescription instances. We refrain from attributing the subgroup differences solely to the dynamic reasoning mechanism, since prescription length also affects the difficulty of fixed-K evaluation.

A.2 Training Algorithm

Algorithm 1 summarizes the complete order-free training procedure of CRAFT. At each step, all remaining ground-truth herbs are treated as valid next actions, while STOP becomes the target action once the prescription is complete. This procedure avoids imposing an artificial ordering on the ground-truth herb set.

A.3 Implementation Details

CRAFT is implemented in PyTorch and trained using the Adam optimizer. The symptom and herb embedding dimension is set to 128, with a hidden dimension of 256. We use two graph reasoning layers with four attention heads, a dropout rate of 0.1, and a batch size of 512. The learning rate and weight decay are set to 3×10^{-4} and 1×10^{-3}, respectively. Training runs for at most 200 epochs

Algorithm 1 Order-free Training of CRAFT

Require: Ground-truth herb set $\mathcal{Y}$, maximum steps T
Initialize $\mathcal{P}_0 \leftarrow \emptyset$ and $\mathbf{z}_0$
for $t = 0, \ldots, T-1$ **do**
  $\mathcal{R}_t \leftarrow \mathcal{Y} \setminus \mathcal{P}_t$
  if $\mathcal{R}_t \neq \emptyset$ **then**
    $\mathcal{V}_t \leftarrow \mathcal{R}_t$
  else
    $\mathcal{V}_t \leftarrow \{\text{STOP}\}$
  end if
  Compute $\ell_t(a)$ for all $a \in \mathcal{A}_t$
  Accumulate the set-valued decision loss
  if $\mathcal{R}_t = \emptyset$ **then**
    break
  end if
  $h_t^+ \leftarrow \arg\max_{h \in \mathcal{R}_t} \ell_t(h)$
  $\mathcal{P}_{t+1} \leftarrow \mathcal{P}_t \cup \{h_t^+\}$
  Update $\mathbf{z}_{t+1}$ with h_t^+
end for
Compute $\mathcal{L}_{\text{seq}}$ and $\mathcal{L}_{\text{pres}}$
$\mathcal{L} \leftarrow \lambda_{\text{seq}}\mathcal{L}_{\text{seq}} + \lambda_{\text{pres}}\mathcal{L}_{\text{pres}}$

with an early-stopping patience of 15, and gradients are clipped to a maximum norm of 5.0. The maximum number of dynamic decision steps is 30, and λ_{seq} and λ_{pres} are both set to 1.0. We use a fixed random seed of 2021 for reproducibility. Table 6 summarizes the main hyperparameters used in our experiments. Unless otherwise specified, the same configuration is used for all experiments to ensure consistent evaluation.

Table 6: Hyperparameter settings of CRAFT.

Hyperparameter	Value
Embedding dimension	128
Hidden dimension	256
Graph layers	2
Attention heads	4
Dropout	0.1
Batch size	512
Learning rate	3×10^{-4}
Weight decay	1×10^{-3}
Maximum epochs	200
Early-stopping patience	15
Gradient clipping	5.0
Maximum decision steps	30
λ_{seq}	1.0
λ_{pres}	1.0
Random seed	2021