算法调研

Neural Factorization Machines for Sparse Predictive Analytics

核心问题：在稀疏数据（**通过one-hot编码生成的高维稀疏特征**）中，如何高效建模特征交互以提升预测性能。传统FM 仅能建模二阶线性交互，无法捕捉真实数据中的非线性结构；而深度神经网络（DNNs）如Wide&Deep和DeepCross虽能建模非线性交互，但面临优化困难（如梯度消失、过拟合）和训练复杂度高的问题。

主要贡献：

NFM 将 FM 在二阶特征交互建模中的线性性与神经网络在高阶特征交互建模中的非线性无缝结合

引入Bi-Interaction Pooling操作，高效编码二阶特征交互，为后续神经网络层提供更丰富的信息基础

Attentional Factorization Machines:Learning the Weight of Feature Interactions via Attention Networks

核心问题：传统因子分解机（FM）在建模特征交互时存在一个根本性缺陷——对所有特征交互赋予相同权重，这将会导致两个问题

- 噪声干扰：无用特征交互被平等对待，引入噪声降低预测精度
- 次优预测：关键交互无法通过更高权重体现其重要性

主要贡献：引入注意力加权机制——替换FM的固定权重和，通过神经网络动态学习交互权重

参考设计

- **Bi-Interaction pooling（NFM）**：把 pairwise 交互编码为向量集合（element-wise product），再聚合以得到更丰富的二阶信息；把 CVSS、EPSS、PoC、Trend、CWE、Age、ENV 等按语义先投影到同一维度的“领域嵌入”，再计算这些特征的逐元素交互（能表达“严重度×可利用性”“PoC×Trend”等协同）
- **Attention on interactions（AFM）**：不同 pairwise 交互的重要性不同，用 attention 给每对交互分配权重
- **NFM 的浅层非线性思想**：用少量非线性变换（而不是深层黑箱网络）即可显著增强表达力

模型公式

$$z = \eta x_{\text{env}} \cdot (w_C C + w_I I + w_T T) + b, \quad R = \sigma(z).$$

总体框架

归一化特征经**可解释投影（规则化 embedding）** → 构建 **Deterministic Bi-Interaction (DBI)** 向量集合 → 以 **Rule-Attention** 对交互加权 → 聚合并作 **可解释非线性投影** 得到交互贡献 I；同时保留线性核心项 C 与时序项 T，并用 **HiC-Gate（层次化门控）** 根据环境 x_{env} 在三者间重新分配权重，最终归一化输出风险 $R \in (0, 1)$

可解释投影（规则化 embedding）$v_i = \Phi_i(x_i)$

把异构标量或类别映射到同一维度 k，使后续逐元素乘积 $vi \odot vj$能表达不同"基"上的耦合

对标量项（CVSS、EPSS、PoC、Trend、Age、ENV）采用如下构造:

$$vi = [x_i, x_i^{\rho}, log(1+\alpha x_i), cwe\_vec]$$

Deterministic Bi-Interaction（DBI）：交互向量 $u_{ij} = v_i \odot v_j$

把 NFM 中的"Bi-Interaction pooling"解析化：不训练 perceptron 层，而直接把每个有意义的特征对在 embedding 维上做元素乘积，从向量角度表达"逐基交互"。例如 CVSS×EPSS 在不同维上体现"严重度在弱幂/对数尺度上与利用概率的耦合"

自定义交互集合P

- (CVSS,EPSS) — 严重度 × 利用概率
- (CWE,PoC) — 漏洞类别 × PoC 成熟度
- (PoC,Trend) — PoC × 社区热度（武器化）
- (CVSS,ENV) — 严重度在当前环境下的匹配性

对每对 $(i,j) \in P$进行计算： $u_{ij} = v_i \odot v_j \in R^k, u_{ij}^{(\ell)} = v_i^{(\ell)} \cdot v_j^{(\ell)}.$

Rule-Attention（规则化注意力）a_{ij}

使用可解释规则函数 g_{ij}计算每对交互的重要性分数 s_{ij},再用 softmax 归一化得到权重 a_{ij}

基本形式

$$S_{ij} = g_{ij}(C_i, L_j, \mathrm{CWE}, L_{\mathrm{enu}}, L_{\mathrm{age}}, \ldots) \qquad a_{ij} = \frac{\exp(TS_i)}{\sum_{p,a}\exp(TS_{pa})}$$

交互聚合 U 与可解释非线性投影 I(交互影响项)

聚合 $U = \sum_{i,j} d_{ij} u_{ij} \in \mathbb{R}^{\mathbb{k}}$　每个 U_ℓ 是所有交互在第 ℓ个基上的加权和

非线性投影 $I = \sum_{\ell=1}^{k} \beta_\ell \tanh(\gamma_\ell U_\ell).$

- β_ℓ 控制每维对最终交互得分的权重（可设均匀 $\beta_\ell = 1/k$)
- γ_ℓ 控制非线性斜率
- $tanh$保证有界且中心化（对称），便于后续证明与解释

线性核心项 C与时序项 T

核心项：保留并解释性扩展线性项，表示每个指标的一阶贡献

$C = W_1 x_{\mathrm{cvss}} + W_2 x_{\mathrm{epss}} + W_6 x_{\mathrm{poc}} + W_7 x_{\mathrm{trend}} + \ldots$

时序项：捕捉生命周期与热度　$T = \lambda_1 e^{-\mu \cdot \mathrm{Age}} + \lambda_2 x_{\mathrm{trend}}^{\beta}.$

层次化门控

子项权重由 ENV 决定

定义 closed-form attention： $\tilde{s}_C = \exp(\rho_C x_{\mathrm{env}}), \quad \tilde{s}_I = \exp(\rho_I x_{\mathrm{env}}), \quad \tilde{s}_T = \exp(\rho_T x_{\mathrm{env}}),$

$$w_C = \frac{\tilde{s}_C}{\tilde{s}_C + \tilde{s}_I + \tilde{s}_T}, \ \mathrm{etc.}$$

参数 ρ_*控制 ENV 对子项的偏好（例如把 ρ_I 设高，表示在高环境匹配时更重视交互项）