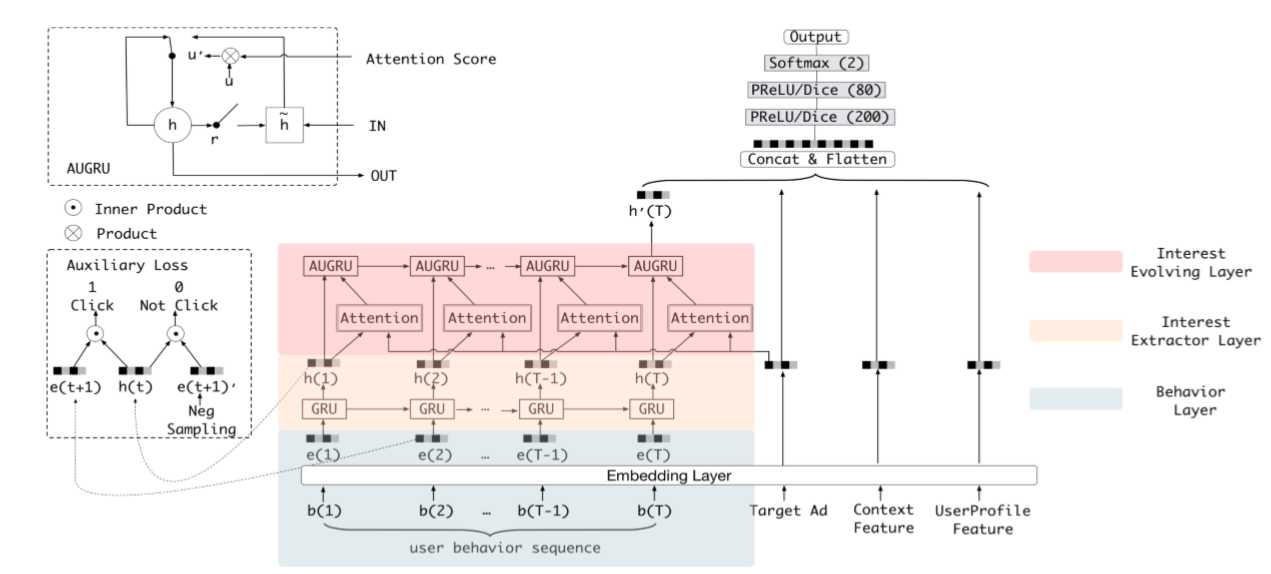
Deep Interest Evolution Network for Click-Through Rate Prediction

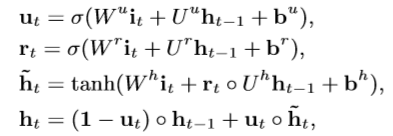
该论文本质上是DIN模型的一种扩展形式，其输入形式跟DIN一致，利用Embeeding的思想将用户、物品的属性信息特征映射成为对应的嵌入向量，在目标物品与用户历史行为之间建立attention机制，关联目标物品跟历史行为物品，从而捕捉到用户的兴趣信息。



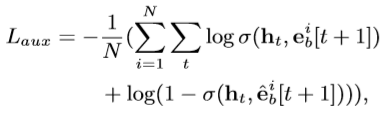
DIEN在此基础上，将DIN的attention层进一步扩展成Interest Extractor Layer（兴趣提取层）和Interest Evolving Layer（兴趣进化层），能够更好地对历史行为序列进行建模，本质上类似于self-attention，能够将历史行为序列里面包含的兴趣转移信息提取出来进行建模，更好地提升推荐效果。

·Interest Extractor Layer

该层用于提取用户在历史行为序列上各个购买或者点击行为的兴趣信息，使用的结构时GRU模型，该模型能够避免RNN的梯度消失问题，还能够比LSTM的复杂度低，能够有效提取历史序列中蕴含的用户潜在兴趣信息以及保留了顺序信息，可以在一定程度上保留用户的旧兴趣以及新爱好，跟nlp的句子GRU建模类似。



该论文为了让该GRU模型更好地提取特征，使用了auxiliary loss（辅助损失），这是因为如果不添加额外的损失，在对该GRU模型只能针对当前的点击或者购买行为来更新参数，而没有用到序列里面蕴含的信息，不利于序列建模。因此在构造训练集的时候，会针对每个样本产生对应的T个正样本和负样本，用于训练GRU，且损失函数即普通的二类交叉熵损失函数，其中隐层h与目标物品的嵌入向量使用点乘进行计算。



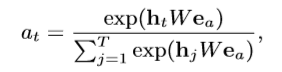


该损失将以alpha的比例与最后的损失函数进行加权组合。



·Interest Evolving Layer

该层用于结合GRU序列信息以及attention信息，但是attention的计算没有沿用DIN的前馈网络，而是直接使用简单的加权內积加上softmax进行计算。



往后，为了获取用户的兴趣转移信息，论文将attention信息与GRU模型进行结合，产生了三种GRU模型：AIGRU，AGRU以及AUGRU，区别在于将attention应用于GRU的不同地方。

* GRU with attentional input (AIGRU)  
  attention信息直接嵌入到输入中变成GRU的新输入



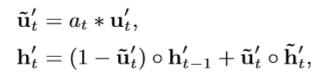
* Attention based GRU(AGRU)

attention信息代替GRU的更新门



* GRU with attentional update gate(AUGRU)

attention信息与GRU的更新门进行融合。



最后论文采用最后一种方法，因为在第一种方法下，即便attention为0，隐层状态还是会进行更新，影响模型的表现力；而第二种方法能够被第三种含括，因此选用第三种。

·效果对比

论文基于公开的数据集以及工业数据集进行测试，除此以外，还对模型进行A/B测试，效果不错。

