实验部分The samples we used is designed for the online conversational scenarios. We randomly select a part of component services of a mashup as the context services, and build a positive sample with a label of 1 by combining them with the mashup and a random unselected component service. A negative sample with a label of 0 is composed of the context services, the mashup and a random service not invoked by the mashup. We randomly select the context services multiple times to enlarge the sample set and to simulate the situations where developers have choose different component services in their sessions. 数据不均衡的处理？

* TF-IDF：The content-based method make a prediction of a service over a new mashup using the TF-IDF-based cosine similarity between their content information.
* *AFUP* (Jain *et al.*, 2015) [11]. Two probabilities that a mashup invokes a service are first calculated according to their content similarity and the historical interaction of neighbor mashups and the service, and then integrated by Bayes’ theorem to .
* then multiplies them based on Bayes’ theorem, and finally ranks candidates according to their popularity.
* *SFTN*(Samanta *et al.*, 2017)[12]. The authors improve their previous work, AFUP, by using the hierarchical Dirichlet process (HDP) [36] and probability matrix factorization to process the content information and usage history.
* *PNCF* (Chen *et al.*, 2018) [14]. The framework compresses all sparse features of users and items in an embedding layer and then uses an MLP to model their interaction. However, its feature extraction component does not apply to extract textual features, and we implement two variants for this scenario: PNCF-HDP, which applies HDP adopted in SFTN, and PNCF-Deep, which uses our feature extraction strategy.

NI：

我们历史信息空间的模型实例中，设计的基于近邻和node2vec表示新mashup的策略，是创新，效果很好。

实验部分：

1. 多种方法的横向对照 包括我们的CI，NI，混合 综合指标 123综合

内容相似度 TF-IDF 非常差 加了pop也非常差

WVSM+POP 效果好 0.38左右

试试HDP等的方法？

~~User-based CF~~

User-CF

IsRec\_best

PasRec

一种基于user-based CF方法。使用HIN有效组织各种信息，多种meta-path的相似度，更准确地衡量mashup之间的综合相似度。

使用pairwise的优化思路，优化不同路径的相似度的权重。

缺陷：

1.内容：只依靠近邻mashup对待测服务的评分作出预测。然而用户的选择行为很大程度上从功能角度衡量。

2.近邻：借用近邻服务的方式过于简单，直接用标量的评分不够精确，没有深入挖掘它们的隐含潜在的交互关系。 跟我们的NI部分相比，使用node2vec得到的表示已经包含丰富交互信息，然后使用DNN去捕捉交互信息。

3.没有考虑已选择服务的信息。

混合

AFUP SFTN 内容相似度\* （基于MF和近邻思想得到的score）0.23

1.内容特征质量不高。词袋模型，忽略词序。

2.使用内容相似度的embedding表示。

2.1.1内容质量不高。

2.1.2相似度计算方法也不好，只利用了内容信息，没有考虑已选择的服务的信息。

2.2 pmf（本身质量一般，只能捕捉线性关系，在稀疏情况下质量不好）。

3.两个概率独立的假设，将其相乘作为评分，但是具体交互情况复杂未知，这种假设没有证据。

The reasons are two-fold. First, their feature extraction methods ignored word orders and lost some semantic information. Second, the two probabilities derived from functionality and usage history were multiplied with the assumption that they are conditionally independent of each other. However, it is hard for the single multiplication to capture the way how the content information and usage history jointly affect service recommendation. Instead, MISR aims to capture such complicated interactions by a DNN.

深度模型：

相同的特征（相当于backbone）；不同的交互方式

现有的deepCTR等工作主要针对多种离散特征的情况，离散特征先进行embedding操作，多种特征之间再进行充分的交互。Embedding和交互方式的定义（模型结构）是统一的整体，相互影响。

1. 我们的特征只有文本和历史信息。文本只能提取特征；对于历史信息，可以使用其ID形式，但是新ID没有embedding，所以只能使用我们提出的表示方法。所以相当于特征已经确定。总之，我们的特征表示策略使得基于文本和历史信息的基于深度学习的在线对话式推荐可行。

所以使用相同的特征表示，突出交互模式的对比。

2.同种特征之间交互即可，对于文本和历史的特征，交互意义不大，反而增加交互学习的难度。

PNCF m和s的信息输入到模型中 相当于不加选择服务信息

Embedding+MLP 加入行为数据，平均池化

DINRec

我们的三种

PNCF:

双塔模型，首先将mashup和待测服务的各种特征整合，然后使用一个MLP学习交互。

1模型部分，也没有考虑已选择服务的信息。

2 全交互。

Embedding+MLP 加入行为数据，平均池化

1.没有突出已选择服务的重要性不同。

2. 全交互。各种特征整合后，统一交互。增大了学习难度。见DIN分析。

PNCF借鉴了双塔模型的设计，它首先得到mashup和api的总体的特征表示，然后在交互层使用一个MLP学习它们的交互，最终做出预测。无脑交互。

这里，我们把基于内容信息的实例得到的mashup和api的表示作为它们的内容特征，把基于历史信息的实例中采用的表示方法得到的表示，作为他们的信息的表示。然后两种特征的拼接作为mashup/服务的表示。

我们的效果优于PNCF。因为它没有考虑到用户新选择的服务。同时这种所有特征拼接之后再交互的方式，常用于特征和任务高度相关，基于embedding的模型。而我们这里的两种特征，内容和历史之间，不需要交互，只需要同种信息的交互。

DIN:

1.DINRec中使用的全部是离散特征，需要embedding化，比如ID。在线对话式场景中无法使用。我们设计了特征表示方法。

2.Mashup，已选择服务和待测服务的各种种类的特征拼接，之后做一次attention激活处理。相当于对每个已选择服务对候选只有一个影响权重。当特征比较复杂繁多时，**从各个空间内的多个特征出发考虑的影响可能不一致**，也**增加了交互学习的难度**。我们的混合模型相当于在多个空间都进行了attention，从各个特征空间考虑已选择和待测，mashup之间的交互，增强了模型对交互关系的建模能力。全交互。

3.相对于DIN，我们扩展性更强。当引入新的特征时，只需要在新特征空间学习交互，将交互加入到原来的交互向量，然后重新学习MLP即可。不fine-tuning效果也很好。而DIN则需要整个模型，所有参数更新一遍。

1. 加入已选择的服务的效果（不加，以及几种不同的方式） 在1 2 3上分别对比
   1. 加入有提升
   2. Attention的优势。当选择的服务越多时，相对提升越高
2. 相似度计算方法的对照：
   1. PasRec全
   2. IsRec全
   3. IsRec\_best（实际上只使用了文本和tag特征）
   4. 只使用文本和tag特征：PasRec两条路径
   5. 只使用文本和tag特征：我们的特征
3. 加入已选择信息的，整体的DIN?