一、上次的讨论：

- 从论文中找原因，弄清楚实验数据中，出发的网格id格式、目标网格id、网格变化是如何处理的来的
- 对预计时间进行随机高斯采样

二、原文中提及实验的部分：

一、本文主要目标：

> Assuming that there are three candidate grids Di (0-2), the main problem is to choose a most appropriate grid for drivers to hunt passengers. Based on the above, we take the hot degree of Di, the road condition of O to Di, and the average earning of Di into consideration comprehensively, and produce a final score for every grid to express the selection degree. **Finally the grid Di scoring the most will be chosen and recommended to taxi drivers.**

二、对地图的网格划分：

> In the module of offline mining and training, we first handle the map (map meshing) and the GPS trajectory big data (trajectories segmenting), respectively. Unlike road networks used in most of existing approaches, in this paper, we use the grid map generated from map meshing. After trajectories segmenting, we carry out the task of map matching. That is, each GPS point is attached to the corresponding road. Thereby, we can easily achieve the whole occupied or idle trips.(p4)

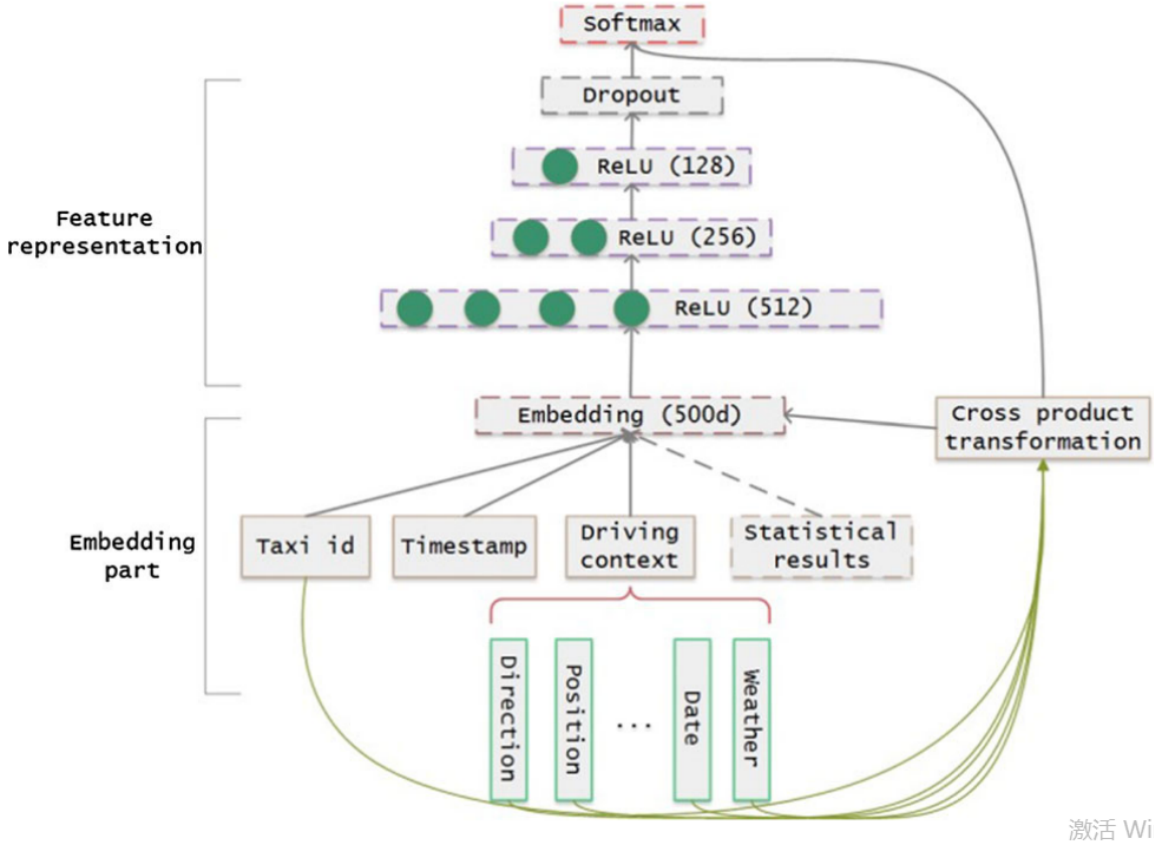
三、对数据的规范化预处理：

> After this step, we utilize statistical learning approaches to handle the grip map and trips obtained in the front and produce three types of complete information: (1) picking up points, (2) road information, and (3) the earning of each grid. These three types of information are the **main input data of the wide & deep models.** It is worth noting that the trips obtained in the front cannot be straightly used in the wide & deep models. And we need to preprocess these trips and extract the discriminative features from them. Then these data pass through **normalization (for continuous features)** and **encoding operations (for categorical features)** to guarantee effectiveness and reasonableness. After this step, we employ the word embedding methods, such as word2vec or GloVe, to transform these features to lowdimensional vectors. Followed by word embedding, two efficient algorithms (sparse autoencoder and t-SNE) are used for dimensionality reduction.

四、wide & deep model

key information：

- crossed features are applied as higher-order features by Cartesian product.



五、 Evaluation for TRec

how they evaluate their model?

> TRec combines these three components with weights and calculate a score for every grid in the map. Then TRec recommends the grid which has the highest score to taxis. What the taxi drivers most concern about is their earnings eventually.
>
> we use the following formula to calculate a score for every grid g in the map（p11）：

$$\Upsilon(g) = \frac{\sqrt{w_{pp} \times w_{ge}}}{w_{rc}} \times \frac{\Psi \times \Gamma}{D} \times \lambda^{\left\lfloor \frac{k}{10} \right\rfloor}.$$

example:

> In order to evaluate our recommendation system TRec, we choose 1568 trips which are all produced between 5 am and 11 pm. These trips are chosen following the normal distribution. According to the charge standard of taxis in Shanghai, we utilize the following equation to achieve the earnings of a specific trip T, where |T| is the length of T.