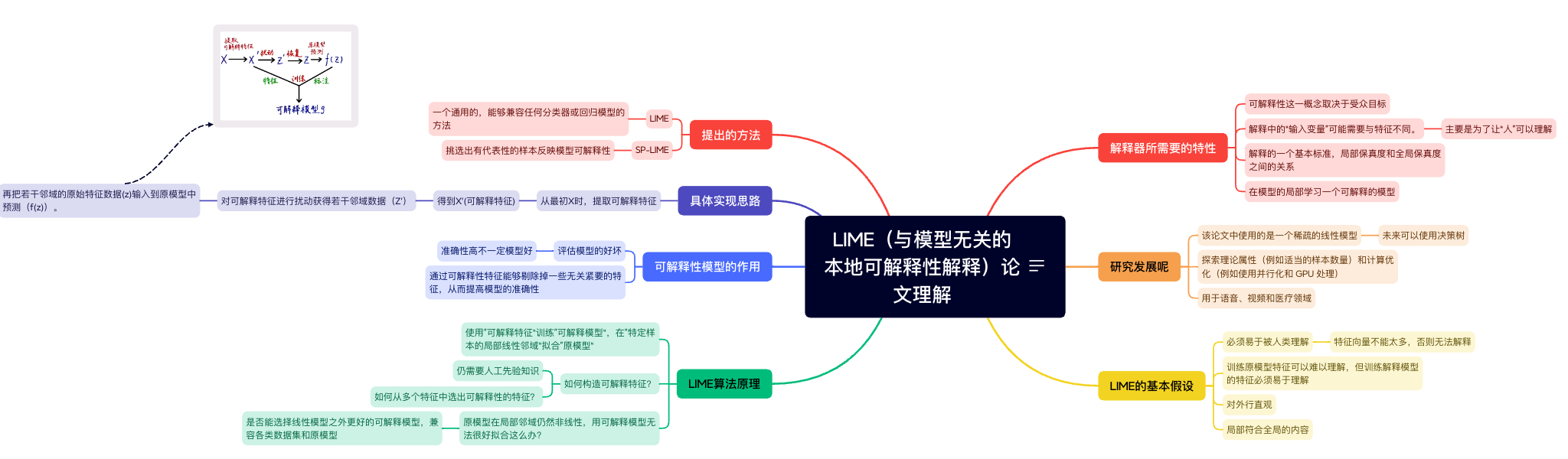
**研究方向进展报告**

汇报人：熊俐凯

日期：2025年2月28日

### ****一、本周核心工作****

#### ****1. 文献研读****

* ✅完成研读LIME框架的论文《“Why Should I Trust You?” Explaining the Predictions of Any Classifier》形成思维导图：
* 

#### ****模型可解释性实验****

实验数据：据集包含2013年9月欧洲持卡人使用信用卡进行的交易。该数据集展示了两天内发生的交易，在284,807笔交易中，我们有492笔欺诈交易。数据集高度不平衡，正类（欺诈）占所有交易的0.172%。由于保密问题，提供的特征没有具体名称，特征名V1, V2, … V28 是通过PCA获得的主成分，唯一没有通过PCA转换的特征是“Time”和“Amount”。

实验过程：采用XGBoost分类器进行分类，最终代码得到的AUC Score：0.937784。由于实验数据类别不平衡采用SMOME方法处理数据。

实验结果：分别采用SHAP可解释方法对全局模型进行解释，采用LIME方法对局部的数据进行解释分析**。**

|  |  |
| --- | --- |
| **shap_sum_1** | **shap_sum_2** |

图 1SHAP全局分析

**图中展示的是在全局模型中对模型判断影响最高的几个特征向量。**

|  |
| --- |
|  |

图 2 LIME局部分析

**图2展示的是使用LIME方法对第10个数据尽心局部分析的结果，展示了V12，V14和V11是有较大贡献的特征。这个局部的信息也是符合全局的整体信息。**

### ****后续计划****

* + 1. **对LIME和SHAP的方法进行更深入的理解和应用。**
    2. **寻找更加复杂的数据集进行实验，寻找创新点。**
    3. **查找前沿的论文进行创新点规整。**