

目标：

- 1. 确定数据价值评价特征维度
 - 1.1 5 大类细分 (44+X)D
- 2. 获取每个特征维度的值
 - 2.1 获悉每个特征维度的获取方法，分为机器获取和元数据获取
- 3. 归纳数据评价模型，作出数据价值评价
 - 3.1 目前收集的常见的评价模型（1.AHP 2.DeepLearning）

影响因素：

表 3-6 数据资产价值的影响因素

Table 3-6 The Seven Characteristics of Data Asset Value

影响因素	对数据资产价值的影响
数据质量	数据资产的价值会随着数据质量的 <u>提高而增加</u>
数据规模	数据资产的价值会随着数据规模的 <u>增长而增加</u>
可访问性	数据资产的价值随着访问、使用的便捷程度的 <u>增多而增加</u>
鲜活性	数据资产的价值随着时间的 <u>推移而降低</u>
关联性	数据资产的价值 <u>随着关联的数据资产数量增加而增加</u>
使用效果	数据资产的价值随着使用效果的增强而增加
价值密度	数据资产的价值随着价值密度的增大而增加
数据类型	数据资产的价值随数据类型的增多而增加
共享性	数据资产的价值会随着多用户的共享而增加
再生性	数据资产的价值不会随着使用而消耗

概括性地给出影响因素对数据资产的影响，主要分为正反馈和负反馈

在得到了影响因素之后，使用特征维度进一步概括。

特征维度与影响因素的对应关系：

表 3-7 各特征维度对应的价值特性与影响因素

Table 3-7 The Influencing Factors and Value Characteristics of Each Feature Dimension

特征维度	价值特性与影响因素
颗粒度	数据质量、共享性
多维度	可访问性、多样性
活性度	活性、再生性、使用效果
规模度	数据规模、价值密度
关联度	关联性

特征维度的细分：（即 44 个指标特征维度）

颗粒度（数据质量、共享性）

表 3-8 颗粒度的细分维度

Table 3-8 The Subdivision Dimension of Graininess

维度名称	具体含义
数量	该指标反应该项数据资产拥有的数据属性的数量
类型	该指标反应该项数据资产拥有的数据属性的数据类型
精度	该指标反应该项数据资产拥有的数据属性的数据精度
准确度	该指标反应该项数据资产拥有的数据属性的数据准确程度
长度	该指标反应该项数据资产拥有的数据属性的字段长度
完整度	该指标反应该项数据资产拥有的数据属性的字段完整程度

Necessary

维度名称	具体含义
合规性	该指标反应该项数据资产拥有的数据属性的合规率
维护频率	该指标反应该项数据资产拥有的数据属性的维护情况
格式	该指标反应该项数据资产拥有的数据属性的格式规范
编码方式	该指标反应该项数据资产拥有的数据属性的编码规范
标准	该指标反应该项数据资产拥有的数据属性的取值范围
命名规则	该指标反应该项数据资产拥有的数据属性的命名规范

Unknown

多维度（可访问性、多样性）

表 3-9 多维度的细分维度

Table 3-9 The Subdivision Dimension of Diversity

维度名称	具体含义
来源种类	该指标反应该项数据资产拥有多少不同的来源渠道种类
来源数量	该指标反应该项数据资产拥有多少个不同的来源
来源方式	该指标反应该项数据资产的产生方式
来源类型	该指标反应该项数据资产对应来源的类型
覆盖范围	该指标反应该项数据资产能够覆盖多少个不同系统
重复率	该指标反应该项数据资产不同来源获取的重复情况
一致情况	该指标反应该项数据资产不同来源相互校验的准确程度
采集方式	该指标反应该项数据资产的采集手段

Necessary

活性度 (活性、再生性、使用效果):

表 3-10 活性度的细分维度

Table 3-10 The Subdivision Dimension of Activity

维度名称	具体含义
更新间隔时间	该指标反应该项数据资产平均多长时间更新一次
访问间隔时间	该指标反应该项数据资产平均多长时间被访问一次
存在时间	该指标反应该项数据资产的存在时间
更新差异度	该指标反应该项数据资产每次更新会造成变化的程度
访问系统数量	该指标反应该项数据资产会由多少个系统使用
常用属性数量	该指标反应该项数据资产的常用数据属性占比
累积访问次数	该指标反应该项数据资产累积的使用次数
累积更新次数	该指标反应该项数据资产累积的更新次数

Necessary

规模度（数值规模、价值密度）：

表 3-11 规模度的细分维度

Table 3-11 The Subdivision Dimension of Scale

维度名称	具体含义
数据条数	该指标反应该项数据资产拥有数据的条数
资产大小	该指标反应该项数据资产的存储体积
增长速度	该指标反应该项数据资产的增长速度
使用范围	该指标反应该项数据资产的应用使用范围
获取难易程度	该指标反应该项数据资产获取的难易程度
独占程度	该指标反应该项数据资产的企业独占情况

该项指标描述不够详细，或者说不够精确

关联度 (关联性):

表 3-12 关联度的细分维度

Table 3-12 The Subdivision Dimension of Relevance

维度名称	具体含义
流入数据数量	该指标反应该项数据资产流入数据的数量
流出数据数量	该指标反应该项数据资产流出数据的数量
流入数据频率	该指标反应该项数据资产流入数据的频率
流出数据频率	该指标反应该项数据资产流出数据的频率
流入数据大小	该指标反应该项数据资产流入数据的大小
流出数据大小	该指标反应该项数据资产流出数据的大小
流入数据关联强度	该指标反应该项数据资产流入数据的关联强度
流出数据关联强度	该指标反应该项数据资产流出数据的关联强度
数据依赖程度	该指标反应该项数据资产对于其他数据资产的依赖程度
数据独立程度	该指标反应该项数据资产自身的独立存在程度

Necessary

5 个特征维度对应的详细的输入维度：

表 4-1 模型输入维度示例

Table 4-1 The Example of Input Dimension of EDAAM

特征维度	输入维度
颗粒度	数据属性数量、数据属性类型、数据属性精度、数据属性准确性、数据属性长度、数据属性完整性、数据属性合规性、维护频率、数据属性格式、编码方式、标准、命名规则
多维度	数据来源种类、数据来源数量、数据来源方式、数据来源类型、数据覆盖范围、数据重复率、数据一致性、数据采集方式
活性度	数据更新频率、数据访问频率、数据存在时间、数据更新差异大小、访问系统数量、常用属性数量、累积访问次数、累积更新次数
规模度	数据量、数据使用范围、数据大小、数据增长速度、数据获取难易程度、数据独占程度
关联度	流入数据数量、流出数据数量、流入数据频率、流出数据频率、流入数据大小、流出数据大小、流入数据关联强度、流出数据关联强度、数据依赖程度、数据独立程度

- 1. 44 个指标输入维度，从概念上来看是对 100 个原始输入维度的初步筛选，
- 2. 实际的监控指标的属性-->原始输入维度的标准

数据资产价值分析模型框架：

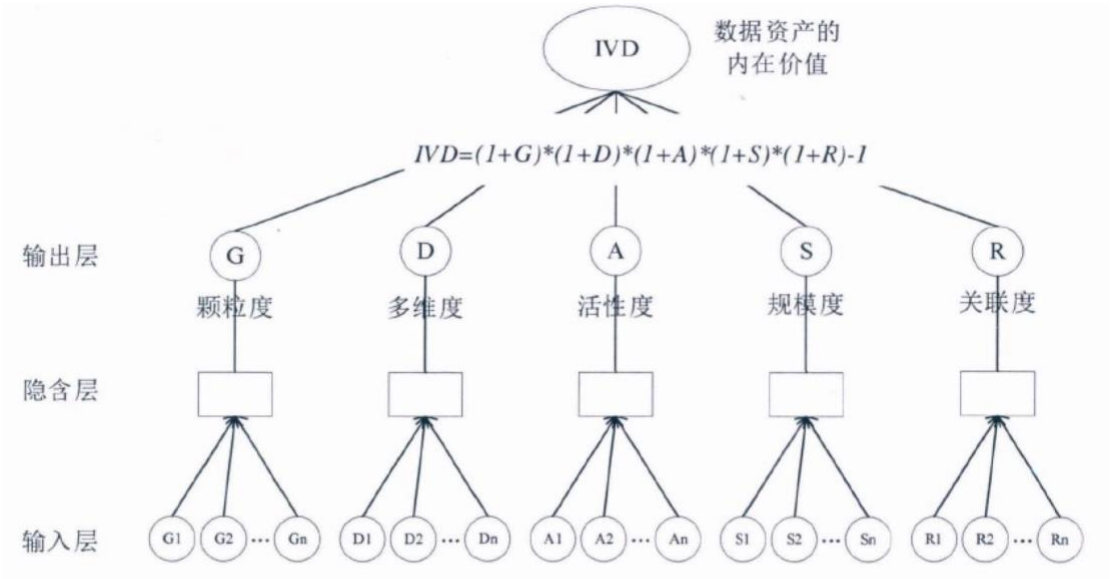


图 3-2 基于特征维度的数据资产价值分析模型

Figure 3-2 The Value Analysis Model of Data Assets Based on Feature Dimension

简述：

144 个维度经过 DL 得出 5 个特征维度的值，再计算出数据资产的内在价值

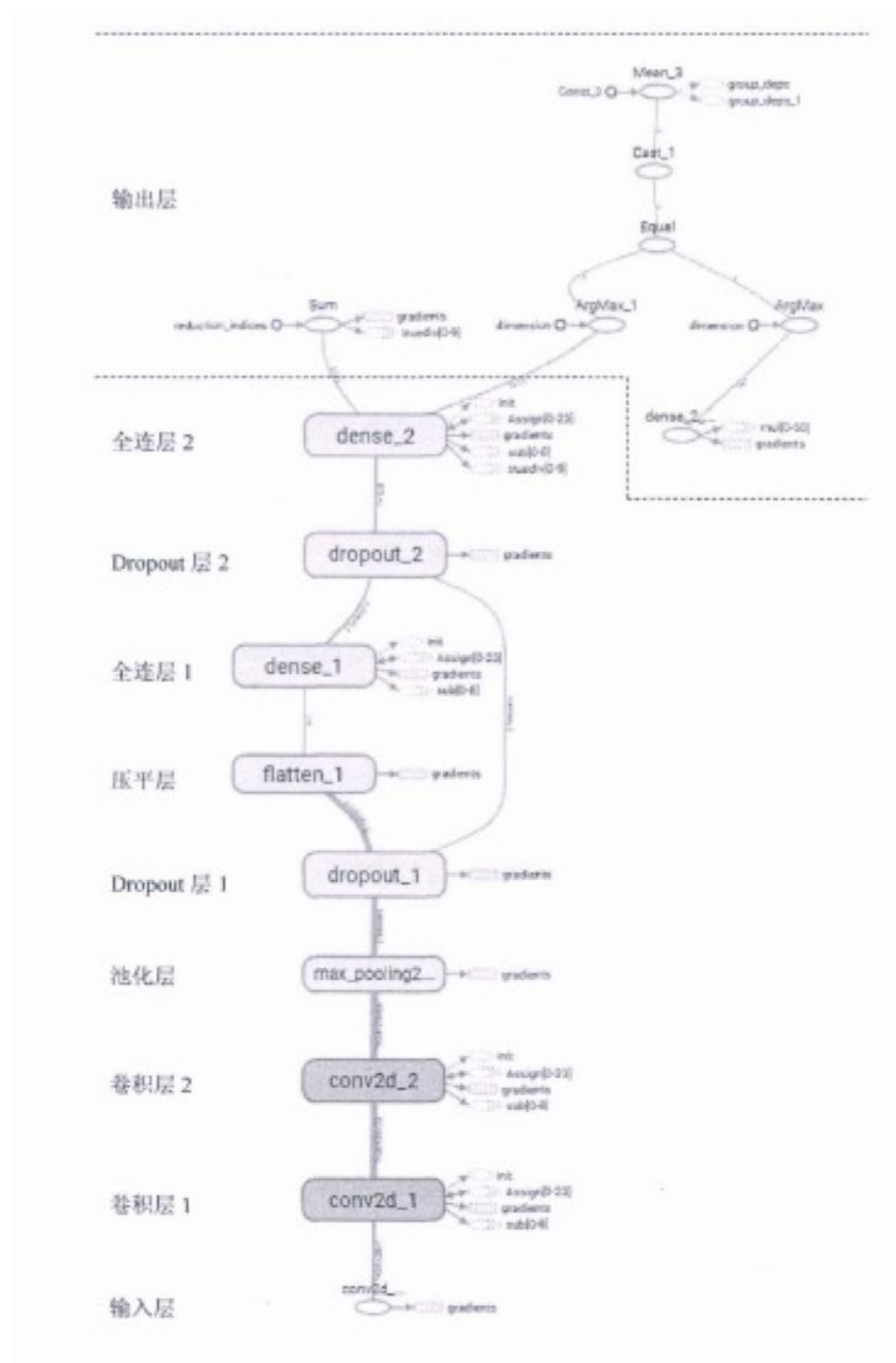


图 4-2 数据资产价值分析模型结构

Figure 4-2 The Structure of the Model

DL 网络的模型

表 4-7 EDAAM 模型的训练结果

Table 4-7 The Result of EDAAM

		颗粒度	多维度	规模度	活性度	关联度
进口业务	过磅数据	8.5	6.1	5.2	9.1	3.7
	客户装卸量数据	4.3	8.7	3.6	5.8	8.4
	客户出库单数据	9.2	3.4	9.4	8.7	6.3
	客户交易情况数据	7.7	9.9	6.5	7.2	3.5
	客户需求量数据	8.8	8.5	4.7	5.6	6.6
	客户库存量	3.8	6.5	6.5	7.8	4.4
出口业务	过磅数据	8.2	5.8	5.3	9.5	3.5
	客户装卸量数据	4.6	7.9	4.0	6.0	8.8
	客户出库单数据	9.1	3.5	9.5	9.5	6.3
	客户交易情况数据	7.6	9.8	6.2	7.9	3.8
	客户需求量数据	8.0	8.3	4.3	5.6	6.1
	客户库存量	3.6	6.8	6.3	7.1	4.4

这里的数字是依据什么公式计算出来的？

是否根据 **正确率*10** 的得到这个数字？文章没有详细给出