图2.1 编辑距离计算示意图 7

图2.2 MPN算法流程图 11

图2.3 多趟优先队列扫描算法过程示意图 12

图2.4 神经元模型 12

图2.5 三层BP神经网络 14

图3.1 滑动窗口扫描过程 23

图3.2 不同大小的滑动窗口的归并过程图 24

图3.3 自适应大小的滑动窗口示意图 26

图3.4 OMPN算法流程图 30

图3.5 SNM、MPN、OMPN算法查全率折线图 36

图3.6 SNM、MPN、OMPN算法查准率折线图 36

图3.7 SNM、MPN、OMPN算法运行时间折线图 37

图3.8 SNM、MPN、OMPN算法F1-measure折线图 38

图4.1 进行相似重复记录检测的BP网络结构 40

图4.2 BP网络进行重复记录检测的算法流程图 41

图4.3遗传神经网络检测相似重复记录流程图 43

图4.4 A-OMPN算法流程图 47

图4.5 BP-OMPN算法流程图 48

图4.6 GA-ANN、OMPN、A-OMPN、BP-OMPN算法查全率折线图 51

图4.7 GA-ANN、OMPN、A-OMPN、BP-OMPN算法查准率折线图 51

图4.8 GA-ANN、OMPN、A-OMPN、BP-OMPN算法运行时间折线图 52

图4.9 A-OMPN、BP-OMPN算法训练时间折线图 53

图5.1 六大功能模块示意图 56

图5.2 航天情报信息管理系统架构图 57

图5.3实体关系结构图 58

图5.4 系统web端数据检索界面图63

图5.5 系统web端数据分析界面图 63

图5.6 系统移动端界面图 64

图5.7 数据清洗业务流程图 64

表2.1 学生信息表中的重复记录 5

表3.1 四条相似重复记录示例表 18

表3.2排序关键字方案一 19

表3.3排序关键字方案二 19

表3.4排序关键字方案三 19

表3.5 基于字段区分度的排序关键字表 21

表3.6 缺失数据及不完整数据示例表 27

表3.7 febrl数据记录字段说明 30

表3.8 febrl公开接口的参数说明 31

表3.9 测试数据集参数取值说明 33

表3.10 实验环境配置表 34

表3.11 SNM、MPN、OMPN算法查全率对比表 34

表3.12 SNM、MPN、OMPN算法查准率对比表34

表3.13 SNM、MPN、OMPN算法运行时间对比表 34

表3.14 SNM、MPN、OMPN算法F1-measure值对比表 37

表4.1 GA-ANN、OMPN、A-OMPN和BP-OMPN算法查全率对比表49

表4.2 GA-ANN、OMPN、A-OMPN和BP-OMPN算法查准率对比表 50

表4.3 GA-ANN、OMPN、A-OMPN和BP-OMPN算法运行时间对比表 50

表4.4 A-OMPN和BP-OMPN算法训练时间对比表 53

表5.1需求分析总结表 55

表5.2 航天器字段设计表 58

表5.3 航天国家字段设计表 59

表5.4航天器故障字段设计表 59

表5.5 卫星平台字段设计表 60

表5.6 航天情报信息系统开发与运行环境 60

表5.7 数据接口设计表 61

表5.8 航天器重复记录举例 62

表5.9 数据预处理规则表 65