DTCC

第十四届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA

数智赋能 共筑未来

IT168.com ChinaUnix ITPUB

北京国际会议中心 | 2023/8/16-18

大数据软件开发项目造价（估算）应用研究初探

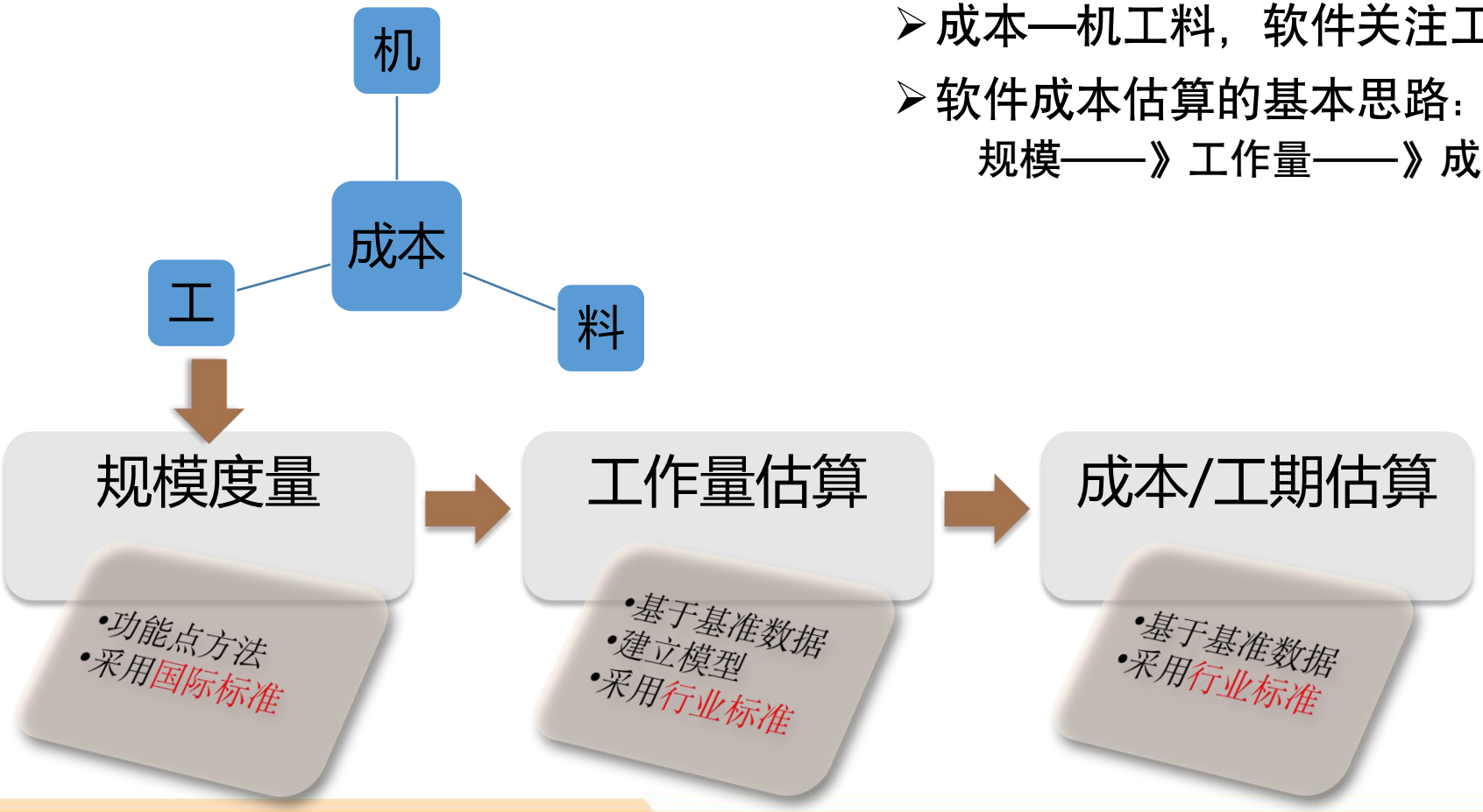
赛宝认证中心 技术经理 门轩庭

通用软件开发造价方法

遵循标准：

- 《GB/T 36964-2018 软件工程 软件开发成本度量规范》

国标普通软件开发造价的基本思路



- 成本—机工料，软件关注工
- 软件成本估算的基本思路：
  规模——》工作量——》成本——》价格

软件规模是最基础的度量估算项

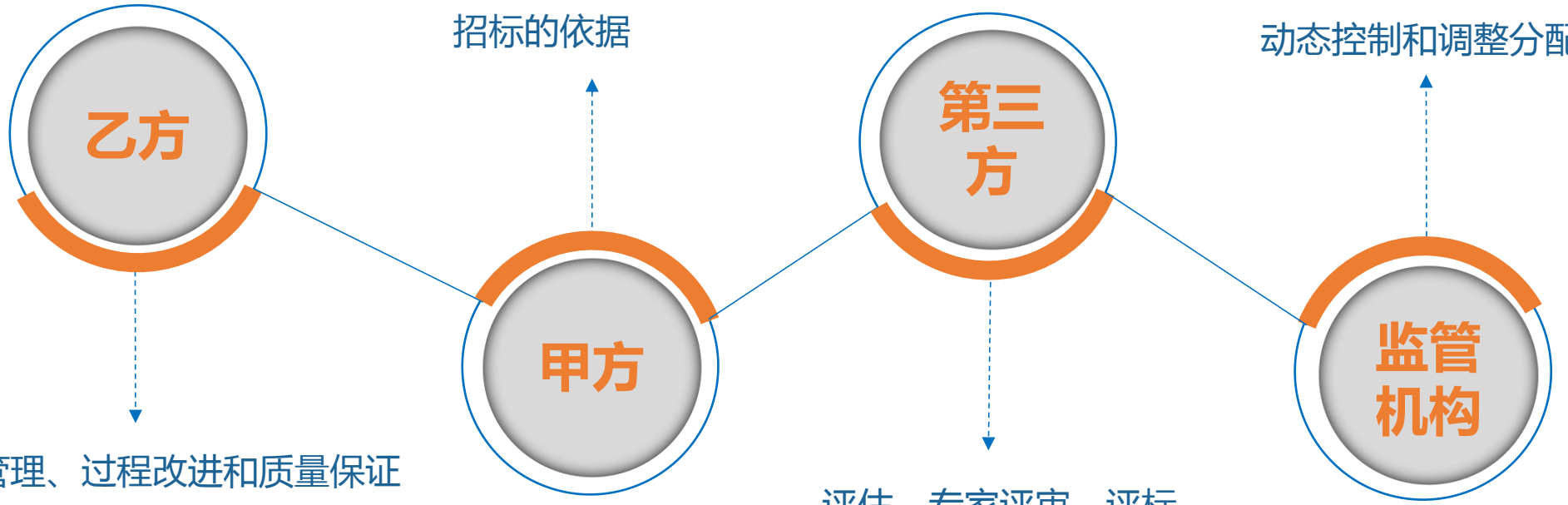
软件度量估算闭环 — 始于规模、终于规模

终：
成本/规模=单位规模费率
工作量/规模=生产率
缺陷数/规模=缺陷率

始：
软件规模*生产率=工作量
软件规模*单位规模费率=成本
软件规模*缺陷率=缺陷数

规模-->成本
规模-->工作量-->成本

造价应用场景-机构



项目管理、过程改进和质量保证
投标依据
良性竞争环境

审批预算的依据
提高预算审批效率
招标的依据

评估、专家评审、评标
培训、咨询
客观、公正依据

监督管理，产业指导
合理规划，促进发展
动态控制和调整分配

造价应用场景-项目阶段

- 各省市行业发布的指导办法；
- 权威行业机构的基准数据和人月费率；
- 潜在投标人所在地域；
- 技术要求和领域的成熟度；
- 如和预算无实质性变更可直接使用预算结果；
- 评标基准采用行业数据中值，或有效报价的平均价；

- 应对任务进行分解估算；
- 采用经验法如专家验证法进行验证，如果有较大差异时分析原因并适当调整；
- 制定计划时需要考虑各项任务的资源约束；

- 委托方、开发方应对变更范围达成一致；
- 估算结果应得到委托方、开发方评审，达成共识；

- 应考虑合理毛利润和维护因素
- 如需求模糊可直接使用基准比对方法，直接估算工作量和成本；
- 需要考虑需求变更、质量要求、工期约束；
- 审核是应考虑项目预算合理性和资金情况；

- 宜将项目规模、工作量和成本与预算进行对比，考察预算执行情况和项目实施效果；
- 将项目规模、生产率、成本纳入基准数据库
- 可与行业基准数据库进行比对分析，发现改进机会；

功能点度量—NESMA的五大功能项

- 功能点国际标准和国家标准

➢数据功能类型（逻辑文件）：系统使用或维护了哪些数据

- 内部逻辑文件ILF：在本系统维护的业务数据
- 外部接口文件EIF：本系统引用，其他系统维护的业务数据

➢事务功能类型（基本过程）：系统如何使用或维护这些数据

- 外部输入EI：对数据进行维护或改变系统状态/行为
- 外部输出EO：对数据加工后呈现或输出
- 外部查询EQ：对已有数据直接呈现或输出

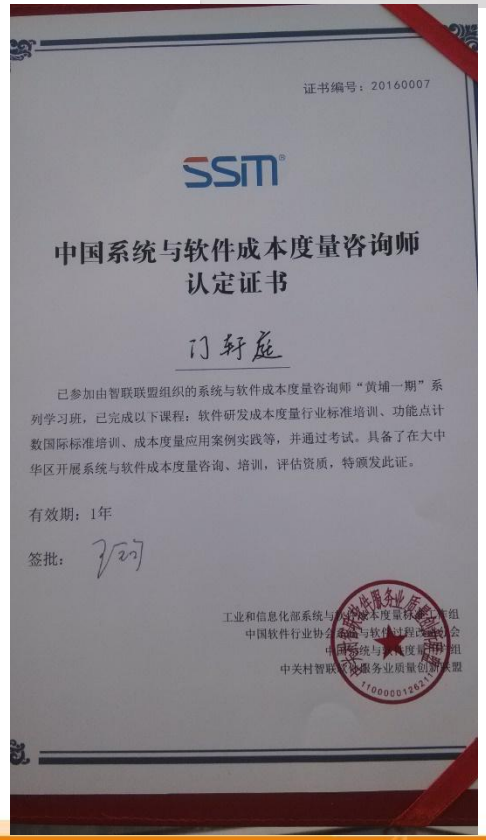
资格证书


持证人参加：
软件工程造价师
职业技术培训，完成培训计划所
规定的内容，经考核，达到相关
职位要求的专业技术水平。
特发此证。
姓 名：门轩庭
身份证号：612301197310290817
证书号码：C1811J46R2700043
2019 年 5 月 8 日


持证人参加：
软件工程造价评估师
职业技术培训，完成培训计划所
规定的内容，经考核，达到相关
职位要求的专业技术水平。
特发此证。
姓 名：门轩庭
身份证号：612301197310290817
证书号码：C1911J46R2900367
2019 年 4 月 30 日


厦门国家会计学院
XIAMEN NATIONAL ACCOUNTING INSTITUTE
结业证书
门轩庭 性别 男，于 2019 年
11 月 21 日至 2019 年 11 月 25 日在本院
参加 科技事业全面实施预算绩效管理 第五期专题培训班 学习，
共计 24 学时，成绩合格，准予结业。
厦门国家会计学院
2019 年 11 月 25 日
编号：XM J201911043055


系统与软件度量
SSMCCAP
Certified Software Cost Appraisal Professional
门轩庭
612301197310290817
Certification ID：CCAP_201512200009
Test Date：December 20, 2015
秘书长签字：
Copyright 2016 by IQalliance.All rights reserved.


证书编号：20160007
SSM
中国系统与软件成本度量咨询师
认定证书
门轩庭
已参加由智联联盟组织的系统与软件成本度量咨询师“黄埔一期”系列学习班，已完成以下课程：软件研发成本度量行业标准培训、功能点计数国际标准培训、成本度量应用案例实践等，并通过考试。具备了在大中华区开展系统与软件成本度量咨询、培训，评估资质，特颁发此证。
有效期：1年
签批：


Cito | International
nesma
CERTIFICATE
The Netherlands Software Metrics users Association, NESMA, certifies that
M.X.T. Men
has achieved the title
CERTIFIED FUNCTION POINT ANALYST
CFPA®
Date of birth: 29-10-1973
Date of exam: 23-11-2016
Score: 8
Result: Pass
Ton Dekkers
Chairman NESMA
Alex Stolwijk
Project Manager Cito
Arnhem, 20-01-2017
now you know


COSMIC
Common Software Measurement International Consortium
Achievement
Award
This award recognises
Xuanting Men, CCPL
Certified COSMIC at Foundation Level
Has achieved Certification and meeting the standard required to be identified as a COSMIC Foundation Level (v4) in April 2019
Jean-Marc Desharnais
Signature

大云智算的融合性

- 大数据将成为“生产资料”，计算会是“生产力”。
- 大数据这一“食材原料”，借由人工智能这把锋利的“菜刀”，加上大规模算力的“火力”，配以IoT、5G、区块链等“佐料”，以云计算和移动互联的“厨艺”，工程师就可以做出一桌智能应用的“美食佳肴”。

- 大数据、算法计算、AI、云计算等等技术能力是天然嵌入式地整合在一个平台上的。

赛宝大数据造价项目　部分案例

- 国家电网基于电力大数据的能源公共服务平台及应用建设项目
- 电网2020年生产域个性化运营管控应用建设设目大数据及算法功能
- 电网数据中心实时数据服务大数据平台
- 电网大数据AI平台与算法计算大数据类软件开发造价研究项目开题报告
- 电力大数据的能源公共服务平台
- 国家电网基于大数据平台的全网谐波数据分析研究和推广应用
- 中国人民银行金融大数据分析及服务平台大规模并行处理分析型分布式数据库系统建设服务项目
- 武汉AI园智能办公-大数据成品软件采购询价项目
- 时代凌宇基于大数据的智慧城市应用及运营解决方案项目
- 大连港集团编码系统大数据元集项目
- 深圳海关大数据智能分析平台通关物流风险监控子系统
- 深圳电子口岸“深关通”大数据交换客户端优化完善项目
- 广东税务大数据基础软件采购项目
- 广东省税务局大数据平台监理服务项目
- 广东税务大数据基础软件采购项目
- 广东税务大数据应用融合项目
- 广东国税大数据管理平台及应用系统
- ……

赛宝认证中心

工业和信息化部电子第五研究所直属机构

首家把认证引入中国的企业

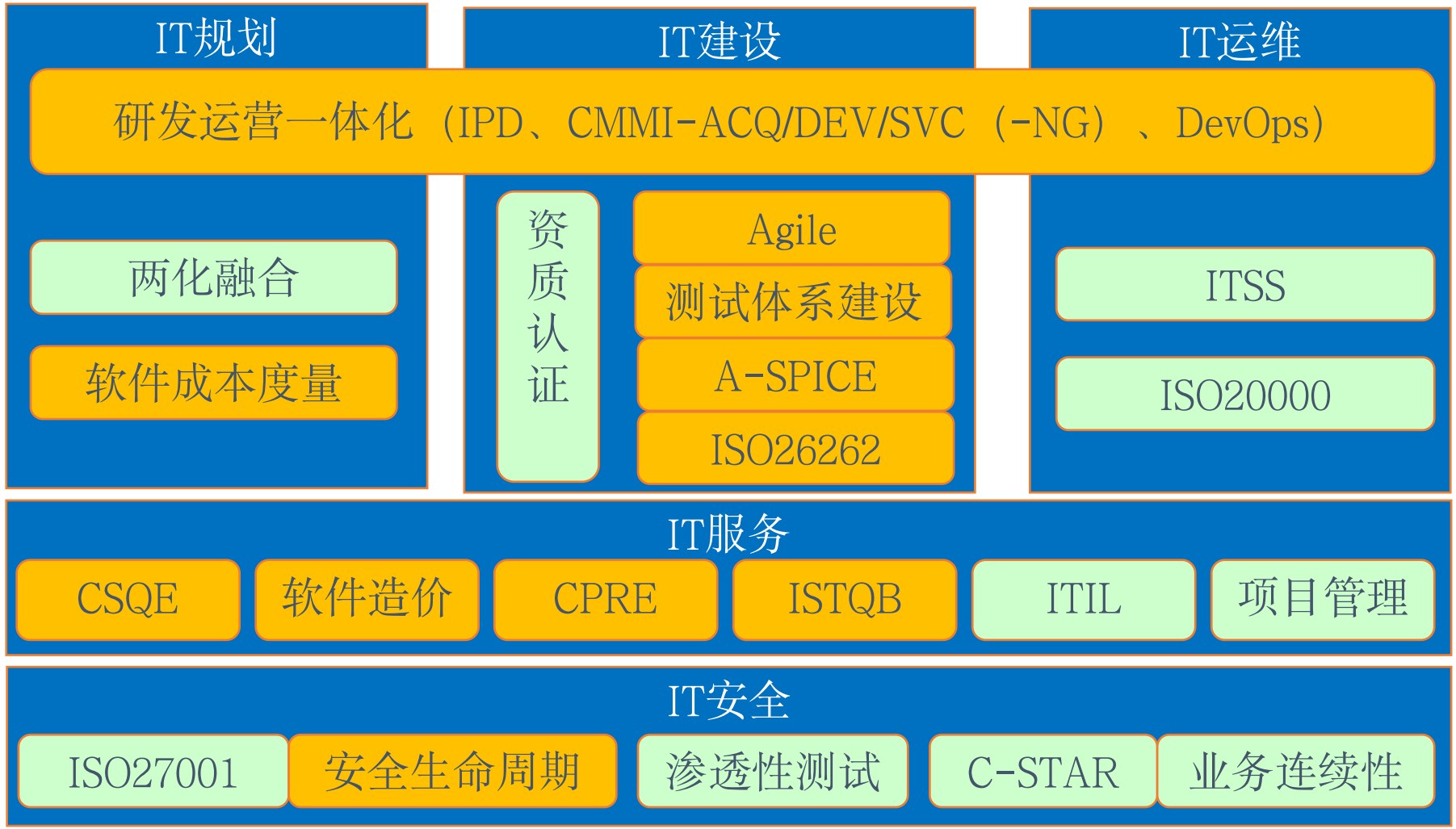
CEPREI

赛宝认证

获得十多家国家主管部门和机构的授权和认可

第三方认证、评估、培训及技术服务权威专业机构

赛宝认证中心IT业务概述



赛宝软件研发管理业务部分客户

研究情况

序号	计划	完成情况	
1	大数据类软件开发现状及需求调研：结合数字化转型发展，充分调研了解近几年大数据类软件开发的现状及需求，**分析大数据类软件开发的特点，了解当前大数据类软件开发估算、概算等测算过程中存在的问题与不足。**	进行了行业内和内部的现状调研，分析大数据类软件开发的特点，提出当前进行大数据类软件开发造价遇到的问题。	已完成
2	**构建大数据类软件开发造价模型及标准**：根据大数据类软件开发的特点，借鉴行业软件成本度量的领先实践，建立一套适合大数据类软件开发的造价模型及测算标准。	针对前期现状调研发现的问题，结合大数据类软件开发特点，参考行业内其他领先实践，构建大数据类软件开发的造价模型及测算方法。	已完成
3	应用验证：**将新的造价模型应用到大数据类软件开发造价测算**，并比较新旧两种造价模型的差异，最后形成大数据类软件开发造价标准。	选取大数据类软件开发项目进行验证，将新造价模型得到的测算结果与原造价模型得到的结果进行对比分析，得出结论。	已完成

现状分析范围

本次分析的范围包括对国内外软件行业造价的发展及现状进行调研分析。

主要包括以下内容：

1) 大数据类软件开发项目的分类特点；
2) 行业内大数据类软件开发造价现状；
3) 大数据类软件开发规模度量方法；

参考文献

- 2A市智慧城市建设项目成本管理研究_王玺博.caj
- 人工智能在电力系统及综合能源系统中的应用综述_杨挺.pdf
- 基于大数据的区域医疗信息共享体系研究_张传文.caj
- 大数据平台管理系统的设计与实现_冯腾霄.caj
- 大数据系统综述_李学龙.pdf
- 1浅谈大数据环境下项目成本管理优化_彭晋谦.pdf
- 大数据_概念_技术及应用研究综述_方巍.pdf
- 人工智能技术的发展与应用_贺倩.pdf
- 大数据时代下数据挖掘技术在电力中的应用分析_杜韫成.pdf
- 网络大数据_现状与展望_王元卓.pdf
- GB T 36073-2018 数据管理能力成熟度评估模型.pdf
- 智能电网_研究综述_鞠平.pdf
- 大数据系统和分析技术综述_程学旗.pdf
- 大数据研究_未来科技及经济社会发_省略_领域_大数据的研究现状与科学思考_李国杰.pdf
- 人工智能技术研究及未来智能化信息服务体系的思考_王志宏.pdf
- 大数据下的机器学习算法综述_何清.pdf
- 人工智能发展综述_朱祝武.pdf
- 重大装备工业大数据平台的实现方案_师超.p
- 大数据研究综述_涂新莉.pdf
- 大数据研究_严霄凤.pdf

参考文献：

[1]王元卓，靳小龙，程学旗. 网络大数据:现状与展望[J]. 计算机学报，2013(06):3-16.
[2]程学旗，靳小龙，王元卓等. 大数据系统和分析技术综述[J]. 软件学报，2014，25(9):1889-1908.
[3]GB/T 36964-2018.软件过程 软件开发成本度量规范[S].北京：中国标准出版社.2018.
[4]中国电子技术标准化研究院.2020 年中国软件行业基准数据[R].北京，2020.
[5]李华北. 软件成本度量及造价分析[M]. 电子工业出版社,2018.
[6]张旸旸.软件成本度量国家标准实施指南:理论.方法与实践[M].电子工业出版社,2020.

大数据类软件开发和传统应用开发的区别

非结构化数据

不仅可以处理结构化数据，也可以处理图像、声音、文件等非结构化数据

传输与存储

线下传输升级为API接口传输

数据规模

和传统应用相比，大数据类软件开发处理的数据规模量极大

处理方式

标签化处理，根据标签抽取数据

价值的不确定

对现象发生过程的全记录，通过数据不仅能够了解对象，还能分析对象，掌握对象运作的规律

大数据软件开发项目的分类特点

- 大数据应用

- 快速的数据流转
- 多样的大数据类型

- 海量的数据规模
- 价值密度不定
- 超出了传统数据库软件工具能力范围的数据集合
- 和传统的流程型、事务性软件相比，在数据采集、清理及分析方面的功能更多

大数据软件的主要功能项识别做法

大数据软件包括数据集、数据模型、数据采集、数据清洗、数据加工、数据计算分析、数据挖掘钻取、数据查询搜索、数据统计等等。

这些功能都可以与5类功能项建立对应关系。

例如：

- 大数据平台的数据集都是业务系统中的数据功能，记逻辑文件；
- 数据采集通常识别为EI；
- 数据清洗有计算处理，计为EO；
- 为报表平台提供报表识别为EO。

数据预处理的功能点识别

大数据项目其中一个特点是数据源的多样性，可以包含各种类型各种版本的数据库、文本文件、网页、日志，甚至包含图片、视频信息，也可能包括传感器、软硬件接口等信息来源。为确保后续工作能够有一个高质量的数据集，在数据采集时往往会进行必要的预处理。

- 针对一个数据源的同一数据对象，如存在多处需要进行预处理的信息，仅识别一个ILF
- 由于预处理工作本身涉及格式转换、协议解析、图形识别等计算过程，因此这些功能应该识别为外部输出EO
- 每一个预处理场景识别一个外部输出EO，而不可依据抓取数据的字段进行识别
- 从结构化数据中获取数据，如从数据库、确定格式的excel、列表文件中采集数据，识别EI

数据采集层

不同的接入结构和模式，需要分别识别为不同EI或EO

序号	采集方法	采集工具	说明
1	离线采集	ETL	包括数据的提取（Extract)、转换(Transform)和加载(Load)。在转换的过程中，需要针对具体的业务场景对数据进行治理，例如进行非法数据监测与过滤、格式转换与数据规范化、数据替换、保证数据完整性等。EO
2	实时采集	Flume/Kafka	实时采集主要用在考虑流处理的业务场景，数据采集会成为Kafka的消费者，就像一个水坝一般将上游源源不断的数据拦截住，然后根据业务场景做对应的处理（例如去重、去噪、中间计算等），之后再写入到对应的数据存储中。EO
3	互联网采集	Crawler, DPI等	Scribe是Facebook开发的数据(日志)收集系统。又被称为网页蜘蛛，网络机器人，是一种按照一定的规则，自动地抓取万维网信息的程序或者脚本，它支持图片、音频、视频等文件或附件的采集。EO
4	其他数据采集方法		对于企业生产经营数据上的客户数据、财务数据等保密性要求较高的数据，可以通过与数据技术服务商合作，使用特定系统接口等相关方式采集数据。EI

1)数据清洗

序号	场景	处理方法	功能点识别
1	值缺失	平均值、最大值、最小值或更为复杂的概率估计代替缺失	EO
2	错误值检测	统计分析的方法识别	EO
3	重复记录检测	重复检测	EO
4	不一致检测	完整约束定义	EO

存储层

2)数据识别

类别	场景	说明	功能点识别
实体识别	同名异义	字段名相同，但实际是不同的数据	EO
	异名同义	字段名虽然不同，但是实际是相同的数据	EO
	单位不统一	同一个字段里，用的多个单位	EO
冗余属性识别	同一属性多次出现	识别出属性数据的重复	EO
	同一属性命名不一致导致多次重复	识别因命名不一致导致的重复	EO

3)数据变换

序号	场景	处理方法	功能点识别
1	简单函数变换	平方、开方、取对数、插分运算	EO
2	规范化（归一化）	离差标准化、标准差标准化、小数定标规范化	EO
3	连续属性离散化	常用的离散化方法、等宽法、等频法、基于聚类分析的方法	EO
4	属性构造	构造出新的属性并添加到属性中	EI
5	小波变换	时频分析变换	EO

4)数据规约

序号	场景	处理方法	功能点识别
1	属性规约	寻找出最小的属性子集，并确保新数据子集的概率分布尽可能地接近原来数据集的概率分布	EO
2	数值规约	回归、对数线性模型、直方图、聚类、抽样	EO

5)数据融合

序号	场景	处理方法	功能点识别
1	数据汇聚	将非结构化的数据进行梳理总结，汇聚在一起形成有意义的数据信息	EO
2	数据合并	将不同数据表中的数据经过汇总，合并保存为同一个数据	EO

6)数据存储

序号	功能说明	说明	功能点识别
1	数据信息	对采集到的大数据信息进行预处理后存储在系统中的处理后数据，每一个有独立存在意义的数据实体可识别为一个ILF	ILF
2	数据信息-增	涉及到的数据新增功能	EI
3	数据信息-删	涉及到的数据删除功能	EI
4	数据信息-改	涉及到的数据修改功能	EI
5	数据信息-查	涉及到的数据查询功能	EO
5	数据信息-查	涉及到的数据查看功能	EQ

分发层

数据分发主要指数据通过网络传递到不同节点的过程。

分发层主要为应用层提供服务及数据的分发能力。

- 为完成数据分发目的，专门创建的对外发送的文件可以识别为内部逻辑文件ILF
- 数据接口应按EO进行识别
- 对外提供数据的每一个接口均可识别为一个外部输出EO
  - 这里的接口必须为在本系统内开发，给其他系统调用的接口，
  - 而其他系统开发本系统仅仅是使用该接口的情况，则不能计数为本项目的功能点。

数据分析挖掘

数据分析挖掘是大数据应用体系中的关键支撑环节，是指从大数据中发现潜在未知信息和模型的分析计算过程。

针对数据分析挖掘的相关成本度量规则如下：

- 分析建立的模型记录信息，识别为内部逻辑文件ILF
- 各类数据分析挖掘算法，包括决策树分类、K均值聚类、支持向量机分类等算法，使用到每一种算法可以识别为一个外部输出EO
- 多个数据对象进行同一个算法分析，不可重复填报算法分析EO

服务层

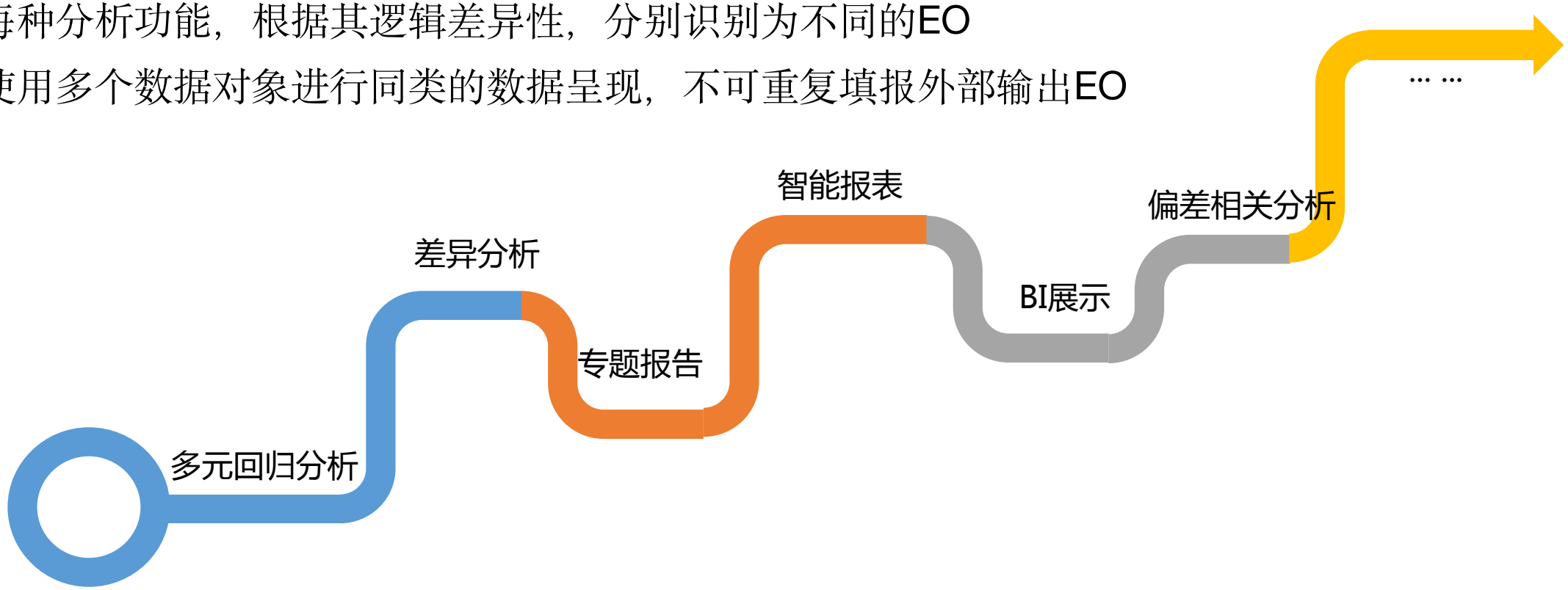
大数据的核心有约32个算法，对已经列出的算法进行分析，确定每一个算法赋予的功能点数

序号	类型	说明	功能点
1	数据挖掘	每种挖掘技术分别识别为EO，如数据网络挖掘、特异群组挖掘、图挖掘等新型数据挖掘技术； 每种行为分析分别识别为EO：如用户兴趣分析、网络行为分析、情感语义分析等。	EO
2	模型预测-预测模型	对模型的相关操作识别为对应的EI、EO等	ILF
3	模型预测-数据建模	对现实世界各类数据的抽象组织，确定数据库需管辖的范围、数据的组织形式等直至转化成现实的数据库	EI
4	模型预测-机器学习	对大量数据进行分析，寻找统计规律，建模，并使用模型对新数据进行预测和分析	EO
5	模型预测-建模仿真	采取数学手段描述事务，并通过参数输入等方法模拟真实情况并进行展示	EO

数据呈现应用层

数据呈现类工作，以获取并处理好的大数据为基础，通过智能报表、专题报告、BI展示、平台接口等方式展示应用数据。

- 每种分析功能，根据其逻辑差异性，分别识别为不同的EO
- 使用多个数据对象进行同类的数据呈现，不可重复填报外部输出EO



建立大数据软件开发规模估算指南

通过规模估算指南确定系列识别规则，确保能够针对大数据类软件开发方便的识别和计数，

并确保其实现工作量与通用的功能点识别规则是相当的

大数据应用方面

进行系统层级梳理，对十余种常见的大数据平台开发工作进行分解分类

根据分解分类后的典型工作项，分别分析该功能项的识别方法

选取项目对评估方法及造价模型进行试点应用

项目验证

例如：城市驾驶舱大数据展示

按照本项目研究出的评估方法，对本次验证项目进行功能点识别，识别的原始功能点数为7527个功能点

城市运行全景图	城市运行指数展示	EO
	社会保障指数展示	EO
	社会治理指数展示	EO
	生态环境指数展示	EO
	公共安全指数展示	EO
	城市设施指数展示	EO
	市容市貌指数展示	EO
	城市秩序指数展示	EO
	城市执法指数展示	EO
	民情民意态势	EO
	社区管理服务态势	EO
	城市运行管理态势	EO
	人口管理态势	EO
	卫生健康态势	EO
	民生服务态势	EO
	教育关爱态势	EO
	特殊群体社会关爱态势	EO
	公共资源态势	EO
	生态质量指数	EO
	城市建设态势	EO
	城市生命线建设态势	EO
	水体整治态势	EO
	交通拥堵指数态势	EO
	非机动车态势	EO
	城市停车态势	EO
	大客流态势	EO
	待处理事件数量	EO
	告警处置率	EO
	自然灾害预警	EO
	治安安全预警	EO
	市场监管预警	EO
	交通安全预警	EO
	火警警情预警	EO
	食品安全预警	EO
	药品安全预警	EO
社会治理一张图	网格治理监测指标	EO
	城市管理监测指标	EO

项目验证

项目名称	原评估工作量	新规则指导评估工作量	实际总工作量	偏差率	
				无指导原评估	新指导规则
数据资产管控	509	655	668	-24%	2%
数据集市建设	1508	2077	2034	-26%	-2%
监控指挥中心	2064	2586	2706	-24%	4%
城市驾驶舱	1043	1194	1231	-15.27%	-2.98%
档案数据中台	2156	2315	2417	-14.58%	-4.18%

对比分析验证

定义详细的识别规则，通过详细的识别规范及对应的示例，减少不规范识别功能项及上报功能点数的问题，使原始功能点的识别更加正式、规范。

验证结论：由验证项目示例可以看出，新的评估规则方法更加科学、细致，在对大数据软件开发项目进行评估时，更能够依据项目实际情况反应其建设规模、工作量和成本

THANKS