

데이터 전처리 Data Preprocessing

修: 닦을 수

1. 닦다, 익히다, 연구하다(研究-)
2. 꾸미다, 엮어 만들다
3. 고치다, 손질하다
4. 다스리다, 정리하다(整理-)
5. 갖추다, 베풀다
6. (도덕, 품행)을 기르다
7. 길다, 높다
8. 뛰어나다
9. 행하다(行-), 거행하다(舉行-)
10. 뛰어난 사람

安: 편안 안

1. 편안(便宜)
2. 편안하다(便宜-)
3. 편안(便宜)하게 하다
4. 안존하다(安存--: 아무런 탈 없이 평안히 지내다)
5. 즐거움에 빠지다
6. 즐기다, 좋아하다
7. 어찌
8. 이에(乃), 곧
9. 어디에
10. 안으로, 속으로

Site: <http://suanlab.com>

E-mail: suanlab@gmail.com



- **Data preprocessing** is an important step in the data mining process.
- The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects.
- Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: -100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc.
- Analyzing data that has not been carefully screened for such problems can produce misleading results.
- Thus, the representation and quality of data is first and foremost before running an analysis.
- Often, data preprocessing is the most important phase of a machine learning project, especially in computational biology.



Syllabus

1. 데이터 구조와 종류
2. 데이터 전처리 개요
3. 데이터 수집
4. 데이터 정제
5. 데이터 통합
6. 데이터 축소
7. 데이터 변환
8. 중간고사
9. 연관 규칙 Association Rules
10. 분류 Classification
11. 클러스터링 Clustering
12. 특징 생성
13. 특징 선택
14. 실전 데이터 전처리
15. 기말고사



<http://suanlab.com/dp.html>

Home – Lecture – Data Preprocessing

Q & A