



Promoting Data Mining

Gregory Piatetsky-Shapiro
KDnuggets



DM Conferences

- KDD-07, 08:
 - over 500 research submissions
 - 60-80 Industry Track
- ICDM – over 500 research submissions

Conferences in 2008

Europe (8):

- Jan, EGC'2008 (Knowledge extraction and management), France
- May, Wessex Data Mining 2008, Cadiz, Spain.
- Jul, Industrial Conference on Data Mining ICDM 2008, Leipzig, Germany
- Jul, IADIS European Conference On Data Mining 2008, Amsterdam
- Sep, ECML/PKDD-2008, Antwerp, Belgium
- Oct, Discovery Science (DS-2008), Budapest, Hungary
- Oct, (**new!**) RecSys'08: Recommender Systems, Lausanne, Switzerland.
- Dec, ICDM '08: Pisa, Italy.

North America (3):

- Feb, (**new!**) WSDM 2008, Web Search and Data Mining, Stanford
- Apr, SIAM (SDM) 2008, Atlanta
- Aug, KDD-2008, Las Vegas

Asia

- May, PAKDD 2008, Osaka, Japan.
- Jul, ADMA 2008, Chengdu, China.

Australia, Brasil, Latin America

Vendor conferences: SPSS, SAS



KDD Data Repository

- Currently, no large public data repository
 - UCI KDD has only 6 med-size datasets
 - KDD Cup data
- Create infrastructure for large data
 - NCDM in Chicago will have facilities
- DM conferences and journals to encourage authors to also submit data

Common Meta-data Format in DMML : Data Mining Meta-Data Language



- Different data formats will remain
 - Open: csv, txt, arff
 - Proprietary: SAS
 - Database
- Common meta-data will improve interoperability
- Meta-data generator for simple formats



Data Mining Meta-data Language

- DMML – should be simple, easy to use
 - Subset of PMML
- Converters from common types
- Common ontology for semantic types can be developed
 - (similar to GO: Gene Ontologies)
- US and EU and Asian organizations can fund development



Data Mining Curriculum

- Broad applications
 - Data Mining but also
 - E-commerce, Web search, social networks, recommender systems, web analytics

- Attract more students from CS, Engineering, Stats, Math, Physics

Improve Public Image of Data Mining in US



Current Image is Negative

- Stories about US Government data mining
- Big Business data mining (e.g. Pharmacy data mining)
- Not a problem in Europe or Asia, but a problem for US researchers
- Generate press on positive contribution of DM to society
- New prize – DM contributions to society?