The Bright and Dark Side of Data Mining Research

Limsoon Wong
Prepared for Panel Discussion
at PAKDD 2006
11 April 2006 @ Hilton Hotel
Question 1
Share an experience that illustrates the bright or dark side of data mining research

• Bright side:

Results

Conventional Tx:
• intermediate intensity to everyone
  ⇒ 10% suffers relapse
  ⇒ 50% suffers side effects
  ⇒ costs US$150m/yr

Our optimized Tx:
• high intensity to 10%
• intermediate intensity to 40%
• low intensity to 50%
• costs US$100m/yr

• High cure rate of 80%
• Less relapse
• Less side effects
• Save US$51.6m/yr
Question 2
In your opinion, what are the bright areas in data mining research today?

• **Marriage of statistics and data mining**
  – Efficient, sound, & complete mining of patterns satisfying more sophisticated statistical properties

• **Combining IE & data mining, giving more attention to feature generation & selection**

• **Impt application areas:**
  – Biomedicine
  – Economic crimes?
Question 3
What are the common barriers/traps faced by data mining researchers and what advice would you give to avoid them?

• **Traps**
  – Is this yet another algorithm?
  – Is this frequent pattern really useful?

• **Advice**
  – Will my work make an appreciated lasting difference?