



Evaluation of Adaptive Systems

Stephan Weibelzahl



What has to be done, to guarantee the success of an adaptive system and to improve the effects of adaptivity?

Empirical evaluations of adaptive systems are hard to find. However, evaluations are necessary unambiguously. What has to be evaluated and which criteria can be applied?



Each information processing step has to be evaluated on its own to identify false representations and wrong adaptations!

With a so called Layered Evaluation (Karagiannis and Sampson, 2000) each part of the user modeling system is evaluated separately. The steps are interdependent: e.g., evaluation of step 1 is prerequisite to a correct interpretation of step 2, because a wrong inference might either result from inaccurate input data or from a faulty inference mechanism.

Layered Evaluation

1 Evaluation of data acquisition

Example: mouse clicks, ratings, selections

Is the input reliable?

Is it of high external validity?

Criteria: retest-reliability, split-half-reliability, congruence with expert observations, ...

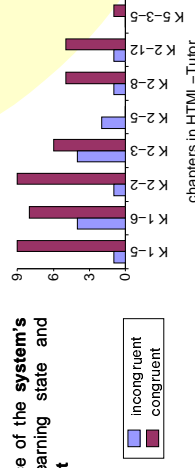
2 Evaluation of inference

Example: inference of preferences or learning status

Is the inference reliable?

Are the user properties correct?

Criteria: congruence with expert ratings, comparison of properties with external test, ...



4 Evaluation of total interaction

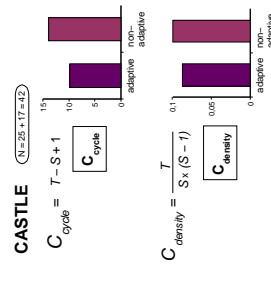
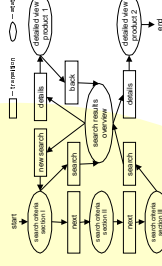
Does the system behavior change?

Does the user behavior change?

How is the interaction quality?

Criteria: impact of adaptivity, user satisfaction, customer satisfaction, duration, usability, behavioral complexity, ...

Adaptivity aims at simplifying the interaction. If interaction is seen as individual state-transition network, it is possible to compute the complexity of the interaction.



3 Evaluation of adaptation decision

Example: link annotation, recommendation of items

Does the system choose the right adaptation decision?

Criteria: comparison of different adaptation decisions, ...