From expert-driven to data-driven adaptation
- Prof. dr. Paul De Bra
- Web-Engineering Group

October 6, 2014
Some wise words to start with (Plato)

• Wise men speak because they have something to say; Fools because they have to say something.

• More wise words, by me: I will let the audience decide which of the above statements applies…
Warning: this lecture touches on religion

- Christians 33.32%
- Muslims 21.01%
- Hindus 13.26%
- Buddhists 5.84%
- Other religions 12.48%
- Non-religious 14.09%

Source: wikipedia.org
The most adaptive system is the human

- we adapt to circumstances that do not fit at all
- sometimes this is because of an (arbitrary) standard
- sometimes it is because of dictatorship (HiPPO)
- democracy can tip the balance
Experts are wrong more often than we think

Doctors Doing Harm Since Hippocrates

‘Explosive’
British Medical Journal

source: book cover

source: James Gillray
Relying on data requires *scientific approach*

Pierre Louis, 1836, had to treat people with pneumonia
- one of the first clinical trials; at random he applied
  - early, aggressive bloodletting, or
  - less aggressive measures
- at the end counted the bodies
  - they were stacked higher over by the bloodletting sink
- on the Web doing controlled experiments to find out what works best is *cheap*
Teaching versus Tutoring

- **classroom model:**
  - one *expert* (teacher) preaches to the whole class
  - one size fits all approach, no or little room for feedback

- **tutoring model:**
  - one-on-one, interactive session with tutor and student
  - easy to do personalization
  - prohibitively expensive

- **only feasible solution:** *automatic* personalization in *on-line learning*

- **note:** a MOOC is *not* the solution, because it still uses one size fits all
User Modeling – Adaptation Loop

- System
  - Collects
  - Processes
  - Processes

- Data about user
  - User Modelling
  - Adaptation

- User Model
  - Adaptation effect
Loop extended with **scrutability**
Simple AHAM Reference Model

- Run-time Layer
  - Presentation Specifications
    - Adaptation model
      - Domain Model
      - User Model
    - Storage Layer
  - Anchoring
    - Within-Component Layer
The Domain Model: the *what* question

- *what* is the application about?
  - *concepts, topics, objects*, often predefined
    - concepts to learn in a course
    - products to buy/sell in a store
    - topics for discussion groups
  - *topics* often *defined* if not predefined
    - topics of news sites
  - *topics* may also be *dynamically deduced*
    - e.g. what do people talk about on Twitter?
The Domain Model: connecting concepts
Connections in the Adaptation Model

- connections in the domain model represent facts
  - e.g. Pieter Lastman was teacher of Rembrandt van Rijn
  - what does that mean for adaptation?
- connections in the adaptation model represent discourse or pedagogical relations:
  - “A is a prerequisite for B”
    \[\approx \text{“A should be read or studied before B”}\]
  - clear intended meaning, relationship defined by expert
- question: how to go from expert-defined adaptation rules (relationships) to data-driven ones?
We need to know **who the user really is**

- well... better think again!

"On the Internet, nobody knows you're a dog."

source: The New Yorker
User Modeling; adapt to whom?
Adaptation: *how* to adapt
How to adapt: *adaptation techniques*

Latest taxonomy of adaptation techniques [Knutov et al, 2009]
Discussion and Impact

Adaptive hypermedia has great potential. There seems to be no end to the amount of information available on the World Wide Web today. When authors, publishing this huge amount, would have the ability to write things differently or write different things, based on their knowledge of the reader, would they use that power? One can never answer that question without having realized the technical context to support it.

Many adaptive hypermedia systems have been created. However, the added complexity of authoring adaptive hypermedia and the limited functionality and performance issues of adaptive hypermedia systems have limited the amount of material produced. This thesis, and accompanying prototype GALE, is based on 15 years of experience of authors writing adaptive hypermedia and engineers developing systems to support them. It answers the questions raised in the introduction.

Research Questions

The research questions that motivated the work described in this thesis are:

- What does a usable generic adaptive hypermedia system look like?
- What performance considerations affect its design?
- How can the design be improved to support the global scale of the World Wide Web?
Example of content and link adaptation

How It Works

The HTTP request URL (that can be seen in the address bar) contains the name of the concept requested. For the current concept 'How It Works' that is: http://gabe.win.tue.nl/thesis/howitworks. A file on the server describes how accessing this concept should affect the user model. Every time a link is followed, the user model is updated. Before the page is sent, the server will use information in the user model to change the page and make it more suitable to your needs. Various techniques are used to change the information in the page.

The links that lead to topics that are part of this thesis, can change in color. You might remember the various colors used:

- blue, when the link is recommended
- purple, when the link is recommended but you have already read its content
- blueish purple, when the link is recommended, you have already read its content, but in the meantime something has changed on the page
- black, when the link is not recommended
- orange, when the link refers to external material

Recommending links is based on properties found in the user model. I try to determine if a concept is suitable, based on your knowledge of prerequisite concepts (explained a bit further down). I keep track of how often you have visited concepts that were suitable. These user model properties are used as a basis for the link colors. Some expression determines what the CSS class of the link should be. The thesis.css file contains sections for the link classes 'good', 'neutral', 'bad', and others.
Example of layout adaptation

The HTTP request URL (that can be seen in the address bar) contains the name of the concept requested. For the current concept ‘How It Works’ that is: http://gale.win.tue.nl/thesis/howitworks. A file on the server describes how accessing this concept should affect the user model. Every time a link is followed, the server will use information in the user model to change the page and make it ‘more suitable’ to your needs. Various techniques are used to change the information in the page.
Expert-driven adaptation

- simplest case: define a *system reaction* for each *user action*
  - this is simple *automation*, not *adaptation*
- standard case: define a *system reaction* for each combination of *user action* and *user model state*
  - of course an adaptation rule only looks at some *specific variables* in the *user model*
- advanced case: define how the *system should learn* to react on a combination of *user action* and *user model state*
Data-driven adaptation: *item-based*

- content-based recommendations based on similarity in the *domain model*: challenge is to find similarity
Data-driven adaptation: *user-based*

- recommendation based on similarity in the *user model*: challenge is to *cluster* users for *collaborative filtering*

- the similarity could actually be between the users or the items
Bad approach: fire all the guns you have

- there is just so much data available…
- content-based filtering works well if there is quality domain-model data available
- collaborative filtering works well if there is a large quantity of event data available
- sometimes combining both improves the recommendation quality, but sometimes it does not!
  - picking the right method or combination of methods is a science by itself (good reason for OpenML research)
Adaptation danger: *over-specialization*

- when adaptation or recommendations become “too good” they turn into *censorship*
- example: adaptive news sites fail to provide important news that does *not fit your profile*
- recommenders need to “think outside the box”
- *adaptation scrutability* is even harder than *user model scrutability*
- especially when you find the user model is correct but the adaptation is wrong
Where did it start?

- photo: Peter and Paul, early apostles of adaptive web-based systems
The future: meta-adaptation

- data-driven adaptation example:
  - “if many people read A before B and few people read B before A then A must be a prerequisite for B”

- data-driven meta-adaptation example:
  - “optimize the thresholds for “many” and “few” in the above example
  - “if you observe a certain navigation pattern often you can start recommending that pattern to others”

- we can start giving the system more and more freedom in how to optimize how it adapts
Where will it end?

• when you base everything on *observed behavior* you may end up with an illusion of:
  • “wisdom of the crowds”
  • democracy
• when you leave the reasoning completely up to the system…
  • Hollywood suggests it leads to *doomsday* scenarios where machines decide to rid the world of humans
  • to prevent this… we must restrict what we allow systems to do, on-line or in the physical world
• many people say *data-driven adaptation is the future*
• but we need to implement *data-driven adaptation* in such a way that in the end it does not end up being *expert-driven* adaptation we *think* it is *data-driven*
• *it’s just like management*…
Questions?

- list of some related projects for reference...
  - AHA! (Adaptive Hypermedia for All)
  - GRAPPLE (Generic Responsive and Adaptive Personalized Learning Environment)
  - CHIP (Cultural Heritage Information Presentation)
  - SeeQR (Structural indexing and EfficiEnt Query processing on massive RDF data sets)
  - Data Mining Without Discrimination
  - HaCDAIS (Handling Concept Drift in Adaptive Information Systems)
  - CAPA (Context-Aware Predictive Analytics)
  - OpenML (Open science platform for Machine Learning)
  - Impulse 1 (WiBAF, with Adversitement BV)
  - Impulse 2 (Customer Journey, with Philips)
Upcoming DSC/e events

Lecture Series:

Oct 21    prof.dr. Bettina Speckmann
           Analysis and visualization of complex moving objects

Nov 4     dr.ir. Frans Willems
           The Context-Tree Weighting Method

Nov 13    prof.dr. Chris Snijders
           Experts vs data: intuition, expertise and the robust beauty of
           improper linear models

Nov 25    prof.dr. Johan van Leeuwaarden
           Data-driven critical service systems

www.tue.nl/dsce/lectures
Nov 12 Start MOOC by prof. Wil van der Aalst
Process Mining: Data Science in Action
online course on Coursera
www.coursera.org/course/procmin

Dec 4 The Dutch Data Science Summit 2014
www.tue.nl/dsce/summit

Follow us on Twitter @TUe_DSCE