Supporting the Authoring of Adaptive Hypermedia with Structural Information?

Stephan Weibelzahl Stefan Lippitsch Gerhard Weber

Pedagogical University Freiburg, Germany
[weibelza,lippitsc,webergeh]@ph-freiburg.de

Authoring adaptive hypermedia requires skills that go beyond pure text editing, because relations between concepts have to be specified among other knowledge engineering activities. A recent approach to support authors is to compute several structural measures of the domain model to estimate its adaptivity degree. However, our empirical results of eight different courses and more than 1300 users suggest that subjective adaptivity success and structural adaptivity degree are not related. We argue, that the adaptivity degree is an inherent property of the content and does not imply adaptation quality.

1 Supporting Authors of Adaptive Hypermedia

During the last decade, adaptive hypermedia and intelligent tutoring systems evolved from purely academic toys to productive instruments that are used for teaching and education in many domains. Thus, not only researchers but also authors, that have not been involved in the development of the systems, have to be able to implement courses or environments for their domain or subject.

Existing adaptive learning environments (De Bra and Calvi, 1998; Brusilovsky, Eklund and Schwarz, 1998; Carro, Pulido and Rodríguez, 2001; Sanrach and Grandbastien, 2000; Murray, Shen, Piemonte, Condit and Thibedeau, 2000; Weber, Kuhl and Weibelzahl, 2001) offer different degrees of authoring support. However, the more widely these systems are used, the more obvious is the need for a good authoring tool, because adaptive hypermedia require activities that are beyond text editing, including knowledge engineering.

Most of the current adaptive hypermedia systems require the specification of at least two kinds of relations between concepts or pages: *is prerequisite of* and *infers*. Prerequisite concepts usually have to be learned before the related concept, i.e., understanding a concept requires to know a prerequisite concept beforehand. Concepts are inferred by other concepts if knowing the second concept implies knowing the first. Note, that this terminology is adopted from NetCoach courses. Other authoring systems might use different terms for the same type of relations. Some systems might even offer other kind of relations. However, prerequisite and inference are the most commonly used relations.

Based on this structural information, it is possible to provide adaptive features such as adaptive curriculum sequencing, adaptive annotation, and adaptive link hiding.

The authors want to thank the anonymous reviewers for their useful comments on an earlier draft of this paper.

2 Using Domain Model Characteristics for Authoring Support

Having identified the need for authoring support, several tools with different ways of support have been discussed (Murray, 1999). Recently, a new approach has been introduced by Cini and de Lima (2002). The authors propose that it is possible to estimate the adaptivity degree of adaptive hypermedia by computing several structural measures. This could serve two purposes: on the one hand, it could provide authors with hints, whether they should increase the adaptivity degree of their course, and which adaptivity component requires additional concept relations. E.g., a low adaptivity degree in terms of few concepts that have prerequisites might indicate that the author should add more prerequisites. The higher the adaptivity degree, "the larger will be the amount of users that can use the presentation in a personalized way" (p. 498). On the other hand, the proposed measures are a kind of evaluation criterion, because a higher adaptivity degree should result in better adaptation in terms of navigation support and “satisfaction degree with the presentation” (p. 500).

Cini and de Lima (2002) proposed the following six measures. In order to provide a better overview, we cite the definitions, name each of them and add according textual formulas:

- The adaptivity degree of the user model in the generation of updates: percentile of pages that update other concepts in relation to the total pages of the presentation

$$A_{have\ inferences} = \frac{pages\ with\ inferences}{total\ pages}$$

- The adaptivity degree in restrictions of the adaptation model: percentile of pages that have restrictions for their presentation in relation to the total pages of the presentation

$$A_{have\ prerequisites} = \frac{pages\ with\ prerequisites}{total\ pages}$$

- The user adaptable behavior degree in the presentation: percentile of concepts which can be altered directly by the user and are used as requirements in restrictions for other concepts in relation to the total number of concepts which can be altered directly by the user

$$A_{adaptable} = \frac{adaptable\ concepts\ that\ are\ prerequisites}{adaptable\ concepts}$$

- The content adaptation degree in the pages: percentile of pages that have conditional fragments in relation to the total of pages

$$A_{have\ conditional\ fragments} = \frac{pages\ with\ conditional\ fragments}{total\ pages}$$

- The adaptive navigation degree in the pages: percentile of pages that have conditional links in relation to the total pages of the presentation

$$A_{have\ conditional\ links} = \frac{pages\ with\ conditional\ links}{total\ pages}$$

- The existence of an adaptive navigational map

The rational of the $A_{have\ ...}$ measures is, that only pages (respectively concepts) that *have* any relation to other concepts increase the adaptivity. All other pages are static. However, we argue that adaptivity degree could be interpreted the other way round as well: the more concepts that *are* prerequisite of another concept the more different adaptive suggestions may occur during interaction. Accordingly, the more concepts that are inferred by other concepts, the more pages might be skipped to reach a learning objective. Thus, we propose to consider the following $A_{are\ ...}$ measures as well:

- Percentile of concepts that are prerequisites of pages in relation to the total concepts. The more different concepts are prerequisite of at least one page the more different guiding suggestions may occur

$$A_{are\ prerequisites} = \frac{concepts\ that\ are\ prerequisite}{total\ concepts}$$

- Percentile of concepts that are inferred by pages in relation to the total concepts. The more different concepts are inferred by a page the more changes in the user model may occur

$$A_{are\ inferred} = \frac{concepts\ that\ are\ inferred\ by\ other\ concepts}{total\ concepts}$$

Both, $A_{are\ prerequisites}$ and $A_{are\ inferred}$ can be influenced by making implicit relations explicit without changing the structure. E.g., if A is prerequisite of B and B is prerequisite of C, then A is also prerequisite of C. Adding this last relation would increase $A_{are\ prerequisites}$, but the domain model would remain the same. Thus, for the computation of the above measures we also considered these indirect relations, because NetCoach uses them for the adaptation mechanism as well.

Moreover, we could also compare the number of relations that have been specified by the author. I.e., instead of counting the concepts that are or have prerequisites we could register how many prerequistes there are. The absolute number of relations should be standardized by the number of possible relations.

- Relative amount of prerequisites in relation to the maximum number of possible prerequisites

$$A_{prerequisite\ rate} = \frac{total\ prerequisites}{P_{max}}$$

- Relative amount of inferences in relation to the maximum number of possible inferences

$$A_{inference\ rate} = \frac{total\ inferences}{I_{max}}$$

  For NetCoach courses, the maximum of prerequisites P_{max} and the maximum of inferences I_{max} that can be specified depends on the number of concepts n only, while cyclic prerequisites are disallowed.

$$P_{max} = I_{max} = \frac{n \times (n-1)}{2}$$

The prerequisite measures obviously require that the user is free to navigate trough the course. Otherwise a course with the maximal amount of prerequisites would be completely rigid, and not adaptive at all.

3 Empirical Findings

The previous section lists many different measures, but which of them are useful? Should we urge authors of courses with low adaptivity degree to specify more concept relations to get better adaptivity?

We collected some empirical data from eight different courses in different domains to answer this question. Most of these courses are part of the PSI project (Lippitsch, Weibelzahl and Weber, 2002) which develops adaptive online courses based on the authoring system NetCoach (Weber, Kuhl and Weibelzahl, 2001) to introduce students to pedagogical psychology. The course subjects include interpersonal communication (*Kommunikation*), student assessment (*Leistungsbeurteilung*), empirical methods (*Methoden*), social perception (*Personenwahrnehmung*), cognitive developmental psychology (*Piaget*), problem solving (*Problemlösen*), and psycholgical fields (*Psychologie*).

Students had to complete these courses as part of their curriculum. In addition, all courses, including the HTML-Tutor, which introduces to publishing on the web, are available online for everybody. At the end of each course a questionnaire is presented and the students had to rate the course in terms of several dimensions, including navigation, orientation, adaptation in general, annotation, and page suggestions on a 10-point scale:

- Navigation: *Navigating through the course is ...* (difficult ... easy)
- Orientation: *During interaction I knew my current location (chapter, page) in the course.* (never ... always)
- Adaptation in general: *The course adapted to your learning progress. Do you think this was successful?* (not successful at all ... very successful)
- Annotation: *In the table of contents on the left hand side, chapters were annotated with different colors in accordance with your current knowledge level. The system intended to improve your orientation throughout the course by this.* (not successful at all ... very successful)
- Page suggestions: *The system tried to suggest pages to you that are adequate for your knowledge level. Has this been successful?* (not successful at all ... very successful)

In addition, the learners had to rate their impression of the interaction with the system in respect to four dimensions on a 10-point scale: terrible ... wonderful; difficult ... easy; monotonous ... stimulating; rigid ... flexible. We will call the mean value of these four scales *overall impression* of a course.

The upper part of Table 1 shows the mean values of these ratings for all courses. In addition, we computed six of the structural measures for each course separately. The other measures cannot be applied to NetCoach courses, because there are neither conditional fragments, nor conditional links. Moreover, all pages are adaptable in terms of their knowledge status directly by the user, and thus $A_{adaptable}$ is equal to $A_{are\ inferred}$. Both, conditional fragments and conditional links are specific for AHA! systems, which have been the main targets of Cini and de Lima (2002).

Note, that many of the structural measures that are concerned with inferences are 0. In six of the eight courses it was impossible to specify any inference. Thus, the following results for the inference measures are limited. However, other courses in different domains will have few inferences as well, because the condition of implying a complete concept is hard to fulfill.

Given these data, it is possible to correlate the structural course measures with the subjective ratings, in order to estimate the relation between these variables. The results are shown in Table 2.

Table 1: Means of subjective ratings and structural information of eight NetCoach courses. The sample sizes for the subjective ratings are shown in table 2

		Kommunikation	**Leistungsbeurteilung**	**Methoden**	**Personenwahrnehmung**	**Piaget**	**Problemlösen**	**Psychologie**	**HTML-Tutor**
subjective ratings	**navigation**	7.35	7.24	7.09	7.53	7.06	7.22	6.77	6.86
	orientation	7.1	7.06	7.37	7.78	7.37	7.82	6.98	7.12
	adaptation	6.26	5.64	6.14	6.3	5.72	6.3	6.02	5.95
	suggestions	6.51	5.94	6.32	6.53	5.94	6.55	6.39	5.91
	annotation	6.5	6.07	5.99	6.64	6.2	6.3	5.52	6.09
	overall impression	5.54	4.59	5.30	5.53	4.68	5.78	5.42	6.07
structure	$A_{are\ prerequisites}$	0.63	0.96	0.97	0.92	0.72	0.76	0.96	0.95
	$A_{have\ prerequisites}$	0.97	0.98	0.97	0.95	0.95	0.96	0.96	0.92
	$A_{prerequisite\ rate}$	0.79	1	1	0.95	0.79	0.83	1	0.88
	$A_{are\ inferred}$	0	0	0.15	0	0	0	0	0.37
	$A_{have\ inferences}$	0	0	0.13	0	0	0	0	0.1
	$A_{inference\ rate}$	0	0	0.02	0	0	0	0	0.04

Despite of the very big sample size, all bivariate correlations are quite low ($|r| \leq .1$), i.e., only $r^2 < 1\%$ or less of the variance in one variable can be explained by the other. Taking an effect size of $r = .1$ for granted (which is very low) the correlations have a test power of $1 - \beta > .95$, i.e., even very small effects would have been detected. Nevertheless, only four correlation are significant, all of them are very low. In summary, we found some statistically significant correlations, but the empirical effect size is probably not of importance for educational purposes.

4 Discussion

There are at least three possible interpretations of these results. First, the fact that we failed to find considerable relations between the learners' subjective ratings and the structure of the courses might indicate, that all of the proposed measures are useless for authors. The specified content structure does not provide hints for further improvement of course adaptivity. At least it seems not to be

Table 2: Correlations of structural measures with subjective ratings of course users. The bivariate correlation and the sample size are reported. Statistically significant results are indicated with * ($p < .05$) and ** ($p < .01$). For all correlations the power is $1 - \beta > .95$, given $\alpha = .05$ and an effect size $r = .1$. In addition we report the correlation of overall impression with subjective adaptivity ratings (last column) and with the strutural measures (last row)

	$\mathbf{A}_{have\ prerequisites}$	$\mathbf{A}_{are\ prerequisites}$	$\mathbf{A}_{prerequisite\ rate}$	$\mathbf{A}_{have\ inferences}$	$\mathbf{A}_{are\ inferred}$	$\mathbf{A}_{inference\ rate}$	**overall impression**
navigation	.032	-.056*	-.048	-.023	-.035	-.034	.386**
	1379	1379	1379	1379	1379	1379	1240
orientation	-.006	-.023	-.039	-.005	-.018	-.016	.333**
	1412	1412	1412	1412	1412	1412	1269
adaptation	.010	-.010	.002	.003	-.008	-.006	.471**
	1377	1377	1377	1377	1377	1377	1237
suggestions	.035	-.006	.019	-.027	-.047	-.045	.455**
	1345	1345	1345	1345	1345	1345	1205
annotations	.049	-.046	-.031	-.079**	-.100**	-.099**	.394**
	1261	1261	1261	1261	1261	1261	1127
overall impression	-.097**	.036	.024	.099**	.052	.094**	—
	1384	1384	1384	1384	1384	1384	

related to the subjective impression of the users. Nevertheless, adaptivity degree might be useful for authors to get a kind of summary of their presentation.

However, and this is the second interpretation, the subjective ratings might have been useless to indicate what the structural measures should detect. The learners' answers in the questionnaire might have be influenced by the overall impression of the system regardless of the factual adaptivity success. As shown in Table 2, the overall impression correlates highly with the subjective adaptivity success measures, but not with the structural measures. However, a partial correlation with control for overall impression improves the relation between subjective ratings and course structure only slightly. The highest bivariate correlation ($A_{have\ prerequisite}$ ~ annotation) is raised to $.095$. All partial correlations with other structural measures are negative. This is, in fact, an implicit problem of the evaluation of adaptivity. The perfect adaptation is not even noticed by the user and can thus not be reported.

Third, we have to consider the fact, that adaptation is never independent of the content. Opposed to the idea, that more concept relations and a higher adaptivity degree result in a better course, each content might have its own ideal structure. Adaptivity degree might be an inherent property of a

content that cannot been influenced. While some contents have many internal dependencies, others might have only very few. Increasing the adaptivity degree by specifying additional relations will not improve the adaptivity any more, or might even yield mal-adaptations.

Thus, the proposed structural measures might be interesting to compare the degree of possible adaptivity across contents, but our data does not support the claim, that they are useful for authoring support and evaluation.

A better way of supporting authors in specifying relations might be to visualize the domain (e.g., as a network or a matrix) or to check the relations for consistency automatically (Wu, Houben and De Bra, 1999), in order to avoid circles and other failures that would disturb the adaptation process.

References

Brusilovsky, P., Eklund, J., and Schwarz, E. (1998). Web-based education for all: A tool for developing adaptive courseware. In *Computer Networks and ISDN Systems. Proceedings of Seventh International World Wide Web Conference, 14-18 April 1998*, volume 30, pages 291–300.

Carro, R. M., Pulido, E., and Rodríguez, P. (2001). TANGOW: a model for internet based learning. *International Journal of Continuing Engineering Education and Life-Long Learning*, *11*(1-2).

Cini, A. and de Lima, J. V. (2002). Adaptivity conditions evaluation for the user of hypermedia presentations built with AHA! In *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, Málaga, Spain, AH2002*, pages 497–500, Berlin. Springer.

De Bra, P. and Calvi, L. (1998). AHA! An open adaptive hypermedia architecture. *The New Review of Hypermedia and Multimedia*, *4*, 115–139.

Lippitsch, S., Weibelzahl, S., and Weber, G. (2002). Adaptive learning courses in pedagogical psychology. The PSI project and the authoring system NetCoach. *Künstliche Intelligenz*.

Murray, T. (1999). Authoring Intelligent Tutoring Systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10*, 98–129.

Murray, T., Shen, T., Piemonte, J., Condit, C., and Thibedeau, J. (2000). Adaptivity in the MetaLinks hyper-book authoring framework. In *Workshop Proceedings of Adaptive and Intelligent Web-Based Education Systems workshop at ITS 2000*.

Sanrach, C. and Grandbastien, M. (2000). ECSAIWeb: A web-based authoring system to create adaptive learning systems. In Brusilovsky, P., Stock, O., and Strapparava, C. (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems. International Conference, AH 2000*, pages 214–226. Berlin: Springer.

Weber, G., Kuhl, H.-C., and Weibelzahl, S. (2001). Developing adaptive internet based courses with the authoring system NetCoach. In Reich, S., Tzagarakis, M., and de Bra, P. (Eds.), *Hypermedia: Openness, Structural Awareness, and Adaptivity*, pages 226–238, Berlin. Springer.

Wu, H., Houben, G.-J., and De Bra, P. (1999). Authoring support for adaptive hypermedia applications. In *Proceedings of the Ed-Media 99 Conference*, pages 364–369.