

Designing Adaptive Hypermedia for Internet Portals: A Personalization Strategy Featuring Case Base Reasoning With Compositional Adaptation

Syed Sibte Raza Abidi

Faculty of Computer Science, Dalhousie University, Halifax B3H 1W5, Canada

Email: sraza@cs.dal.ca

Abstract. We propose that the *Case Based Reasoning (CBR)* paradigm offers an interesting alternative to developing adaptive hypermedia systems, such that the inherent analogy-based reasoning strategy can inductively yield a ‘representative’ user model and the case adaptation techniques can be used for dynamic adaptive personalization of generic hypermedia-based information content. User modeling is achieved by applying *CBR retrieval techniques* to collect a set of similar past cases, which are used to form a global user-model. Adaptive personalization is accomplished by a *compositional adaptation technique* that dynamically authors a personalized hypermedia document—a composite of multiple fine-grained *information ‘snippets’*—by selectively collecting the most relevant information items from matched past cases and systematically amalgamating them to realize a component-based personalized hypermedia document.

1. Introduction

Web-mediated information portals routinely suffer from their inability to satisfy the heterogeneous needs of a broad base of information seekers. For instance, web-based education systems present the same static learning content to learners regardless of their individual knowledge of the subject; health information portals deliver the same generic medical information to consumers with different health profiles; and web e-stores offer the same selection of items to customers with different preferences.

A solution to this overly-simplified approach for ‘generic’ information delivery is the development of *adaptive hypermedia systems*—web-based systems that belong to the class of user-adaptive software systems—that have the ability to adapt their behavior to the goals, tasks, interests and needs of individual users and group of users [1]. An adaptive hypermedia system involves two distinct activities: (a) development of a user model and (b) adaptation of static generic information content to user-specific personalized content [2].

In this paper, we argue that the *Case Based Reasoning (CBR)* paradigm [3] offers an interesting alternative to developing adaptive hypermedia systems [4], such that the inherent analogy-based reasoning strategy can inductively yield a ‘representative’ user model and the case adaptation techniques can be used for dynamic adaptive personalization of generic hypermedia-based information content [5]. In our work, user modeling is achieved by applying *CBR retrieval techniques* to collect a set of similar past cases, which are used to form a global user-model. Adaptive personalization is accomplished via a novel *compositional adaptation technique* that dynamically authors a personalized hypermedia document—a composite of multiple fine-grained *information ‘snippets’*—by selectively collecting the most relevant information items from matched past cases (i.e. not the entire past solution) and systematically amalgamating them to realize a component-based personalized hypermedia document. We present an adaptive hypermedia system designed to dynamically author personalized healthcare information content based on an individual’s current health profile, as personalized health maintenance information is deemed to have a significant impact in ensuring wellness maintenance both at the individual and community level [5, 6, 7].

2. CBR-Mediated Adaptive Personalization

Our CBR-mediated adaptive hypermedia system development approach builds on a corpus of past *cases* specified by medical practitioners. Each case depicts a situation-action construct, such that (a) the situation component defines the local user-model—i.e. an individual’s *Health Profile (HP)*—in terms of attribute-value pairs (ideally originating from the individual’s electronic medical record); and (b) the action component comprises a corresponding *Personalized Healthcare Information Prescription (PHIP)* that is composed of a number of fine-grain, *Problem-focused (hypermedia) Documents (PD)*. Each PD is designed to contain health maintenance information pertaining to a specific medical problem/issue. Note that the PHIP is a composite of multiple PDs, whereby each constituent PD is prescribed by a medical practitioner in response to some facet (i.e. an attribute-value) of an individual’s HP.

2.1. Problem Specification

We argue that one limitation of traditional CBR approaches is that the recommended solution to a new problem-situation—i.e. a new case—is taken as the entire solution of the matched past case. In a healthcare information delivery context where information accuracy is paramount it would be rather naive to assume that heterogeneous individuals may have a similar HP or user model! Hence, it is argued that the entire PHIP associated with matched past cases (i.e.

existing user-profiles) cannot be regarded as an accurate inferred solution to a new user-model. In this scenario, adaptive personalization is characterized as the problem of selective collection of only the relevant information ‘snippets’ from the multiple matched past PHIPs, as opposed to selecting the entire PHIP. We believe that a *component-based* information representation and compilation strategy will ensure that the healthcare content disseminated to an individual is specifically focused towards the individual’s prevailing healthcare needs, akin to the kind of personalized service one enjoys from a visit to a medical practitioner [5].

2.2. Our Compositional Adaptation Strategy

We have devised a case adaptation strategy—based on notions of compositional adaptation [8, 9]—that is applicable to the adaptation of a specialized class of cases in which the case solution is a composite of individual *sub-solutions*; where each sub-solution addresses a particular problem-defining attribute of a case. Our compositional adaptation strategy is applicable to dynamic adaptive personalization of hypermedia documents, as it allows the tailoring of personalized documents via user-profile driven selection of ‘generic’ information snippets (analogous to sub-solutions) from an ensemble of past-compiled hypermedia documents. The systematic amalgamation of ‘relevant’ information snippets yields a unified personalized document corresponding to a particular user-model. Figure 1 shows our CBR-mediated compositional adaptation strategy for adaptive hypermedia personalization.

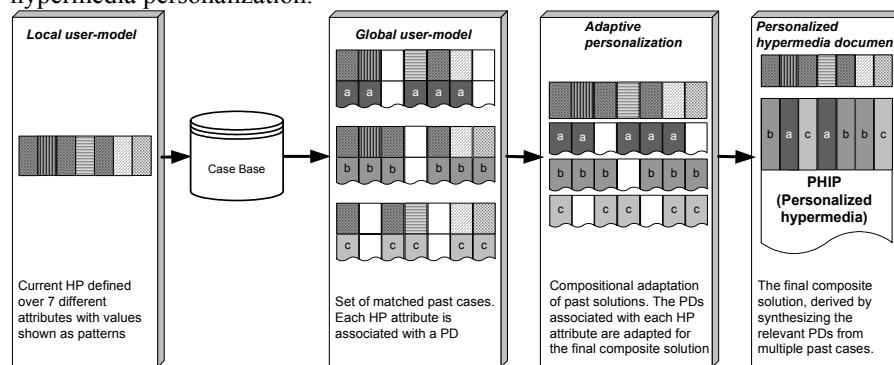


Fig. 1. A pictorial illustration of our CBR-Mediated compositional adaptation based strategy for generating adaptive personalized hypermedia documents.

The rationale for our approach is grounded in the principle that since inter-case similarity is determined at an attribute-level, therefore fine-grained solution adaptation should also be conducted at the attribute-level. By adapting the attribute-specific sub-solutions based on the attribute’s similarity measure we

ensure that the best matching attribute values impact the most on a selected segment of the solution—i.e. the sub-solution component associated with the attribute—as opposed to impacting the entire solution component [9]. In this way we are able to generate a solution that contains components that reflect the ‘best’ features—i.e. most relevant information—of similar past solutions.

3. An Algorithm for CBR-Mediated Adaptive Personalization

We will now discuss our compositional adaptation algorithm for performing adaptive personalization of hypermedia documents.

3.1 Case Representation Scheme

The HP depicts a ‘local’ user-model defined in terms of a list of health specific attributes as shown in Table 1. The HP, deemed as the problem description in a CBR content, contains multi-valued attributes, where the domain of attribute-values is determined from standard medical resources. In a CBR-context, the PHIP is deemed as the solution component of a case. Structurally, the PHIP is a composite of multiple PDs. Conceptually, each HP attribute is related to at least one PD in the solution component.

Table 1. An exemplar HP illustrating the 7 information groups and their corresponding values.

Acute Disease (AD)	Short-Term Illness (SI)	Current Symptoms (S)	Current Drugs (D)	Allerg-ies (A)	Demograp-hic Data (DD)	Lifestyle Data (LD)
Diabetes-Mellitus Hypertension	Fever	High Temp. Cough Rashes	Panadol Bendryl	Allergic Rhinitis	Age : 56 y Sex : Male Edu.: High	Fitness: N Diet : H Smoke: Y

3.2 User-Modeling: Case Retrieval Procedure

In CBR terms, user modeling involves the generation of a *global user-model* derived based on the similarity between the local user-model (i.e. the HP) and a set of past user-models. Given a local user-model, we retrieve a set of similar past user-models based on similarity measures—referred as **Total Weighted Distance (TWD)**. Table 2 shows a current HP that has 3 values for the AD attribute. Each HP attribute-value code is derived as a combination of the *class-code*, *sub-class-code* and the *element-code*.

Table 2. Current HP and 4 matching past cases. Note that only the AD attribute is shown.

Current Case	HP			
		AD ₁ = 1-1-002 ₁ AD ₂ = 1-3-035 ₂ AD ₃ = 2-1-004 ₃		
Past Cases (PC)	PC ₁	PC ₂	PC ₃	PC _{total=4}
	ad ₁ = 1-1-002 ₁ ad ₂ = 1-3-035 ₂ ad ₃ = 2-1-004 ₃	ad ₁ = 1-2-021 ₁ ad ₂ = 2-1-003 ₂ ad ₃ = 1-1-002 ₃	ad ₁ = 1-1-020 ₁ ad ₂ = 1-3-035 ₂ ad ₃ = 2-1-004 ₃	ad ₁ = 3-1-002 ₁ ad ₂ = 3-1-004 ₂

Table 3. Similarity Matrix used to determine DS between the current and past HP attributes.

Degree of Similarity (DS)	Class Code	Sub-Class Code	Element Code	Numeric Value for DS
Perfect Match	√	√	√	1
Close Match	√	√	×	75
Weak Match	√	×	×	25
No Match	×	×	×	100

A domain-specific *similarity matrix* (as shown in Table 3) is used to determine the attribute-level *Degree of Similarity* (DS)—the DS spans from *perfect match* to *close match* to *weak match* and *no match*—between the current and past HP attribute-values belonging to the same attribute. For instance, the attribute values 1-2-2001 and 1-2-2002 will result in a DS of ‘close match’ as the class and sub-class codes match, whereas the DS between the attribute values 1-2-2001 and 1-3-3004 is a ‘weak match’ because only the class code is similar. We trace below the steps involved in the calculation of TWD between a current HP and a set of past HPs, leading to the retrieval of similar past cases.

Step 1: Determine attribute-level *Distance*

The idea is to establish equivalence between the current HP and a past case’s HP at the attribute level. We calculate the DS between each current HP attribute-value with respect to corresponding attribute-value(s) in each past case’s HP. Since each HP attribute can have multiple values, we need to individually determine the DS for each current HP attribute-value. The pseudo code shows the calculation of DS.

```

For P = 1 to PCtotal {total is the no. of past cases}
  For J = 1 to ADn {N = no. of AD values in current HP}
    For K = 1 to adm {m = no. of ad values in a past HP}
      compare each ADj with all adk in PCp using the
      similarity matrix given in table 4 such that
        DS[ADj, adkp] = similarity_matrix(ADj, adkp)
    
```

Step 2: Find the best matching attribute-value in the past HP

For each current HP attribute-value, we find the best matching attribute-value(s) in the past cases based on the value of $DS(AD_x, ad_y)$ by determining the *Distance* (D).

For $P = 1$ to PC_{total}
For $J = 1$ to AD_N
For $K = 1$ to ad_m

$$D_{AD_J}^{ad_K^P} = \min(DS[AD_J, ad_K^P])$$

where $D_{AD_J}^{ad_K^P}$ implies that AD_J best matches with the attribute-value ad_K in the past case P , and the variable D holds the distance measure between AD_J and ad_K which would be the minimum for all ad values in the past case P . Note that we individually calculate $D_{AD_J}^{ad_K^P}$ for all the past cases. Using the current HP and the set of past cases given in Table 2, we present the calculation of DS in Table 4.

Step 3: Calculate the Total Distance for each current HP attribute

For each current HP attribute, we calculate its distance with the corresponding attribute in a specific past case. Since each attribute can have multiple values, the TD is derived via averaging the individual matching D 's associated with the multiple attribute-values. We calculate a separate TD (shown in Table 5) for each current HP attribute for all past cases as follows:

For $P = 1$ to PC_{total}

$$TD_{AD}^P = \sum_{K=1}^N D_{AD_K}^{ad^P} / N$$

where TD_{AD}^P refers to the total distance of the current HP attribute of AD with the same attribute in the past case P , and N is the number of non-zero D_{AD} . Note that the same procedure is applied to calculate the TD for the other four attributes in the current HP, given as TD_{SI} , TD_S , TD_D and TD_A .

Step 4: Calculate the Total Weighted Distance for each past case

We use the individual TD values for all the current HP attributes with respect to a specific past case to calculate the TWD between the entire current HP and the HP component of a specific past case. The case-level distance is weighted—i.e. the user can modulate the influence of each attribute in the determining the similarity between the current and past HPs.

Step 5: Retrieve similar past cases to form global user model

We retrieve all past cases that have a TWD less than a pre-defined threshold.

Table 4. Calculation of DS and TD for the current HP and the set of past cases. The legend ($AD_1 \rightarrow ad_1$) implies that the attribute value AD_1 matches with value ad_1 .

P	J	K	DS[AD_J, ad_K^P]	TD_{AD}
1	1	1	1 ($AD_1 \rightarrow ad_1$)	1.00
		2	75	
		3	100	
	2	1	75	
		2	1 ($AD_2 \rightarrow ad_2$)	
		3	100	
	3	1	100	
		2	100	
		3	1 ($AD_3 \rightarrow ad_3$)	
2	1	1	75	33.67
		2	100	
		3	1 ($AD_1 \rightarrow ad_3$)	
	2	1	75 ($AD_2 \rightarrow ad_1$)	
		2	100	
		3	100	
	3	1	100	
		2	25 ($AD_3 \rightarrow ad_2$)	
		3	100	

P	J	K	DS[AD_J, ad_K^P]	TD_{AD}
3	1	1	25 ($AD_1 \rightarrow ad_1$)	9.00
		2	75	
		3	100	
	2	1	75	
		2	1 ($AD_2 \rightarrow ad_2$)	
		3	100	
	3	1	100	
		2	100	
		3	1 ($AD_3 \rightarrow ad_3$)	
4	1	1	100	100.00
		2	100	
	2	1	100	
		2	100	
	3	1	100	
		2	100	

Table 5. Calculation of the TWD of the current HP with the HP component of the past cases. The TDs for attribute other than AD are set to 50 for illustration purposes only.

Past Case	TD_{AD}	TD_{SI}	TD_S	TD_D	TD_A	TWD	Case Retrieved (TWD < 55)
PC ₁	1.00	50	50	50	50	40.20	✓
PC ₂	33.67	50	50	50	50	46.73	✓
PC ₃	9.00	50	50	50	50	41.80	✓
PC ₄	100.00	50	50	50	50	60.00	✗

3.3 Adaptive Personalization Via Compositional Adaptation

In the adaptive personalization stage, we personalize the solution component of the retrieved past cases to generate an individual-specific solution—i.e. a PHIP. As per our compositional adaptation approach, for each HP's attribute-value we select the most relevant past sub-solution (which manifests as a specific PD) from the entire solution of the retrieved past cases. The processing sequence is as follows: (i) Each attribute-value of the current HP is mapped to a set of matching attribute-values in the retrieved past cases; (ii) the PD associated with the matching past case's attribute value is selected; and (iii) the set of selected PDs are systematically amalgamated to yield the most representative PHIP. We explain below our compositional adaptation technique, building upon the case retrieval mechanism described earlier.

Step A: Calculate the Relative Distance of each matched current HP value

We determine the *Relative Distance* (RD) of each current HP attribute-value with respect the attribute-level distance (calculated earlier as D) and case-level distance (calculated earlier as TWD) for each retrieved past case as follows:

For P = 1 to PC_{retrieved}

For K = 1 to N {N = total no. of matched AD value}

$$RD_{AD_K}^{ad_x^P} = (D_{AD_K}^{ad_x^P} * W_{Field} + TWD^P * W_{TWD}) / (W_{Field} + W_{TWD})$$

where $RD_{AD_K}^{ad_x^P}$ is the relative distance between the current HP attribute-value AD_K and the corresponding attribute-value ad_x in the retrieved past case P (shown in Table 7). Here, we introduce two user-specified weights $W_{Attribute}$ and W_{TWD} to impact the influence of attribute-level and case-level similarity, respectively.

Table 7. Calculation of RD of each AD attribute-value with the corresponding attribute-values in the three retrieved cases.

K	P	TWD ^P	ad _x ^P	D _{AD_K} ^{ad_x^P}	RD _{AD_K} ^{ad_x^P}	AD _K	Temp	NRD ^P _{AD_K}
1	1	40.20	1-1-002	1	28.44	1	0.092	0.38
	2		1-1-002	1	33.01			0.33
	3		1-1-020	25	36.76			0.29
2	1	46.73	1-3-035	1	28.44	2	0.086	0.40
	2		1-2-021	75	55.21			0.21
	3		1-3-035	1	29.56			0.39
3	1	41.80	2-1-004	1	28.44	3	0.092	0.38
	2		2-1-003	25	40.21			0.26
	3		2-1-004	1	29.56			0.36

Step B: Calculate the Normalized Relative Distance of current HP values

To acquire a uniform range of RD's over the entire set of current-HP attribute values we calculate the *Normalized Relative Distance* (NRD) of a specific current HP attribute-value over the entire set of retrieved past cases (i.e. PC_{retrieved}):

For K = 1 to AD_N

$$Temp_{AD_K} = \sum_{P=1}^{PC_{retrieved}} 1 / RD_{AD_K}^{ad_x^P}$$

Next, the NRD for the attribute-value AD for a retrieved past case P is calculated as follows: $NRD_{AD_K}^{ad_x^P} = 1 / (Temp_{AD_K} * RD_{AD_K}^{ad_x^P})$, where $NRD_{AD_K}^{ad_x^P}$ is the normalized relative distance between the current HP attribute-value AD_K and the attribute-value ad_x in the past case P (as shown in Table 7).

Step C: Determine the appropriateness of available solution components

Since each current HP attribute-value can match with one or more past case’s attribute-value, a current HP attribute-value can be linked with multiple PDs. Hence, we select the most appropriate PDs for each current HP attribute-value by determining the *Appropriateness Factor* (AF) of all the available PDs via the aggregation of their NRD over the entire set of retrieved cases:

For $I = 1$ to AD_N

$$AF_{AD_I}^{ad_x^P} = \sum_{P=1}^{PC_{retrieved}} NRD_{AD_I}^{ad_x^P}$$

where $AF_{AD_I}^{ad_x^P}$ is the appropriateness factor for the PD associated with the attribute-value ad_x in the past case P with respect to the current HP attribute-value of AD_I . Next, we compare the AF for each PD against a pre-defined threshold; IF the AF of a PD exceeds the threshold then it is included in the final solution.

The Output: A Personalized Document Comprising Multiple Sub-Documents

Table 8 shows the AF for the 9 candidate PDs (note that there are only 6 distinct PDs). For attribute AD_1 , we have two distinct candidate PDs: PD 1-1-002 from two past cases—i.e. PC_1 and PC_2 ; and PD 1-1-020 from PC_3 . Since, PD 1-1-002 is recommended by two past cases it has a stronger bias for being included in the final solution, as is reflected by its AF value.

The final solution clearly illustrates an adaptive personalization affect whereby the composed PHIP comprises three PDs, one each for AD_1 , AD_2 and AD_3 . The solution for AD_1 is collected from past cases 1 and 2, whereas the solution for AD_2 is collected from past cases 1 and 3. This is in accordance with our compositional adaptation approach that posits the collection of the most appropriate sub-solutions from all the retrieved past cases as opposed to the selection of the entire solution of the most similar past case.

Table 8. Selection of the most appropriate PDs based on their AF values. The selection criteria is $AF_{PD} > 0.35$. The selected PDs represent the final solution component (i.e. PHIP).

AD_N	$ad^P \rightarrow PD^P$	NRD	AF_{PD}	PD selection for the FINAL SOLUTION
AD_1 (1-1-002)	$ad^1 = 1-1-002$	0.38	0.71	√ (selected) (1-1-002)
	$ad^2 = 1-1-002$	0.33		
	$ad^3 = 1-1-020$	0.29	0.28	× (not selected)
AD_2 (1-3-035)	$ad^1 = 1-3-035$	0.40	0.79	√ (1-3-035)
	$ad^3 = 1-3-035$	0.39		
	$ad^2 = 1-2-021$	0.21	0.21	×
AD_3 (2-1-004)	$ad^1 = 2-1-004$	0.38	0.74	√ (2-1-004)
	$ad^3 = 2-1-004$	0.36		
	$ad^2 = 2-1-003$	0.26	0.26	×

5. Concluding Remarks

In this paper, we have presented an interesting compositional adaptation technique that is applied to problem of adaptive hypermedia design. We conclude that our compositional adaptation approach is well-suited for personalized hypermedia document generation, if the hypermedia document is a composite of multiple fine-grained *information 'snippets'*. In this scenario, we design a personalized hypermedia document by selecting the most appropriate sub-solutions (or information snippets) from all the retrieved past cases. From our experiments, we have determined that (a) the higher the frequency of occurrence of a particular sub-solution across the various retrieved past cases, the higher its appropriateness towards the current solution; and (b) the appropriateness of a particular sub-solution is more accurately determined by taking into account both its individual appropriateness factor and the similarity measure of the entire past case with the current problem description.

References

1. Brusilovsky, P., Kobsa, A. and Vassileva, J. (Eds): Adaptive Hypertext and Hypertext, Kluwer Academic Publishers, Dordrecht, 1998.
2. Fink, J., Koenemann, J., Noller, S., and Schwab, I.: Putting Personalization into Practice, Communications of the ACM 45:5, 2002.
3. Aamodt A., Plaza E.: Relating Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches, AI Communications, 7:1, 1994
4. Bradley, K., Rafter, R. and Smyth, B.: Case-Based User Profiling for Content Personalization. In Brusilovsky, P. et al (Eds): Adaptive Hypertext and Adaptive Web-based Systems. Lecture Notes in Computer Science, 1892, Springer Verlag, 2000.
5. Wilke W, Bergmann R.: Techniques and Knowledge Used for Adaptation During Case Based Problem Solving. Lecture Notes in Artificial Intelligence, Vol. 1416. Springer-Verlag, Berlin Heidelberg New York, 1998, pp. 497-505
6. Bental, D., Cawsey, A., Pearson, J., and Jones, R.: Adapting Web-based Information to the Needs of Patients with Cancer. In Proc. Intl. Conf. On Adaptive Hypertext and Adaptive Web-based Systems, Trento, Italy, 2000
7. Abidi S.S.R., Han, C.Y. and Abidi, S.R.: Patient Empowerment via 'Pushed' Delivery of Personalised Healthcare Educational Content Over the Internet. In 10th World Congress on Medical Informatics (MedInfo'2001), London, 2001.
8. Abidi S.S.R., Goh A.: A Personalized Healthcare Information Delivery System: Pushing Customized Healthcare Information Over the WWW. In: Hasman A., Blobel B., Dudeck J., Engelbrecht R., Gell G., Prokosch H. (eds.): Medical Infobahn for Europe. IOS Press, Amsterdam, 2000, pp. 663 – 667
9. Arshadi N., Badie K.: A Compositional Approach to Solution Adaptation in Case-based Reasoning and its Application to Tutoring Library, Proceedings of 8th German Workshop on Case-Based Reasoning. Lammerbuckel, 2000.