

Hedaquin: A reputation-based health data quality indicator

Ton van Deursen^{a,b}, Paul Koster^a, Milan Petković^a

^a *Philips Research Eindhoven, The Netherlands*

^b *Université du Luxembourg, Luxembourg*

Abstract

A number of applications based on personal health records (PHRs) are emerging in the field of health care and wellness. PHRs empower patients by giving them control over their health data. Health data for PHRs can be supplied by patients, wellness providers and health care providers. Health care providers may use the PHRs to provide medical care. Unfortunately, the quality of the health data cannot be guaranteed in all cases (e.g. consider cases where non-professionals such as patients and wellness providers supplied the data).

To address this problem, we present in this paper Hedaquin, a system that provides health care professionals with an indication of the quality of health data in a PHR. This indication is based on the reputation of the supplier and on metadata provided by measurement devices. The proposed reputation system mimics the way trust in health data and their suppliers is built in the real world.

Hedaquin uses the Beta reputation system as a starting point and extends it in several directions to cover specific PHR requirements. Firstly, Hedaquin supports the automatic calculation of a rating based on a repeated measurement. Secondly, certificates for the user such as diplomas are taken into account. Thirdly, Hedaquin calculates reputation for different scopes in order to discriminate among different tasks the suppliers of health data can perform. Finally, the time difference between the ratings and the calculation of the reputation influences the weight that is given to a rating.

Keywords: Trust management, reputation system, personal health record, quality indication

1 Introduction

For years, patient information has been maintained in paper-based records. Paper-based systems bring many disadvantages such as lack of availability and loss or misunderstanding of important health data [7]. Because of these disadvantages, paper-based systems are being replaced by electronic health record (EHR) systems. These EHRs are maintained by health care providers. Next to these EHR systems, personal health record (PHR) systems are being developed. The PHR is a health record maintained by the patient instead of the health care provider. The patient can manage and share his health data in his PHR at his own discretion. After sharing, the health data in the PHR can be used by health care providers and wellness providers to improve the patient's health.

Health care providers are the only parties providing health data for EHRs. However, health care providers are not the only parties that can provide health data for

*This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: www.elsevier.nl/locate/entcs*

the patient’s well-being. Patients (but also people that are not ill, but are concerned about their health) may want to collect health data for their health records. Think for example of weight, heart rate and blood pressure information. Furthermore, as wellness providers such as fitness clubs and weight control clubs are professionalizing, they may want to use and provide relevant health data for the patient’s health record. This data is valuable information that can help a health care provider when treating the patient. The health data supplied by the patient, wellness providers and health care providers is stored in the patient’s personal health record (PHR).

In contrast to health data in EHRs, the health data in a PHR can be of varying quality. This is the result of the varying medical knowledge of the suppliers of the health data. The goal of this paper is to allow health care providers to use also the health data created by patients and wellness providers. This results in reduced costs and improved quality of care.

If quality of health data would be expressed with a value, this value would be different for different users. The reason for this is that quality is a subjective measure. In order to have an indication of the quality of health data in a PHR, we design a reputation system, called Hedaquin, that mimics the way trust in health data and their suppliers is built in the real world. To determine the quality of health data, the reputation of that user is used as a quality indication of the health data. In our approach, to determine the reputation we consider the following three factors: credentials of the health data supplier, ratings for the health data supplier and metadata supplied by measuring devices. Ratings for the health data supplier can be subdivided into three categories: ratings supplied by the user calculating the reputation, ratings supplied by other users and automatically calculated ratings based on the comparison of health data.

2 Related work

The eBay feedback forum¹ is one of the earliest reputation systems and is often said to be eBay’s main success factor. Kamvar *et al.* designed EigenTrust [5], a reputation system that calculates a global trust value for a user in a peer-to-peer network. Xiong and Liu designed PeerTrust [11], a reputation system that calculates reputation by taking a weighted average of the ratings. P2PRep [1] uses a fuzzy technique to decide whether to interact with another peer. Caballero *et al.* [2] present a reputation system that performs decision-making based on interaction patterns. Liu and Issarny [6] use a fuzzy technique incorporating time dimension and context dimension. REGRET [8] is a reputation system that models trust based on reputation and calculates the reliability of this reputation.

2.1 Beta reputation system

We use the Beta reputation system [3] as a starting point for Hedaquin. Next to the reputation, this system also calculates a measure of uncertainty of this reputation. This uncertainty is also indicated to the health care provider in order to make a more informed decision.

¹ <http://www.ebay.com/>

The Beta reputation system is based on the beta probability density function which can be used to represent probability distributions of binary events. The beta-family of probability density functions is indexed by two parameters α and β . The beta distribution $f(x|\alpha, \beta)$ can be expressed using the gamma function Γ :

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \text{ with } 0 \leq x \leq 1, \alpha > 0, \beta > 0$$

The beta distribution can be used as a reputation system by setting α and β to:

$$\alpha = R + 1 \quad \beta = S + 1$$

where R and S are the number of positive and negative experiences. In [4], Jøsang presents a logic for uncertain probabilities. Jøsang suggests using a metric called an opinion to represent the belief of one user in another. An opinion is a tuple (b, d, u) with $b, d, u \in \{x | x \in \mathbb{R} \wedge 0 \leq x \leq 1\}$ with the constraint that b, d and u add up to 1 ($b + d + u = 1$). An opinion represents the belief (b), disbelief (d) and uncertainty (u) of a user in another user. The opinion (b, d, u) can be computed using the reputation (R, S) as follows:

$$b = \frac{R}{R+S+2} \quad d = \frac{S}{R+S+2} \quad u = \frac{2}{R+S+2}$$

When the sum $R + S$ increases (i.e. the number of transactions increases), the uncertainty u decreases. Using an opinion is much more natural than using a pair (R, S) , because opinions are normalized. When reasoning about the reputation of a user it is good practice to first transform a reputation (R, S) to an opinion (b, d, u) .

3 Architecture

Figure 1 depicts the architecture of Hedaquin. The reputation engine takes ratings (local, global, aggregation and rule ratings) as input and calculates a reputation based on these ratings. This reputation is then shown to the health care provider and is used as a quality indication for health data provided by the user. The aggregation engine calculates ratings based on the comparison of measurements and the rule engine creates ratings based on certificates of the supplier of health data.

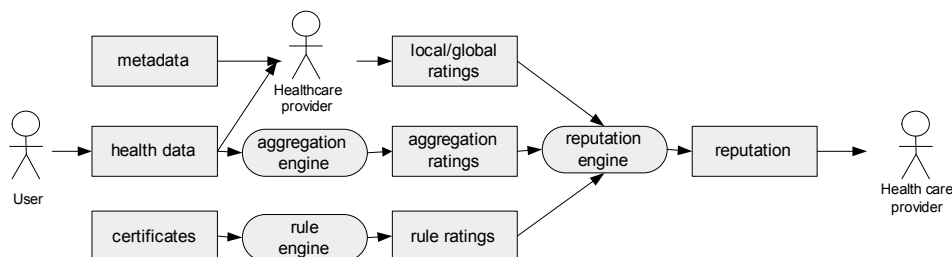


Fig. 1. Hedaquin architecture

A rating is a tuple $(r_{x,y,sc,tt,t}, s_{x,y,sc,tt,t}, c_{x,y,sc,tt,t})$, where r, s and c are real numbers between 0 and 1 and $r + s = 1$. In this tuple, r represents the positive fraction,

s represents the negative fraction and c represents the certainty of the rating. The subscript x defines the user who provided the rating, y is the user who is rated, sc is the scope of the health data for which the rating is given, tt is the trust type for which the rating is given (either functional or recommendation) and t is the time of creation of the health data. A scope is a tuple representing a measurement kind (e.g. blood pressure) and a device (e.g. sphygmomanometer).

A reputation is a tuple $(R_{x,y,sc,tt,t}, S_{x,y,sc,tt,t})$, where $R, S \in \mathbb{R}^+$ and the subscripts are analog to the subscripts of ratings. The reputation of a user at the time he created the health data is used as a quality indication for that health data.

3.1 Reputation engine

The reputation engine calculates the reputation of a supplier of health data. The reputation is a subjective measure and therefore depends on the health care provider calculating the reputation. The reputation of a user is calculated using four different kinds of ratings:

- Local ratings: A local rating is a rating provided by a health care provider x after he checked the quality of the health data that y supplied.
- Global ratings: To gain a broader view about a user z , a user x can ask other users y about their ratings for z . This mechanism is called transitivity of ratings. If y has ratings on z and x trusts y for providing recommendations, then x can use these ratings.
- Rule ratings: A rule rating is based on the user's possession of certificates issued by independent organizations. Rule ratings are calculated by the rule engine.
- Aggregation ratings: An aggregation rating is based on the comparison of two measurements carried out on the same person by different users. Aggregation ratings are calculated by the aggregation engine.

In essence, a global rating is a local rating that is given less weight. Ratings given by other users should be discounted (i.e. given less weight) because a user's own ratings are always more reliable than another user's ratings. The ratings of other users are discounted by the recommendation reputation of the user y (the supplier of the rating). The recommendation reputation of a user y is a pair $(R_{x,y,sc,R,t}, S_{x,y,sc,R,t})$. Clearly, ratings from users with high recommendation reputation should be given more weight than users with low recommendation reputation. Therefore a rating of a user y can be easily discounted by his recommendation reputation by discounting the certainty. The discounted rating can then be calculated as follows:

$$\begin{aligned} r_{x,z,sc,F,t} &= r_{y,z,sc,F,t} \\ s_{x,z,sc,F,t} &= s_{y,z,sc,F,t} \\ c_{x,z,sc,F,t} &= \frac{R_{x,y,sc,R,t}}{R_{x,y,sc,R,t} + S_{x,y,sc,R,t} + 2} \cdot c_{y,z,sc,F,t} \end{aligned}$$

Every user in the system has two different kinds of reputations. Functional reputation is the reputation for providing measurements and recommendation reputation is the reputation for providing ratings. Although they are different, they can be calculated in the same way (with the exception that aggregation ratings

cannot be calculated for recommendation reputation). In this paper we will focus on the calculation of the functional reputation.

Functional reputation

The functional reputation is a function over the ratings for health data (local, global and aggregation ratings) and rule ratings. After calculating the reputation part based on ratings for health data ($RH_{x,y,sc,F,t}$, $SH_{x,y,sc,F,t}$) and the reputation part based on rule ratings ($RR_{x,y,sc,F,t}$, $SR_{x,y,sc,F,t}$), the functional reputation can be calculated as follows:

$$R_{x,y,sc,F,t} = \omega_H \cdot RH_{x,y,sc,F,t} + \omega_R \cdot RR_{x,y,sc,F,t}$$

$$S_{x,y,sc,F,t} = \omega_H \cdot SH_{x,y,sc,F,t} + \omega_R \cdot SR_{x,y,sc,F,t}$$

where ω_H and ω_R are the weights given to the different parts with $\omega_H, \omega_R \in \mathbb{R}^+$.

Combining ratings for health data

The reputation part based on ratings for health data can be calculated by combining the ratings for a user. The positive fraction R of the reputation part is calculated by summing all positive fractions r of the local ratings. The positive fractions r of the ratings are scaled by several factors:

- The certainty c : a rating with a high certainty should be given more weight than a rating with a low certainty. The certainty is supplied by the health care provider.
- The order function $f(i, j)$: a function that gives more weight to more recent ratings. Users may start behaving better (or worse) over time. Therefore, recent ratings should be given more weight than older ratings. This approach is similar to the one of the Beta reputation system [3].
- The time function $g(i, j)$: a function that gives more weight if the health data was created closer in time. As the time between the rating and the calculation of the reputation part increases, the rating should be given less weight. After all, if a user performed good (or bad) a long time ago, there is no guarantee that he will do so now. Wixted suggest that humans forget memories according to a power function [10]. Therefore, a power function is used to implement $g(i, j)$.
- The similarity in scope function $SS(sc_i, sc_j)$: if not enough ratings are available for calculating the reputation, ratings for other scopes can be used. For the ratings to be useful, the scope for which the reputation is calculated has to be related to the scope of the ratings. The scope function gives high values (i.e. close to 1) if the two scopes are similar and low values (i.e. close to 0) if the two scopes are not similar. For example, the similarity between taking blood pressure measurements and taking weight measurements is close to 0, but the similarity between taking blood pressure measurements with highly similar measurement devices is close to 1.

For the negative fraction of the reputation SH the same factors are used. This leads to the following calculation for RH and SH (where it is assumed that the ratings in the sets H_j are ordered by time and indexed by i):

$$\begin{aligned}
 RH_{x,y,sc,tt,t} &= \sum_{j=1}^{|SC|} \left(SS(sc, sc_j) \cdot \sum_{i=1}^{|H_j|} g(t_i, t) \cdot f(i, |H_j|) \cdot c_{x,y,sc_j,tt,t_i}^{(i)} \cdot r_{x,y,sc_j,tt,t_i}^{(i)} \right) \\
 SH_{x,y,sc,tt,t} &= \sum_{j=1}^{|SC|} \left(SS(sc, sc_j) \cdot \sum_{i=1}^{|H_j|} g(t_i, t) \cdot f(i, |H_j|) \cdot c_{x,y,sc_j,tt,t_i}^{(i)} \cdot s_{x,y,sc_j,tt,t_i}^{(i)} \right)
 \end{aligned}$$

Combining rule ratings

Suppose R is a set of rule ratings $(r_{y,sc,tt,t}^{(i)}, s_{y,sc,tt,t}^{(i)}, c_{y,sc,tt,t}^{(i)})$ indexed by i . The rule part of the functional reputation can be computed as follows:

$$RR_{y,sc,tt,t} = \sum_{i=1}^{|R|} c_{y,sc,tt,t_i}^{(i)} \cdot r_{y,sc,tt,t_i}^{(i)} \quad SR_{y,sc,tt,t} = \sum_{i=1}^{|R|} c_{y,sc,tt,t_i}^{(i)} \cdot s_{y,sc,tt,t_i}^{(i)}$$

3.2 Aggregation engine

An aggregation rating is calculated by the aggregation engine by comparing measurements from different suppliers with a small time difference. If two users (e.g. a doctor and a patient) take the same measurement on the same person and these measurements are similar, then the reputation of both users can be increased. If two users take the same measurement on the same person and the measurements are not similar, then the reputation of both users must be decreased. The amount with which the reputation should be increased or decreased depends on the reputation of the users that take the measurements.

Health data is denoted by $hd_{x,y,sc,t}$, where x is the person that created the health data, y is the subject of care, sc is the scope and t is the time of creation. A scope sc is a pair (m, d) where m is the measurement kind and d is the measurement device. Suppose $hd_{y,z,(m,d),t}$ is a measurement and D is a set of measurements $hd_{y_i,z,(m,d_i),t_i}$ (indexed by i) of the same kind and on the same person. An aggregation rating for $hd_{y,z,sc,t}$ is calculated as follows:

$$\begin{aligned}
 r_{x,y,(m,d),F,t} &= SH(hd_{y,z,(m,d),t}, D) \\
 s_{x,y,(m,d),F,t} &= 1 - SH(hd_{y,z,(m,d),t}, D) \\
 c_{x,y,(m,d),F,t} &= \frac{R_c}{R_c + S_c + 2}
 \end{aligned}$$

where $SH(hd_{y,z,(m,d),t}, D)$ is the function that compares the measurement $hd_{y,z,(m,d),t}$ to the measurements in the set D and (R_c, S_c) is the combined reputation of the users that took the measurements in D . The combined reputation (of the users that supplied the measurements in D) is calculated as follows:

$$R_c = \sum_{i=1}^{|D|} ST(t, t_i, m) \cdot R_{x,y_i,(m,d_i),F,t_i} \quad S_c = \sum_{i=1}^{|D|} ST(t, t_i, m) \cdot S_{x,y_i,(m,d_i),F,t_i}$$

where $(R_{x,y_i,(m,d_i),F,t_i}, S_{x,y_i,(m,d_i),F,t_i})$ is the reputation of the supplier of $hd_{y_i,z,(m,d_i),t_i}$ and $ST(t, t_i, m)$ is the similarity in time between $hd_{y_i,z,(m,d_i),t_i}$ and the measurement for which the aggregation rating is calculated. The reputation of the users that supplied measurements close in time should be given more weight than the reputation of users that supplied measurements that were not close in time. Therefore, the similarity in time is used as a scaling factor.

The certainty c of the aggregation rating is high if R_c is high and S_c is low. This is the case if (1) the reputation of the suppliers of the health data in D is high, and (2) the measurements were taken close in time.

Similarity of health data

The aggregation engine uses a function $SH(hd_{y,z,(m,d),t}, D)$, where $hd_{y,z,(m,d),t}$ is the measurement for which the aggregation rating needs to be computed and D is the set of values against which $hd_{y,z,(m,d),t}$ has to be compared.

First, the most probable value for the measurement $hd_{y,z,(m,d),t}$ is calculated. Therefore, a set $D' = D \cup \{hd_{y,z,(m,d),t}\}$ is constructed. The most probable value is a weighted average of the measurements in D' where the weights depend on the similarity in time:

$$mpv(hd_{y,z,(m,d),t}, D) = \frac{\sum_{i=1}^{|D'|} ST(t, t_i, m) \cdot hd_{y_i,z,(m,d_i),t_i}^{(i)}}{\sum_{i=1}^{|D'|} ST(t, t_i, m)}$$

After that, a function based on the normal distribution is used to determine the similarity between a measurement and the most probable value:

$$SH(hd_{y,z,(m,d),t}, D) = e^{-\frac{(hd_{y,z,(m,d),t} - \mu)^2}{2 \cdot \sigma_{SH,m}^2}}$$

where $\mu = mpv(hd_{y,z,(m,d),t}, D)$ and $\sigma_{SH,m}$ is a system parameter, representing the standard deviation for measurement kind m .

$ST(t_1, t_2, m)$ is a function that returns the similarity in time between measurements of kind m at time t_1 and t_2 . The similarity in time between two measurements can be seen as the probability that the physical state of a patient has not changed between the two measurements. A possible implementation can be found in [9].

3.3 Rule engine

The rule engine computes rule ratings that can be used by the reputation engine. The computation relies on available certificates of the user for which the reputation is calculated. The rule engine uses a predefined mapping to find the rule ratings associated with a certificate. A certificate is represented as a tuple (x, p, t) stating property p about user x , where p can be any property leading to a rule rating (e.g. ‘completed medical school’ or ‘successfully completed online tutorial for measuring blood pressure’). In this tuple, t is the time of creation of the certificate.

4 Conclusion

In this paper, we proposed to use a reputation system and metadata provided by measurement devices to give a quality indication for health data. Therefore, health care providers can also use health data supplied by non-professionals, such as patients and wellness providers. This results in reduced costs and higher quality of health care.

The purpose of a reputation system is to build trust in online environments. As a side-effect, reputation systems provide an incentive for good behavior. Therefore, using a reputation system to make a quality indication of health data is a natural choice. It also mirrors real practice, in which a health care provider builds trust in patients and the health data they supply.

The most important advantage of Hedaquin is that it can calculate ratings for health data automatically. Furthermore, certificates for the user such as diplomas are used in the calculation of the reputation. Hedaquin calculates reputation for different scopes in order to discriminate among different tasks the users can perform. Finally, the time difference between the ratings and the calculation of the reputation influences the weight that is given to a rating.

All in all, Hedaquin gives health care providers the opportunity to make an informed decision on the quality of health data that is supplied by patients and wellness providers. For patients and wellness providers there is no overhead in using the system. The overhead for health care providers is minimal.

References

- [1] Aringhieri, R., E. Damiani, S. D. C. di Vimercati, S. Paraboschi and P. Samarati, *Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems.*, JASIST **57** (2006), pp. 528–537.
- [2] Caballero, A., J. A. B. Blaya and A. F. Gómez-Skarmeta, *A new model for trust and reputation management with an ontology based approach for similarity between tasks.*, in: *MATES*, 2006, pp. 172–183.
- [3] Ismail, R. and A. Jøsang, *The beta reputation system*, in: *Proceedings of the 15th Bled Conference on Electronic Commerce*, 2002.
- [4] Jøsang, A., *A logic for uncertain probabilities.*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **9** (2001), pp. 279–212.
- [5] Kamvar, S. D., M. T. Schlosser and H. Garcia-Molina, *The eigentrust algorithm for reputation management in p2p networks.*, in: *WWW*, 2003, pp. 640–651.
- [6] Liu, J. and V. Issarny, *Enhanced reputation mechanism for mobile ad hoc networks.*, in: *iTrust*, 2004, pp. 48–62.
- [7] Commission on Systemic Interoperability, “Ending the Document Game,” U.S. Government Printing Office, 2005.
- [8] Sabater, J. and C. Sierra, *Regret: reputation in gregarious societies.*, in: *Agents*, 2001, pp. 194–195.
- [9] van Deursen, T., “Hedaquin: A reputation-based health data quality indicator for a personal health record,” Master’s thesis, Technische Universiteit Eindhoven (2007).
- [10] Wixted, J. T., *On common ground: Jost’s (1897) law of forgetting and Ribot’s (1881) law of retrograde amnesia*, Psychological review (2004).
- [11] Xiong, L. and L. Liu, *Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities.*, IEEE Trans. Knowl. Data Eng. **16** (2004), pp. 843–857.