Evaluating an ordered list of recommended physical activities within health care system

Igor Kulev[1], Elena Vlahu-Gjorgievska[2], Saso Koceski[3], Vladimir Trajkovik[1]

[1]Faculty of Computer Science and Engineering,
University "Ss Cyril and Methodious", Skopje, Macedonia
{igor.kulev,trvlado}@finki.ukim.mk
[2]Faculty of administration and information systems management,
University "St.Kliment Ohridski", Bitola, Macedonia
elena.vlahu@uklo.edu.mk
[3]Faculty of Computer Science,
University "Goce Delcev", Stip, Macedonia
saso.koceski@ugd.edu.mk

Abstract. Information and communication technologies make it possible to bridge the gap and time barriers in the flow of health information and knowledge, allowing every involved part in the health process to have access to the information. This approach provides the knowledge of the individual to contribute effectively to the improvement in human health. But also, helps the collective knowledge effectively to solve health problems on individual level. In this paper we are evaluating the algorithm that generates recommendation for users. We are using simulations on generic data to see how different types of activities are affecting the accuracy of the algorithm. On the basis of the performed activities and blood glucose measurements, our recommendation algorithm should determine list of activities that have bigger influence on the change of the blood glucose levels. Generic data for our simulations are based on modeling of food intake and physical activity influence over the blood glucose level.

Keywords: recommendation algorithm, blood glucose level, evaluation

1 Introduction

The advances in communication and computer technologies have revolutionized the way health information is gathered, stored, processed, and communicated to decision makers for better coordination of healthcare at both the individual and population levels [1]. Pervasive health care takes steps to design, develop, and evaluate computer technologies that help citizens participate more closely in their own healthcare [2], on one hand, and on the other to provide flexibility in the life of patient who lead an active everyday life with work, family and friends [3].

adfa, p. 1, 2011.

Life style with moderate eating habits and increased physical activity plays a key role in disease management. Some clinical conditions (like diabetes, metabolic syndrome, chronic heart failure) can be prevented by proper diet and regular physical activity. There are number of studies that have shown that increased physical activity and diet modification (termed as 'lifestyle interventions'), independent of other risk factors, has a protective effect against the development of chronic diseases as diabetes and metabolic syndrome [4, 5]. Guidance and interactive training regarding appropriate choices of diet and exercise plans combined with encouragement and monitoring of progress, can empower patients to make beneficial lifestyle modifications [6].

The recommendation algorithm, evaluated in this paper, is part of the Collaborative Health Care System Model – COHESY [7]. COHESY is "a tool" for personal heathcare. It is deployed over three basic usage layers. The first layer consists of the bionetwork (that reads parameters' values from various body sensors) and a mobile application (that collects users' bio data and parameters of performed physical activities). The second layer is presented by the social network and the third layer enables interoperability with the primary/secondary health care information systems. The usage of social network and its' recommendation algorithm are the main advantages of COHESY. The social network enables different collaboration within the end user community. It allows communication between users, exchange of their experiences and gathering of large amount of data about their health parameters, food intake and performed activities. The recommendation algorithm takes into account the effects of food and physical activity on health parameters (e.g. blood glucose level), and based on prior knowledge (data gathered from social network and clinical centers) recommend physical activity that will improve the users' health.

The next section gives a brief overview of the recommendation algorithm. In the third section experimental methodology and result will be discussed. The fourth section is the conclusion of the paper.

2 Recommendation algorithm

The main purpose of this algorithm is to find the dependency of the users' health condition, food intake and physical activities they perform. The algorithm incorporates collaboration and classification techniques in order to generate recommendations and suggestions for preventive intervention. To achieve this we consider the data read by the bionetwork (parameters values), the data about the user's physical activities, the user's medical record (obtained from a clinical centre) and the data contained in the user profile on the social network (so far based on the knowledge of the social network). We use classification algorithms on these datasets to group the users by their similarity. Use of classified data when generating the recommendation provides more relevant recommendations because they are enacted on knowledge for users with similar medical conditions and reference parameters.

Generally in our algorithm we use a similarity metrics in order to find the most similar users to the active user according to their medical history. We assume that if two users had the same combination of parameter values in the past, there is bigger probability that similar latent factors affect their health condition. For each user from the set of similar users we keep the details about the physical activities he performed and the measurements of his health parameters. Further, we use only data from the active user and from the users most similar to him, and we calculate the usefulness of each type of physical activity. We analyze the history of activities and measurements of each user and we want to find the type of influence of each type of activity on each of the health parameters. For this purpose two measurements (value of the parameter) are selected for each activity – the most recent measurement before the execution of the activity and a measurement performed a particular time period after the execution of the activity. We do not choose the first measurement after the activity because a time is needed for the activity to show its effect. The difference between the next and the previous measurement approximates the influence of the activity on the parameter change. After this, we use the information about the usefulness of each activity in order to generate recommendations. For each user from the set of similar users (plus the active user) we obtain the most useful activity that could potentially improve his health condition. The activity which is declared as the most useful to most of the users is recommended to the active user.

Simulations made for the evaluation presented in next section are for one user that has food intake 3 times per day. Although the algorithm has more steps, which are in details explained in [8], for the evaluation covered in this paper the step for calculating the benefits of performing the activity is important. This step is presented below.

Find the benefits of performing an activity. The benefit of performing activity *a* by the user *u* for parameter *p* is calculated by Eq.(1).

$$V_{u,a,p} = \frac{importance_p \cdot \sum_{a_u}\left(\frac{|next_p(a_u) - prev_p(a_u)|}{timeSpan(duration(a_u))} \cdot validity(a_u) \cdot dir(a_u, p) \cdot intensity\ (a_u)\right)}{num(a_u)}, \qquad (1)$$

$$validity(a_u) = validityPrev(a_u) \cdot validityNext(a_u),$$

$$\forall a, a \in A', \forall p, p \in P''$$

- *A'* – set of different activities;
- *P'* – set of health parameters;
- $next_p(a_u)$ - function that returns the value of the parameter after performing the activity *a*;
- $prev_p(a_u)$ - function that returns the value of the parameter before performing the activity *a*;
- $duration(a_u)$ - the duration of the activity *a*;
- $intensity\ (a_u)$ - the value of the intensity of performed activity *a*, calculated by an appropriate formula;
- $num\ (a_u)$ - the number of reviewed readings of activity *a*;
- $validity(a_u)$ - function that returns the validity of the measured activity *a*;
- $importance_p$ – the importance of this parameter for the user (the bigger importance of the parameter for the health of the user - the higher its coefficient is);
- $timeSpan(x)$ – logarithmic function.

$dir(a_u, p)$= -1 when we are calculating the benefits of an activity that decreases the value of the parameter. When the activity increases the value of the parameter the value of this function is 1.

3 Methodology

We assume that there are two factors that affect the parameter value: food intakes and activities. Food intakes tend to increase and activities tend to decrease the parameter value. We assume that after consuming the food, there is a short period of time where it causes rapid increase of the parameter value, and after that there is a short period of time where it causes decrease of the parameter value. After this "unstable" period of increase and decrease, as a result there is a small increase of the parameter value. The final change of the parameter value caused by the food intake happens 6-9 hours after the food intake and depends on the type of the food intake. Three food intakes happen each day: breakfast, lunch and dinner. These food intakes affect the parameter value with different intensity. Lunch has the largest effect and causes the biggest increase after the unstable period. Dinner causes smaller increase than lunch and breakfast causes the smallest increase. Breakfast, lunch and dinner happen around 09:00, 12:30 and 18:00 accordingly. The exact moment of occurrence of breakfast, lunch and dinner is determined in each day of the simulation by using a Gaussian distribution. The parameter function affected only by food intakes that happen in one day is given on Fig. 1. We define the maximal increase to be the largest increase caused by food intake and the stable increase to be the increase of the parameter value after the "unstable" period. The stable increases for breakfast, lunch and dinner are 0.2, 0.4 and 0.3 accordingly.

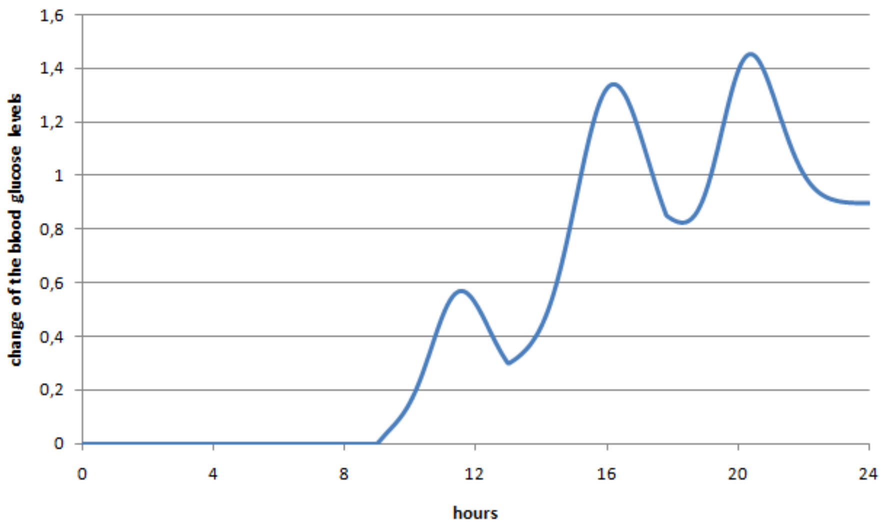

Fig. 1. Change of the blood glucose level during one day under the influence of breakfast, lunch and dinner and under no other kind of influence

Activities have opposite effect on the parameter value. After the activity there is a short period of time where it causes rapid decrease of the parameter value, and after that there is a short period of time where it causes increase of the parameter value. After this "unstable" period of decrease and increase, as a result there is a small decrease of the parameter value. The final change of the parameter value caused by the activities happens 3 days after the activity. Minimal decrease and stable decrease could be defined for activities in a similar way as for food intakes (Fig 2).

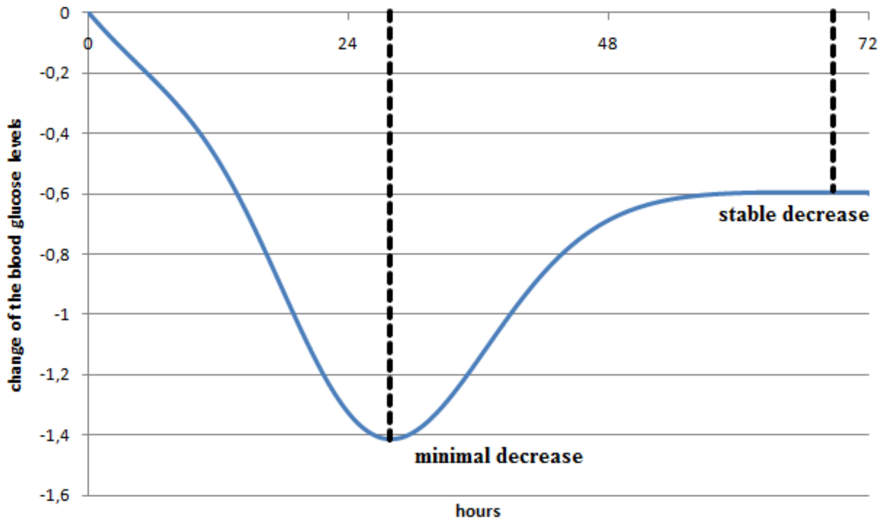

Fig. 2. Change of the blood glucose levels generated by one activity

We simulate 40 days in which the parameter value is changed by three different types of food intakes defined above and N different types of activities. The minimal decrease and the stable decrease caused by each type of activity are defined by the following formulas:

$$md_i = (-0.9) - \frac{(N-1)}{2} diffMD + (i-1) \cdot diffMD \quad (2)$$

$$sd_i = (-0.6) - \frac{(N-1)}{2} diffSD + (i-1) \cdot diffSD \quad (3)$$

where $1 \leq i \leq N$. Our simulator has three parameters: N, $diffMD$ and $diffSD$. $diffMD$ represents the difference between minimal decreases of two consecutive types of activities and $diffSD$ represents the difference between stable decreases of two consecutive types of activities. We generate the same number of activities of each type. When we choose the number of activities of each type we assume that the expected change of the parameter value in the end of the simulation is zero (or as close to zero as possible). We calculate the number of activities of each type according to the formula:

$$numberOfActivitiesOfEachType = \frac{-(simulationLengthInDays \cdot (0.2+0.4+0.3))}{N \cdot (-0.6)} \quad (4)$$

The activities and 40 measurements are generated at random moments during one simulation. The change of the parameter value during one simulation is shown on Fig. 3a. The data that is provided to the recommendation algorithm (activities and measurements) is shown on Fig. 3b.

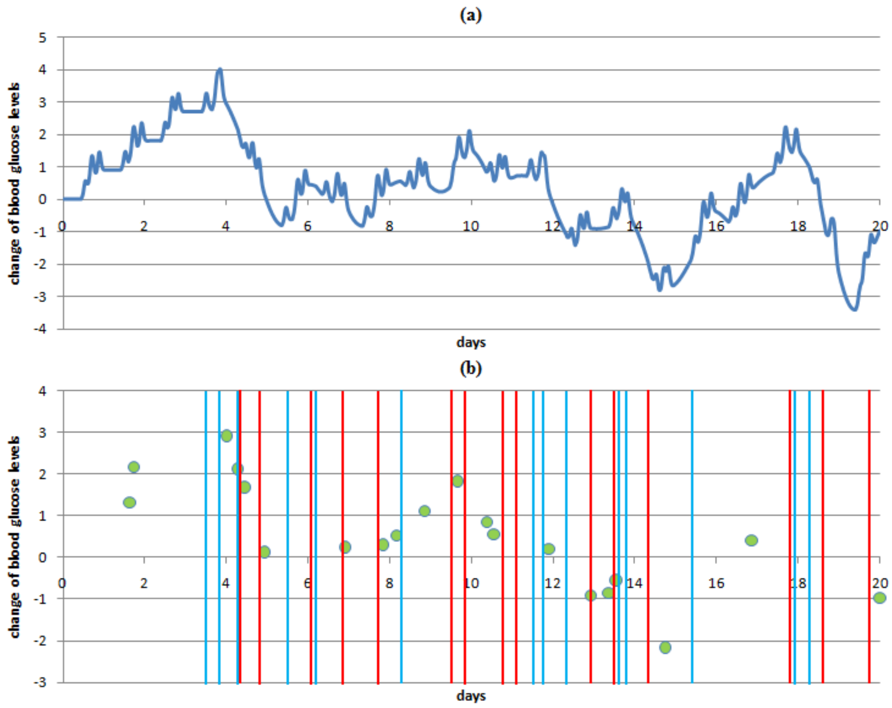

Fig. 3. a) Change of the blood glucose levels during one simulation; b) Data provided to the recommendation algorithm. The moments when the first type of activity has occurred are denoted with blue vertical lines. The moments when the second type of activity has occurred are denoted with red vertical lines. The measurements are denoted with green circles.

On the basis of the activities and measurements, our recommendation algorithm should determine the usefulness of all types of activities and should provide an ordered list of the types of activities according to their usefulness. If some type of activity has higher rank than other type of activity, this means that the algorithm has concluded that the first type of activity has bigger stable decrease than the second one. In this paper we want to evaluate how well the algorithm ranks the types of activities.

4 Results and analysis

We have performed three different experiments. In the first experiment we have changed the value of $diffMD$ from 0.025 to 0.1 (in time intervals of length 0.025), in the second experiment we have changed the value of $diffSD$ from 0.025 to 0.1 (in time intervals of length 0.025) and in the third experiment we have changed both the values of $diffMD$ and $diffSD$ from 0.025 to 0.1 in the same time (in time intervals of length 0.025). In each experiment we have evaluated the quality of the ordered list using three evaluation metrics: Normalized Discounted Cumulative Gain (NDCG), Precision and Recall, and the Number of inversions.

4.1 Normalized Discounted Cumulative Gain

Normalized discounted cumulative gain (NDCG) measures the performance of a recommendation system based on the graded relevance of the recommended entities. It varies from 0.0 to 1.0, with 1.0 representing the ideal ranking of the entities. This metric is commonly used in information retrieval and to evaluate the performance of web search engines [9]. All three experiments are performed for different N from 2 to 12. The results of the evaluations using NDCG for all three experiments are shown on Fig. 4. It can be seen that the ranked list is relevant because the normalized discounted cumulative gain for each combination of values of N and $diffMD/diffSD$ is higher than the normalized discounted cumulative gain of random ordering of a list. Although we can conclude that the generated ordered list is relevant and better than random list, we cannot say how good the ordering is. Additionally, we cannot compare two NDCGs of results from experiments with different N. From Fig. 4 we can conclude that when we decrease $diffMD$ we get better results. Same happens when we decrease $diffSD$. In both cases we increase the absolute difference between $diffMD$ and $diffSD$. However, when we decrease $diffMD$ and $diffSD$ in the same time, we get worse results than if we decrease only one of the two parameters: $diffMD$ or $diffSD$ (except in the case when $diffMD = -0.1$ and $diffSD = -0.1$). This is conclusion stands for all different values of N.

4.2 Precision and Recall

If we separate the types of activities in two groups: relevant and not relevant, then we can use the Precision and Recall measure to evaluate how much the relevant types of activities are ranked higher by the recommendation algorithm. Precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also known as sensitivity) is the fraction of relevant instances that are retrieved [10]. We have chosen that $N = 10$ in all simulations. We have marked the first 5 activities that have the highest $diffMD$ and $diffSD$ as relevant and we consider the others as not relevant. We have performed all three experiments and the results are shown on Fig. 5. These results show that the ordered list is relevant and they affirm the results obtained by the NDCG measure. Analyzing the Precision and Recall curves, we can also affirm the conclusion that the algorithm gives lower accuracy when we change $diffMD$ and $diffSD$ in the same time.

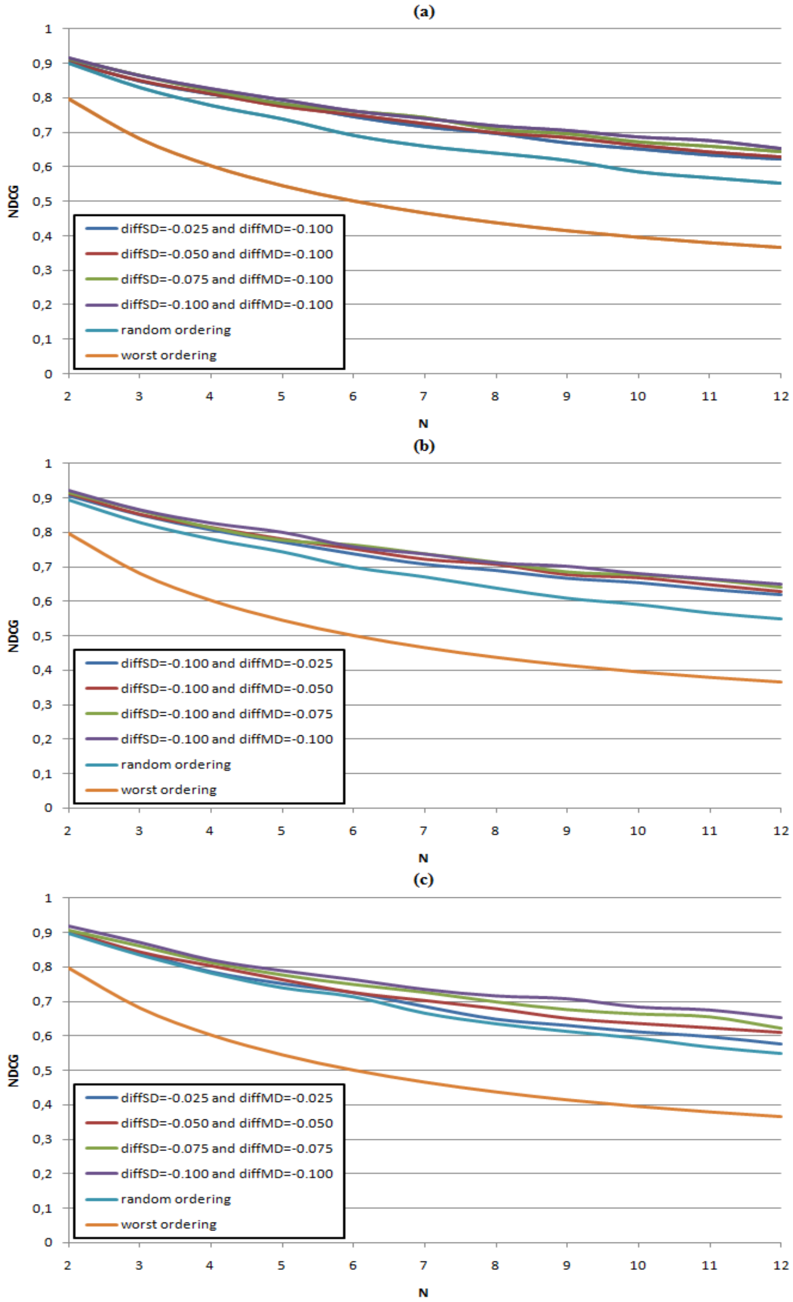

Fig. 4. Normalized Discounted Cumulative Gain for experiments with different number of activities

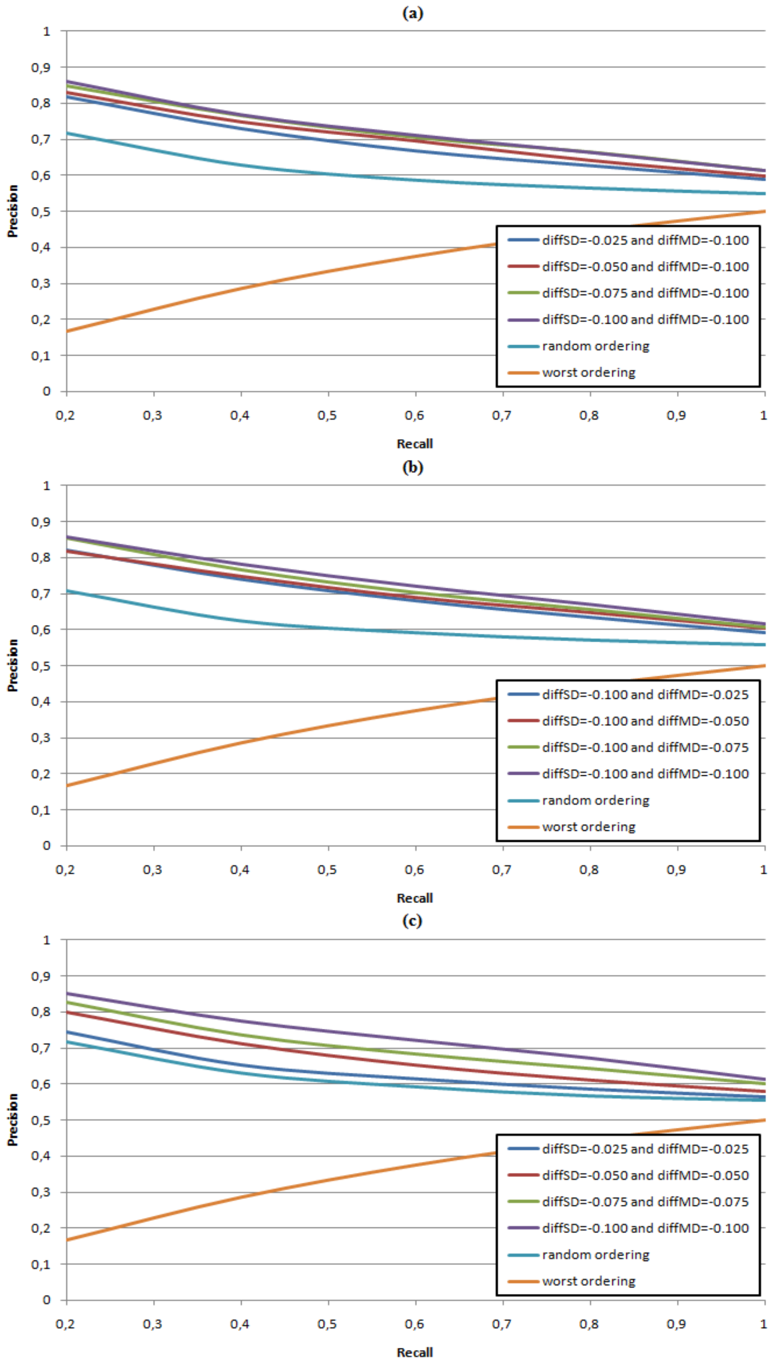

Fig. 5. Precision and Recall curves for experiments with 10 types of activities

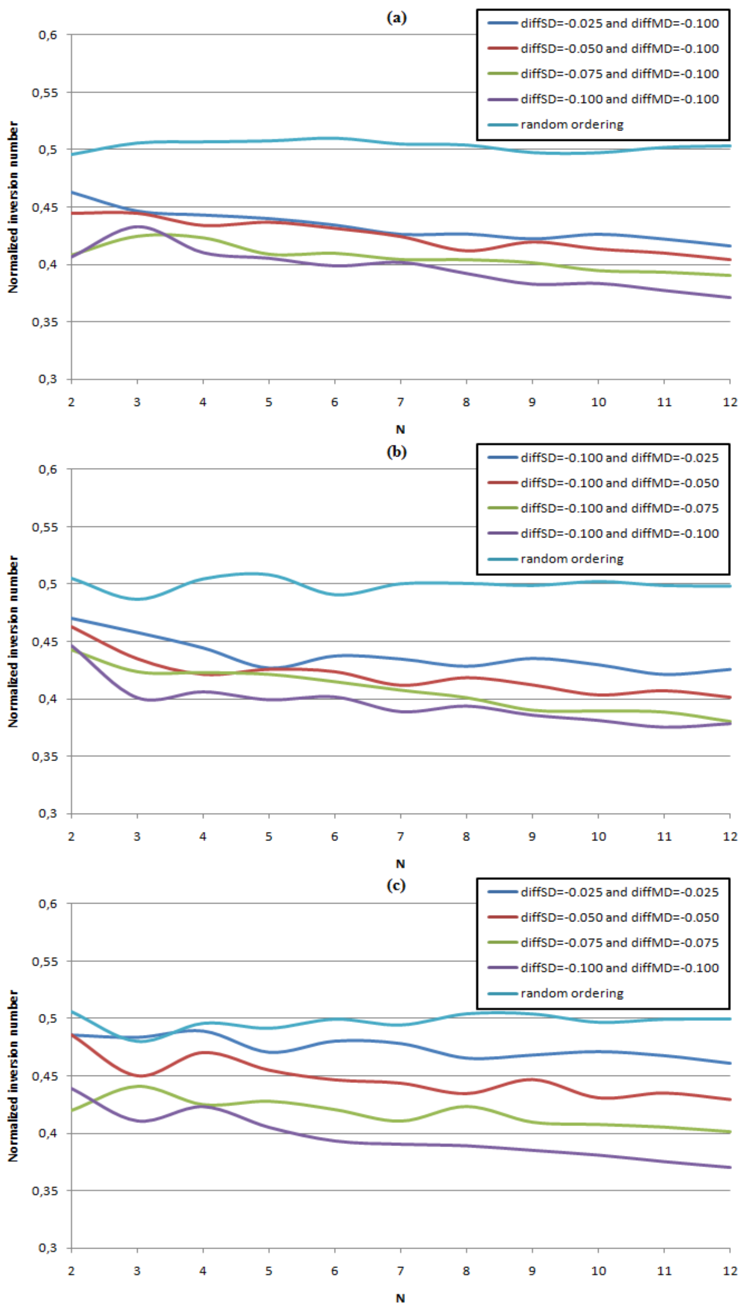

Fig. 6. Normalized inversion numbers for experiments with different number of activities

4.3 Number of Inversions

In computer science and discrete mathematics, an inversion is a pair of places of a sequence where the elements on these places are out of their natural order. The inversion number of a sequence is one common measure of its sortedness [11,12]. All three experiments are performed for different N from 2 to 12. The inversion number is normalized – the best ordering should have normalized inversion number 0, the worst ordering should have normalized inversion number 1 and the random ordering should have normalized inversion number 0.5. The results of our experiments are shown on Fig. 6. We can confirm the results obtained by the NDCG measure and the Precision and Recall measure. Additionally, using the normalized inversion number we can compare the performance of the algorithm for different N. We can see a linear decrease of the normalized inversion number in all curves obtained from our experiments. This means that as we increase the number of different types of activities we get better ordered list. The explanation of this result could be that as we increase the number of different types of activities, the difference between the type of activities with the biggest and smallest $diffMD/diffSD$ increases and it could be easier for the algorithm to conclude which of these activities produce a bigger/smaller stable decrease. Before the evaluation we weren't sure how the algorithm would behave when we increase N. This was because when the total number of activities increases, the global parameter function becomes more complicated so this could mean the algorithm might behave worse.

5 Conclusion

Our evaluation shows that the recommendation algorithm gives relevant ranking of the types of activities according to their usefulness. We have confirmed this conclusion by using three evaluation metrics: Normalized Discounted Cumulative Gain (NDCG), Precision and Recall, and the Number of inversions. We also conclude that the quality of the generated ordered list increases if the difference between the minimal decrease ($diffMD$) and the stable decrease ($diffSD$) is bigger or when the magnitude of both $diffMD$ and $diffSD$ is bigger. Increasing the number of different types of activities results in more complicated parameter function, but according to the simulation results it does not mean that the algorithm gives imprecise recommendations.

Acknowledgements. This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius" University. The authors would also like to acknowledge the contribution of the COST Action IC1303 - AAPELE, Architectures, Algorithms and Platforms for Enhanced Living Environments.

References

1. Fichman, R.G., Kohli. K., Krishnan. R.: Editorial Overview-The Role of Information Systems in Healthcare: Current Research and Future Trends. Information Systems Research. 22(3), 419-428 (2011)

2. Ahamed, S.I., Haque, M.M., Khan, A.J.: Wellness assistant: a virtual wellness assistant using pervasive computing. In: Symposium on Applied Computing, pp. 782-787. USA: ACM (2007)
3. Ballegaard, S.A., Hansen, T.R., Kyng, M.: Healthcare in everyday life: designing healthcare services for daily life. In: Conference on Human Factors in Computing Systems, pp. 782-787. USA: ACM (2008)
4. Nachman, L., Baxi, A., Bhattacharya, S., Darera, V.: Jog Falls: A Pervasive Healthcare Platform for Diabetes Management. Pervasive Computing. 6030, 94–111 (2010)
5. Orozco, L.J., Buchleitner, A.M., Gimenez-Perez, G., Roqué i Figuls, M., Richter, B., Mauricio, D.: Exercise or exercise and diet for preventing type 2 diabetes mellitus. Cochrane Database of Systematic Reviews. 3(Art. No.: CD003054) (2008)
6. Davidson, J.: Strategies for improving glycemic control: effective use of glucose monitoring. The American Journal of Medicine.118(Suppl 9A):27S–32S (2005)
7. Vlahu-Gjorgievska, E., Trajkovik, V.: Towards Collaborative Health Care System Model – COHESY. In: 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, , pp:1-6. IEEE Computer Society, Washington, DC (2011)
8. Kulev, I., Vlahu-Gjorgievska, E., Trajkovik, V., Koceski, S.: Development of a novel recommendation algorithm for collaborative health - care system model. Computer Science and Information Systems. 10(3), 1455-1471 (2013)
9. Kalervo, J., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS). 20(4): 422-446 (2002)
10. Nandish, C., Goyani, M.: Enhanced Multistage Content Based Image Retrieval. (2013)
11. Philippe, F., Vitter, J. S.: Average-case analysis of algorithms and data structures. (1987)
12. Wilhelm, B., Mutzel, P., Jünger, M.: Simple and efficient bilayer cross counting. Journal of Graph Algorithms and Applications 8(2): 179-194(2004)