

Dynamic Time Warping for Retrospective Patient Case Matching

Sadiq Sani¹, Nirmalie Wiratunga¹, Stewart Massie¹, and Markus Steiner²

¹ School of Computing and Digital Media,
Robert Gordon University,
Aberdeen AB10 7GJ, Scotland, UK
{s.sani,n.wiratunga,s.massie}@rgu.ac.uk

² University of Aberdeen
School of Medicine
Department of Child Health
Aberdeen AB25 2ZG, Scotland, UK
m.steiner@abdn.ac.uk

Abstract. Phase 4 or Post-marketing trials are conducted in order to prepare a new drug for the mass market, after a license has been granted. An important study in Phase 4 trials is case-control where pairs of patients with similar pre-trial characteristics are selected and only one patient of the pair (the trial patient) receives treatment using the new drug. The outcome the trial patient is then compared to that of the other (control) patient over the period of the trial. This is done in order to further understand any benefits or risks associated with the drug. Any difference in outcome between the two patients is then attributed to the effects of the new treatment. In this paper, we present a CBR system for finding pairs of patients with similar pre-trial characteristics. Patients are often characterised by recorded observations over a period of time before the trial. Thus, our main contribution is the definition of a similarity measure based on Dynamic Time Warping, that takes into account stretching and compression of patient attributes in the time dimension. This allows us to more accurately capture similarity between patients based on the general pattern of an observed characteristic between the two patients over time, rather than the exact matching of patient characteristics at the same point in time.

Keywords CBR, Similarity, Dynamic Time Warping, DTW, Clinical Studies, Medical Application.

1 Introduction

Phase 4 or Post-marketing trials are conducted after a new drug has gone through previous trials and a license has been granted. According to a recent survey, the global average cost for Phase 4 trials across the pharmaceutical and biomedical industry is six thousand US dollars per patient per study [4]. The cost is found to range from \$1,000, at the lower end, to \$ 12,000 at the higher end. Definitely, any

cost savings that can be achieved by automating some of the processes involved in phase 4 trials will be warmly welcomed by these industries.

Reasons for running Phase 4 trials are to find out:

- More about the side effects and safety of a drug
- What the long term risks and benefits are
- How well the drug works in real world settings.

An important study in Phase 4 trials is case-control where the aim is to evaluate, given a group of patients receiving the new treatment, the outcome achieved, compared to if the treatment had not taken place. However, it is not possible to measure both outcomes simultaneously on the same group of patients. Hence, a control group of patients with similar characteristics to the trial group is chosen such that, the distribution of important pre-trial characteristics in the two groups is matched. Accordingly, any difference in outcome between the two groups can be attributed to the effects of the new treatment. However, choosing an appropriate and convincing control group is not a simple task.

There is recently a growing interest in using observational (non-randomised) methods for finding control groups [5]. In this approach, selection of individuals for the control group is determined by some measured characteristics of these individuals. In order to account for systematic bias, statistical approaches are used for the selection of individuals. Propensity matching is a popular approach in this category where a probability of group membership (i.e. treatment or control) for each individual is computed based on observed characteristics (covariates) that are predictive of the membership. Thus, propensity scores attempt to mitigate any selection bias by equating the two groups based on these covariates. However, correctly applying propensity scoring can be quite involved, particularly where temporal aspects are involved. Propensity scoring is also based on probabilities. Thus, there is a need to carefully choose covariates that are suitable predictors of whether or not a patient should be added to the control group.

In this paper, we present a Case-Based Reasoning (CBR) system for pairing patient in order to create trial and control groups for phase 4 clinical trials. Our system is based on a sample dataset collected by NHS Scotland containing all dispensed drugs which have been prescribed by General Practitioners (GPs) for asthma. This data is routinely collected for the aim of reimbursing pharmacists for the dispensed drugs. The Prescribing Information Service (PIS) has a 95% completeness in terms of linking the dispensed drug to an individual person, using the Community Health Index (CHI) which is unique for every person in Scotland registered with a GP. From the prescription data, we are able to estimate the amount of prescribed drugs taken daily by a patient, which allows us to infer important information such as: the severity of the condition of the patient; the level of control of their condition i.e. whether the patient's condition is stable, improving or getting worse; as well as the monthly drug intake of the patient over a certain time period, e.g. one year. While some of this information might not be suitable for propensity score matching due to changes over time, we can easily model this in a CBR system where each of these pieces of information

is considered an attribute of a patient case description. In this paper, we describe the development of such a CBR system. We also present a similarity measure for time dependent attributes based on Dynamic Time Warping (DTW) [2]. DTW is a popular algorithm used for the comparison of time-related data, which takes into account distortions in the time dimension. This is particularly important for our system because it allows us to effectively estimate similarity between patients that may have say, similar severity patterns even if their conditions may not be equally severe at the same point in time.

This paper is organised as follows: Section 2 introduces related work on applications of CBR to medical studies, as well as the use of DTW for the comparison of medical data. Section 3 gives a description of our dataset, while Section 4 presents our case structure and gives a detailed description of the similarity metrics used for each attribute. Section 5 describes preliminary experiments on our system and Section 6 presents interesting challenges we are looking to address in future work. Conclusions are presented in Section 7.

2 Related Work

Case based reasoning has a history of successful applications in health sciences [9] [3] [1]. For example, a CBR framework for eligibility screening of patients using their electronic health records is presented in [7]. The system is designed to reduce the amount of work involved in screening a large pool of patients for eligibility for a clinical trial by screening an initial small sample of patients. Once this small sample of trial patients has been identified, their information is used to derive a general representation of a prototypical trial patient. Additional patients from the pool are included into the trial sample based on similarity to the prototypical trial patient. Another CBR system which uses patient similarity for Comparative Effectiveness Research (CER) is described in [6]. The idea of CER is to evaluate the effectiveness of alternative methods of treatment. One of the important aspects of CER is the focus on the identification and the most effective strategy for treating the individual patients in the study. A case-based approach has also been proposed for assigning appropriate treatments to patients, by reusing the solution (treatment) of similar patients from a case-base [6]. The paper however, only describes a theoretical CBR system for this task, along with the challenges that might be faced e.g. the definition of “similarity” and “sufficiently similar patients”. No actual CBR system was developed.

A limitation of the CBR frameworks discussed so far is that they do not take into account the time-dependence of observations. However, often enough, patient data for case representation is derived from readings that are time dependent. DTW allows us to find an optimal alignment between two time-dependent sequences, providing both a distance measure that is insensitive to local compressions and stretches, as well as the warping, which deforms one time series to resemble, as much as possible, the other [8]. Originally, DTW was proposed in speech recognition to account for differences in speaking rates when compar-

ing speech patterns. Since then, DTW has been widely applied in data mining and information retrieval to automatically account for time deformations and differences in speeds when comparing time-dependent data.

In clinical studies, DTW has been used for comparing patient records which are derived from time-dependent observations. For example, an approach that uses DTW for the comparison of multi-modal medical data is proposed in [10]. The data is called multi-modal because it has been collected using different modalities, each representing a different temporal observation with its own individual characteristics. Successful matching of patients therefore, relies on the effective use of all modalities provided. In this approach, similarity between patients is assessed by estimating multiple DTW distances, one for each modality, and a global similarity between patients is obtained using a 'fusion', which is a sum of all the calculated DTW distances. This has similarities to our proposed CBR approach where each modality is considered as a separate attribute in a patient case. However, in the approach of [10], only time series attributes are considered. This is unlike our approach where we consider a number of different attribute types, including temporally independent data for patient case representation. Hence, determining a global similarity between patients in our approach is slightly more challenging than computing a sum of individual feature similarities.

3 Data Description

Our data consists of single dispensed asthma prescriptions collected from GPs, with information on the dispensed date, the name and code of the drug, the dosage, and the dispensed amount. As asthma is a chronic condition where patients most likely use their medication at regular intervals, it is possible to estimate daily intake from the dispensation information. The number of drugs considered in the data collection include short-acting beta2-agonists (SABA), long-acting beta2-agonists (LABA), inhaled cortico-steroids (ICS), LABA/ICS, short- and long-acting antimuscarinic bronchodilators (SAMA, LAMA), theophyllines (MethylX), leukotriene receptor antagonists (LRTA), and cromoglicerate. In total, eleven drugs were recorded.

In addition to the drugs, the severity of each patient's condition is calculated for each month in the data. Severity levels are assigned according to a scale from 1 to 5, defined by the British Thoracic Society. The definition of each severity level (BTS Step) is provided in Table 1. The asthma severity level of each patient is recorded monthly over a one year period preceding the change to a new drug.

Also, the control level of the patient's condition is recorded. The control level of asthma is a value between 1 and 3 which indicates how controlled a patient's condition is, defined by how often a patient has occurrences of certain types of symptoms. Description of the three control levels is given in Table 2. A value of 1 is used to represent Controlled, 2 is partly Controlled and 3 is Uncontrolled. The control level of each patient is recorded for each of the 12 months of the year. Finally, we also record the age in years of the patient.

STEP	DESCRIPTION
1	Mild intermittent asthma
2	Introduction of regular preventer therapy
3	Initial add-on therapy
4	Persistent poor control
5	Continuous or frequent use of oral steroids

Table 1: Description of BTS levels

Characteristic	Controlled	Partly Controlled	Uncontrolled
Daytime Symptoms	2 x week or less	> 2 x week	3 or more features of Partly Controlled asthma present in any week
Limitations of Activities	None	Any	
Nocturnal symptoms	None	Any	
Need for reliever	2 x week	> 2 x week	
Lung function	Normal	<80% predicted or personal best (if known)	

Table 2: Description of the 3 control levels of asthma symptoms.

4 Case Structure

Our case representation consists of four attributes, Age, Drugs, BTS, and Control as shown in Figure 1. Age is a single integer value representing the age of the patient in years. The Drugs attribute consists of all 11 drugs from the data, where each drug is represented as a vector of length 12 representing the 12 months of the year, and each dimension contains a boolean value (1 or 0) indicating whether the patient had taken that drug in that month. BTS and Control are also represented as vectors of length 12 where each dimension represents a month of the year. The values in each dimension of the BTS vector range from 1 to 5 representing the BTS severity level of the patient while the Control vector ranges from 1-3 representing the asthma control level as described in Section 3.

To compare between cases, we define local similarity functions for each attribute except for the Age attribute. For this, we define a filter with an interval of one year. This is because the specifications of this application require that the age difference between patients in a pair be not greater than one year. Accordingly, the filter is applied such that any patient in the casebase whose age difference compared to the query patient is greater than one year is excluded from further similarity computation. This filter is the first step applied during retrieval which has the additional benefit of making our similarity computation very efficient as only a small proportion of patients cases from the casebase are ever considered for similarity computation.

The Drugs attribute is not recorded as a quantitative observation because it simply records a binary presence/absence of a particular drug in a particular month. This makes this attribute not suitable for DTW or any other metric that computes similarity as a function of distance on some numeric scale. Accordingly,

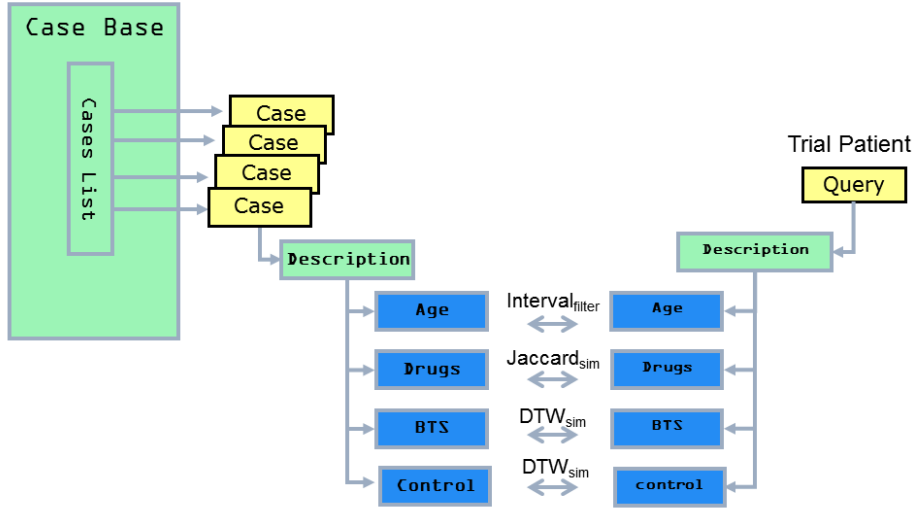


Fig. 1: Case structure.

we use the Jaccard similarity coefficient which is convenient for computing similarity based on the size of overlap between two sets. Given two sets A and B , the Jaccard similarity between the two is computed as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Treating each month for each drug as a separate element in the set of all months and all drugs, we are able to find the similarity between two patients based on the number of the same drugs they took at exactly the same month of the year.

Both BTS and Control are recorded as time-dependent ordinal data with meaningful interval between values. Hence, we define similarity measures for these two attributes based on DTW. The main advantage of DTW over standard linear distance measures like Euclidean, is that it takes into account non-linear distortions in the time dimension. This means that the similarity between two time series with similar shape and patterns, but within acceptable offsets, can be more accurately estimated using DTW.

The difference in alignment between a linear distance measure and DTW on two time series is illustrated in Figure 2, where the image on the left shows alignment between the two time series X and Y using a linear distance measure and the image on the right shows alignment using DTW. Note that while the two time series have very similar shapes, the exact peaks and drops in values are not linearly aligned. The linear distance metric aligns points on the two time series based on their positions in time and hence, is likely to estimate the two time series as being less similar. This is a potential problem for measuring

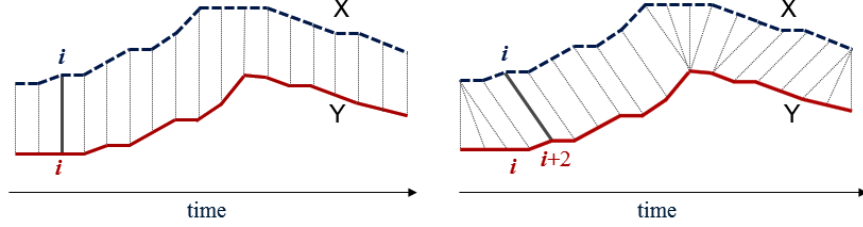


Fig. 2: Difference in alignment between a linear distance measure and DTW on two time series.

similarity in asthma severity between two patients because, patients may have similar severity patterns although their conditions may not be equally severe at the same point in time. DTW addresses this limitation by finding matching points in the neighbourhood that are closer in distance. Thus, we see the point i in time series X has been matched with point $i + 2$ in times series Y .

Given two time dependent sequences, $X := (x_1, x_2, \dots, x_N)$ of length N , and $Y := (y_1, y_2, \dots, y_M)$ of length M , sampled at equal distances in time. We need to define a local distance metric d for comparing two points x_i and y_j such that:

$$d(x_i, y_j) \rightarrow R \geq 0 \quad (2)$$

The alignment between time series by DTW is done using the local distance metric d where, the higher the value of d , the greater the dissimilarity between the two points and vice versa. Typically, Manhattan distance is used for local similarity, i.e., the absolute difference between the points:

$$d(x_i, y_j) = |x_i - y_j| \quad (3)$$

Finding the local distance between all pairs of points x_i and y_j produces a distance matrix D , called a cost matrix, where $D(i, j) = d(x_i, y_j)$. The objective of DTW then becomes finding the path of minimum cost, $P = (p_1, p_2, \dots, p_L)$, through the cost matrix D , where $p_l = D(i, j)$. The path P is required to fulfill the following conditions:

1. Boundary condition: the path starts at the bottom left of the cost matrix and ends at the top right.
2. Monotonicity condition: the path can never turn back itself i.e., $p_1 \leq p_2 \leq \dots \leq p_L$
3. Continuity condition: the path can only advance one step at a time i.e. both indices i, j in the matrix D along the path P can only be increased by a single step at a time.

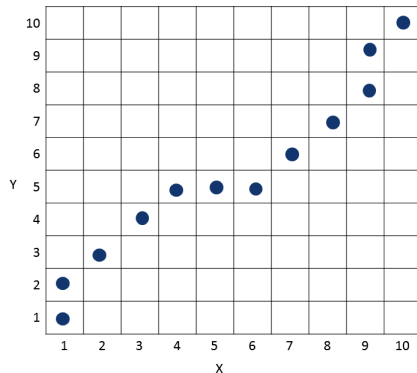


Fig. 3: Example of a valid minimum cost path through a cost matrix

An example of a valid minimum cost path through a cost matrix is illustrated in Figure 3. Thus, an optimal path through the distance matrix D is a valid path P' having minimal total distance d among all possible paths, where d is also the DTW distance between the time series X and Y , i.e.:

$$\text{dist}_{DTW}(X, Y) = d \tag{4}$$

However, DTW is a distance metric not a similarity measure. Thus we convert the DTW distance into a similarity measure (sim_{DTW}) by taking the inverse as shown in Equation 5.

$$\text{sim}_{DTW} = \frac{1}{\text{dist}_{DTW}} \tag{5}$$

5 Experiments

In this section, we discuss preliminary experiments looking at the performance of our system. To this end, we have identified trial/control pairs that our system has judged to be very similar and also, pairs that are considered to have low similarity. The aim, is to further analyse these pairs using visualisations in order to see if our system is matching pairs in a meaningful fashion. Particularly, we are interested in analysing our DTW similarity computation.

We selected a patient pair with high overall similarity according to our system, i.e. 0.94 out of a maximum possible similarity of 1.0. Figure 4a and 4b show a visualisation of the similarity calculation for the BTS attribute of these two patients. Figure 4a shows a three-way plot where the query (trial patient) time series is plotted at the bottom and the control patient time series is plotted on the left. Note that both time series have very similar shape which indicates a small distance between them. Indeed, the DTW distance between these two time series is 1.0. The main part of Figure 4a shows the shape of the alignment path

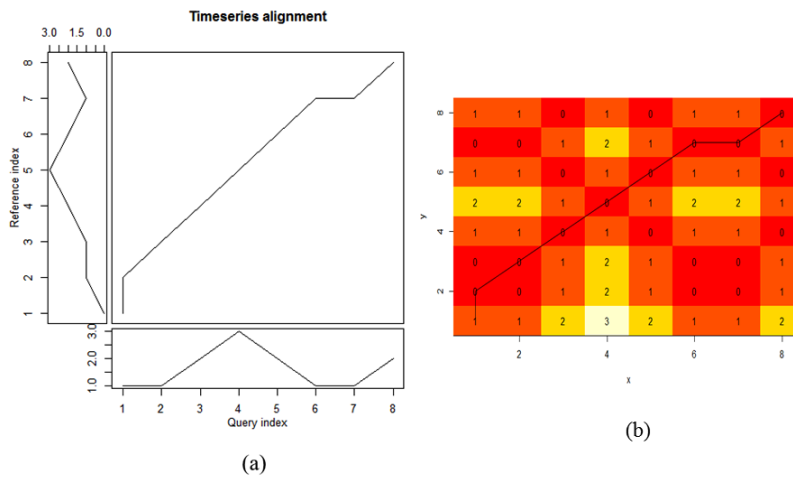


Fig. 4: Visualisation of similarity for BTS Attribute of similar patient pairs.

through the distance matrix. Figure 4b specifically shows the distance matrix and the actual distances in each cell of the matrix. Note that the cells generally contain small distance values because of the similar shape of the two time series. However, the alignment path is traced through cells with 0 distances meaning that the DTW has found the optimal alignment and hence the best similarity between the two time series.

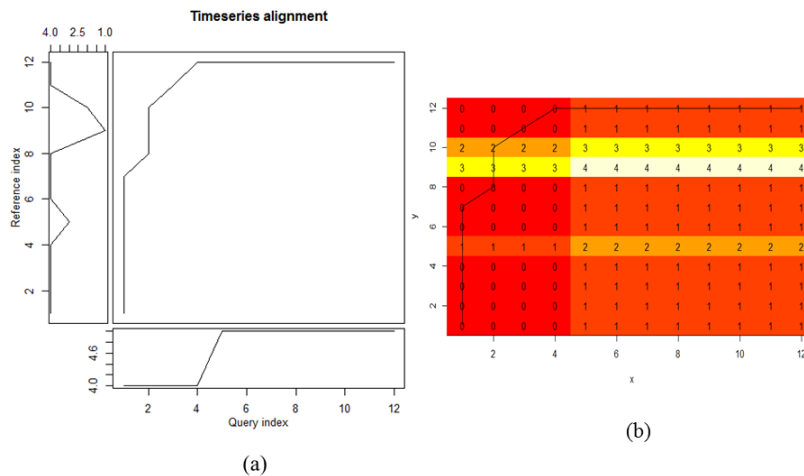


Fig. 5: Visualisation of similarity for BTS Attribute of dissimilar patient pairs.

Figures 5a and 5b show the BTS time series of two dissimilar patients with low overall similarity of 0.5. Observe from Figure 5a that the two time series have very dissimilar shapes. The DTW distance between these two time series is 14.0, which reflects the level of dissimilarity between the two vectors which is also indicated by the difference between the shapes of the two time series. Hence, we can say at this point that the system appears to be effectively measuring similarity between patient pairs as the similarity values produced by the system reflect an intuitive view based on the data. Next, we would like to have a domain expert assign relevance judgments to pairs of patients in order to evaluate our system based on how closely the pairings produced match with the supplied relevance judgments.

6 Challenges

Two main challenges have been identified which will form part of future work for this project. The first is the problem of the time relevance of DTW similarity between two time series. Figure 6 shows three times series, A, B and C, where both B and C are being compared to A for similarity.

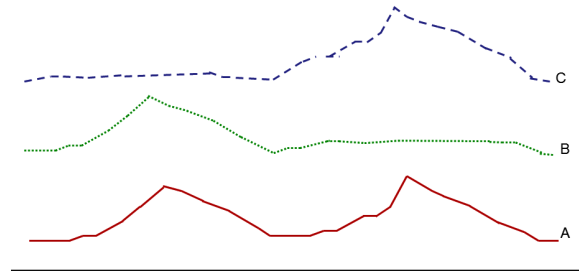


Fig. 6

Observe that both B and C are similar to different parts of A (i.e. B is more similar to the first parts of A and C to the second). With standard DTW, one would be led to believe that B and C are at similar distances from A. However, what if we would like similarity to be biased by recency in which case, we want C to be more similar to A than B because the similarity in pattern between C and A is more recent in time than that between B and A. Relating this to our data, the peaks in the time series could be increases in asthma severity levels where patient B had an increase in their severity level earlier in the year and now appears stable. This can be the sign of a patient responding positively to a particular medication. However, both patients C and A have a more recent increase in their asthma severity level which could mean that they are both reacting negatively to treatment, making A more similar to C than B. Thus, our future work will look to extend DTW in order to bias similarity computation by

recency. Our initial idea is to introduce some type of monotonically decreasing weighting function on the time dimension where alignments that are found earlier on the time dimension are given less importance compared with alignments found later on.

Another challenge we are looking to address in future work is that of optimising the pairing of patients in the trial and control group. Recall that our system pairs a trial patient with the most similar patient from the control group. However, a situation can arise where a control patient is most similar to more than one trial patient. This situation is illustrated in Figure 7 where we have two trial patients P_1 and P_2 and two control patients C_1 and C_2 . The dotted arrows between the patient pairs indicates similarity while the solid lines show actual assignment. Note that C_1 is the most similar control patient to both P_1 and P_2 in the casebase. However, because similarity computation progresses linearly, the similarity between P_2 and C_1 will never get calculated because C_1 would have already been paired off with P_1 even though, the optimal pairing would have been to pair P_1 with C_2 (which has approximately the same similarity as C_1) and then pair P_2 with C_1 . Achieving this may require all pairwise similarities to be computed first, before applying some optimisation algorithm after that for the pairing.

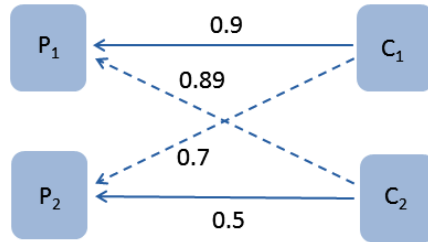


Fig. 7: Example identifying non-optimal pairing

7 Conclusion

In this paper we presented a CBR system for finding patient pairs for use as trial and control patients for Phase 4 clinical studies. Our system works with data which is readily available to the National Health Service (NHS) in Scotland. This means that our system introduces minimal overhead in terms of data collection and preparation. However, the system has the potential to introduce significant cost savings when conducting phase 4 trials by automating the pairing of patients for case-control studies.

Some of the attributes we use for representing a patient case are time dependent. Similarity between patients with respect to such attributes, e.g. the severity

level of the condition of the patient over a period of 12 months, depend more on the general pattern of these severity levels, rather than the conditions of the two patients being equally severe at the same point in time. Hence we introduced a similarity measure based on Dynamic Time Warping (DTW) to cater for any stretching or compression in the time dimension that may otherwise affect similarity computation. Our initial observations of using DTW combined with other typical similarity metrics for pairing patients appears to be very promising.

Finally, we introduced two interesting challenges that will form part of future work. One is the challenge of taking into account recency during DTW computation. This is particularly important in the situation where two different halves of a query time series matches two other different time series. The objective of recency is to give preference to the time series that matches on the more recent half of the time series. This can be perhaps addressed by introducing weights along the time dimension and also exploring ensemble approaches to time windowing. Another challenge is optimising the pairing of patients such that the system always produces a pairing that maximises the sum of pairwise similarities.

References

1. Ahmed, M.U., Begum, S., Olsson, E., Xiong, N., Funk, P.: Case-based reasoning for medical and industrial decision support systems. In: Successful Case-based Reasoning Applications-I, pp. 7–52. Springer (2010)
2. Bellman, R., Kalaba, R.: On adaptive control processes. *Automatic Control, IRE Transactions on* 4(2), 1–9 (1959)
3. Bichindaritz, I., Marling, C.: Case-based reasoning in the health sciences: What’s next? *Artificial intelligence in medicine* 36(2), 127–135 (2006)
4. Goldfarb, N.M.: Phase iv clinical trials: Best practices in post-marketing study management. *Journal Of Clinical Research Best Practices* 7(2) (2012)
5. Hosman, C., Hitinder, G.S.: Using propensity score matching in clinical investigations; a discussion and illustration. *International Journal of Statistics in Medical Research* 4(2), 208–216 (2015)
6. Markatou, M., Don, P.K., Hu, J., Wang, F., Sun, J., Sorrentino, R., Ebadollahi, S.: Case-based reasoning in comparative effectiveness research. *IBM Journal of Research and Development* 56(5), 4–1 (2012)
7. Miotto, R., Weng, C.: Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. *Journal of the American Medical Informatics Association* 22(e1), e141–e150 (2015)
8. Müller, M.: *Information Retrieval for Music and Motion*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
9. Nilsson, M., Sollenborn, M.: Advancements and trends in medical case-based reasoning: An overview of systems and system development. In: FLAIRS Conference. pp. 178–183 (2004)
10. Tsevas, S., Iakovidis, D.K.: Dynamic time warping fusion for the retrieval of similar patient cases represented by multimodal time-series medical data. In: *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*. pp. 1–4. IEEE (2010)