Modeling the Dynamics of Multiple Disease Occurrence by Latent States

Marcos L. P. Bueno[1,2,5(✉)], Arjen Hommersom[1,3], Peter J. F. Lucas[1,4], Mariana Lobo[2], and Pedro P. Rodrigues[2]

[1] Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands
{mbueno,arjenh,peterl}@cs.ru.nl
[2] CINTESIS - Centre for Health Technology and Services Research, Porto, Portugal
{marianalobo,pprodrigues}@med.up.pt
[3] Faculty of Management, Science and Technology, Open University, Heerlen, The Netherlands
[4] Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands
[5] Department of Computer Science, Federal University of Uberlândia, Uberlândia, Brazil

Abstract. The current availability of large volumes of health care data makes it a promising data source to new views on disease interaction. Most of the times, patients have multiple diseases instead of a single one (also known as multimorbidity), but the small size of most clinical research data makes it hard to impossible to investigate this issue. In this paper, we propose a latent-based approach to expand patient evolution in temporal electronic health records, which can be uninformative due to its very general events. We introduce the notion of *clusters of hidden states* allowing for an expanded understanding of the multiple dynamics that underlie events in such data. Clusters are defined as part of hidden Markov models learned from such data, where the number of hidden states is not known beforehand. We evaluate the proposed approach based on a large dataset from Dutch practices of patients that had events on comorbidities related to atherosclerosis. The discovered clusters are further correlated to medical-oriented outcomes in order to show the usefulness of the proposed method.

Keywords: Machine learning · Unsupervised learning
Hidden Markov model · Clustering · Electronic health records
Multimorbidity

1 Introduction

With the availability of large volumes of health care data, additional data sources become available for understanding disease interaction, in particular in multimorbidity research, i.e. when multiple diseases occur in people [2,16,19]. Influenced by factors such as the population aging, multimorbidity is the rule, not the

D. Ciucci et al. (Eds.): SUM 2018, LNAI 11142, pp. 93–107, 2018.
https://doi.org/10.1007/978-3-030-00461-3_7

exception in patients. The small size of most clinical research data makes it hard to impossible to investigate this issue, thus leaving room for alternative data sources that could support, e.g., the discovery of previously unnoticed disease interactions. Recently, machine learning techniques applied to large electronic health records (in the order of billion data points) have been able to provide accurate predictions [18], which shows that it is possible to take advantage of such datasets, despite their great differences compared to traditional clinical data.

In spite of its volume-related advantages, health care data are noisy, incomplete, and mostly not completely research-tailored, making analysis hard. For example, a patient visit to their general practitioner is often represented as a single event containing a main diagnosis, which can be very generic telling no more than which chronic or non-chronic disease was involved. Hence, more specific data such as symptoms and signs are often not available in patient data. As a consequence, a deeper understanding of patient situation beyond such episodic information is very challenging to be obtained, which could potentially help understand how the involved diseases interact. Uncertainty also plays a central role because future events are typically not completely determined by the current patient status. Much research has been dedicated to the analysis of health care data, but it tends to focus on managerial aspects such as patient flow, hospital resources, etc. [3,12] more often than on understanding diseases dynamics [7,14].

In this paper, we hypothesize that latent information next to the diagnostic data can increase our understanding of disease dynamics. By using as a basis hidden Markov models (HMMs) [17], multiple latent states can be associated to a given diagnostic event (where an event could be a visit due to, e.g., type 2 diabetes mellitus or a myocardial infarction). Based on this, we introduce the notion of *clusters of hidden states*, where a cluster contains all the states that produce the same observation (i.e. the same event). Although apparently simplistic, states within a cluster can have quite different dynamics in terms of transitioning patterns (i.e. how a state can be reached by/left from). By looking at these transition patterns, we will be able to give multiple meanings to each event, which sheds light on the influence of such event in the whole care process, as well as on the comorbidity interaction. Besides the structural differences of states within a cluster, we show that these states are associated in different ways to medical outcomes. The identification of latent information has been shown valuable for gaining a better understanding of health care data [6,7], although we pursue a different angle on what to cluster than previous research.

The contributions of this paper are as follows. We first define the notion of clusters of states from the perspective of electronic health records. This is followed by the identification of general transition patterns that might emerge in clusters of hidden states. We then introduce a case study based on data collected from Dutch practices amounting to 32,227 patients that had visits related to atherosclerosis. Atherosclerosis is a medical condition that can be seen as an umbrella to many other diseases, thus it is suitable for illustrating clusters and the role of their states in real-world data. Once an HMM is learned

from the atherosclerosis data, we provide application-oriented interpretation to the clusters of states by looking at a medical outcome (the number of total diseases that were registered in patients) correlated to states of clusters.

2 Multimorbidity Event Data

2.1 Representation

In the considered electronic health records (EHRs), patient visits to their general practitioner are recorded such that each patient visit is assigned a diagnosis code. This diagnosis code can be related, e.g., to a chronic condition (e.g. due to diabetes mellitus) or not (e.g. a fracture). The time interval between two any visits is arbitrary. Next to the diagnosis data, additional data might be available, such as medication prescription and lab exams.

In general, let us consider that there are n possible diagnoses, each one represented by a random variable X_i taking values on $\{0, 1\}$. The full set of diagnosis variables is denoted by $\mathbf{X} = \{X_1, \ldots, X_n\}$. As a patient visit in the considered EHRs will be often assigned a single diagnosis code (sometimes called the main diagnosis), each visit will be represented by an instantiation of $\mathbf{X}$, such that $X_i = 1$ and $X_1 = \cdots = X_{i-i} = X_{i+1} = \ldots = X_n = 0$, where X_i corresponds to the main diagnosis. An alternative representation would be using a single variable taking values on a domain with n values, which could be seen as the state space of a Markov chain. However, we prefer using individual diagnosis variables because it is more general and flexible enough for easily allowing one to add more information into events. This could be the case, for example, when a secondary diagnosis is also available, or when one wants to lump together multiple visits into a single event (e.g. based on a pre-determined time interval or number of visits, if such knowledge is available).

2.2 Modeling

Multimorbidity event data tends to be fine grained, in the sense that each event will likely reflect information limited to the current patient visit. This differs, e.g., from longitudinal clinical trials [21], which are often characterized by repeated measurements of symptoms and signs associated to one or more conditions. As a consequence, data from such clinical trials normally allows for a more complete assessment of patient evolution, as opposed to multimorbidity event data. This suggests that unmeasured patient information could be searched for in such data in the form of latent information, such that when combined with observable data could provide a richer characterization of patients.

Hidden Markov models are often used to capture the interaction between observable and latent variables in a sequential process. In the multimorbidity context, the diagnosis variables $\mathbf{X}$ would be the observable variables, and we assume that there is a latent variable S for representing the hidden or latent states, with domain $\text{val}(S) = \{s_1, \ldots, s_K\}$. The hidden states would, thus, compensate for the mentioned difficulties present in temporal EHRs. In order to fully

specify an HMM over a time horizon $\{0, \ldots, T\}$, we specify a factorization of its joint distribution:

$$P(\mathbf{S}^{(0:T)}, \mathbf{X}^{(0:T)}) = P(S^{(0)}, \mathbf{X}^{(0)}) \prod_{t=0}^{T-1} P(S^{(t+1)} \mid S^{(t)}) P(\mathbf{X}^{(t+1)} \mid S^{(t+1)}) \quad (1)$$

By assuming that the probabilistic interaction between the observables is mediated by the states, we factorize the observables (also known as the emission distribution) as:

$$P(\mathbf{X}^{(t)} \mid S^{(t)}) = \prod_{i=1}^{n} P(X_i^{(t)} \mid S^{(t)}) \quad (2)$$

where $t = 0, \ldots, T$.

3 Identifying Transition Patterns

3.1 Clusters of States

The events constructed from multimorbidity data imply that in order to fully comply with the data concerning n diagnoses, the hidden states should be constrained to emit one out of n different observations at each moment. In spite of this apparent simplicity, the underlying process being modeled could still be quite complex (e.g. by having multiple stages at different moments). In order to properly capture such distribution, more states could be needed, which can lead to the situation where multiple states are associated to the same diagnosis (e.g. if one decides to model more states than observable variables). From these considerations, we define a *cluster of states* as a set of states that have the same emission distribution.

3.2 Transition Patterns

Modeling state transitions in a probabilistic way, e.g. as in Markov chains, implies that a state can often be reached by different ways and can lead to different future states. When clusters of states are considered, such dynamics are further enriched, because such past-present-future transitioning can occur by multiple ways. For example, consider two states s_i and s_j belonging to a cluster C, as shown in Fig. 1. This suggests that s_i will likely be reached earlier for the first time than s_j, and it also suggests that both states can lead to quite different incoming and outgoing states. Of course, such multiple *roles* of a given diagnosis (represented by the cluster C) stem from the complexity of the underlying process, where the a given diagnosis could be associated to different medical outcomes when one looks at the whole care process. For example, the cluster states could be associated to different levels of severity or worsening of patient health that could happen at different moments.

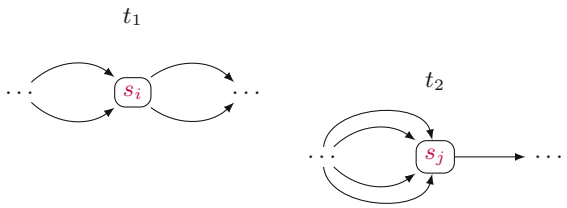


Fig. 1. Cluster of states $C = \{s_i, s_j\}$, where s_i can be reached from two states and transition to two states, while s_j can be reached from four states and can transition to a single state.

In order to better understand the roles of states in clusters, we discuss transition patterns that might arise. This characterization involves states and transitions from and to them, and is provided at a high level, because it is intuitively unfeasible to anticipate all the possible ways by which the states of clusters can interact without having an actual model at hand.

Internal Patterns. A state is associated to an *internal transition pattern* if most of the probability mass of its incoming and outgoing probabilities associates to states from the same cluster. The most trivial internal pattern occurs when a state has a loop probability close to 1, which we call a *recurrent pattern*. A more formal description is that a state s has a recurrent pattern if s has a transition probability $P(s \rightarrow s) \geq \alpha$, where α will typically be close to 1.

A more complex internal pattern would occur when there is a cycle involving two or more states from the same cluster. In this case, at any moment it is very likely that the system (e.g. a patient) is switching between the same diagnosis represented by different states. We call such patterns *internal feedback patterns*.

External Patterns. By taking a closer look at the internal feedback patterns, one would probably conclude that it does not seem to make sense to have different states within a cluster if they switch only among themselves. It would probably suffice to have a single state with a recurrent pattern instead. By opposition, we consider *external feedback patterns*, in which there are states from multiple clusters.

In the context of multimorbidity, these patterns would mean that transitions could involve different diagnoses, as opposed to internal patterns. Hence, if a cluster is involved in both an internal and an external pattern, then the same diagnosis could lead to different multimorbidity future events. In other words, the same diagnosis could play distinct roles.

Example 1. Suppose two clusters of states $C1 = \{s_1, s_2\}$ and $C2 = \{s_3, s_4, s_5\}$, where $C1$ and $C2$ are associated to two different diagnosis codes, as shown in Fig. 2. It holds that state s_1 is involved in a recurrent pattern due to its high

self-transition probability (for $\alpha = 0.95$). States s_4 and s_5 are involved in an internal feedback pattern, while states s_2 and s_3 are involved in an external feedback pattern.

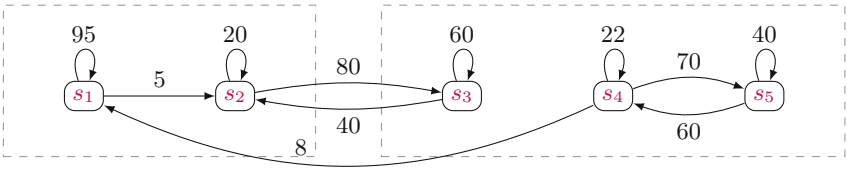


Fig. 2. An example with two clusters of states $C1$ (left) and $C2$ (right) for depicting transition patterns.

4 Case Study

In order to illustrate the value of the proposed methods, we consider the Primary Care Database from the NIVEL institute (Netherlands Institute for Health Services Research), a Dutch institute that maintains routinely electronic health records from health care providers to monitor health in Dutch patients [1]. In the NIVEL data, patient visits are assigned an ICPC code (International Classification of Primary Care) indicating a diagnosis for the visit. Each patient visit is assigned an ICPC code (International Classification of Primary Care), which indicates the diagnosis for the visit.

4.1 Variables and Observations

Atherosclerosis is a cardiovascular condition that has complex associations to a number of other conditions. Although in the literature atherosclerosis has been known to be associated to chronic diseases like diabetes [8], there is still active research on its implications and associations [13,15,20]. In our data preprocessing steps, we first selected ICPC codes related to atherosclerosis, then groups of codes that refer to a given medical symptom or condition were built based on medical experts. As a result, each group of codes gave rise to a variable (i.e. an observable), as shown in Table 1. The variables constructed based on Table 1 can be seen as comorbidities that might occur in patients with atherosclerosis.

In order to construct the event data from the raw NIVEL data, we first ordered the raw data in ascending dates. Then, whenever a patient visit having as diagnosis one of the ICPC codes from Table 1 was found, a new observation was created, where the variable associated to the ICPC code would have a 1 and the remaining variables would have zeros. The visits that were not associated to any of such ICPC codes were ignored.

Table 1. ICPC codes related to atherosclerosis, and their mapping into variables of the model.

ICPC code, description	Variable (model)
K02.00, Pressure/tightness of heart K74.00, Angina pectoris K74.02, Stable angina pectoris K76.01, Coronary sclerosis	*Angina*
K75.00, Acute myocardial infarction K76.02, Previous myocardial infarction (> 4 weeks earlier)	*Myocardial infarction*
K89.00, Transient cerebral ischemia/TIA K90.00, Cerebrovascular accident K90.03, Cerebral infarct	*Cerebrovascular accident*
K92.01, Intermittent claudication	*Claudication*
K99.01, Aortic aneurysm	*Aortic aneurysm*
K91.00, Atherosclerosis	*Atherosclerosis*

4.2 Sample

We considered a sample of 32,227 patients that had visits between 1st of January, 2003 and 31st of December, 2011. To be included, a patient must have had at least one visit related to one of the diagnoses listed in Sect. 4.1. The data construction procedure previously discussed resulted in a dataset with 216,580 observations, where the average number of observations per patient is 6.7 (SD = 10.9), 11,932 patients have only one observation, whereas 20,295 have two or more.

4.3 Number of Hidden States

In order to select an appropriate number of states when learning HMMs, the Akaike Information Criterion (AIC) shown in Eq. 3 was minimized. Models are evaluated by increasing number of states until the addition of states does not improve the score substantially, which is an indication of model overfitting.

$$\mathrm{AIC}(M) = 2\log p - 2\log(\hat{L}) \tag{3}$$

where M is a candidate model, p is the number of parameters of M, and $\hat{L}$ is the log-likelihood of M based on maximum likelihood estimates of the parameters.

During the learning of HMMs, the expectation-maximization algorithm is used, which is quite sensitive to its initial parameters, especially with larger number of states. In order to reduce such effect, the best initial model was selected out of 30 candidates randomly generated.

4.4 Clinical Interpretation of Clusters

If clusters of states are identified in the learned model, one would expect that states within a cluster are indeed necessary, i.e. they should not be replaced by a single state, at the cost of, e.g., worsening model fit. States can be distinguished based on the transition patterns in which they are involved, which provides a dynamics-based description of their differences. Moreover, states of a cluster can be distinguished at a medical level by looking at associations with other data available in patient data. In this case study, we consider as a medical outcome the total number of distinct diagnoses that were registered for each patient (which might include other events than those listed in Sect. 4.1), which provides an approximation to the number of diseases that have occurred in the patient. Such result can indicate medical significance to the cluster states.

Let us consider a state j and a patient i. We first compute the chances that this patient is in such state at some instant t based on the full observations of the patient:

$$\gamma_i^t(j) = P\left(S^{(t)} = j \mid \mathbf{X}^{(0:T_i)}\right) \tag{4}$$

where T_i refers to the last observation of patient i. When the patient has more than one observation, this will result in a sequence of probabilities for a state j. As we will associate the states to the total number of diseases, the average of such probabilities is taken:

$$\overline{\gamma}_i(j) = \frac{1}{T_i + 1} \sum_{t=0}^{T_i} P\left(S^{(t)} = j \mid \mathbf{X}^{(0:T_i)}\right) \tag{5}$$

Once the quantities in Eq. 5 are computed, they are grouped based on the total number of diseases. Then, the average of such quantities is taken per group. Grouping is used to adjust for the near-zero probabilities that might occur in $\overline{\gamma}_i(j)$. As a result, pairs with number of diseases and group averages are obtained, which we use for computing associations (e.g. the Pearson correlation coefficient).

5 Experimental Results

5.1 Model Dimension

Figure 3 shows the model selection scores, which served as a basis for selecting an HMM with 9 states as the suitable model. All the states of the model were associated to fully deterministic emission distributions, such that only one diagnosis variable had a probability equal to 1 in each state, while the other variables had probabilities equal to zero. This means that the property that models should produce events with only one active variable (representing the main diagnosis) was met.

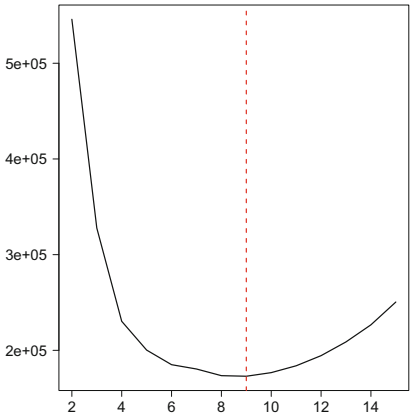


Fig. 3. Model selection scores. X axis: number of hidden states, Y axis: AIC score. The vertical line indicates the number of states where the AIC was minimal.

5.2 Clusters

Figure 4 shows the learned HMM, where each state is named according to the observable that is active (i.e. the observable that has probability equal to 1). Figure 4 shows that three non-unitary clusters were obtained, suggesting that visits associated to angina, myocardial infarction and cerebrovascular accident were suitably represented by 2 states each. Intuitively, it is relevant to model a visit to, e.g., angina by means of 2 different states, hence such a could lead to two different patient courses. As expected, determining which of the two states a visit is associated to depends, e.g., on what is known so far about the patient in terms of past visits.

5.3 Transition Patterns

Figure 4 shows the state transitions of the learned HMM. For each cluster, there is clearly a state that will very likely take a self-transition, which are CVA6, Angina7 and MI3. Thus, such states produce internal patterns as recurrent transition patterns. The HMM shows external patterns as well. In particular, angina seems to be a central event in this model: when moving from the two other clusters, it is more likely that this transition will reach angina (i.e. Angina5). Once in angina, a transition to the other clusters is also possible, with probability larger than 0.05. Hence, such external patterns can be thought of as feedback transition patterns.

5.4 Clinical Interpretation of Clusters

The average probabilities defined in Eq. 5 are summarized in Fig. 5. The histograms suggest that within each cluster there are states that are substantially

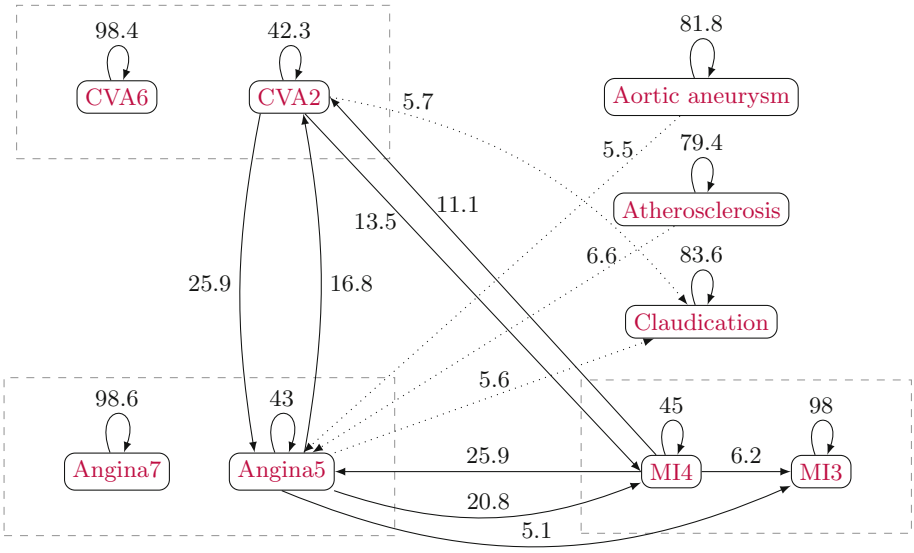
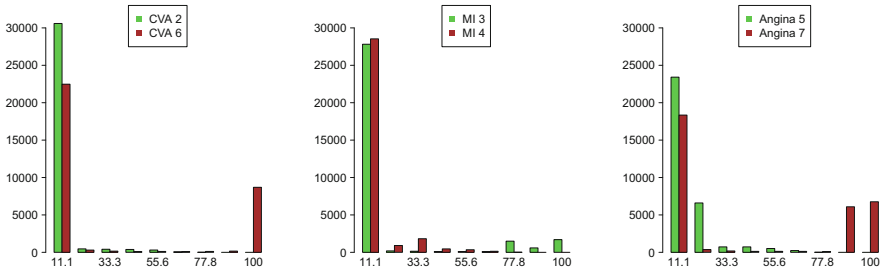


Fig. 4. Clusters of hidden states. Arcs denote state transitions, with labels indicating probability (in %). For the sake of visualization, only transitions with probability greater than or equal to 5% are shown and some transitions were shown dashed.

more prevalent than others, and such separation is more or less uniform depending on the cluster. In general, the recurrent states were usually more likely than their counterparts, which suggests that patients likely had several visits due to the same diagnosis before a diagnosis associated to a different comorbidity was registered.

Figure 6 shows the total number of diseases in patients against the group probabilities. Visual inspection shows that up to 50 diagnoses the trend is substantially more stable than that of all the groups; indeed, around 97% of the patients had at most 50 diagnoses, thus we focus on such groups for obtaining a better understanding of the general trend.

We first note that whenever a patient has a given event (e.g. a CVA), then it holds that the patient will be in one of the states of the corresponding cluster (e.g. either in CVA 2 or CVA 6 states). Figure 6 suggests that, in general, the states of clusters are correlated to the number of diseases in different ways. For the CVA case, patients with only a few diseases are more likely in state CVA 6 (internal patterns) rather than CVA 2 (external patterns). However, as the number of diseases increases, the chances to be in CVA 6 decreases while the chances to be in CVA 2 increases, although such trends occur at different paces. Analogously, for an MI event, it is likely the patient will be in state MI 3 (internal patterns) if the case involves only a few diseases, but a probability decrease is expected for cases involving more diseases. On the other hand, not much can be said about MI 4, as the correlation is very low. Intuitively, one would indeed



(a) **CVA cluster**. (b) **MI cluster**. (c) **Angina cluster**.

Fig. 5. Histograms of average probabilities of states (in %). X axis: average probability of state j in patient i (i.e. $\overline{\gamma}_i(j)$), Y axis: number of patients.

expect that with more diseases patients will transit more between the clusters, which partially explains the trends of the CVA and MI clusters.

As opposed to the previous clusters, Fig. 6 suggests that the Angina cluster has a less straightforward dynamics. In this cluster, both of its states become more prevalent as the number of diseases increases (up to 50), which might suggest the increasing importance of angina that might work as a proxy for the comorbidities considered in this paper, as well as for other chronic and non-chronic events not explicitly considered.

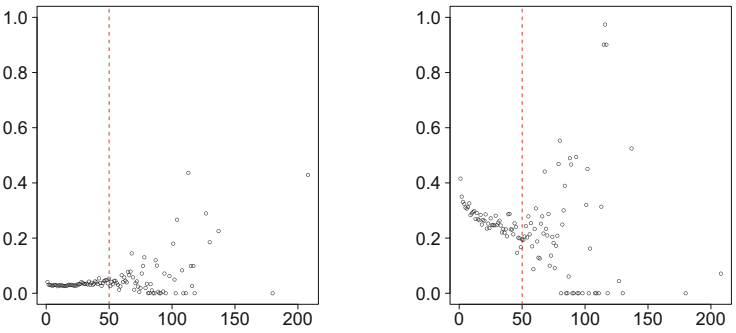
5.5 Are the Clusters Needed? A Comparison to Markov Chains

The need for the clusters learned in the HMM can be assessed by comparing the model fit of the HMM with that of a Markov chain (MC). The state space of such MC is $\mathbf{X}$, i.e., the six comorbidities listed in Sect. 4.1, hence learning this MC amounts to estimating the initial and transition probabilities involving the variables in $\mathbf{X}$. This comparison can illustrate whether the multiple states associated to a given comorbidity (in this paper, the multiple states of CVA, MI and Angina) are indeed necessary for delivering a better model.

Table 2 shows the AIC scores computed for the 9-state HMM and for the MC, which indicates a superior model fit for the HMM. Besides such advantage, with the MC it is no longer possible to identify that the occurrence of a certain event such as angina, can be correlated to different patient characteristics (we used in this paper the total amount of diseases, but other medical outcomes could be devised as well).

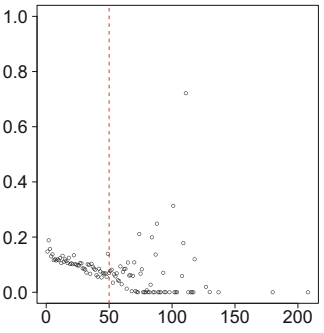
Table 2. AIC scores of the HMM and the Markov chain learned from the multimorbidity data. The smaller the AIC, the better the model fit is.

Model	State clusters	AIC
9-state HMM	3 clusters	172,942.8
Markov chain	No clusters	185,013.5

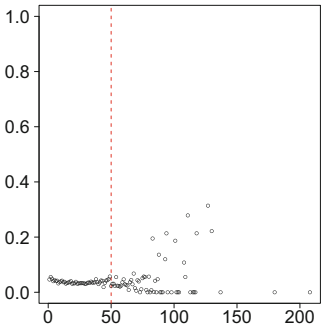


(a) **CVA2**: R($\leq$ 50) = 0.65

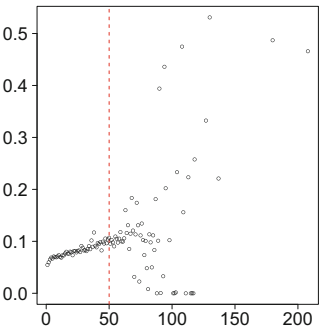
(b) **CVA6**: R($\leq$ 50) = -0.83



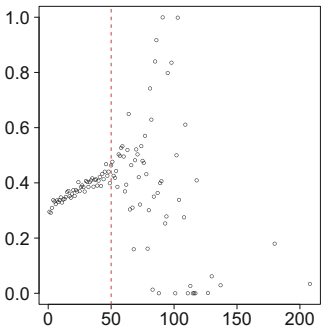
(c) **MI3**: R($\leq$ 50) = -0.79



(d) **MI4**: R($\leq$ 50) = -0.18



(e) **Angina5**: R($\leq$ 50) = 0.91



(f) **Angina7**: R($\leq$ 50) = 0.93

Fig. 6. Association of cluster states to clinical outcome. X axis: number of distinct diagnoses, Y axis: group probability. The vertical line is drawn at $X = 50$. R indicates the Pearson coefficient, calculated considering only the groups with at most 50 diagnoses (which amounts to 97% of all the patients).

6 Related Work

The notion of clustering states in hidden Markov models has not been investigated so far to the best of our knowledge. A related approach is clustering

applied to timed automata [5,22], where state sequences are clustered based on their distance by means of hierarchical clustering methods. Based on Bayesian HMMs that use topic modeling, clustering of patient journeys has been proposed [6], which uses the full set of events associated to unstable angina. By opposition, in our case the clusters are determined based on the states, which shifts the focus towards the dynamics that involve states within clusters. Despite their differences, our methods and those from the literature share the goal of moving towards explainable artificial intelligence [4,11], as we aimed not only to obtain a model with suitable fit, but also to understand more about the patient situation by looking at the structure of the HMM. An example in our case is the deterministic emissions, which can facilitate interpreting models like HMMs to a great extent, at the same time obeying constraints of the multimorbidity problem.

In the context of electronic health records of multimorbidity, a cohort of the NIVEL data used in this paper had been used for learning graphical models based on Bayesian networks, in static [9] and temporal [10] contexts. In those cases, however, the goal was to model differences in practices, hospitals, or regions, without taking into account latent variables.

7 Conclusions

In this paper we proposed a modeling methodology for representing multimorbidity data collected from Dutch practices as single-event observations. Due to the fine-grained nature of such data, we used latent variable-based models, namely HMMs, for extracting additional information that are not directly measured. By using such models, we showed that clusters of states could be discovered, which are states associated to the same observation (or diagnosis, in the case of multimorbidity data), which can, however, be reached from and lead to different transitions. Based on this, we defined the notion of transition patterns.

For the experiments, we considered data concerning variables that are associated to atherosclerosis. The learned model had 9 states, in which clusters involving angina, myocardial infarction and cerebrovascular accident were identified. This suggests that these diagnoses are too complex to be managed by a single latent state, hence a model with better fit is obtained when such diagnoses are allowed to be each represented by multiple states (or roles), as we did with the used HMMs.

Suggestions for future work include investigating the effect of modeling medication and lab exams, which are available to some patients in the NIVEL data. These could be added into the model as inputs (i.e. covariates), for example, which would allow to capture switching regimes for the transitions. Regarding the methodological aspect, it might be worth to extend the learning algorithm to impose deterministic model emissions. Although we were able to obtain a model that satisfied this problem requirement, it might not be always the case that unconstrained learning would be suitable, as there might not be enough data for learning enough states. As a consequence, non-deterministic emissions could be

obtained, which although could lead to better model fit than their deterministic counterpart, would violate the problem property that a single event should be produced at each moment.

Acknowledgments. This work has been partially funded by NWO (Netherlands Organisation for Scientific Research), project Careful (62001863), by FAPEMIG, and by NORTE 2020 (project NanoSTIMA). Project "NORTE-01-0145-FEDER-000016" (NanoSTIMA) is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

1. NIVEL Primary Care Database. https://www.nivel.nl/en/dossier/nivel-primary-care-database. Accessed 30 Apr 2018
2. Barnett, K., Mercer, S., Norbury, M., Watt, G., Wyke, S., Guthrie, B.: Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet **380**, 37–43 (2012). https://doi.org/10.1016/S0140-6736(12)60240-2
3. Côté, M.J., Stein, W.E.: A stochastic model for a visit to the doctors office. Math. Comput. Model. **45**(3), 309–323 (2007). https://doi.org/10.1016/j.mcm.2006.03.022
4. Gunning, D.: Explainable Artificial Intelligence (XAI) (2016). http://www.darpa.mil/program/explainable-artificial-intelligence
5. Hammerschmidt, C.A., Verwer, S., Lin, Q., State, R.: Interpreting finite automata for sequential data. In: Interpretable Machine Learning for Complex Systems: NIPS 2016 Workshop Proceedings (2016)
6. Huang, Z., Dong, W., Wang, F., Duan, H.: Medical inpatient journey modeling and clustering: a Bayesian hidden Markov model based approach. In: AMIA Annual Symposium Proceedings, pp. 649– 658 (2015)
7. Huang, Z., Dong, W., Bath, P., Ji, L., Duan, H.: On mining latent treatment patterns from electronic medical records. Data Min. Knowl. Discov. **29**(4), 914–949 (2015). https://doi.org/10.1007/s10618-014-0381-y
8. Hyvärinen, M., et al.: The impact of diabetes on coronary heart disease differs from that on ischaemic stroke with regard to the gender. Cardiovasc. Diabetol. **8**(1), 17 (2009). https://doi.org/10.1186/1475-2840-8-17
9. Lappenschaar, M., Hommersom, A., Lucas, P.J., Lagro, J., Visscher, S.: Multilevel Bayesian networks for the analysis of hierarchical health care data. Artif. Intell. Med. **57**(3), 171–183 (2013). https://doi.org/10.1016/j.artmed.2012.12.007
10. Lappenschaar, M., et al.: Multilevel temporal Bayesian networks can model longitudinal change in multimorbidity. J. Clin. Epidemiol. **66**(12), 1405–1416 (2013). https://doi.org/10.1016/j.jclinepi.2013.06.018
11. Liu, S., Wang, X., Liu, M., Zhu, J.: Towards better analysis of machine learning models: a visual analytics perspective. Vis. Inform. **1**(1), 48–56 (2017). https://doi.org/10.1016/j.visinf.2017.01.006
12. Marshall, A., Vasilakis, C., El-Darzi, E.: Length of stay-based patient flow models: recent developments and future directions. Health Care Manag. Sci. **8**(3), 213–220 (2005). https://doi.org/10.1007/s10729-005-2012-z

13. Mytton, O.T., et al.: Association between intake of less-healthy foods defined by the United Kingdom's nutrient profile model and cardiovascular disease: a population-based cohort study. PLOS Med. **15**(1), 1–17 (2018). https://doi.org/10.1371/journal.pmed.1002484
14. Najjar, A., Reinharz, D., Girouard, C., Gagn, C.: A two-step approach for mining patient treatment pathways in administrative healthcare databases. Artif. Intell. Med. **87**, 34–48 (2018). https://doi.org/10.1016/j.artmed.2018.03.004
15. O'Donovan, G., Lee, I., Hamer, M., Stamatakis, E.: Association of "weekend warrior" and other leisure time physical activity patterns with risks for all-cause, cardiovascular disease, and cancer mortality. JAMA Internal Med. **177**(3), 335–342 (2017). https://doi.org/10.1001/jamainternmed.2016.8014
16. Prados-Torres, A., et al.: Multimorbidity patterns in primary care: interactions among chronic diseases using factor analysis. PLOS ONE **7**(2), 1–12 (2012). https://doi.org/10.1371/journal.pone.0032190
17. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **72**, 257–286 (1989)
18. Rajkomar, A., et al.: Scalable and accurate deep learning with electronic health records. NPJ Digit. Med. **1**(1), 18 (2018). https://doi.org/10.1038/s41746-018-0029-1
19. Sinnige, J., Korevaar, J.C., Westert, G.P., Spreeuwenberg, P., Schellevis, F.G., Braspenning, J.C.: Multimorbidity patterns in a primary care population aged 55 years and over. Family Pract. **32**(5), 505–513 (2015)
20. Stamatakis, E., de Rezende, L.F.M., Rey-López, J.P.: Sedentary behaviour and cardiovascular disease. In: Leitzmann, M., Jochem, C., Schmid, D. (eds.) Sedentary Behaviour Epidemiology, pp. 215–243. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61552-3_9
21. Wijkstra, J., et al.: Treatment of unipolar psychotic depression: a randomized, double-blind study comparing imipramine, venlafaxine, and venlafaxine plus quetiapine. A. Psych. Scandinavica **121**(3), 190–200 (2009)
22. Zhang, Y., Lin, Q., Wang, J., Verwer, S.: Car-following behavior model learning using timed automata. IFAC-PapersOnLine **50**(1), 2353–2358 (2017). 20th IFAC World Congress. https://doi.org/10.1016/j.ifacol.2017.08.423