

A DISCRETE REGULARIZATION METHOD FOR HIDDEN MARKOV MODELS EMBEDDED INTO REPRODUCING KERNEL HILBERT SPACE

Hidden Markov models are a well-known probabilistic graphical model for time series of discrete, partially observable stochastic processes. We consider the method to extend the application of hidden Markov models to non-Gaussian continuous distributions by embedding a priori probability distribution of the state space into reproducing kernel Hilbert space. Corresponding regularization techniques are proposed to reduce the tendency to overfitting and computational complexity of the algorithm, i.e. Nyström subsampling and the general regularization family for inversion of feature and kernel matrices. This method may be applied to various statistical inference and learning problems, including classification, prediction, identification, segmentation, and as an online algorithm it may be used for dynamic data mining and data stream mining. We investigate, both theoretically and empirically, the regularization and approximation bounds of the discrete regularization method. Furthermore, we discuss applications of the method to real-world problems, comparing the approach to several state-of-the-art algorithms.

Keywords: hidden Markov model, data stream mining, reproducing kernel Hilbert space, online algorithm, regularization.

Introduction

Development of proper models for time series of stochastic semi-observable processes is crucial for solving a wide variety of problems in the learning theory. Most of the observed data from the system does not depict the true states but rather noisy variates of them. Moreover, the observed state space is generally only a subset of the true state space, as the sensory equipment of most systems is limited.

Hidden Markov models (HMM) are applied to various learning problems, including prediction, classification, clustering, identification, segmentation, reinforcement learning, pattern recognition, time series change point detection, and as an online algorithms they are widely used for dynamic data stream mining [1; 2]. A basic assumption for HMM is that to obtain a current hidden state we need only a fixed number of preceding hidden states (Markovian property for the transition model), and an observation depends conditionally on its corresponding hidden state (the observation model). Accordingly, HMM has a bunch of disadvantages, among which a large number of unstructured parameters, limitations caused by Markov property for the first order HMMs, and the most critical is that only a small portion of distributions may be represented by HMM due to the assumption of discrete number of hidden states.

A well established concept that extends the ideas of HMMs to continuous domains is the Kalman filter (KF), which assumes the linear system dynamics and represents the state as a Gaussian random variable.

Considering non-linear system dynamics by means of its sequential linearization leads to Extended Kalman filter (EKF), assuming zero mean multivariate Gaussian noises for transition and observation models. As a further step to address complex problems with non-linear models and non-Gaussian noise, the particle filter has been proposed. The particle filter is a technique for implementing recursive Bayesian filter by Monte Carlo sampling representing the posterior density by a set of random particles with associated weights; thereby, estimates are computed based on these samples and weights. Despite its undoubted ability to represent arbitrary densities and deal with non-Gaussian noise, there is a list of disadvantages of the particle filter, among which high computational complexity, difficulties while determining the optimal number of particles, the number of particles increase with the increasing model dimension, a vital role of the proper importance density choice, and the necessity of resampling to avoid a potential risk of degeneracy and the loss of diversity. Various modifications of the particle filter have been proposed; nevertheless, there is still a research challenge to develop an optimal algorithm with a reduced complexity.

In our study we consider a nonparametric HMM that extends traditional HMMs to structured and non-Gaussian continuous distributions by means of embedding HMM into Reproducing Kernel Hilbert Space (RKHS). Much recent progress has been made for Hilbert space embedding in probabilistic distributions and their application to HMM [3–7]. Due to interference and ill-posedness of the inverse problem arising at

learning embedded HMM into RKHS, regularization is required. The proposed training algorithms [3; 4; 6] use L_1 , L_2 and truncated spectral regularization to invert the corresponding kernel matrix. In our research, we consider more general regularization techniques [8], a discrete regularization method [9], specifically, Nyström-type subsampling [10]. Moreover, simultaneous regularization by means of Nyström-type subsampling and improved optimization technique enable us to use this approach for online algorithms.

This paper is organized as follows. In Section 2 we develop and study the basic structure and theoretical background of the method. Section 3 describes the experimental framework used to evaluate the performance, the pro's and con's of the method.

Embedding HMM into RKHS

In the standard type of HMM, there is a hidden random variable $x(t)$, $x(t) \in \{x_1^t, x_2^t, \dots, x_n^t\}$ and random variable $y(t)$ of the corresponding observation at time t . HMM is defined by emission probabilities $P(y(t)|x(t))$, transition probabilities $P(x(t)|x(t-1))$ and initial state probabilities. To train HMM generally Viterbi training or expectation-maximization (EM) algorithm are used. A priori assumptions on the distribution model (i.e. Gaussian mixture) lead to narrowing the suite of considered probability densities. Employment of RKHS embedding for probability distributions allows the generalization of machine learning methods to arbitrary probability densities, not only Gaussian ones, by providing a uniform representation of functions and, consequently, probability densities as elements of a RKHS.

Here we briefly remind method for RKHS embedding for distributions and HMM described in [5; 11; 12].

Definition 1. RKHS is a Hilbert space of functions $f: \Omega \rightarrow \mathbb{R}$ with a scalar product $\langle \cdot, \cdot \rangle$ that is implicitly defined by Mercer kernel function $k: \Omega \times \Omega \rightarrow \mathbb{R}$ as $\langle \varphi(x), \varphi(y) \rangle = k(x, y)$, where $\varphi(x)$ is a feature mapping into space corresponding to the kernel function. According to reproducing property $\forall x \in \mathcal{X}, \forall f \in \mathcal{H} \langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$ we have for any element f from RKHS $f(y) = \sum_{i \in I} \alpha_i k(x_i, y)$, $\alpha_i \in \mathbb{R}$.

Kernel functions have been thoroughly explored since initiative paper [13], and they have been defined on various structured objects, such as strings and graphs, although standard Gaussian radial basis function kernel is widely used as well.

Joint and conditional distributions may be embedded into a RKHS and manipulate the probability densities, by means of the chain, sum and Bayes' rule, entirely in Hilbert space.

Remark 1. Given a set of feature mappings $\Phi = [\varphi(x_1), \dots, \varphi(x_m)]$ any distribution $q(x)$ may be embedded as a linear combination $\hat{\mu}_q = \Phi\beta$, with

weight vector $\beta \in \mathbb{R}^m$. The mean embedding of a distribution can be used to evaluate expectation of any function f in the RKHS, e.g. if $f = \Phi\alpha$, then

$$\begin{aligned} \mathbb{E}_q[f(x)] &= \langle \hat{\mu}_q, f \rangle = \langle \Phi\beta, \Phi\alpha \rangle = \\ &= \beta^\top \Phi^\top \Phi \alpha = \beta^\top K \alpha, \end{aligned}$$

where $K = \Phi^\top \Phi$ is Gramian matrix, $K_{ij} = k(x_i, x_j)$. **Theorem 1** ([5]). Assume $k(x, x')$ is bounded. With probability $1 - \delta$

$$\|\hat{\mu}_q - \mu_q\|_{\mathcal{H}} = O\left(m^{-1/2} \sqrt{-\log \delta}\right).$$

Now we are ready to consider RKHS embedding for HMM.

Definition 2. Assuming RKHS \mathcal{F} with kernel $k(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{F}}$ defined on the observations, and RKHS \mathcal{G} with kernel $l(h, h') = \langle \phi(h), \phi(h') \rangle_{\mathcal{G}}$ defined on the hidden states, *observable operator* $\mathcal{A}_x: \mathcal{G} \rightarrow \mathcal{F}$ is defined as

$$\mathcal{A}_x \phi(h_t) = p(X_t = x | h_t) \mathbb{E}_{H_{t+1} | h_t} [\phi(H_{t+1})].$$

The *observation operator* is defined as a conditional operator $\mathcal{C}_{X_{t+1} | H_{t+1}} = \mathcal{C}_{X_t | H_t}$ that maps distribution function over hidden states embedded into \mathcal{G} to a distribution function over emissions embedded in \mathcal{F} .

Straightforward from Theorem 1 we have

Corollary 2. Assume $k(x, x')$ and $l(x, x')$ are bounded. Then with probability $1 - \delta$

$$\|\hat{\mathcal{C}}_{XY} - \mathcal{C}_{XY}\|_{\mathcal{F} \otimes \mathcal{G}} = O\left(m^{-1/2} \sqrt{-\log \delta}\right).$$

For conditional embedding operator use of regularization is needed. Thus, for Tikhonov regularization and given regularization parameter λ we have

Corollary 3. Assume $k(x, x')$ and $l(x, x')$ are bounded. Then with probability $1 - \delta$

$$\|\hat{\mu}_{Y|x} - \mu_{Y|x}\|_{\mathcal{G}} = O\left(\sqrt{\lambda} + \sqrt{\frac{-\log \delta}{\lambda m}}\right).$$

The appropriate value of regularization parameter λ may be selected by means of classical approaches, such as Morozov's discrepancy principle, or by using Linear Functional Strategy considered in [14]. Moreover, other regularization techniques may be successfully applied for regularization, such as Nyström subsampling [10] or regularization family $\{g_\lambda\}$ [15].

Definition 3 ([15], Definition 2.2). A family $\{g_\lambda\}$ is called a regularization on $[0, a]$ if there are constants $\gamma_{-1}, \gamma_{-1/2}, \gamma_0$ for which

$$\begin{aligned} \sup_{0 < t \leq a} |1 - tg_\lambda(t)| &\leq \gamma_0, \\ \sup_{0 < t \leq a} |g_\lambda(t)| &\leq \frac{\gamma_{-1}}{\lambda}, \\ \sup_{0 < t \leq a} \sqrt{t} |g_\lambda(t)| &\leq \frac{\gamma_{-1/2}}{\sqrt{\lambda}}. \end{aligned}$$

It is clear that by taking $g_\lambda(t) = (t + \lambda)^{-1}$ we get widely used regularized matrix inversion. Note that $g_\lambda(t) = \frac{1}{t}$ for $t \geq \lambda$, and 0 otherwise, corresponds to the regularization by means of spectral cut-off scheme. For details on regularization families and corresponding approximation bounds we refer to [15].

Nyström subsampling is a learning scheme applied in RKHS setting for matrix inversion, where the kernel matrix is replaced with a smaller matrix obtained by column subsampling [16; 17]. Note, that arbitrary regularization family may be applied after Nyström subsampling.

This approximation-preserving reduction allows us to extend Corollary 3 to general regularization scheme and gives us an estimation of the emission probability distribution with an approximation error of order $O(\lambda^{1/2} + (\lambda m)^{-1/2})$.

In order to evaluate transition probability distribution, we use Algorithm 1 and Theorem 1 from [5] extended for a general regularization scheme that gives us bound for

$$\|\mu_{X(t+1)|\{X(1), X(2), \dots, X(t)\}} - \hat{\mu}_{X(t+1)|\{X(1), X(2), \dots, X(t)\}}\|_{\mathcal{F}}$$

of order

$$O(t(\lambda^{1/2} + (\lambda m)^{-1/2})).$$

In order to reduce the computational complexity, for each of kernel matrices used in Algorithm 1 [5], we apply Nyström subsampling, considering instead of $m \times m$ matrix $m' \times m$ for $m' \ll m$ that reduces kernel matrix construction complexity from quadratic to sub-quadratic preserving approximation bounds. Note also that embedding into RKHS reduces training and application of HMM to linear operations for the kernel and feature matrices for the fixed sampling basis, which consequently reduces the computational complexity.

Note that for some regularization methods, such as Tikhonov regularization, Moore-Penrose pseudo-inverse is used. Combining this approach with Nyström subsampling allows to adopt unlabeled samples to the kernel matrix construction, enabling semi-supervised learning for a suitable kernel.

Applications

The need for online denoising and data stream segmentation occurs in various real-life problems. For various health-care problems it becomes vital, i.e. for nocturnal hypoglycemia prevention for diabetic patients wearing either continuous glucose monitoring devices or self-monitoring glucometers [18; 19]. In [12] applications of embedded HMM in RKHS to robot vision, slot car inertial measurement and audio event classification were shown as exceeding previous state-of-the-art solutions, including ordinary HMM as

well. We conducted sets of experiments to evaluate the effectiveness of learning embedded HMMs into RKHS for real-world prediction and filtering tasks.

Map Matching. Widely used applications such as traffic sensing, the routing time prediction and recommendations require a reliable online localization algorithm. Due to various errors of sensors, imprecise measurements, and imperfect maps, state-of-the-art online map-matching algorithms employ HMMs [20]. Most existing approaches use Viterbi algorithm, using various sliding windows to improve performance. In HMM-based map-matching algorithms, candidate paths are sequentially generated and evaluated on the basis of their likelihood. When a new trajectory point is acquired, past hypotheses of the map-matched route are extended to account for the new observation.

One of the advantages of the approach is the corresponding likelihood estimation, which can be considered as a way for uncertainty quantification. In this setting, to meet the requirement of HMM on a limited state space, for each observation point only a fixed number of position candidates are considered. It leads to considering only one candidate on each map-graph edge (as a rule, the closest one to the observation point). Moreover, emission and transition probabilities for HMM are predetermined by the authors, and tuning of corresponding parameters is required. It respectively implies distinctive drawbacks, such as cutting-off the angle at crossroads, extra U-turns, back-and-forth jumps on a road segment, etc.

For our experiments, we need accurate dataset with noised and ground-truth positions. To generate it we used traffic simulator SUMO — Simulation of Urban MObility [21] and OpenStreetMap data [22]. It enables us to generate accurate map ground-truth positions (taking into account road network information from OpenStreetMap), corresponding exact GPS positions, and model other observations. Then we added noise to these accurate observations. Noise was modelled according to our assumptions and evaluated on a known dataset [23]. An online algorithm was set as a sequence of trajectory reconstructions for a sliding window of up to 50 previous observation points (to limit the number of layers in corresponding HMM).

Performance was measured in terms of accuracy (hitting the correct road-segment) and RMSE for the point-to-point correspondence of map-matched and ground-truth positions. As a baseline we used [23] algorithm (with heuristics for probability distribution assumptions). For our algorithm, we used 50 trajectories for training HMM, and then applied it for the following trajectories. For training we applied RKHS with Gaussian kernel with various values of $\sigma \in \{0.5, 1, 5, 10\}$, and applied Linear Functional Strategy [14] while converting kernel matrices. For baseline and proposed algorithm outputs we compared the accuracy (hit rate, value from 0 to 1) and RMSE for

point-to-point correspondences in meters. The results are presented in Table 1.

Table 1. Performance of baseline algorithm (BL) and implemented algorithm (EHMM) in terms of accuracy and RMSE

	Accuracy		RMSE	
	BL	EHMM	BL	EHMM
min	0.01	0.07	3.71	0.07
max	0.73	0.98	19.58	16.94
mean	0.26	0.66	13.14	6.08
std	0.18	0.21	3.88	4.34

We observe a dramatic improvement of performance for the proposed algorithm. Here we have to notice that the baseline's performance suffers because of fixed heuristics applied for determining the emission and transition probability distribution of HMM. Re-trained HMM by means of EM algorithm (Gaussian mixture) shows a better performance, although the implemented algorithm (EHMM) outperforms it as well; apparently it is because of an implicit ensemble of Gaussian kernels in the implemented solution.

Seizure prediction on Electroencephalography signal. The electroencephalography (EEG) is a crucial tool for monitoring brain activity in various clinical applications. The typical EEG data contains a set of signals measured with electrodes placed on the human scalp. The brain state recognition from EEG signals requires specific signal processing and pre-processing, features extraction, and classification tools. For more details on experiment setting we refer to [24].

We evaluate our approach on seizure prediction on Electroencephalography (EEG) signal. As a hidden variable we consider seizure risk, and observation is given by EEG signal and processed cumulative features for a given sliding time-interval.

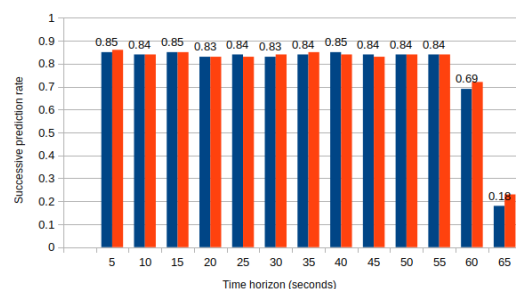


Figure 1. Accuracy of seizure prediction depending on prediction of time-horizon for baseline algorithm (blue) and proposed method (red)

In [24] we applied a ranking algorithm for seizure risk prediction, and achieved the successive prediction rate of approximately 83% for the time horizon up to 1 minute. That algorithm requires a lot of pre-processing

and calibration, therefore it could not be considered as a real-time application. In our current investigation we applied embedding of HMM in RKHS for a corresponding hidden state and observation model. We've achieved the same accuracy for the same time horizon (see Figure 1), whereas performance increased dramatically.

Temporal Audio Segmentation. Embedding of HMM into RKHS with Tikhonov regularization was studied in [3], although without implicit naming. In [3] experimental results were presented both on segmentation of the whole audio track from a TV show and on the speaker diarization within the interview segments. Namely, two soundtracks of the French 1980s entertainment TV-shows ("Le Grand Echiquier") of approximately three hours each, labelled with characteristic, i.e. "applause", "movie", "music", "speech", "speaker turns". After data preprocessing, every 10 ms of audio were matched to a 13-dimensional vector. The experimental results from [3] are presented in Table 2, where the following methods were compared: *regularized kernel Fisher discriminant ratio* (KFDR), which is basically embedded HMM with truncated spectral regularization, *Maximum Mean Discrepancy* (MMD), *Kernel Change Detection* (KCD) algorithms and standard supervised HMM.

Table 2. Best Precision and Recall for benchmarked methods, for both semantic segmentation and speaker segmentation tasks [3]

	Semantic segmentation		Speaker segmentation	
	Precision	Recall	Precision	Recall
KFDR	0.72	0.63	0.89	0.90
MMD	0.71	0.58	0.76	0.73
KCD	0.65	0.63	0.78	0.74
HMM	0.73	0.65	0.93	0.96

The authors mention in [3] that HMM outperforms all the algorithms, but it is explained by a rather unrealistic training procedure, as all speakers and possible labels involved are explicitly modelled beforehand in the speech sections, whereas the proposed method demonstrated competitive performance with a completely unsupervised approach.

Unfortunately, we did not manage to find the mentioned dataset to reproduce the results. Therefore, we applied the same preprocessing technique to dataset Jakobovski/free-spoken-digit-dataset v1.0.7 (Zohar Jackson, César Souza, Jason Flaks, Hereman Nicolas, <https://doi.org/10.5281/zenodo.1136198>). The Free Spoken Digit Dataset consists of 1500 audio records in WAV files at 8kHz of English pronunciations of digits by 3 speakers (50 of each digit per speaker). A corresponding digit is easily labelled, as each file is named in the following format: {digitLabel}_{speakerName}_{index}.wav.

We split the preprocessed dataset into training and testing sets, as it was proposed by its owners, namely 10% (records with indices 0–4 inclusive) of recordings for the training set, and 90% for the testing one (indices 5–49).

In this setting, we compare HMM trained with EM algorithm, embedded HMM, and embedded HMM with regularization by means of Linear Functional Strategy over regularized with Nyström subsampling solutions. Corresponding results are presented in Table 3.

Table 3. Precision and Recall for audio segmentation task

	Speaker segmentation	
	Precision	Recall
HMM	0.82	0.71
EHMM	0.89	0.78
EHMMN	0.88	0.81

Conclusion

We consider a Hilbert space embedding of HMMs as an extension of traditional HMMs to continuous observation distributions. In this setting we apply

more advanced regularization techniques comparing to Tikhonov regularization. Simultaneous regularization by means of Nyström-type subsampling and improved optimization technique enable us to use this approach for online data stream mining. Combining Nyström-type subsampling and Linear Functional Strategy apparently reduce error and its variation, presumably, due to the boosting effect, although a detailed investigation is needed. As further steps of research we consider multi-penalty regularization for multi-dimensional observation models. Note that combining modern kernel methods, regularization techniques, and graphical models significantly improves well-known algorithms, preserving the advantages of each one.

Acknowledgment

The author thanks the International Charitable Foundation for Renaissance of Kyiv-Mohyla Academy for financial support of her research. Apart from that, the author would like to thank Prof. Dr. Sergei Pereverzyev, Johann Radon Institute for Computational and Applied Mathematics (RICAM) of the Austrian Academy of Sciences, for sharing his wisdom and support.

References

- Bargi A. AdOn HDP-HMM: An adaptive online model for segmentation and classification of sequential data / A. Bargi, R. Y. D. Xu, M. Piccardi // IEEE Transactions on Neural Networks and Learning Systems. — 2018. — Vol. PP, no. 99. — P. 1–16.
- Kohlmorgen J. A dynamic HMM for on-line segmentation of sequential data / Jens Kohlmorgen, Steven Lemm // Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. — NIPS'01. — Cambridge, MA, USA : MIT Press, 2001. — P. 793–800.
- A regularized kernel-based approach to unsupervised audio segmentation / Z. Harchaoui, F. Vallet, A. Lung-Yut-Fong, O. Cappe // 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. — [S. l. : s. n.], 2009. — P. 1665–1668.
- Kernel mean embedding of distributions: A review and beyond / Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, Bernhard Schölkopf // Foundations and Trends[®] in Machine Learning. — 2017. — Vol. 10, no. 1-2. — P. 1–141.
- A Hilbert space embedding for distributions / Alex Smola, Arthur Gretton, Le Song, Bernhard Schölkopf // Algorithmic Learning Theory / ed. by Marcus Hutter, Rocco A. Servedio, Eiji Takimoto. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2007. — P. 13–31.
- Hilbert space embeddings of hidden Markov models / Le Song, Byron Boots, Sajid M. Siddiqi [et al.] // Proceedings of the 27th International Conference on International Conference on Machine Learning. — ICML'10. — USA : Omnipress, 2010. — P. 991–998.
- Hilbert space embeddings of conditional distributions with applications to dynamical systems / Le Song, Jonathan Huang, Alex Smola, Kenji Fukumizu // Proceedings of the 26th Annual International Conference on Machine Learning. — ICML 09. — New York, NY, USA : ACM, 2009. — P. 961–968.
- Bauer F. On regularization algorithms in learning theory / Frank Bauer, Sergei Pereverzev, Lorenzo Rosasco // Journal of Complexity. — 2007. — Vol. 23, no. 1. — P. 52–72.
- Nair M. T. A discrete regularization method for ill-posed operator equations / M. T. Nair // The Journal of Analysis. — 2017. — Vol. 25, no. 2. — P. 253–266.
- Kriukova G. Nyström type subsampling analyzed as a regularized projection / Galyna Kriukova, Sergiy Pereverzyev Jr, Pavlo Tkachenko // Inverse Problems. — 2017. — Vol. 33, no. 7. — P. 074001.
- Hilbert space embeddings of conditional distributions with applications to dynamical systems / Le Song, Jonathan Huang, Alex Smola, Kenji Fukumizu // Proceedings of the 26th Annual International Conference on Machine Learning. — ICML '09. — New York, NY, USA : ACM, 2009. — P. 961–968.
- Hilbert space embeddings of hidden Markov models / Le Song, Byron Boots, Sajid M. Siddiqi [et al.] // Proceedings of the 27th International Conference on International Conference on Machine Learning. — ICML'10. — USA : Omnipress, 2010. — P. 991–998.
- Kimeldorf G. Some results on Tchebycheffian spline functions / George Kimeldorf, Grace Wahba // Journal of Mathematical Analysis and Applications. — 1971. — Vol. 33, no. 1. — P. 82–95.
- A linear functional strategy for regularized ranking / Galyna Kriukova, Oleksandra Panasiuk, Sergei V. Pereverzev, Pavlo Tkachenko // Neural Networks. — 2016. — Vol. 73. — P. 26–35.
- Lu S. Regularization theory for ill-posed problems: selected topics / Shuai Lu, Sergei V Pereverzev. — Inverse and Ill-Posed Problems Series. — Berlin : De Gruyter, 2013.
- Smola A. J. Sparse greedy matrix approximation for machine learning / Alex J. Smola, Bernhard Schölkopf // Proceedings of the Seventeenth International Conference on Machine Learning. — ICML '00. — San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2000. — P. 911–918.
- Williams C. Using the Nyström method to speed up kernel machines / Christopher Williams, Matthias Seeger // Advances in Neural Information Processing Systems 13. — [S. l.] : MIT Press, 2001. — P. 682–688.

18. Kriukova G. Application of regularized ranking and collaborative filtering in predictive alarm algorithm for nocturnal hypoglycemia prevention / G. Kriukova, N. Shvai, S. V. Pereverzyev // 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS).— Vol. 2.— [S. l. : s. n.], 2017.— P. 634–638.
19. Real-time decision rules for diabetes therapy management by data stream mining / S. Fong, J. Fiaidhi, S. Mohammed, L. Moutinho // IT Professional.— 2017.— Vol. PP, no. 99.— P. 1.
20. Online map-matching based on hidden Markov model for real-time traffic sensing applications / C. Y. Goh, J. Dauwels, N. Mitrovic [et al.] // 2012 15th International IEEE Conference on Intelligent Transportation Systems.— [S. l. : s. n.], 2012.— P. 776–781.
21. Recent development and applications of SUMO — Simulation of Urban MObility / Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, Laura Bieker // International Journal On Advances in Systems and Measurements.— 2012.— Vol. 5, no. 3&4.— P. 128–138.
22. OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org>. — <https://www.openstreetmap.org>. — 2017.
23. Newson P. Hidden Markov map matching through noise and sparseness / Paul Newson, John Krumm // Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.— GIS '09.— New York, NY, USA : ACM, 2009.— P. 336–343.
24. Distributed system for sampling and analysis of electroencephalograms / O. Sudakov, G. Kriukova, R. Natarov [et al.] // 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS).— Vol. 1.— [S. l. : s. n.], 2017.— P. 306–310.

Крюкова Г. В.

МЕТОД ДИСКРЕТНОЇ РЕГУЛЯРИЗАЦІЇ ДЛЯ ПРИХОВАНИХ МАРКОВСЬКИХ МОДЕЛЕЙ, ЗАНУРЕНИХ У ГІЛЬБЕРТІВ ПРОСТІР ІЗ ВІДТВОРЮЮЧИМ ЯДРОМ

Приховані марковські моделі — добре відомі ймовірнісні графічні моделі для часових рядів дискретних, частково спостережуваних стохастичних процесів. Ми розглядаємо спосіб розширити застосування прихованих марковських моделей до негаусових неперервних розподілів за допомогою занурення апріорного ймовірнісного розподілу простору станів у гільбертів простір із відтворюючим ядром. Відповідні методи регуляризації запропоновано для зменшення схильності до перенавчання та обчислювальної складності алгоритму, наприклад, метод підвибірki Ністрома та узагальнене сімейство регуляризаційних функцій застосовуються під час побудови обернених ядерної та ознакової матриць. Цей метод може бути використаний у різних задачах статистичного виведення, зокрема класифікації, передбачення, ідентифікації, сегментації, а також як онлайн-алгоритм — для динамічної обробки даних та обробки потоку даних. Далі ми наводимо приклад застосування методу до прикладних задач, порівнюємо запропонований підхід із сучасними алгоритмами.

Метою дослідження є розробка методів регуляризації обернених задач, що виникають на стадії навчання ймовірнісних графічних моделей, в яких уявлення про розподіл занурено в гільбертів простір із відтворюючим ядром. Основною методикою реалізації є застосування узагальненого сімейства регуляризаційних функцій та дискретної регуляризації, зокрема метод Ністрома, до відповідних обернених задач обертання матриць ядра та ознак. Задачу вибору відповідних регуляризаційних змінних та параметрів ядра, що визначає гільбертів простір, розв'язано за допомогою методу лінійної функціональної стратегії, тобто ансамблю рішень, побудованих із різними значеннями параметрів. У результаті дослідження отримано теоретичні апроксимаційні оцінки та оцінки складності алгоритму, а також у процесі чисельного експерименту запропонований підхід було порівняно з продуктивністю інших алгоритмів.

Ключові слова: прихована марковська модель, обробка потоку даних, гільбертів простір із відтворюючим ядром, онлайн-алгоритм, регуляризація.

Матеріал надійшов 02.07.2018