Subjectively Interesting Motifs in Time Series
(Extended Abstract)

Junning Deng, Jefrey Lijffijt, Bo Kang, and Tijl De Bie*

Department of Electronics and Information Systems, IDLab, Ghent University

Abstract. This paper introduces an approach to find motifs in time series that are *subjectively interesting*. That is, the aim is to find motifs that are surprising given an informative background distribution, which may for example correspond to the prior knowledge of a user of the tool. We quantify this surprisal using information theory, and more particularly the FORSIED framework. The resulting interestingness function according to which motifs are ranked is then subjective in the statistical sense, enabling us to find subsequence patterns (i.e., motifs and outliers) that are more truly interesting. Although finding the best motif appears intractable, we develop relaxations and a branch-and-bound approach that is implemented in a constraint programming solver. As shown in experiments on synthetic data and two real-world data sets this enables us to mine interesting patterns in small or mid-sized time series.

1 Introduction

Prior work on time series motif detection tends to evaluate a motif's interestingness by assessing its significance against some objectively chosen prior distribution for the time series. In this paper, we introduce an approach to identify motifs that are *subjectively interesting*, in comparison to a prior distribution.

More particularly, given a time series, we are interested in detecting (recurring) subsequences that are highly informative to a user. Here, 'informative' refers to the patterns providing most insight, i.e., patterns strongly contrasting with the user's background knowledge, which are thus surprising and subjectively interesting; while 'recurring' means that we are looking for a set of subsequences that bear significant similarity to each other and they altogether can constitute a *motif*. Alternatively, detecting non-recurring subsequences that are subjectively interesting is also useful, as these subsequences constitute *outliers*.

To achieve such a subjective interestingness measure for motifs and outliers, we define these subsequence patterns as local probabilistic models that we can incorporate into a *background distribution*. The subjective interestingness is then equal to the amount of information (in the formal information theoretic sense)

* This work was supported by the ERC under the EU's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement no. 615517, FWO (project no. G091017N, G0F9816N), and the EU's Horizon 2020 research and innovation programme with the FWO under the MSC Grant Agreement no. 665501.

contained in a local model, *in comparison with the background distribution.* In turn, the background distribution is a maximum entropy model for the data given a set of constraints, such as expected mean, variance, and co-variance between neighbouring points in the time-series. A wide variety of constraints would be possible, which can also contain the prior knowledge of a user, making the information content of the local models dependent on such prior knowledge. Hence, this score indeed quantifies interestingness subjectively.

To find the most informative motifs and outliers efficiently, we develop relaxations with bounds on this relaxed form, and propose a search method implemented in a constraint programming solver. Together with a pruning technique, this enables us to mine subsequence patterns relatively efficiently.

Our specific contributions are:

- Novel definitions of motifs and outliers as probabilistic patterns. [Section 2]
- A quantification of their Subjective Interestingness (SI), based on how much information a user gains when observing this pattern. [Section 2]
- We propose a relaxation of the exact setting and an algorithm to efficiently mine the most informative recurring pattern to a user. [Section 3]
- We empirically evaluate this algorithm on one synthetic data and two real-world data, to investigate its ability to encode the user's prior beliefs and identify informative subsequence patterns. [Section 4]

2 Motif templates and their Subjective Interestingness

Basic definitions. We denote a *time series* as $\hat{\mathbf{X}} \triangleq (\hat{X}_1, \ldots, \hat{X}_n)' \in \mathbb{R}^n$, i.e., an ordered collection of n real numbers $\hat{X}_i \in \mathbb{R}$, where $i \in [n] = [1, \ldots, n]$. We write $\hat{\mathbf{S}}_{i,m}$ for $\hat{\mathbf{X}}$ for the *subsequence* of length $m \leq n$ starting from position i. That is, $\hat{\mathbf{S}}_{i,m} = \hat{X}_i, \ldots, \hat{X}_{i+m-1}$. And we write $\mathbb{T} = \{\hat{\mathbf{T}}_1, \ldots, \hat{\mathbf{T}}_{|\mathbb{T}|}\}$ for a set of subsequences, i.e., $\hat{\mathbf{T}}_i = \hat{\mathbf{S}}_{j,m}$, all of equal length m, of cardinality $|\mathbb{T}|$.

Definition 1 (Motif). A *motif* is a set $\mathbb{T}$ with $|\mathbb{T}| \geq 2$ of 'similar' subsequences.

The criterion by which we judge 'similarity' is explained below shortly below.

Definition 2 (Outlier). An *outlier* is a subsequence $\hat{\mathbf{S}}_{i,m}$ of a time-series for which that time-series does not contain 'similar' non-overlapping subsequences.

Our general aim is to find subjectively interesting 'motifs'. However, what one typically means is not actually a set of subsequences that are similar, but a general subsequence pattern that is re-occurring in a time-series. To avoid working with a set of subsequences, one can use a single exemplar. Here we introduce a probabilistic local model as the target object, the *motif template*:

Definition 3 (Motif template). A *motif template* θ is a description of local model that describes the shape of a motif.

More concretely, we propose a model where we capture the mean and variance statistics of subsequences. We call this template a *mean-variance motif template*:

Definition 4 (Mean-variance motif template). A *mean-variance motif template* θ is a tuple $\theta = (\mu, \sigma)$, where μ is a vector of means and σ a vector of variances, both of cardinality m. The parameters μ, σ are used to define a multivariate Gaussian distribution $\mathcal{N}(\mu, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ is the diagonal matrix with the values of σ as the main diagonal and zero elsewhere.

In this paper, we take μ, σ as the maximum likelihood parameters over a set of subsequences $\mathbb{T}$. We denote the motif template learned from a set of subsequences $\mathbb{T}$ as $\theta_{\mathbb{T}} = (\mu_{\mathbb{T}}, \sigma_{\mathbb{T}})$. Examples are given in Figs 1 and 2.

FORSIED. The FORSIED framework [1,2] stipulates in abstract terms how methods can be derived to mine subjectively interesting patterns. The basic procedure is that a *background distribution* P is defined over the space of all possible data sets Ω, which here would be all possible realizations of a time series $\mathbf{X}$. Since $\mathbf{X} \in \mathbb{R}^n$, the background distribution is a probability density hence we denote it as $p(\mathbf{X})$ instead. It was argued that a good choice for the background distribution is the maximum entropy distribution subject to constraints, and these constraints should be defined to capture the user's prior beliefs.

Subjective Interestingness. Then, given the type of patterns that we want to mine (here motifs), we need to derive the Information Content (IC) of such patterns and define their Description Length (DL). The Subjective Interestingness (SI) of a pattern can then be computed as $\text{SI} = \text{IC}/\text{DL}$ [2].

Description Length. We deem motifs of varying length are beyond the scope of this paper, and as we will see, the IC already brings a balance between the number of instances and the accuracy with which a motif template describes the instances. Hence, the DL can be assumed to be a constant and the SI is proportional to the IC: $\text{SI} \propto \text{IC}$.

Incorporating a motif template into the background distribution. The background distributions p for all prior belief types discussed in this paper are essentially multivariate Gaussian distributions each of which is parametrized by a n-dimensional mean vector μ and a $n \times n$ covariance matrix $\mathbf{\Sigma}$. As mentioned, the motif template can also be described by a multivariate Gaussian distribution, $\mathbf{T} \sim \mathcal{N}(\mu_{\mathbb{T}}, \mathbf{\Sigma}_{\mathbb{T}})$. Once a motif template along with its subsequence instances are identified and showed to the user, the user's belief state changes, and the background distribution needs to be updated. To do this, we simply set the blocks of μ and $\mathbf{\Sigma}$ corresponding to the subsequence instances equal to $\mu_{\mathbb{T}}$ and $\mathbf{\Sigma}_{\mathbb{T}}$, and the off-diagonal elements of $\mathbf{\Sigma}$ corresponding to instances equal to 0.

Information Content. We define the IC of a motif $\mathbb{T}$ as the amount of information gained by it being communicated, which is equivalent to the Kullback-Leibler divergence between the initial background distribution $p(X)$ and the updated background distribution $p_{\mathbb{T}}(X)$:

$$\text{IC}(\mathbb{T}) = -\log p(\mathbb{T}) = \log p_{\mathbb{T}}(\hat{\mathbf{X}}) - \log p(\hat{\mathbf{X}}). \quad (1)$$

Finding the most subjectively interesting motif template. Now we can formalize our goal of finding the most interesting motif template in a time series as an optimization problem with the following objective:

$$\text{Objective 1:}\ \underset{\mathbb{T}}{\operatorname{argmax}} \quad \log p_{\mathbb{T}}(\hat{\mathbf{X}}) - \log p(\hat{\mathbf{X}}).$$

We propose a relaxed version of *Objective 1* which only depends on the probability of subsequences in $\mathbb{T}$. This objective is similar to *Objective 1*, but is more straightforward to optimize efficiently.

$$\text{Objective 2:}\ \underset{\mathbb{T}}{\operatorname{argmax}} \quad \sum_{\hat{\mathbf{T}}\in\mathbb{T}} \log p_{\mathbb{T}}(\hat{\mathbf{T}}) - \sum_{\hat{\mathbf{T}}\in\mathbb{T}} \log p(\hat{\mathbf{T}}).$$

3 How to find the most informative motif template

Let us denote the optimal motif by $\mathbb{T}^*$ and let the set $\mathbb{I} = \{i_1, i_2, \dots, i_c\}$ index the instances of $\mathbb{T}^*$ in $\hat{\mathbf{X}}$, where c is the number of instances in $\mathbb{T}^*$. As all the instances are assumed to be of the same length l, we have $\mathbb{T}^* = \{\hat{\mathbf{S}}_{i_1,l}, \hat{\mathbf{S}}_{i_2,l}, \dots, \hat{\mathbf{S}}_{i_c,l}\}$. The problem of finding all the instances in $\mathbb{T}^*$ is actually to identify all the elements in $\mathbb{I}$. In this work, we adopted a greedy search algorithm to approximately solve this problem. The general idea is to first seed $\mathbb{I}$ by finding a small set of k instances similar wrt to Objective 2 and then greedily grow that set using Objective 1.

The algorithm consists of three major steps:

1. Model the user's prior belief by the original background distribution p_0;
2. Seed by finding a small set of instances which optimizes *Objective 2*;
3. Grow that set by adding an instance which optimizes *Objective 1*, and iterate.

The original background distribution. We wish to define constraints and compute a maximal entropy distribution such that these constraints are preserved in expectation. For the original background distribution p_0, we consider three kinds of constraints. They respectively express the user's prior knowledge about the mean and the variance of each data point, as well as the first order difference in $\mathbf{X}$. Notice these expectation values can be anything, here we equate them to the empirical values. The solution of the maximal entropy distribution with these three constraints is a multivariate Gaussian distribution $p_0(\mathbf{X})$ with a mean vector μ_0 and covariance matrix $\mathbf{\Sigma}_0$. μ_0 and $\mathbf{\Sigma}_0$ can be derived by applying the Lagrange multiplier method. Also, we further improve the computation efficiency by using the property that maximizing the entropy and maximizing the likelihood are the dual of each other in the class of exponential form distributions. The computation details and results are omitted here for brevity.

Finding an initial template set $\mathbb{T}_0$ with k instances. The search starts by finding k non-overlapping optimal instances for $\mathbb{T}_0$. Let $\mathbb{I}_0$ denote index set for $\mathbb{T}_0$. We choose to optimize *Objective 2*, and the problem of finding the most subjectively interesting $\mathbb{T}_0$ can be formulated as

$$\textit{Problem 1} \quad \underset{\mathbb{T}_0}{\operatorname{argmax}} \sum_{\mathbf{T}\in\mathbb{T}_0} \log p_{\mathbb{T}_0}(\mathbf{T}) - \sum_{\mathbf{T}\in\mathbb{T}_0} \log p_0(\mathbf{T}) \equiv$$

$$\underset{\mathbb{I}_0}{\operatorname{argmax}} \sum_{i\in\mathbb{I}_0} \log \mathcal{N}(\hat{\mathbf{S}}_{i,l}|\mu_{\mathbb{T}_0}, \mathbf{\Sigma}_{\mathbb{T}_0}) - \sum_{i\in\mathbb{I}_0} \log \mathcal{N}(\hat{\mathbf{S}}_{i,l}|\mu_0^{(i:i+l-1)}, \mathbf{\Sigma}_0^{(i:i+l-1,i:i+l-1)}) \tag{2}$$

$$\text{where } \mu_{\mathbb{T}_0} = \frac{1}{k}\sum_{i\in\mathbb{I}_0} \hat{\mathbf{S}}_{i,l}, \quad \mathbf{\Sigma}_{\mathbb{T}_0} = \operatorname{diag}(\frac{1}{k-1}\sum_{i\in\mathbb{I}_0}(\hat{\mathbf{S}}_{i,l}^{(1)} - \mu_{\mathbb{T}_0}^{(1)})^2,$$
$$\frac{1}{k-1}\sum_{i\in\mathbb{I}_0}(\hat{\mathbf{S}}_{i,l}^{(2)} - \mu_{\mathbb{T}_0}^{(2)})^2, \ldots, \frac{1}{k-1}\sum_{i\in\mathbb{I}_0}(\hat{\mathbf{S}}_{i,l}^{(l)} - \mu_{\mathbb{T}_0)}^{(l)})^{(2)}))$$

Note the superscript of a symbol is used to denote the corresponding entry in a vector or matrix. Using the expression for the multivariate Gaussian distribution and $\mathbf{\Sigma}_{\mathbb{T}_0}$'s being a diagonal matrix, we can write *Eqs.* (2) as

$$\begin{aligned}
&\sum_{i\in\mathbb{I}_0} \log \mathcal{N}(\hat{\mathbf{S}}_{i,l}|\mu_{\mathbb{T}_0}, \mathbf{\Sigma}_{\mathbb{T}_0}) - \sum_{i\in\mathbb{I}_0} \log \mathcal{N}(\hat{\mathbf{S}}_{i,l}|\mu_0^{(i:i+l-1)}, \mathbf{\Sigma}_0^{(i:i+l-1,i:i+l-1)}) \\
&\quad = -\frac{k*l}{2}\log(2\pi) + \frac{k*l}{2} * \{\log k + \log(k-1)\} \\
&\quad \underbrace{-\frac{k}{2}\sum_{h\in[l]} \log\Big\{\sum_{\substack{i,j\in\mathbb{I}_0\\ i<j}} (\hat{\mathbf{S}}_{i,l}^{(h)} - \hat{\mathbf{S}}_{j,l}^{(h)})^2\Big\}}_{I} - \frac{1}{2}(k-1)*l \\
&\quad \underbrace{-\sum_{i\in\mathbb{I}_0} \log \mathcal{N}(\hat{\mathbf{S}}_{i,l}|\mu_0^{(i:i+l-1)}, \mathbf{\Sigma}_0^{(i:i+l-1,i:i+l-1)})}_{II}
\end{aligned} \tag{3}$$

Note the parts related to the choice of instances in $\mathbb{T}_0$ are underbraced and numbered respectively as I and II. Term I makes the search for optimal instances very computationally demanding, reaching $\mathcal{O}(n^k k^2)$. We thus consider optimizing a relaxed objective of *Problem* 1 based on bounding the term I. Applying Jensen's inequality results in a summation form taken from all the instance pairs. This allows us to convert *Problem* 1 to be the search for a submatrix with maximum sum, mitigating the k^2 factor of the original time complexity. Also, we measure the potential of each candidate and prune the search space so that the search goes through only those highly potential candidates.

Greedily searching for a new instance. The algorithm then continues to search for a new subsequence which optimizes *Objective 1*, until the probability of the time series decreases, i.e. $p_j(\hat{\mathbf{X}}) - p_{j-1}(\hat{\mathbf{X}}) <= 0$. During this step, we also apply a heuristic to prune subsequences which pose little potential, which

Table 1: Run-time to search the initial template set, with pruning factor 99%.

n	l	**Time(s)**	n	l	**Time(s)**	n	l	**Time(s)**
1800	100	9.96	3600	100	50.12	7200	100	369.92
7200	25	328.09	7200	50	350.65	7200	100	369.92

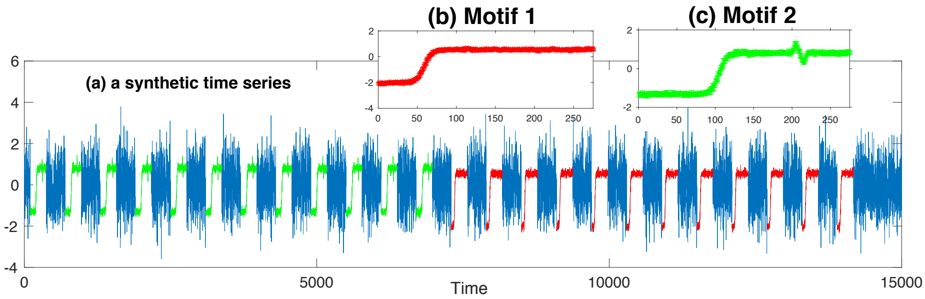

Fig. 1: The algorithm correctly retrieves the two patterns in the synthetic data.

significantly reduces the search space and thus speeds up the computations.

Remark 1. Although the three basic steps described above are for finding a single motif template (i.e. the most informative one to the user), our analysis is not finished with that. A new search for another motif template is started by running step 2 and step 3 again based on an updated background distribution, the one that has already incorporated the user's knowledge of the previous template.

4 Experiments

Data. Synthetic data was generated by embedding two instances from the UCR Trace Data (length 275) of different classes twelve times into a time series of length 15,000, separated by Gaussian noise. The ECG data set is recording #205 in the MIT-BIH Arrhythmia DataBase [7]. We chose a part of 20 seconds (7200 samples) to experiment on that includes normal heartbeats and ventricular tachycardia beats. The Belgian Power Load data set (from the *Open Power System Data* [8]) covers the year 2007 and is resampled at hourly resolution, for a total length of $24 * 365 = 8760$.

Pruning and Scalability. In all experiments we used 99% pruning, as even with such heavy pruning the optimal motif template was still found. Table 1 shows that that the length of the motif template does not influence the computational cost that much, but the influence of the time-series length is more than quadratic.

Synthetic data. We verified whether our method could correctly find the embedded motifs in the synthetic data, the result of which is shown in Fig. 1.

ECG time-series. We analyzed the ECG data using a motif template length of 100, corresponding to a duration of $0.28s$. In this fairly short recording (see Fig. 2a), our algorithm identified three motifs. The first two motifs correspond to normal heartbeats (highlighted with red and green, templates shown in 2b

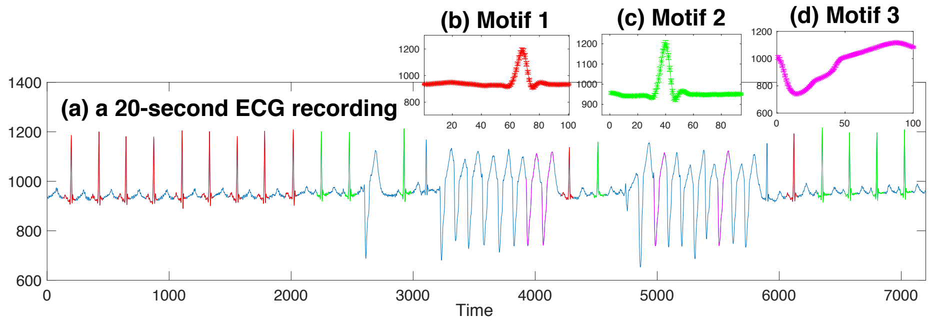

Fig. 2: Three motif templates identified in the 20-second ECG recording.

and 2c). We see their shapes mostly coincide, with a horizontal shift. Another motif identified by the algorithm lies in the area of ventricular tachycardia (pink sections). The instances do not cover all the ventricular tachycardia heart beats, but the small error bars in Fig. 2d indicate that these instances are uncannily similar to each other, and the reason why other ventricular tachycardia subsequences lose the membership for this motif set is their smaller similarity.

Belgium Power Load Data. We analyzed this data searching for motifs of length 24 (one day). The first motif found covers many weekdays, except for Fridays, during winter time. The second motif corresponds to many days between Monday to Thursday as well, but during hot seasons. Interestingly, days in July appear to be different from either. Actually, part of these days (i.e. Monday to Thursday in the last two weeks of July) constitute the third motif. The fourth motif covers many Sundays from middle of April to the beginning of October.

5 Related work

A considerable body of literature has been devoted to techniques to discover recurring patterns in time series. There exist several challenging issues in this pattern discovery problem, including scalability [9,12], the detection of motifs with various lengths [10,6], multi-dimensional time-series [11], and handling distortions [3]. All existing methods are objective, in the sense that they not consider a user's beliefs or expectations and thus operate regardless of context. The novelty of our algorithm is in modeling and using the user's beliefs, and inserting the subjective informativeness into the targeted patterns.

Most related in spirit are algorithms that involve user interaction elements. An efficient method to visually mine patterns in time series is VizTree [4]. This algorithm involves the usage SAX, a symbolic representation of time series which is invariant to time shifting [5]. VizTree allows users to visually evaluate and inspect subsequence patterns. However, VizTree uses a predefined length of the motif to be found. For example, GrammarViz3 [10] allows discovery of variable-length motifs in an interactive setting.

6 Discussion

We propose a new methodology for motif and outlier discovery and a concrete implementation for a specific type of motifs and outliers where the interestingness score can incorporate prior beliefs, and hence they are subjectively interesting. We develop a relaxation of this interestingness score with bounds that can be optimized relatively efficiently using constraint programming. An empirical evaluation demonstrates the potential of the proposed approach.

For future work, it would be useful to develop a motif template that incorporates a form of time warping. Secondly, the length of the subsequences considered is currently a parameter, and could be optimized as well. To make this possible, further speedup techniques should be developed, such as early abandoning.

References

1. De Bie, T.: An information-theoretic framework for data mining. In: Proc. of KDD. pp. 564–572 (2011)
2. De Bie, T.: Subjective interestingness in exploratory data mining. In: Proc. of IDA. pp. 19–31 (2013)
3. Keogh, E., Wei, L., Xi, X., Lee, S., Vlachos, M.: Lb_keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures. In: VLDB. pp. 882–893 (2006)
4. Lin, J., Keogh, E., Lonardi, S., Lankford, J.P., Nystrom, D.M.: Viztree: a tool for visually mining and monitoring massive time series databases. In: International Conference on Very Large Data Bases. pp. 1269–1272 (2004)
5. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing sax: a novel symbolic representation of time series. Data Mining and Knowledge Discovery **15**(2), 107–144 (2007)
6. Linardi, M., Zhu, Y., Palpanas, T., Keogh, E.: Matrix profile X: VALMOD - scalable discovery of variable-length motifs in data series. In: SIGMOD. pp. 1053–1066 (2018)
7. Moody, G.B., Mark, R.G.: The impact of the mit-bih arrhythmia database. IEEE Engineering in Medicine and Biology Magazine **20**(3), 45–50 (2001)
8. Open power system data: Data package time series. version 2018-03-13 (2018)
9. Rakthanmanon, T., Campana, B.J.L., Mueen, A., Batista, G.E.A.P.A., Westover, M.B., Zhu, Q., Zakaria, J., Keogh, E.: Searching and mining trillions of time series subsequences under dynamic time warping. In: The 18th ACM SIGKDD KDD '12. pp. 262–270 (2012)
10. Senin, P., Lin, J., Wang, X., Oates, T., Gandh, S., Boedihardjo, A.P., Chen, C., Frankenstein, S.: Grammarviz 3.0: Interactive discovery of variable-length time series patterns. ACM TKDD **12**(1), 10:1–10:28 (2018)
11. Yeh, C.M., Kavantzas, N., Keogh, E.: Matrix profile VI: meaningful multidimensional motif discovery. In: IEEE ICDM. pp. 565–574 (2017)
12. Yeh, C.M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H.A., Silva, D.F., Mueen, A., Keogh, E.: Matrix profile I: all pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In: IEEE ICDM 2016. pp. 1317–1322 (2016)