Classification of Electroencephalograph Data: A Hubness-aware Approach

Krisztian Buza, Júlia Koller

BioIntelligence Lab, Institute of Genomic Medicine and Rare Disorders, Semmelweis University, Tömő u. 25-29, H-1083 Budapest, Hungary, buza@biointelligence.hu, jkoller@biointelligence.hu

Abstract: Classification of electroencephalograph (EEG) data is the common denominator in various recognition tasks related to EEG signals. Automated recognition systems are especially useful in cases when continuous, long-term EEG is recorded and the resulting data, due to its huge amount, cannot be analyzed by human experts in depth. EEG-related recognition tasks may support medical diagnosis and they are core components of EEG-controlled devices such as web browsers or spelling devices for paralyzed patients. State-of-the-art solutions are based on machine learning. In this paper, we show that EEG datasets contain hubs, i.e., signals that appear as nearest neighbors of surprisingly many signals. This paper is the first to document this observation for EEG datasets. Next, we argue that the presence of hubs has to be taken into account for the classification of EEG signals, therefore, we adapt hubness-aware classifiers to EEG data. Finally, we present the results of our empirical study on a large, publicly available collection of EEG signals and show that hubness-aware classifiers outperform the state-of-the-art time-series classifier.

Keywords: electroencephalograph; nearest neighbor; classification; hubs

1   Introduction

Ongoing large-scale brain research projects – such as the European Human Brain Project, the BRAIN initiative announced by President Obama[1] and the Hungarian National Brain Research Project – are expected to generate an unprecedented amount of data describing brain activity. This is likely to lead to an increased need for enhancement of statistical analysis techniques, development of new methods and computer software that support the analysis of such data.

One of the most wide-spread devices for monitoring and recording the electrical activity of the brain is the electroencephalograph (EEG). EEG is used in clinical practice, research and various other domains. Its numerous applications contain

---

[1] see also https://www.humanbrainproject.eu/ , http://braininitiative.nih.gov/

pre-surgical evaluation [1], diagnostic decision-making [2] and the assessment of chronic headaches [3]. EEG "is an important diagnostic tool for patients with seizures and other paroxysmal behavioral events" [4], it may provide diagnostic information in case of epilepsy [5], Alzheimer's disease [6], [7], schizophrenia [8] or after a brain injury [9]. EEG is used in various brain-computer interfaces [10] which are core components of EEG-controlled devices, such as spelling tools [11] or web browsers [12] for paralyzed patients. EEG was used to study sleepiness in long distance truck driving [13] and there were attempts to predict upcoming emergency braking based on EEG signals [14].

Continuous, long-term EEG monitoring is required in many cases, such as some forms of epilepsy [15], [16], coma, cerebral ischemia, assessment of a medication [17], sleep disorders and disorders of consciousness [18], psychiatric conditions, movement disorders [19], during anesthesia, in intensive care units and neonatal intensive care units [20], [17]. In these cases, EEG signals are recorded for hours or days. Due to the large amount of captured data, the detailed analysis of the entire records is usually not possible by human experts. Therefore, in order to allow for real-time diagnosis and thorough analysis of the data, various techniques were developed to assist medical doctors and other employees of hospitals and to allow for the (semi-)automated analysis of EEG signals.

A common feature of the aforementioned diagnostic problems and EEG-based tools (such as EEG-controlled web browsers or spelling tools) is that they involve recognition tasks related to EEG signals. As EEG signals can be considered as multivariate time-series, these recognition tasks can be formulated as multivariate time-series classification problems, for which state-of-the-art solutions are based on machine learning. For example, Boostani et al. used Boosted Direct Linear Discriminant Analysis for the diagnosis of schizophrenia [21], Sabeti et al. selected best channels based on mutual information and utilized genetic programming in order to select best features [22], while Srinivasan et al. used neural networks for EEG classification [23]. Sun et al. studied ensemble methods [24]. For an excellent survey about EEG-related analysis tasks we refer to [25].

Nearest-neighbor classifiers using dynamic time warping (DTW) as distance measure have been shown to be competitive, if not superior, to many state-of-the-art time-series classification methods such as neural networks or hidden Markov models, see, e.g. [26]. The experimental evidence is underlined by theoretical results about error bounds for the nearest neighbor classifiers. While classic works, such as [27], considered vector data, in their recent work, Chen et al. [28] focused on the nearest neighbor classification of time series and proved error bounds for nearest neighbor-like time-series classifiers. Besides their accuracy, nearest neighbor classifiers deliver human-understandable explanations for their classification decisions in the form of sets of similar instances which makes them preferable to medical applications. As nearest neighbor classifiers are attractive both from the theoretical and practical points of view, considerable research was performed to enhance nearest neighbor classification. Some of the most promising

recent methods were based on the observation that a few time-series tend to be nearest neighbors of surprisingly large amount of other time-series [62]. We refer to this phenomenon as *the presence of hubs* or *hubness* for short, and the classifiers that take this phenomenon into account are called *hubness-aware classifiers*. Hubness-aware classifiers were originally proposed for vector data and image data [29], [30], [31], and only few works considered hubness-aware classification of time series [32], [33], [34], but none of them considered hubness-aware classifiers for EEG data.

In this paper, we focus on hubness-aware classification of EEG signals. As we will show, hubness-aware classifiers lead to statistically significant improvements over the state-of-the-art in terms of accuracy, precision, recall and F-score.

The paper is organized as follows. In Section 2 we introduce basic concepts and notations, while Section 3 is devoted to the presence of hubs in EEG data and hubness-aware classifiers. In Section 4 we present the results of our experiments. Finally, we conclude in Section 5.

2   Basic Concepts and Notations

We use D to denote the set of EEG signals used to construct the recognition model, called *classifier*. D is called *training data* and each signal in D is associated with a class label. For example, in the simplest case of diagnosing epilepsy, there are two classes of signals, one of them contains the EEG signals of healthy individuals, while the second class contains the EEG signals of epileptic patients. The class label of each signal denotes to which class that signal belongs, i.e., in the previous example, the class label of a particular signal denotes whether this signal originates from a healthy or epileptic individual. The class labels of the training data are known while constructing the classifier. The process of constructing the classifier is called *training*. Once the classifier is trained, it can be applied to new signals, i.e., the classifier can be used to predict the class labels of new signals. In order to evaluate our classifier we will use a second set of EEG signals D^{test}, called *test data*. D^{test} is disjoint from D and the class labels of the signals in D^{test} are unknown to the classifier. We only use the class labels of the signals in D^{test} to quantitatively assess the performance of the classifier (by comparing the predicted and true class labels and calculating statistics regarding the performance).

3   Hubness-aware Classification of EEG Data

3.1   Hubs in EEG Data

The presence of hubs, i.e., the presence of a few instances (objects, signals) that occur surprisingly frequently as neighbors (peers) of other instances, while many instances (almost) never occur as neighbors, has been observed for various natural and artificial networks, such as protein-protein-interaction (PPI) networks or the internet [40], [41], [42], [43], [44]. Hubs were shown to be relevant in various contexts, including text mining [45], [46], music retrieval and recommendation [47], [48], [49], [50], image data [51], [52] and time series [34], [53].

In this study, we focus on EEG signals, and we will describe our novel observation that nearest neighbor graphs built from EEG signals contain hubs.

In context of EEG classification, informally, the hubness phenomenon means that some (relatively few) EEG signals appear as nearest neighbors of many EEG signals. Note that, throughout this paper, an EEG signal is *never* treated as the nearest neighbor of itself. Intuitively speaking, very frequent neighbors, or hubs, dominate the neighbor sets and therefore, in the context of similarity-based learning, they represent the centers of influence within the data. In contrast to hubs, there are signals that occur rarely as neighbors and therefore they contribute little to the analytic process. We will refer to them as *orphans* or *anti-hubs*.

In order to express hubness more precisely, for an EEG dataset D one can define the *k-occurrence* of a signal t from D, denoted by $N_k(t)$, as the number of signals in D having t among their *k*-nearest neighbors:

$$N_k(t) = |\{t_i | t \in \mathcal{N}_k(t_i)\}| \tag{1}$$

where $\mathcal{N}_k(t_i)$ denotes the set of *k*-nearest neighbors of t_i. With the term *hubness* we refer to the phenomenon that the distribution of $N_k(t)$ becomes significantly skewed to the right. We can measure this skewness with the third standardized moment of $N_k(t)$:

$$\mathcal{S}_{N_k(t)} = \frac{E[(N_k(t) - \mu_{N_k(t)})^3]}{\sigma_{N_k(t)}^3} \tag{2}$$

where $\mu_{N_k(t)}$ and $\sigma_{N_k(t)}$ are the mean and standard deviation of the distribution of $N_k(t)$ and the notation E stands for the expected value of the quantity between the brackets. When the skewness is higher than zero, the corresponding distribution is skewed to the right and starts presenting a long tail. It should be noted, though, that the occurrence distribution skewness is only one of indicator statistics and that the distributions with same or similar skewness can still take different shapes.

In the presence of class labels, we distinguish between *good hubness* and *bad hubness*: we say that an EEG signal t' is a *good k-nearest neighbor* of the signal t, if (i) t' is one of the k-nearest neighbors of t, and (ii) both have the *same* class labels. Similarly: we say that the signal t' is a *bad k-nearest neighbor* of the signal t, if (i) t' is one of the k-nearest neighbors of t, and (ii) they have *different* class labels. This allows us to define *good (bad) k-occurrence* of a signal t, $GN_k(t)$ (and $BN_k(t)$ respectively), which is the number of other signals that have t as one of their good (bad respectively) k-nearest neighbors. For EEG signals, both distributions of $GN_k(t)$ and $BN_k(t)$ are skewed, as it is exemplified in Fig. 1, which depicts the distributions of $GN_1(t)$, $BN_1(t)$ and $N_1(t)$ for a publicly available EEG dataset from the UCI Machine Learning repository. We describe this dataset in more detail in Section 4. As shown, the distributions have long tails.
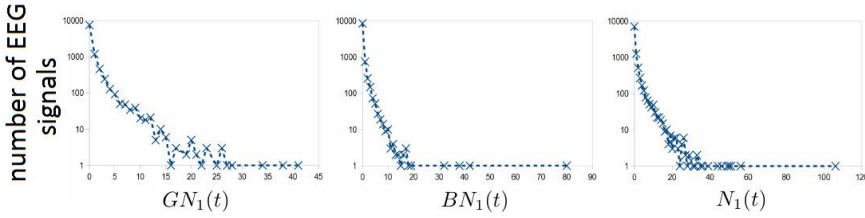

Figure 1

Distribution of $GN_1(t)$, $BN_1(t)$ and $N_1(t)$ for the EEG dataset from the UCI Machine Learning repository. Note that the scale is logarithmic on the vertical axis.

We say that a signal t is a good (or bad) hub, if $GN_k(t)$ (or $BN_k(t)$ respectively) is exceptionally large for t. For the nearest neighbor classification of time series, such as EEG signals, the skewness of good occurrence is of major importance, because some few time series are responsible for large portion of the overall error: *bad hubs* tend to *misclassify* a surprisingly large number of time series [34]. Therefore, one has to take into account the presence of good and bad hubs in EEG datasets.

In the light of the previous discussion, the total occurrence count $N_k(t)$ of an EEG signal t can be decomposed into good and bad occurrence counts: $N_k(t) = GN_k(t) + BN_k(t)$. More generally, we can decompose the total occurrence count into the class-conditional counts:

$$N_k(t) = \sum_{C \in \mathcal{C}} N_{k,C}(t),$$ (3)

where $\mathcal{C}$ denotes the set of all the classes and $N_{k,C}(t)$ denotes how many times t occurs as one of the k-nearest neighbors of signals belonging to class C, i.e.,

$$N_{k,C}(t) = |\{t_i | t \in \mathcal{N}_k(t_i) \ \wedge \ y_i = C\}|,$$ (4)

where y_i denotes the class label of t_i.

3.2   Hubness-aware Classifiers

In this section, we give a detailed description of classifiers that work *under the assumption of hubness*. In our experiments in Section 4, we will examine how these algorithms perform on EEG signals. The algorithms are general, in the sort of sense that they can be applied to any kind of data, provided that an appropriate distance measure between the instances of the dataset is available. In case of EEG-data, we use multivariate DTW as distance measure as described in [36]. As in our case instances are EEG-signals, we will mostly use the term *EEG-signal* instead of *instance* while describing hubness-aware classifiers.

In order to predict how hubs will affect classification of non-labeled signals (e.g. signals arising from observations in the future), we can model the influence of hubs by considering the training data. The training data can be utilized to learn a neighbor occurrence model that can be used to estimate the probability of individual neighbor occurrences for each class. There are many ways to exploit the information contained in the occurrence models. Next, we will review the most prominent approaches. While describing these approaches, we will consider the case of classifying the signal t^*. We will denote its unknown class label as y^* and its nearest neighbors as t_i, where i is an integer number in the range from 1 to k. We assume that the test data is not available when building the model, and therefore $N_k(t)$, $N_{k,C}(t)$, $GN_k(t)$ and $BN_k(t)$ are calculated on the training data.

3.2.1   hw-*k*NN: Hubness-aware Weighting

The weighting scheme proposed by Radovanović et al. [54] is one of the simplest ways to reduce the influence of bad hubs. In this approach, lower voting weights are assigned to bad hubs in the nearest neighbor classifier. In hw-*k*NN, the vote of each neighbor t_i is weighted by $e^{-h_b(t_i)}$, where

$$h_b(t_i) = \frac{BN_k(t_i) - \mu_{BN_k(t)}}{\sigma_{BN_k(t)}}$$

(5)

is the standardized bad hubness score of the neighbor signal t_i in $\mathcal{N}_k(t^*)$, while $\mu_{BN_k(t)}$ and $\sigma_{BN_k(t)}$ are the mean and standard deviation of $BN_k(t)$.

In hw-*k*NN all neighbors vote by their own label. As this may be disadvantageous in some cases [51], in the algorithms considered below, the neighbors do not always vote by their own labels, which is a major difference to hw-*k*NN.

3.2.2   h-FNN: Hubness-based Fuzzy Nearest Neighbors

Consider the relative class hubness $u_C(t_i)$ of each nearest neighbor t_i:

$$u_C(t_i) = \frac{N_{k,C}(t_i)}{N_k(t_i)}$$

(6)

where C denotes one of the classes. The above $u_C(t_i)$ can be interpreted as the fuzziness of the event that t_i occurred as one of the neighbors. Integrating fuzziness as a measure of uncertainty is usual in k-nearest neighbor methods and h-FNN [30] uses the relative class hubness when assigning class-conditional vote weights. The approach is based on the fuzzy k-nearest neighbor voting framework [55]. Therefore, the probability of each class C for the signal $t*$ is estimated as:

$$u_C(t^*) = \frac{\sum_{t_i \in \mathcal{N}_k(t^*)} u_C(t_i)}{\sum_{t_i \in \mathcal{N}_k(t^*)} \sum_{C' \in \mathcal{C}} u_{C'}(t_i)}.$$

(7)

Special care has to be devoted to anti-hubs. Their occurrence fuzziness is estimated as the average fuzziness of points from the same class. Optional distance-based vote weighting is possible.

3.2.3   NHBNN: Naive Hubness Bayesian k-Nearest Neighbor

For each class C, Naive Hubness Bayesian k-Nearest Neighbor (NHBNN) estimates $P(y^* = C \mid \mathcal{N}_k(t^*))$, i.e., the probability that t^* belongs to class C given its nearest neighbors. Then, NHBNN selects the class with highest probability.

NHBNN follows a Bayesian approach to assess $P(y^* = C \mid \mathcal{N}_k(t^*))$. For each training EEG signal t of the training dataset, one can estimate the probability of the event that t appears as one of the k-nearest neighbors of any training instance belonging to class C. This probability is denoted by $P(t \in \mathcal{N}_k | C)$.

Assuming conditional independence between the nearest neighbors given the class, $P(y^* = C \mid \mathcal{N}_k(t^*))$ can be assessed as follows:

$$P(y^* = C | \mathcal{N}_k(t^*)) \quad \propto \quad P(C) \prod_{t_i \in \mathcal{N}_k(t^*)} P(t_i \in \mathcal{N}_k | C)$$

(8)

where $P(C)$ denotes the prior probability of the event that an instance belongs to class C. From the labeled training data, $P(C)$ can be estimated as $|\boldsymbol{D}_C|/|\boldsymbol{D}|$, where $|\boldsymbol{D}_C|$ denotes the number of EEG signals instances belonging to class C in the training data, and $|\boldsymbol{D}|$ is the total number of EEG signals in the training data. The maximum likelihood estimate of $P(t_i \in \mathcal{N}_k | C)$ is the fraction

$$P(t_i \in \mathcal{N}_k | C) \approx \frac{N_{k,C}(t_i)}{|\boldsymbol{D}_C|}.$$

(9)

Estimating $P(t_i \in \mathcal{N}_k | C)$ according to Eq. (9) may simply lead to zero probabilities. In order to avoid it, we can use a simple Laplace-estimate for $P(t_i \in \mathcal{N}_k | C)$ as follows:

$$P(t_i \in \mathcal{N}_k | C) \approx \frac{N_{k,C}(t_i) + m}{|\boldsymbol{D}_C| + mq}$$

(10)

where $m > 0$ and q denotes the number of classes. Informally, this estimate can be interpreted as follows: we consider m additional pseudo-instances from each class and we assume that t_i appears as one of the k-nearest neighbors of the pseudo-instances from class C. We use $m=1$ in our experiments.

Even though k-occurrences are highly correlated, as shown in [61], NHBNN offers improvement over the basic k-NN. This is in accordance with other results from the literature that state that Naive Bayes can deliver good results even in cases with high independence assumption violation [56].

3.2.4   HIKNN: Hubness Information k-Nearest Neighbor

In h-FNN, as in most kNN classifiers, all neighbors are treated as equally important. The difference is sometimes made by introducing the dependency on the distance to t^*, the signal to be classified. However, it is also possible to deduce some sort of global neighbor relevance, based on the occurrence model, which is the basic idea behind HIKNN [29]. It embodies an information-theoretic interpretation of the neighbor occurrence events. In that context, rare occurrences have higher self-information, see Equation (11). The more frequently an EEG signal t occurs as nearest neighbor of other EEG signals, the less surprising is the occurrence of t as one of the nearest neighbors while classifying a new signal.

The EEG signals that rarely occur as neighbors are, therefore, more informative and they are favored by HIKNN. The reasons for this lies hidden in the geometry of high-dimensional feature spaces. Namely, hubs have been shown to lie closer to the cluster centers [57], as most high-dimensional data lies approximately on hyper-spheres. Therefore, hubs are points that are somewhat less 'local'. Therefore, favoring the rarely occurring points helps in consolidating the neighbor set locality. The algorithm itself is a bit more complex, as it not only reduces the vote weights based on the occurrence frequencies, but also modifies the fuzzy vote itself so that the rarely occurring points vote mostly by their labels and the hub points vote mostly by their occurrence profiles. Next, we will present the approach in more detail.

The self-information I_{t_i} associated with the event that t_i occurs as one of the nearest neighbors of an EEG signal to be classified can be calculated as

$$I_{t_i} = \log \frac{1}{P(t_i \in \mathcal{N}_k)} \quad , \qquad P(t_i \in \mathcal{N}_k) \approx \frac{N_k(t_i)}{|\mathcal{D}^{train}|}. \tag{11}$$

Occurrence self-information is used to define the relative and absolute relevance factors in the following way:

$$\alpha(t_i) = \frac{I_{t_i} - \min_{t_j \in \mathcal{N}_k(t_i)} I_{t_j}}{\log |\mathcal{D}^{train}| - \min_{t_j \in \mathcal{N}_k(t_i)} I_{t_j}}, \qquad \beta(t_i) = \frac{I_{t_i}}{\log |\mathcal{D}^{train}|}. \tag{12}$$

The final fuzzy vote of a neighbor t_i combines the information contained in its label with the information contained in its occurrence profile. The relative relevance factor is used for weighting the two information sources. This is shown in Eq. (13).

$$P_k(y^* = C | t_i) \approx \begin{cases} \alpha(t_i) + (1 - \alpha(t_i)) \cdot u_C(t_i), & y_i = C \\ (1 - \alpha(t_i)) \cdot u_C(t_i), & y_i \neq C \end{cases} \tag{13}$$

Hubness-aware classifiers are illustrated by an elaborated example in [61].

3.2.5   On the Computational Aspects of the Implementation of Hubness-aware Classifiers

When classifying EEG signals, i.e., multivariate time series, with hubness-aware classifiers, the computationally most expensive step is the computation of the nearest neighbors of training instances, which is used to determine hubness-scores such as $N_k(t)$, $N_{k,C}(t)$, $GN_k(t)$ and $BN_k(t)$. On the one hand, approaches known to speed-up nearest neighbor classification of time series can be used to reduce the computational costs of hubness-aware classifiers. Such techniques include: speeding-up the calculation of the distance of two time series (by, e.g. limiting the warping window size), indexing and reducing the length of the time series used. For more details we refer to [32] and the references therein. On the other hand, we note that distances between different pairs of training instances can be calculated independently, therefore, computations can be parallelized and implemented on a distributed supercomputer (cloud).

4   Experimental Evaluation

In this section, first, we describe the data we used in our experiments. Next, we provide details of the experimental settings. Subsequently, we present our experimental results.

4.1   Data

In order to evaluate our approach, we used the publicly available EEG dataset[2] from the UCI machine learning repository. This collection contains in total 11028 EEG signals recorded from 122 people. Out of the 122 people, 77 were alcoholic patients and 45 were healthy individuals. Each signal was recorded using 64 electrodes at 256 Hz for 1 second. Therefore, each EEG signal is a 64 dimensional

---

[2] http://archive.ics.uci.edu/ml/datasets/EEG+Database

time series of length 256 in this collection. In order to filter noise, as a simple preprocessing step, we reduced the length of the signals from 256 to 64 by binning with a window size of 4, i.e., we averaged consecutive values of the signal in non-overlapping windows of length 4.

As noted before, the examined EEG dataset exhibits remarkable *hubness*, as the neighbor occurrence frequency is significantly skewed. For instance, if we set $k=1$, there exists a hub signal that acts as a nearest neighbor of 113 other signals from the data. For $k = 10$, the top neighbor occurrence frequency peaks at 707. This illustrates the significance of hub signals in practice. They influence many classification decisions.

As shown in Figure 2, the distribution of such hub signals, as well as detrimental (bad) hubs and anti-hubs differ between the two classes of the EEG dataset and do not follow the prior class distribution. In particular, most hub signals emerge among the Alcoholic class, while most anti-hubs appear among the signals from the Healthy class. Many anti-hubs are in fact known to be outliers and points that lie in borderline regions, far away from local cluster means − and they are, therefore, more difficult to handle and properly associate with a particular class in a prospective study. This suggests that the two classes might not be equally difficult for *k*-NN classification.
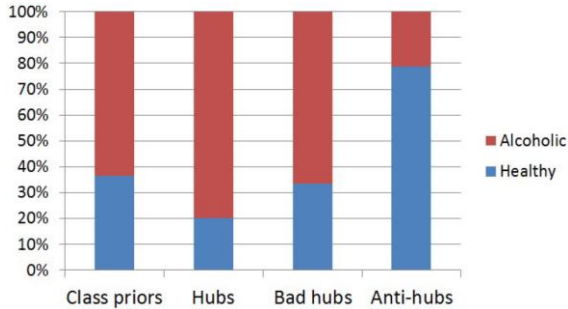

Figure 2
Distribution of hub signals, bad hubs and anti-hubs in the Healthy and Alcoholic class

4.2   Experimental Settings

In our experiments we examined the performance of hubness-aware classifiers. We compared these algorithms to *k*-NN. Both in case of *k*-NN and the hubness-aware classifiers, we used multivariate DTW as distance measure as described in [36]. We set $k = 10$ for the hubness-aware classifiers. This value was chosen, since most hubness-aware methods are known to perform better in cases when k is somewhat larger than 1, because more reliable neighbor occurrence models can be inferred from more occurrence information, see also [29]. In case of the baseline, *k*-NN, we experimented with both $k = 1$ and $k = 10$.

Based on the EEG signals, we aimed to recognize whether a person is affected by alcoholism or not. In other words: the class label of an EEG-signal reflects whether this signal originates from an alcoholic patient or a healthy individual. Both for hubness-aware classifiers and the baseline, we make use of the information that we know which signals originate from the same person: we classify a person as healthy (or alcoholic, respectively) if majority of the signals originating from that person were classified as healthy (alcoholic, respectively).

In all the experiments, we used the *10x10-fold crossvalidation* protocol to evaluate hubness aware classifiers and the baseline. With 10-fold crossvalidation we mean that we partition the entire dataset into 10 disjoint random splits and we use 9 out of these splits as train data, while the remaining split is used as test data. We repeat the experiment 10 times, in each round we use a different split as test data. With 10x10-fold crossvalidation we mean that we repeat the above 10-fold crossvalidation procedure 10 times, each time beginning with a different random partitioning of the data. While partitioning the data, we pay attention that all the signals belonging to the same person are assigned to the same split, and therefore each person either appears in the training data or in the test data, but not in both. On the one hand, this allows to simulate the real-world scenario in which the recognition system is applied to new users; on the other hand, EEG signals are somewhat characteristic to individuals, see e.g. person identification systems using EEG [58], therefore, if the same person would appear in both the train and test data, this could lead to overoptimistic results.

4.3   Performance Metrics

As primary performance measure we used accuracy, i.e., the number of correctly classified persons divided by the number of all the persons in the dataset. We performed t-test at significance level of 0.05 in order to decide whether the differences are statistically significant.

Additionally, we measured precision, recall and F-score for the class of alcoholic patients. Precision and recall regarding class C are defined as $Prec(C) = TP(C) / (TP(C) + FP(C))$ and $Recall(C) = TP(C) / (TP(C) + FN(C))$ respectively, where $TP(C)$ denotes the true positive signals, i.e., signals that are classified as belonging to class C and they really belong to this class; $FP(C)$ denotes false positive signals, i.e., signals that are classified as belonging to class C, but they belong to some other class in reality; and $FN(C)$ denotes false negative, i.e., signals that are *not* classified as belonging to class C, but they belong to class C in reality. F-score is the harmonic mean of precision and recall: $F(C) = 2 \ Prec(C) \ Recall(C) / (Prec(C) + Recall(C))$.

4.4   Results

The results of our experiments are summarized in Tab. 1 and Tab. 2. Tab. 1 shows accuracy of the examined methods averaged over 10x10 folds, while Tab. 2 shows precision, recall and F-score for the identification of alcoholic patients. This experiment simulates the medically relevant application scenario in which EEG is used to diagnose a disease. In both tables, we provide standard deviations after the ± sign. Additionally, in the last two columns of Tab. 1 we provide the results of statistical significance tests (t-test at significance level of 0.05) in the form of a symbol ± where + denotes significance, and − its absence when comparing to 1-NN and 10-NN respectively. In both tables, we underlined those hubness-aware classifiers that outperformed both baselines (in terms of accuracy and F-score).

Table 1

Accuracy ± standard deviation of hubness-aware classifiers and the baselines

Method	Accuracy	Significant difference compared to	
		1-NN	10-NN
1-NN	0.650 ± 0.055		
10-NN	0.662 ± 0.053		
h-FNN	0.690 ± 0.060	+	+
NHBNN	0.780 ± 0.112	+	+
HIKNN	0.663 ± 0.050	+	−
hw-kNN	0.660 ± 0.050	+	−

Table 2

Precision, recall and F-score ± its standard for the class of alcoholic patients

Method	Precision	Recall	F-score
1-NN	0.65 ± 0.04	0.99 ± 0.04	0.78 ± 0.03
10-NN	0.65 ± 0.04	1.00 ± 0.00	0.79 ± 0.03
h-FNN	0.67 ± 0.05	1.00 ± 0.00	0.80 ± 0.03
NHBNN	0.81 ± 0.10	0.87 ± 0.11	0.83 ± 0.09
HIKNN	0.65 ± 0.04	1.00 ± 0.00	0.79 ± 0.03
hw-kNN	0.65 ± 0.04	1.00 ± 0.00	0.79 ± 0.03

4.5   Discussion

Hubness-aware classifiers yield significant overall improvements over k-NN. However, some hubness-aware methods perform better than others.

In particular, all of the hubness-aware classifiers significantly outperformed 1-NN in terms of classification accuracy, whereas two hubness-aware classifiers, namely h-FNN and NHBNN, outperformed 10-NN significantly. Although in terms of accuracy, HIKNN appears to have outperformed 10-NN on average, the difference

is not significant statistically. The simple weighting approach (hw-kNN) did not outperform 10-NN of the examined task of EEG signal classification. The highest improvements in accuracy were achieved by NHBNN, which seems to be very promising for this task.

In medical applications, as we have to deal with class-imbalanced data in many cases, precision and recall are often more important than accuracy. Therefore, in order to further assess the performance of hubness-aware classifiers, we measured their precision, recall and F-score on the class of alcoholic patients. We observed similar trends as in case of accuracy: the performance of the simple hw-kNN was comparable to the baselines, while NHBNN, h-FNN and HIKNN showed clear advantages. Again, NHBNN showed the best overall performance: NHBNN achieved the highest F-score as the relatively low recall of NHBNN was compensated by precision.

In order to interpret these improvements, we have analyzed how different signal types were handled by the tested classifiers. According to [59], we distinguish between four different types of signals: *safe* signals, that lie in class interiors and have all or most of their neighbors belong to the same class, *borderline* signals, that lie in borderline regions between different classes, *rare* signals that are somewhat unusual and distant from the class prototypes and *outliers*. Apart from safe signals, all other signal types are difficult to properly classify.

Figure 3 shows that the two classes in this EEG dataset are formed of different signal type distributions. Most signals of healthy individuals seem to be either borderline, rare or outliers. On the other hand, most signals of alcoholic patients seem to be safe in terms of k-NN classification. This indicates that there is probably a common pattern to most alcoholic EEG signals, while the healthy group might be less coherent and comprise different subgroups.
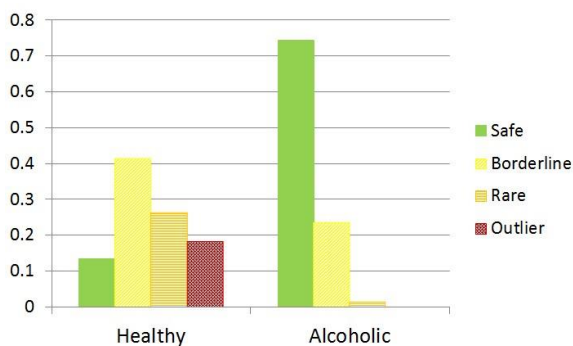

Figure 3

Distribution of different signal types in the Healthy and Alcoholic classes. The two classes have different signal type distributions: compared to the Healthy class, the Alcoholic class seems to be composed of more compact clusters, where most signals lie in class/cluster interiors.

The examined hubness-aware classifiers that improve over 10-NN achieve their improvement by increasing precision of classification for difficult signal types, i.e., borderline, rare and outlier signals, see Fig. 4. This is in concordance with prior observations in other class-imbalanced classification studies [60].

Finally, in order to clarify why hubness-aware classifiers might be well suited for EEG signal classification, we briefly discuss the merits of using neighbor occurrence models on this EEG dataset. Namely, unlike the baseline k-NN, hubness-aware classifiers are based on building neighbor occurrence models that learn from prior occurrences on the training set. Predicting the occurrence profiles of individual points requires us to consider reverse neighbor sets, in contrast to the direct k-NN sets in the k-NN baseline. With reverse neighbors of a signal x, we mean the set of signals that have x as one of their k-nearest neighbors. As Fig. 5 suggests, the average entropy of the reverse neighbor sets in the EEG dataset is lower that the entropy of the direct k-NN sets. This means that less uncertainty is present on average in the reverse neighbor sets.
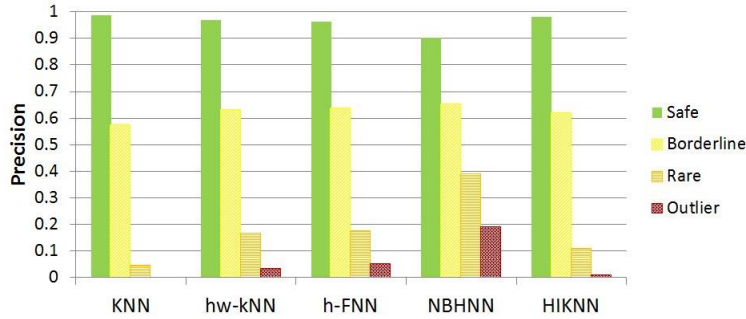

Figure 4

Precision of hubness-aware classifiers and k-NN on different signal types. Performance decomposition indicates clear improvements in case of the difficult signal types (borderline and rare signals, outliers).
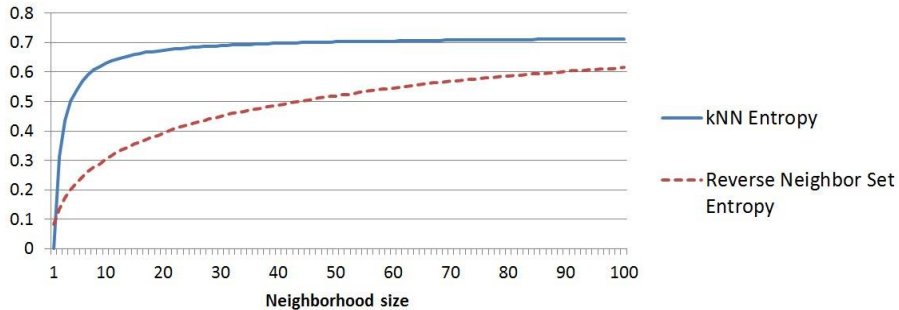

Figure 5

Average entropy (vertical axis) of k-nearest neighbor sets and reverse k-neighbor sets for various neighborhood sizes (horizontal axis). The lower uncertainty of reverse neighbor sets may explain why hubness-aware classifiers outperform k-NN.

Conclusions and Outlook

Classification is a common denominator across biomedical recognition tasks. We examined the effectiveness of hubness-aware classifiers in case of EEG signals.

Hubness-aware classification methods have recently been proposed for classifying complex and intrinsically high-dimensional datasets, under the assumption of *hubness*, which is the skewness of the neighbor occurrence distribution and characterizes many high-dimensional datasets. We have demonstrated that EEG data indeed exhibits significant hubness and that some recently proposed hubness-aware classification methods can be successfully used for signal class recognition.

These recent advances had not been applied to EEG data before and this study attempts to evaluate their usefulness in this context, as well as familiarize domain experts with the potential that these methods seem to hold for these data types.

We have experimentally compared several recently proposed hubness-aware classifiers on a large, publicly available EEG dataset. Our experiments demonstrate significant improvements over the baseline. Naive Hubness-Bayesian *k*-Neareset Neighbor classifier (NHBNN) showed very promising performance. As future work, we will consider different possibilities for boosting hubness-aware methods or combining them into classification ensembles.

Acknowledgement

References

[1]   S. Knake, E. Halgren, H. Shiraishi, K. Hara, H. Hamer, P. Grant, V. Carr, D. Foxe, S. Camposano, E. Busa, T. Witzel, M. Hmlinen, S. Ahlfors, E. Bromfield, P. Black, B. Bourgeois, A. Cole, G. Cosgrove, B. Dworetzky, J. Madsen, P. Larsson, D. Schomer, E. Thiele, A. Dale, B. Rosen, S. Stufflebeam, The Value of Multichannel Meg and Eeg in the Presurgical Evaluation of 70 Epilepsy Patients, Epilepsy Research 69 (2006) pp. 80-86

[2]   J. Askamp, M. J. van Putten, Diagnostic Decision-Making after a First and Recurrent Seizure in Adults, Seizure 22 (2013) pp. 507-511

[3]   U. Kramer, Y. Nevo, M. Y. Neufeld, S. Harel, The Value of EEG in Children with Chronic Headaches, Brain and Development 16 (1994) pp. 304-308

[4]   J. Alving, S. Beniczky, Diagnostic Usefulness and Duration of the Inpatient Long-Term Video-EEG Monitoring: Findings in Patients Extensively Investigated before the Monitoring, Seizure 18 (2009) pp. 470-473

[5]   E. Montalenti, D. Imperiale, A. Rovera, B. Bergamasco, P. Benna, Clinical Features, EEG Findings and Diagnostic Pitfalls in Juvenile Myoclonic Epilepsy: a Series of 63 Patients, Journal of the Neurological Sciences 184 (2001) pp. 65-70

[6]   K. Bennys, G. Rondouin, C. Vergnes, J. Touchon, Diagnostic Value of Quantitative EEG in Alzheimers Disease, Neurophysiologie Clinique/Clinical Neurophysiology 31 (2001) pp. 153-160

[7]   J. Dauwels, F. Vialatte, T. Musha, A. Cichocki, A Comparative Study of Synchrony Measures for the Early Diagnosis of Alzheimer's Disease Based on Eeg, NeuroImage 49 (2010) pp. 668-693

[8]   M. Sabeti, S. Katebi, R. Boostani, Entropy and Complexity Measures for EEG Signal Classification of Schizophrenic and Control Participants, Artificial Intelligence in Medicine 47 (2009) pp. 263-274

[9]   Large-Scale Brain Dynamics in Disorders of Consciousness, Current Opinion in Neurobiology 25 (2014) pp. 7-14

[10]   M. Schreuder, A. Riccio, M. Risetti, S. Dhne, A. Ramsay, J. Williamson, D. Mattia, M. Tangermann, User-centered Design in Braincomputer Interfacesa Case Study, Artificial Intelligence in Medicine 59 (2013) pp. 71-80, Special Issue: Brain-computer interfacing

[11]   N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, H. Flor, A Spelling Device for the Paralysed, Nature 398 (1999) pp. 297-298

[12]   M. Bensch, A. A. Karim, J. Mellinger, T. Hinterberger, M. Tangermann, M. Bogdan, W. Rosenstiel, N. Birbaumer, Nessi: an EEG-controlled Web Browser for Severely Paralyzed Patients, Computational Intelligence and Neuroscience (2007)

[13]   G. Kecklund, T. Åkerstedt, Sleepiness in Long Distance Truck Driving: an Ambulatory EEG Study of Night Driving, Ergonomics 36 (1993) pp. 1007-1017

[14]   S. Haufe, M. S. Treder, M. F. Gugler, M. Sagebaum, G. Curio, B. Blankertz, Eeg Potentials Predict Upcoming Emergency Brakings during Simulated Driving, Journal of neural engineering 8 (2011) 056001

[15]   E. Rodin, T. Constantino, J. Bigelow, Interictal Infraslow Activity in Patients with Epilepsy, Clinical Neurophysiology (2013)

[16]   A. Serafini, G. Rubboli, G. L. Gigli, M. Koutroumanidis, P. Gelisse, Neurophysiology of Juvenile Myoclonic Epilepsy, Epilepsy & Behavior 28, Supplement 1 (2013) S30-S39

[17]   M. L. Scheuer, Continuous EEG Monitoring in the Intensive Care Unit, Epilepsia 43 (2002) pp. 114-127

[18]   U. Malinowska, C. Chatelle, M.-A. Bruno, Q. Noirhomme, S. Laureys, P. J. Durka, Electroencephalographic Profiles for Differentiation of Disorders of Consciousness, Biomedical Engineering online 12 (2013) p. 109

[19]   W. O. Tatum IV, Long-Term EEG Monitoring: a Clinical Approach to Electrophysiology, J. Clinical Neurophysiology 18 (2001) pp. 442-455

[20]   B. McCoy, C. D. Hahn, Continuous EEG Monitoring in the Neonatal Intensive Care Unit, J. Clinical Neurophysiology 30 (2013) pp. 106-114

[21]   R. Boostani, K. Sadatnezhad, M. Sabeti, An Efficient Classifier to Diagnose of Schizophrenia Based on the EEG Signals, Expert Systems with Applications 36 (2009) pp. 6492-6499

[22]   M. Sabeti, S. Katebi, R. Boostani, G. Price, A New Approach for EEG Signal Classification of Schizophrenic and Control Participants, Expert Systems with Applications 38 (2011) pp. 2063-2071

[23]   V. Srinivasan, C. Eswaran, N. Sriraam, Artificial Neural Network-based Epileptic Detection Using Time-Domain and Frequency-Domain Features, Journal of Medical Systems 29 (2005) pp. 647-660

[24]   S. Sun, C. Zhang, D. Zhang, An Experimental Evaluation of Ensemble Methods for EEG Signal Classification, Pattern Recognition Letters 28 (2007) pp. 2157-2163

[25]   D. P. Subha, P. K. Joseph, R. Acharya, C. M. Lim, EEG Signal Analysis: A Survey, Journal of Medical Systems 34 (2010) pp. 195-212

[26]   X. Xi, E. Keogh, C. Shelton, L. Wei, C. A. Ratanamahatana, Fast Time Series Classification using Numerosity Reduction, in: Proceedings of the 23rd International Conference on Machine Learning, ICML '06, ACM, New York, NY, USA (2006) pp. 1033-1040

[27]   L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer Verlag (1996)

[28]   G. H. Chen, S. Nikolov, D. Shah, A Latent Source Model for Nonparametric Time Series Classification, in: Advances in Neural Information Processing Systems 26 (2013) pp. 1088-1096

[29]   N. Tomašev, D. Mladenić, Nearest Neighbor Voting in High Dimensional Data: Learning from Past Occurrences, Computer Science and Information Systems 9 (2012) pp. 691-712

[30]   N. Tomašev, M. Radovanović, D. Mladenić, M. Ivanović, Hubness-based Fuzzy Measures for High-Dimensional k-nearest Neighbor Classification, International Journal of Machine Learning and Cybernetics (2013)

[31]   N. Tomašev, M. Radovanović, D. Mladenić, M. Ivanović, A Probabilistic Approach to Nearest Neighbor Classification: Naive Hubness Bayesian k-nearest Neighbor, in: Proceeding of the CIKM conference

[32]   K. Buza, A. Nanopoulos, L. Schmidt-Thieme, Insight: Efficient and Effective Instance Selection for Time-Series Classification, in: Proceedings of the 15th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining -Volume Part II, PAKDD'11, Springer-Verlag (2011) pp. 149-160

[33]   K. Buza, A. Nanopoulos, L. Schmidt-Thieme, J. Koller, Fast Classification of Electrocardiograph Signals via Instance Selection, in: First International Conference on Healthcare Informatics, Imaging and Systems Biology, IEEE Computer Society, Washington, DC, USA (2011) pp. 9-16

[34]   M. Radovanović, A. Nanopoulos, M. Ivanović, Time-Series Classification in Many Intrinsic Dimensions, in: Proceedings of the 10th SIAM International Conference on Data Mining (SDM) pp. 677-688

[35]   H. Sakoe, S. Chiba, Dynamic Programming Algorithm Optimization for Spoken Word Recognition, Acoustics, Speech and Signal Processing 26 (1978) pp. 43-49

[36]   K. A. Buza, Fusion Methods for Time-Series Classification, Peter Lang Verlag (2011)

[37]   T. F. Smith, M. S. Waterman, Identification of Common Molecular Subsequences, Journal of molecular biology 147 (1981) pp. 195-197

[38]   W. R. Cohen, P. S. Ravikumar, P. S. Fienberg, A Comparison of String Distance Metrics for Name-Matching Tasks, in: Proceedings of the IJCAI-03 Workshop on Information Integration on the Web (2003) pp. 73-78

[39]   V. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions, and Reversals 10 (1966) pp. 707-710

[40]   A. Barabási, Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life, Plume (2003)

[41]   J. B. Axelsen, S. Bernhardsson, M. Rosvall, K. Sneppen, A. Trusina, Degree Landscapes in Scale-Free Networks, Physical Review E-Statistical, Nonlinear and Soft Matter Physics 74 (2006) 036119

[42]   A.-L. Barabási, E. Bonabeau, Scale-Free Networks, Sci. Am. 288 (2003) pp. 50-59

[43]   X. He, J. Zhang, Why Do Hubs Tend to Be Essential in Protein Networks?, PLoS Genet 2 (2006)

[44]   N. N. Batada, L. D. Hurst, M. Tyers, Evolutionary and Physiological Importance of Hub Proteins, PLoS Comput Biol 2 (2006) e88

[45]   A. Nanopoulos, M. Radovanović, M. Ivanović, How does High Dimensionality Affect Collaborative Filtering?, in: Proceedings of the third ACM conference on Recommender systems, RecSys '09, ACM, New York, NY, USA (2009) pp. 293-296

[46]   N. Tomašev, J. Rupnik, D. Mladenić, The Role of Hubs in Cross-Lingual Supervised Document Retrieval, in: Proceedings of the PAKDD Conference, PAKDD (2013)

[47]   J. Aucouturier, F. Pachet, Improving Timbre Similarity: How High is the Sky?, Journal of Negative Results in Speech and Audio Sciences 1 (2004)

[48]   Flexer A., Schnitzer D., Schlüter, J., A Mirex Meta-Analysis of Hubness in Audio Music Similarity, in: Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR'12

[49]   Schedl M., Flexer A., Putting the User in the Center of Music Information Retrieval, in: Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR'12

[50]   D. Schnitzer, A. Flexer, M. Schedl, G. Widmer, Using Mutual Proximity to Improve Content-based Audio Similarity, in: ISMIR'11, pp. 79-84

[51]   N. Tomašev, R. Brehar, D. Mladenić, S. Nedevschi, The Influence of Hubness on Nearest-Neighbor Methods in Object Recognition, in: Proceedings of the 7th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), pp. 367-374

[52]   N. Tomašev, D. Mladenić, Image Hub Explorer: Evaluating Representations and Metrics for Content-based Image Retrieval and Object Recognition, in: Proceedings of the ECML/PKDD Conference, Springer (2013)

[53]   N. Tomašev, D. Mladenić, Exploring the Hubness-related Properties of Oceanographic Sensor Data, in: SiKDD conference (2011)

[54]   M. Radovanović, A. Nanopoulos, M. Ivanović, Nearest Neighbors in High-Dimensional Data: The Emergence and Influence of Hubs, in: Proceedings of the 26rd International Conf. on Machine Learning, ACM, pp. 865-872

[55]   J. E. Keller, M. R. Gray, J. A. Givens, A Fuzzy k-nearest-neighbor Algorithm, in: IEEE Transactions on Systems, Man and Cybernetics, pp. 580-585

[56]   I. Rish, An Empirical Study of the Naive Bayes Classifier, in: Proc. IJCAI Workshop on Empirical Methods in Artificial Intelligence (2001)

[57]   N. Tomašev, M. Radovanović, D. Mladenić, M. Ivanović, The Role of Hubness in Clustering High-Dimensional Data, Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science 6634 (2011) pp. 183-195

[58] M. Poulos, M. Rangoussi, N. Alexandris, A. Evangelou, Person Identification from the Eeg using Nonlinear Signal Classification, Methods of Information in Medicine 41 (2002) pp. 64-75

[59] K. Napierala, J. Stefanowski, Identification of Different Types of Minority Class Examples in Imbalanced Data, in: In Proceedings of Hybrid Artificial Intelligence Systems Conference, Springer Berlin (2012) pp. 139-150

[60] N. Tomašev, D. Mladenić, Class Imbalance and the Curse of Minority Hubs, Knowledge-Based Systems 53 (2013) 157-172

[61] N. Tomašev, K. Buza, K. Marussy, P. B. Kis, Hubness-aware Classification, Instance Selection and Feature Construction: Survey and Extensions to Time-Series, Feature selection for data and pattern recognition (2015) pp. 231-262

[62] N. Tomašev, D. Mladenić, Hub co-occurrence Modeling for Robust High-Dimensional Knn Classification, Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg (2013) pp. 643-659