

ATTI Della Società toscana Di Scienze naturali

MEMORIE • VOLUME CXXV • SUPPLEMENTO • ANNO 2018

TECHNOLOGIES AND INNOVATION FOR SUSTAINABLE MANAGEMENT OF AGRICULTURE, ENVIRONMENT AND BIODIVERSITY





INDICE - CONTENTS

5

pag.

>>

>>

R. BARBUTI, S. CHESSA, R. FRESCO, P. MILAZZO. *Preface*

S. NIN, W.A. PETRUCCI, M. DEL BUBBA, E. GIOR-DANI, Innovative technologies for improved wild strawberry production in marginal Tuscan areas. *Tecnologie innovative per il miglioramento della produzione di fragoline di bosco in aree marginali in Toscana.*

D. PRISA, Italian chabazitic-zeolitite and Effective Microorganisms for the qualitative improvement of olive trees.

Zeolitite-chabasite italiana e microrganismi effettivi per il miglioramento qualitativo degli alberi di olivo.

S. LOMBARDO, D. SARRI, M. VIERI & G. BARAC-CO, Proposal for spaces of agrotechnology co-generation in marginal areas.

Proposta di spazi di co-generazione agrotecnologica in aree marginali.

S. BALDACCI, S. MAIO, A. ANGINO, P. SILVI, M. SIMONI, S. LA GRUTTA, G. VIEGI, F. RUGGIERO, G. BEDINI, F. NATALI, L. CECCHI, U. BERGER, M. PRENTOVIC, I. ANNESI MAESANO, A. MOUSTAFA, M. THIBAUDON, S. MONNIER, S. ORLANDINI, The AIS LIFE project: the Italian enrollment phase results.

Progetto AIS LIFE: risultati della fase di arruolamento in Italia.

F. NATALI, M. NAPOLI, A. DALLA MARTA, G. AR-GENTI, L. CECCHI, S. ORLANDINI, S. BALDACCI, S. MAIO, A. ANGINO, P. SILVI, S. LA GRUTTA, G. VIEGI, F. RUGGIERO, G. BEDINI, U. BERGER, M. PRENTOVIC, I. ANNESI MAESANO, A. MOUSTAFA, M. THIBAUDON, S. MONNIER, AIS LIFE project: pollen allergy risk maps in Tuscany. *Progetto AIS LIFE: mappe di rischio allergico in*

Toscana

		Modelli basati su algebre di processi per l'agricol- tura di precisione sostenibile: il caso di una vigna.		20
*	7	R. BARBUTI, P. BERNI, P. MILAZZO. A mathema- tical model for the study of the impact of small commercial fishing on the biodiversity of artifi- cial reefs.	»	27
		Un modello matematico per lo studio dell'impatto della piccola pesca commerciale sulla biodiversità delle scogliere artificiali.	»	45
>	13	R. FRESCO, G. FERRARI. Enhancing precision agriculture by Internet of Things and cyber physical systems. <i>Miglioramento dell'agricoltura di precisione tra</i> -		53
		mite Internet aeue cose e sistemi cyber-fisici.	*))
>	19	R. PUCCI, A. MICHELI, S. CHESSA. Wild animals' biologging through machine learning models. <i>Biologging di animali selvatici tramite modelli di</i>		61
			>>	01

C. BODEI, P. DEGANO, G.-L. FERRARI, L. GAL-

LETTA. Sustainable precision agriculture from a

nno agos alaghasia na

» 33

» 25

ROBERTO BARBUTI, STEFANO CHESSA, ROBERTO FRESCO, PAOLO MILAZZO

PREFACE

The technological innovation in biology and agriculture often leveraging on innovation in computer science and engineering, pushed forward the process of integration among these disciplines. In particular, information technology (IT) provides common methodologies and tools for the automatic acquisition and analysis of the data that concern the management and optimization of the natural and territorial resources.

In agriculture, applications of IT enable the integration of interventions concerning its sustainability and productivity, by offering methods and tools to monitor, control, analyse and optimize the production while keeping it respectful of the environment. Similarly, the best practices for bio sustainability, for the management of bio-diversity and for the bioremediation of the environment (including soil, water etc...) are also progressively adopting IT, which enable more focused (and thus more effective) applications.

In this context, the conference "Technologies and innovation for sustainable management of Agriculture, Environment and Biodiversity" (TI4AAB), was held in July 2016 at the Natural History Museum of the University of Pisa located in the Calci Charterhouse (Calci, province of Pisa) in order to encourage the sharing of emerging knowledge about the above topics.

In fact, the conference was dedicated to fostering innovative cross-disciplinary research and applications and to stimulating the exchange of strategies and experiences, among academic and company experts from different disciplines (agriculture, biology, computer science and engineering and environmental decision making), in order to encourage a common, interdisciplinary discussion about the adoption and perspectives of IT in modern agriculture, environmental management, biodiversity and bio-sustainability in general.

The conference was held under the auspices of the municipality of Calci, the University of Pisa and of the "Ordine dei Dottori Agronomi e Dottori Forestali". It was also attended and supported by some leading national and worlwide industries, like CAEN RFID, OSRAM, STMicroelectronics, EBV Elektronik, Qprel Srl, AEDIT Srl, EMipiace Srl, and Zefiro Ricerca & Innovazione Srl, and by the Italian National Forestry Authority.

This volume constitutes a selection of the contributions presented at the conference and cover the aspects of innovation in agriculture, biology, and applied information technology. In particular, concerning innovation in agriculture, the paper by Nin et al. studies new soilless cultivation systems for wild strawberry growing in the Tuscan Appennine mountains. The paper by Prisa describes experimental research concerning the use of zeolites in combination with effective microorganisms, in order to improve the quality of olive trees. Finally, the paper by Lombardo et al. describes collaborative approaches to innovation in agriculture (co-generation of technology).

Concerning innovation in biology, the paper by Baldacci et al. describes the results of the preliminary phases of the AIS-LIFE project, which aims at developing aerobiological information systems in order to improve pollen-related allergic respiratory disease management. Still concerning the AIS-LIFE project, the paper by Natali et al. aims to describe the strategy used in AIS-LIFE project, to evaluate daily pollen concentration in the atmosphere produced by many allergic plant species. The use of data and GIS system are shown as an approach to assess allergy risk maps.

Concerning innovation in computer science applied to agriculture and biology, two contributions focus on modeling approaches, and two contributions provide a survey of information technology applied to agriculture and biology. Specifically, the paper by Bodei et al. describes the application of the IOT-LYSA formal modelling framework to a possible scenario of grape cultivation, in order to assess water consumption, and the paper by Barbuti et al. proposes a mathematical model of artificial reefs, in order to study the dynamics of algal coverage and of populations of fish in some Italian

artificial reefs. Finally, the paper by Fresco et. al. explores the current challenges and IT solutions in order to realize a digital agriculture framework, intended as an evolution from Precision Farming to connected knowledge-based farm production systems, and the paper by Pucci et al. provides a survey on biologging methodologies for the collection of knowledge about animals' behaviour, making a review of some related common data analysis techniques.

All papers have been carefully reviewed by experts in the specific fields. Here is the list of the reviewers, that we thank for the collaboration.

ROBERTO BARBUTI Dipartimento di Informatica, Università di Pisa, *roberto.barbuti@unipi.it*

PAOLO BARSOCCHI Istituto di Scienza e Tecnologie dell'Informazione del CNR (ISTI), paolo.barsocchi@isti.cnr.it

GIULIO CARAVAGNA School of Informatics, University of Edinburgh, *giulio.caravagna@ed.ac.uk*

STEFANO CHESSA Dipartimento di Informatica, Università di Pisa, *stefano.chessa@unipi.it*

BARBARA FACCINI Dipartimento di Fisica e Scienze della Terra, Università di Ferrara, *fccbbr@unife.it*

GIANLUGI FERRARI Dipartimento di Ingegneria e Architettura, Università di Parma, *gianluigi.ferrari@unipr.it*

PAOLO FONTANELLI Meccanica Agraria DiSAAA, Università di Pisa, *marco.fontanelli@unipi.it*

ROBERTO FRESCO Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), *roberto.fresco@crea.gov.it*

PAOLO LUSCHI Dipartimento di Biologia, Università di Pisa, *paolo.luschi@unipi.it*

PAOLO MILAZZO Dipartimento di Informatica, Università di Pisa, *paolo.milazzo@unipi.it*

STEFANIA NIN Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), *stefania.nin@crea.gov.it*

ELIO PASSAGLIA Dipartimento di Scienze Chimiche e Geologiche, Università degli Studi di Modena e Reggio Emilia, *elio.passaglia@unimore.it*

DOMENICO PRISA Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), *domenico.prisa@crea.gov.it*

DANIELE SARRI GESAAF Dip. Gestione dei sistemi Alimentari Agrari e Forestali, *daniele.sarri@unifi.it*

ROBERTO TRASARTI Istituto di Scienza e Tecnologie dell'Informazione del CNR (ISTI), *roberto.trasarti@isti.cnr.it*

RITA PUCCI, ALESSIO MICHELI, STEFANO CHESSA (**)

WILD ANIMALS' BIOLOGGING THROUGH MACHINE LEARNING MODELS

ABSTRACT: R. PUCCI, A. MICHELI, S. CHESSA. Wild animals' biologging through machine learning models.

In recent decades the biodiversity crisis has been characterised by a decline and extinction of many animal species worldwide. To aid in understanding the threats and causes of this demise, conservation scientists rely on remote assessments. Innovation in technology in the form of microelectromechanical systems (MEMs) has brought about great leaps forward in understanding of animal life. The MEMs are now readily available to ecologists for remotely monitoring the activities of wild animals. Since the advent of electronic tags, methods such as biologging are being increasingly applied to the study of animal ecology, providing information unattainable through other techniques.

In this paper, we discuss a few relevant instances of biologging studies. We present an overview on biologging research area, describing the evolution of acquisition of behavioural information and the improvement provided by tags. In second part we will review some common data analysis techniques used to identify daily activity of animals.

KEYWORDS: Biologging, Machine Learning, Pattern recognition, Animal activity recognition.

RIASSUNTO: R. PUCCI, A. MICHELI, S. CHESSA. *Biologging di animali* selvatici tramite modelli di machine learning.

Negli ultimi decenni, la crisi della biodiversità è stata caratterizzata dal declino e dall'estinzione di molte specie animali nel mondo. Per studiare le minacce ambientali e le cause delle estinzioni, gli studiosi si basano su valutazioni eseguite a distanza. L'innovazione tecnologica nell'ambito dei sistemi microelettromeccanici (MEMs) ha portato a grandi avanzamenti nella comprensione della vita degli animali. Al giorno d'oggi gli ecologi hanno facile accesso a dispositivi MEMs, e possono usarli per monitorare da remoto le attività degli animali selvatici. Con l'avvento dei tag elettronici basati su MEMs, metodi quali il biologging hanno trovato sempre maggiore applicazione allo studio dell'ecologia degli animali, fornendo informazioni impossibili da ottenere attraverso altre tecnologie.

In questo articolo, discutiamo alcune studi significativi sul biologging. Presentiamo una panoramica di questa area di ricerca, descrivendo l'evoluzione dei metodi di acquisizione di informazioni e il miglioramento ottenuto grazie ai tag. Nella seconda parte, passiamo in rassegna alcune tecniche di analisi di dati comunemente usate per identificare automaticamente le attività giornaliere degli animali.

PAROLE CHIAVE: Biologging, Apprendimento Automatico, Machine Learning, Pattern Recognition, Riconoscimento di Attività Animali

1. ACQUISITION OF BEHAVIOURAL INFORMATION BY BIOLOGGING

The acquisition of information about animal behaviour is one of the best ways to learn about their habits and needs, and to understand how to preserve the biodiversity of our planet (Wilson A.D. 2015). Visual observation (direct observation) is the simplest technique to survey the behaviour of wild animals. While the direct observation of an animal allows biologists to obtain a clear description of the animal life,≤ it is difficult in the case of most wildlife due to their high mobility, nocturnal life, and the danger to observers in accessing wild habitats. Even more challenging a habitat is the hydrosphere, since all the aforementioned reasons that make the observation of wildlife difficult are exacerbated due to it being underwater. The hydrosphere, however, is of vast importance to biodiversity preservation because it hosts 71% of Earth's fauna.

While in the seas surface it is possible to use data logger float package (Whitney N.M. 2016) or remote controlled systems to at least record images of wildlife (Miller 2015), (Whitney 2007) in deep seas submersible remote operated vehicles (ROV) can be used. These tools, however, can not operate in the field for very long and their observation is only composed of sets of "snapshots".

The follow-up observation of an animal is a more interesting way to obtain information. In this case technological tools (hereafter referred to as tags) are attached or implanted onto the animals to collect data between a release point and a recapture point, operating solely as data loggers. Tags can collect both body and geo-local information which both contribute to discern the behaviour and attitude of the

^(**) Università di Pisa, Dipartimento di Informatica, Largo B. Pontecorvo, 3, 56127 Pisa.

animals. Follow-up observation allows a deep understanding of animal life, but the attaching and detaching procedure could be stressful for the animals under observation.

The follow-up observation and the log of signals by means of tags was first called *biologging* in 2004 by Ropert-Coudert and Wilson in (Wilson 2004), distinguishing it as separate from biotelemetry (Boyd 2004). In this paper we consider biologging to differ from biotelemetry in the sense that data are stored locally in the memory of the devices and not transferred via radio waves¹. In addition, we consider biologging to be more focused on the research subject's physiological information.

In this work we review some of the most relevant works on biologging in order to examine its evolution and to highlight some important breakthroughs. In particular, we focus on the devices and sensors used in biologging studies in Section 2, and we focus on the methods of data analysis in sections 3 and 4. Section 5 draws the conclusions.

2. A BRIEF OVERVIEW ON BIOLOGGING WITH TAGS

Data loggers have been used since the 1960s to observe areas previously inaccessible, such as the hydrosphere. The first device used to obtain information about underwater activity was in 1965, in the form of a tag equipped with a time-depth sensor. It was used in one hour stints to measure the diving capacity of Weddell Seals (Leptonychotes weddelli) in Antarctica (Kooyman, G. L. 1965) (Kooyman G. L. 1966). The tags were attached to seals that were known to frequently visit the same holes, thus allowing easy retrieval of the tags. Between the 60s and 70s there was a slow evolution of biologging. In these two decades the primary focus was to increase the operational length of the device, this was evident with the time-depth sensors. By the 80s, a revised architecture of these sensors enabled recording stints of up to three months. For example, in (Le Boeuf B. J. 1988) and later in (Boehlert G. W. 2001), Le Boeuf and Boehlert were able to take advantage of this revised architecture to monitor elephant seals (Mirounga angustirostris) at the Rookery in Año Nuevo, California.

In this case a tag was left on an animal until it returned to the place where the tag was attached. It is worth noting that at this point this technology was only possible to be used with large animals on which large devices could be attached without hindering their movement.

As biologging techniques matured, they were slowly also miniaturised. Smaller sensors and efficient battery, combined with the ability to design and package devices specific to individual species, opened up the possibility to begin research project with small animals. An example of the application on small animals is the study of the sea turtle (Caretta caretta), by mean of Global Positioning System (GPS), (Schofield 2007). Turtles in Buck Island Reef, U.S. Virgin Islands, were equipped with tags and set free to roam about in a diving enclosure for a period of time, after which the tags were retrieved. The tags in this case were a miniaturized data-logger, less than 10cm in length.

At the same time, the follow-up style of observation became popular in medicine for humans, advancing research in activity recognition by sensors (Foerster F. 1999). The medical interest in human activity recognition (HAR) research gave a strong push to the technological devices used in these studies.

Recent major reviews on biologging (such as (Cooke S.J. 2004), (Cooke J.S. 2008), (Bograd S.J. 2010), (Wilmers C.C. 2015)) explain further refinements and advances in biologging equipment (storage capacity, lifetime, and number and types of sensors on board). The increase in data-logger performance makes it possible to use sensors with a high speed sampling rate such as accelerometers and magnetometers. The use of accelerometers introduced the possibility to detect specific movements of an animal's body. A three-dimensional accelerometer provides more accurate diving information about elephant seals, as described in (Mitani Y. 2009). In (Viviant M. 2014) accelerometer data was used to detect when southern elephant seals (*Mirounga leonine*) open and close their mouths to monitor foraging activity. In both (Mitani Y. 2009) and (Viviant M. 2014) further studies were conducted on seals to predict foraging success. In these studies, the seals were released from a point on the coast close to the capture site. Ropert-Courdert, and Kato in (K. A. Ropert-Coudert Y. 2014) observed free-raging animals in Antarctica using biologging technologies. In this study tags were used to observe habits of predators such as seabirds, penguins, and seals in remote marine locations. Examples are presented in (Yoda K. 2001), (G. D. Ropert-Coudert Y. 2004), (Gallon S. 2013), (Carroll G. 2014), (Volpov B.L. 2015), (Xavier J.C. 2016), and (Descamps S. 2016).

Table 1: In the table the first column lists animals that are subjects of studies in first column, the second column lists sensors used to collect behavioural data (time depth, accelerometer, gyroscope, and tilt sensor), and the third column shows of the activities recognised.

Animal	Sensors	Activities
Elephant seals	Time-depth, 3D acc.	Depth, movement of body, time immersion, opening mouth.
Weddell seals	Time-depth	Depth, time immersion.
Antarctic fur seals	Time-depth	Foraging activity
Hawksbill turtles	Time-depth	Depth, time immersion.
Testudo H. turtles	2D acc.	Digging.
Little Penguins	3D acc.	Prey captures activity.
Adelie Penguins	3D acc.	Walk, toboggan, stand on land, lay on land and rest
Seabirds	Time-depth, 3D acc., tilt sensor	Take-off, flap, flight, plunge dive, and land.
Red Foxes	3D acc., magnetometer	Hunting movement, mag- netic alignment.
Leopards	3D acc., gyro- scope	Energy consumed.
Domestic dogs	3D acc.	Walk, run, sit, lie-down, and stand.

¹ Note that this distinction is no universally valid but in (e.g. (Rutz 2009), (Bograd 2010)) presented biologging systems which relay data through radio signals.

The confidence in biologging technologies in the deep seas promoted their application on land animals. Biologging rapidly became an important tool in understanding the behaviour of mammals and reptiles. While it is certainly easier to observe land animals than it is to observe sea animals, observing land animals has its fair share of dangers and difficulties for researchers. This is given more credence after noting that biologging techniques are often used in monitoring predators. Pumas (Puma concolor), for example, have proven difficult to study specifically due to their speed. The first study of pumas with tags was performed in California, where tags were in the form of collars equipped with accelerometers (Williams T.M. 2014). The objective of this study was to monitor movement during hunting activities to infer the expenditure of energy. Similar tags were also used in (Painter M.S. 2016) to study the hunting behaviour of semi-domestic red foxes (Vulpes vulpes) in Prásily, Czech Republic. These tags, however, also contained a magnetometer. The tags were retrieved at the end of the experiment. It is been found that domestic dogs can be used as analogue of certain predators. An example of using dogs in this manner is presented in (Campbell H.A. 2013), in which authors presented an exploratory study recording daily activities of dogs, also using collars with embedded accelerometers.

Table 1 reports a summary of these studies. In particular it shows the sensors used and the activities monitored for each animal under study.

The third column of Table 1 shows the activities observed for each animal. Activities are chosen based on the subject of study and the aim of the research. Researchers have always been interested in studying animal body movement and behaviour, however until recently this has been difficult. The introduction of accelerometers, magnetometers, and gyroscopes in tags, has made this information much more easily accessible. In the early studies, was not possible to transmit data due to power requirements.

Recorded data was therefore usually stored in built-in memory storage inside the tags and was only accessible when the tags were rescued. This is a noteworthy limitation from two points of view: the first is that retrieval of the tags is necessary, which means that the animal must be limited to a certain space or easy to find; the second is that the analysis of data can be done only at the end of each recording period and not in real time. More recently, these two issues brought the identification of animal behaviour in two main directions. One is towards providing a standard, autonomous behavioural annotation system able to recognise activity from data in an automatic way, (Resheff Y.S. 2014), and (Gao L. 2013). The second is towards the development of autonomous systems able to analyse data embedded on tags. That avoids the necessity of retrieving tags from animals, as in (Barbuti R. 2016).

3. TRADITIONAL METHODS OF ANALYSIS OF BIOLOGGING DATA

The use of tags allows scientists to collect copious amounts of data on many aspects of an animal's life. Such a huge amount of data, however, needs to be analysed in order to infer any useful information. Conventional approaches to data analysis have, in the past, ranged from the direct observation of data streams by researchers, to signal analysis techniques. Inspiration for these techniques came from the field of HAR, which advanced hastily due to the need for automatic tools to analyse data in e-health applications.

In both HAR and biologging, conventional methods of data analysis require steps of filtering and feature extraction before applying specific methods of analysis.

In (Mitani Y. 2009) data were analysed for diving behaviour in elephant seals including stroking rates and three-dimensional movements. The signal streams were observed to identify spikes that elucidated the rotation movement performed by elephant seals to hunt their prey. This procedure was one of the first approaches and it was not automated, therefore requiring great effort from a human analyser. This approach, however, soon became obsolete and was replaced by automatic signal analysis.

In (Gallon S. 2013), authors presented an example of signal analysis of behavioural data in seals. This study used tags with accelerometer and depth sensors mounted on the animal's head. An initial set of data was used to identify the sensors' data profiles. These profiles were identified for each activity. The accelerometer and depth data were compared with dive depth profiles and accelerometer profiles. The relationships between the behaviour and acceleration profiles were then used to identify the activities of animals. In (Yoda K. 2001) depth and accelerometer profiles were used in a similar way to identify Adelie penguins' activities to observe their daily life.

In all these studies the profiles were calibrated in an aquarium (and thus in a controlled environment) with a restricted set of samples. The disadvantage of using an aquarium as an analogue for the ocean is that the behaviour of the test subject may not accurately reflect the behaviour of subjects in the wild. The classification may achieve a good result, however there is a potential for the classification system to be not general enough for use in the wild.

Statistics analysis is a method commonly used in biologging to extract information. Tools such as R (Venables W.N. 2016), a programming environment for data analysis and graphics, provides statistics about individual behaviour of animals and can be used to write customised data analysis routines. An example of this technique is presented in (Volpov B.L. 2015), in which the study subject was fur seals. A custom routine, written in *R*, inspired by (Viviant M. 2014) was used to classify the biologging data. The same technique has been used recently to classify biologging data form pumas, (Williams T.M. 2014).

In this paper we refer to all these methods as *traditional*. However, recent trends in analysis of biologging data adopt methods based on Machine Learning (ML) to study complex interaction between biotic and abiotic systems. ML can often outperform the traditional methods of analysis in classification tasks as reported in (Thessen A.E. 2016). In the following section we outline the advantage of machine learning in biologging.

4. MACHINE LEARNING MODEL FOR BIOLOGGING

Machine learning (ML) methods have been used to analyse data recorded by MEMs. The robustness and the ability for universal approximation of these methods (e.g. Artificial Neural Networks) provide the basis for the flexibility of the approach for pattern recognition. In particular, they allow the approximation of arbitrary classification functions from experimental data, despite not having a theory of pattern characteristics. This is particularly advantageous for non-linear classification/pattern recognition tasks, which can be difficult to address by traditional approaches.

A ML classifier is inferred by data: it is able to classify new data after being properly trained with known instances. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) are ML models capable of performing well with noisy data, which is a common problem with animal behaviour data.

In the ML approach, the designer of the application has to teach a model the background of a problem using a dataset, hereafter called the training dataset. In the supervised class of tasks, the training dataset is a collection of reference data recorded through sensors and labelled appropriately with ground truth information. During the training phase the ML model is automatically tuned by the learning algorithm to be able to classify data, (e.g. recognizing the pattern in the activity recognition scenario) according to the provided labelled examples. Several models are applied in data analysis within the field of biologging. Each model referred to in this paper is described in (Haykin S. 2009) and in (Thessen A.E. 2016).

In this context, ML models are becoming more frequently applied, achieving high classification capabilities. In (Painter M.S. 2016), a k-Nearest Neighbours (K-NN) model is applied to identify magnetic alignment responses during hunting in red foxes. In this case the training dataset consists of the k closest examples used as seeds to classify future data streams. A data stream is classified by a majority behaviour referred by its neighbours' seeds. The activity identified from a current data stream is assigned considering the most common activity identified by its knearest neighbours. This model is applied in (Nathan R. 2012) and in (Bidder O.R. 2014), where the ML models are proposed to classify the daily activity of several species. In such applications, the K-NN provides good performance across the dataset. This model, however, is prone to high variance, course of dimensionality, and overfitting issues, all of which can limit its generalization capability. To manage this problem, the focus changes from the K-NN to SVM. The SVM uses a set of seeds that corresponds to the support vectors used to classify each data stream, but also reduces the effect of the aforementioned issues by utilising maximum margin classifiers (Vapnik V. 2013). The SVM model was applied to identify prey capture in little penguins (Eudyptula minor) as explained in (Carroll G. 2014). The model is trained over a sample dataset recorded from penguins in Taronga Zoo. The tuned model will be able to identify prey capture activity in wild penguins.

In (Campbell H.A. 2013) the SVM model is applied to identify the activity of domestic dogs. The model distinguishes between five different activities and classifies each activity individually. In both cases, data are analysed after the recording phase. Indeed, it is necessary that the tag is rescued to allow for data analysis in the laboratory.

To analyse data in real-time is a more interesting challenge and will modify the way in which one analyses the data. Analyse data in real time means moving the analysis stage to the tag on board. Both K-NN and SVM methods need to maintain information about the seeds. This information is necessary to classify the data streams but requires a specific amount of memory space. For this reason different ML methods should be considered for the classification of data streams.

This is the main aim recently addressed with the feasibility analysis in (Barbuti R. 2016). Authors present ML methods for classification of nest digging activity in tortoises (Testudo h. hermanni). The tag was equipped with an accelerometer, a light sensor, and a temperature sensor, all of which are used to identify the activity of the tortoises, and monitoring the surrounding environment. The ML proposed is a customised Input Delay Neural Network (IDNN). This model provides a solution that was designed to find a trade-off between the generality of the ML model in classification, and memory space needed. The result obtained is a model with high performance accuracy and that is embeddable on a tag. The possibility to install the system on the tag is a big improvement for behavioural analysis. In this way tags on animals need not be retrieved after the recording season.

5. DISCUSSIONS AND CONCLUSIONS

In this paper we analysed the use of traditional methods and ML models applied to biologging. In particular, we focused our discussion on the analysis of data stream recorded by tags attached to animals.

As we observed, the introduction of accelerometers and magnetometers in tags caused a dramatic increase in dataset size, which, in turn, created a need for automatic analysis. This allowed biologists to infer new and richer information from data in order to make supportive programs more effective. This new information paved the way for the automatic recognition of animal activities (such as foraging, predating, nesting, etc.), and thus brought biologging from simple telemetry towards the research area of activity recognition, which was previously restricted to human cases (HAR).

It is worth noting that activity recognition in humans is deeply different from activity recognition in animals. This difference is made evident through three main aspects: (i) the sensors used to record movement information may be the same (accelerometers, and magnetometers) but the device and the location on the body could be different both due to physiological differences (e.g. the lack of paws for fish, the lack of wings for humans); (ii) the sensors themselves may be different (e.g., depth sensors are meaningful for fish, and marine mammals but usually not for humans); and (iii) the observed activities may be completely different. Furthermore, the data collecting procedure may be completely different for humans and animals, because even domestic animals may not follow instructions (in simulated scenario) to the same level of accuracy as a human subject for data gathering purpose. Finally, the collection campaign with animals is more complex to perform due to a lack of volunteers and environmental hazards. The result is that the analysis of data from animal activity represents a new challenge for activity recognition due to the heterogeneity of datasets, and to the variable quality of data.

Table 2: In table the first column lists the animals that are the subjects of study at hand, the second column the analysis method, and the third one the year of publication.

<u>Animal</u>	<u>Data analysis</u>	Year
Weddell seals	Traditional method	1966
Hawksbill turtles	Traditional method	2000
Elephant seals	Traditional method	2001
Adelie Penguins	Traditional method	2001
Seabirds	Traditional method	2004
Leopards	Traditional method	2014
Antarctic fur seals	Traditional method	2015
Little Penguins	ML models	2015
Domestic dogs	ML models	2015
Testudo H. turtles	ML models	2016
Red Foxes	ML models	2016

Table 2 shows the methods (either traditional or MLbased) that were used in some meaningful biologging studies. Traditional techniques provide ad hoc solutions for given task, often basing on specific statistical assumption. On one hand the statistical analysis used in these methods provides good results over data samples. On the other hand, the solutions reached are often not general and thus difficult to apply in wild habitats. By contrast ML-models provide methodologies to build a model directly from (real) data. Moreover, they are intrinsically more flexible and can be used to automatically deal with a nonlinear relationship between data and outcomes of classification.

Table 2 shows that biologists have been slow to adopt ML models as a way to analyse biologging data. We argue that this is due to many biologists' familiarity with traditional statistical methods and, consequently, to their lack of familiarity with computer science methodologies.

It is worth nothing that both ML methods and traditional methods are not necessarily applicable to all cases. Some research projects may well be more suited to a specific type of biologging analysis and individual projects should be considered on a case-by-case basis. That being said, ML models in many cases provide better results than the traditional methods for classification and identification of activities.

An interesting study, for example, could be to re-apply ML methods on datasets previously analysed using traditional methods to observe the differences and better assess

the advantages that the former methods can give (in this respect, the production of open datasets is key). Traditional methods are in fact still widely used in recent works. For example, in (Gallon S. 2013), a 2-D accelerometer and a pressure sensor were used to record the rapid movement and the depth of seals to identify foraging activity. In this study, both depth and accelerometer information were analysed by means of thresholds to classify the signal as foraging or diving. In (Volpov B.L. 2015), accelerometer data was analysed with a customised function. This function is composed of a filter that computes six features over each axes of accelerometer signal, and a classifier that identifies the signal using a custom matching function between the new signal and a dataset of previously labelled signals on a common time vector (synced). In both studies the ML were models conveniently trained to identify the foraging event, thus allowing a better classification of the dataset, as it was observed, for different datasets in (Carroll G. 2014).

A different analysis could be done of the research presented in (Williams T.M. 2014). Accelerometer data were used to compute the energy expenditure of a puma during hunting. This problem was not an identification of an activity. ML models may be applied in this task as done in (Bacciu D. 2015) to compute the energy expenditure associated with the physical activities.

The methods and models discussed in this paper are used in biologging for analysing the behavioural information recorded by tags. Some next steps could be to promote the use of ML models, and to open new developments in recognition of more and more activities of animals.

Many possible solutions can be inferred from ML models to open the view to new challenges only partially explored.

REFERENCES

- Bacciu D., Chessa S., Gallicchio C., Micheli A., Ferro E., Fortunati L., Palumbo F., et al., 2015. Smart environments and context-awareness for lifestyle management in a healthy active ageing framework. In: *Portuguese Conference on Artificial Intelligence*, Lecture Notes in Computer Science 9273: 54-66.
- Barbuti R., Chessa S., Micheli A., Pucci R., 2016. Localizing Tortoise Nests by Neural Networks. *PlosOne* 11: e0151168.
- Bidder O.R., Campbell H.A., Gmez-Laich A., Urg P., Walker J., Cai Y., et al., 2014. Love thy neighbour: automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm. *PlosOne* 9: e88609.
- Boehlert G. W., Costa D.P., Crocker D. E., Green P., O'Brien T., Levitus S., Le Boeuf B. J., 2001. Autonomous Pinniped Environmental Samplers: Using Instrumented Animals as Oceanographic Data Collectors. *Journal of atmospheric and oceanic technology* 18(11): 1882-1893.
- Bograd S.J., Block B.A., Costa D.P., Godley B.J., 2010. Biologging technologies: new tools for conservation introduction. *Endangered Species Research* 217: 1-7.
- Boyd I.L., Kato A. Ropert-Coudert Y., 2004. Bio-logging science: sensing beyond the boundaries. *Memoirs of National Institute Polar Research*, *Special Issue*, 2004.
- Campbell H.A., Gao L., Bidder O.R., Hunter J., Franklin C.E., 2013. Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species. *Journal of*

Experimental Biology 2013: 4501-4506.

- Carroll G., Slip D., Jonsen I., Harcourt R., 2014. Supervised accelerometry analysis can identify prey capture by penguins at sea. *Journal of Experimental Biology* 217: 4295-4302.
- Cooke J.S., 2008. Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and iucn red list threat assessments. *Endangered species research* 4: 165-185.
- Cooke S.J., Hinch S.G., Wikelski M., Andrews R.D., Kuchel L.J., Wolcott T.G., Butle P.J., 2004. Biotelemetry: a mechanistic approach to ecology. *Trends in ecology & evolution* 19: 334-343.
- Descamps S., Tarroux A., Cherel Y., Delord K., Godø O.R., Kato A., Krafft B.A., et al., 2016. At-Sea Distribution and Prey Selection of Antarctic Petrels and Commercial Krill Fisheries. *PlosOne* 11(8): e0156968.
- Foerster F., Smeja M., Fahrenberg J., 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior* 15(5): 571-583.
- Gallon S., Bailleul F., Charrassin J.B., Guinet C., Bost C.A., Handrich Y., Hindell M., 2013. Identifying foraging events in deep diving southern elephant seals, Mirounga leonina, using acceleration data loggers. *Deep Sea Research Part II* 88: 14-22.
- Gao L., Campbell H.A., Bidder O.R., Hunter J., 2013 A Web-based semantic tagging and activity recognition system for species' accelerometry data. *Ecological Informatics* 13: 47-56.
- Haykin S, 2009. Neural networks and learning machines. Pearson Upper Saddle River, NJ.
- Kooyman G. L., 1965. Techniques used in measuring diving capacities of Weddell Seals. *Polar Record* 12(79): 391-394, Cambridge University Press.
- Kooyman G. L., 1966. Maximum diving capacities of the Weddell seal, Leptonychotes weddeili. *Science* 151(3717): 1553-1554.
- Le Boeuf B. J., Costa D.P., Huntley A.C., Feldkamp S.D., 1988. Continuous, deep diving in female northern elephant seals, Mirounga angustirostris. *Canadian Journal of Zoology* 66(2): 446-458.
- Mitani Y., Andrews R.D., Sato K., Kato A., Naito Y., Costa D.P., 2009. Three-dimensional resting behaviour of northern elephant seals: drifting like a falling leaf. *Biology letters*, Royal Society Publishing, rsbl20090719.
- Nathan R., Spiegel O., Fortmann-Roe S., Harel R., Wikelski M., Getz W.M., 2012. Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *Journal of Experimental* 215(6): 986-996.
- Painter M.S., Blanco J.A., Malkemper E.P., Anderson C., Sweeney D.C., Hewgley C.W., Çervený J., et al., 2016. Use of bio-loggers to characterize red fox behavior with implications for studies of magnetic alignment responses in free-roaming animals. *Animal Biotelemetry* 4(1): 20.
- Resheff Y.S., Rotics S., Harel R., Spiegel O., Nathan R.,2014. AcceleRater: a web application for supervised learning of behavioral modes from acceleration measurements. *Movement ecology* 2: 1.
- Ropert-Coudert Y., Grémillet D., Kato A., Ryan G.P., Naito Y., Le Maho Y., 2004. A fine-scale time budget of Cape gannets provides insights into the foraging strategies of coastal seabirds. *Animal Behaviour* 67: 985-992.
- Ropert-Coudert Y., Kato A., 2014. Where the wild things are. International Innovation, 141: 100-102.
- Rutz C., Hays G.C., 2009. New frontiers in biologging science. Royal Society Publishing, rsbl20090089.
- Schofield G., Bishop C.M., Katselidis K.A., Dimopoulos P., Pantis J.D.,

Hays G.C., 2007. Microhabitat selection by sea turtles in a dynamic thermal marine environment. *Journal of Animal Ecology* 78(1): 14-21.

- Sisak M.M., Hillis-Starr Z., Philips B., Pemberton Jr R.A., Crow J.W., 2000. Use of a miniature data storage tag on a juvenile Hawksbill turtle (Eretmochelys imbricata) at Buck island reef nm, U.S. Virgin Islands. In: 18th International Sea Turtle Simposium. 256.
- Smolowitz R.J., Patel S.H., Haas H.L., Miller S.A, 2015. Using a remotely operated vehicle (ROV) to observe loggerhead sea turtle (Caretta caretta) behavior on foraging grounds off the mid-Atlantic United States. *Journal of Experimental Marine Biology and Ecology* 471: 84-91.
- Thessen A.E., 2016. Adoption of machine learning techniques in Ecology and Earth Science. *One Ecosystem* 1: e8621.
- Vapnik V., 2013. The nature of statistical learning theory. Springer science & business media, New York.
- Venables W.N., Smith D.M, 2016. An Introduction to R:A Programming Environment for Data Analysis and Graphics. R Development Core Team. Available online: <u>http://cran.r-project.org</u>
- Viviant M., Monestiez P., Guinet C., 2014. Can we predict foraging success in a marine predator from dive patterns only? validation with prey capture attempt data. *PLosOne* 9: e88503.
- Volpov B.L., Hoskins A.J., Battaile B.C., Viviant M., Wheatley K.E., Marshall G., Abernathy K., Arnould J.P. 2015. Identification of prey captures in Australian fur seals (Arctocephalus pusillus doriferus) using head-mounted accelerometers: field validation with animal-borne video cameras. *PlosOne* 10: e0128789.
- Whitney N.M., Lear K.O., 2016. Bringing data to the surface: recovering data loggers for large sample sizes from marine vertebrates. *Animal Biotelemetry* 4: 1.
- Whitney F., Yahel G., Reiswig H.M., Eerkes-Medrano D.I., Leys S.P., 2007. In situ feeding and metabolism of glass sponges (Hexactinellida, Porifera) studied in a deep temperate fjord with a remotely operated submersible. *Limnology and Oceanography*, 52(1): 428-440.
- Williams T.M., Wolfe L., Davis T., Kendall T., Richter B., Wang Y., Bryce C., et al., 2014. Instantaneous energetics of puma kills reveal advantage of felid sneak attacks. *Science* 9: 81-85.
- Wilmers C.C., Nickel B., Bryce C.M., Smith J.A., Wheat R.E., Yovovich V., 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. *Ecology* 96: 1741-1753.
- Wilson A.D., Wikelski M., Wilson R.P., Cooke S.J., 2015. Utility of biological sensor tags in animal conservation. *Conservation biology* 29: 1065-1075.
- Wilson, R.P. & Ropert-Coudert, Y., 2004. Subjectivity in bio-logging science: do logged data mislead? *Memoirs of the National Institute of Polar Research, Special Issue*, 58: 23-33.
- Xavier J.C., Brandt A., Ropert-Coudert Y., Badhe R., Gutt J., Havermans C., Jones C., et al., 2016. Future challenges in Southern Ocean ecology research. *Frontiers in Marine Science* 3: 94.
- Yoda K., Naito Y., Sato K., Takahashi A., Nishikawa J., Ropert-Coudert Y., Kurita M., et al., 2001. A new technique for monitoring the behaviour of free-ranging Adelie penguins. *Journal of Experimental Biology* 204(4): 685-690.

(ms. pres. 15 gennaio 2017; acc. 15 settembre 2017)

Edizioni ETS Piazza Carrara, 16-19, I-56126 Pisa info@edizioniets.com - www.edizioniets.com Finito di stampare nel mese di marzo 2018