

DISS. ETH NO. 19395

# Semi-Supervised Online Learning for Acoustic Data Mining

A dissertation submitted to  
ETH ZURICH

for the degree of  
Doctor of Sciences

presented by  
TSER LING YVONNE MOH  
Dipl.-Inf., RWTH-Aachen  
born 21. April 1977  
citizen of Singapore

accepted on the recommendation of  
Prof. Dr. Joachim M. Buhmann, examiner  
Prof. Dr. Juraj Hromkovic, co-examiner

2010

## Abstract

The main focus of this thesis is semi-supervised online learning for acoustic data analysis. In semi-supervised learning, both labeled and unlabeled instances are used for learning. Online learning addresses the inference scenario where labeled instances are processed in a time-sequential manner. In our scenario, only a subset of the data that arrives over time is labeled. Instead of naively discarding the unlabeled data, semi-supervised online learning tries to exploit these unlabeled data during learning. This setting is motivated by online adaptation of hearing devices using the user feedback obtained during operation. A default classifier is dispatched with the hearing device. A user may wish to personalize the classifier during operation of the device. Clearly, online learning is vital, since storage of the time-sequential data is limited. Furthermore, it is obvious that the user is not dedicated to labeling all the instances while operating the device, resulting in a semi-supervised learning problem. We see that semi-supervised online learning is a practical approach to this setting. We look at classification and relevance feedback problems in semi-supervised learning.

For the classification problem, we view semi-supervised online learning from two angles. The first approach introduces a general plug-and-play structure. Here, any online algorithm can be augmented with a semi-supervised learning algorithm, that suggests labels for the unlabeled instances which the online algorithm uses for its learning process. Many semi-supervised learning algorithms are transductive, i.e., they do not generalize for unseen samples. Through this two-step approach, the information captured by the unlabeled data is exploited by the online learning algorithm, that is able to generalize and adapt over time. The second approach considers the passive aggressive online algorithm where online learning is posed as a convex optimization problem. The semi-supervised information is injected directly into the objective function as a regularization term. We provide a formulation that yields again a convex optimization problem, and show that the updates to the classifier can be analytically computed. Experimental results show that using semi-supervised information improves learning compared to when the unlabeled data is ignored, especially in cases where the labels are biased.

Online relevance learning for content-based audio retrieval is interesting for large unlabeled data sets. The user presents an example snippet of the desired audio signal (e.g., applause), and the algorithm should adaptively retrieve relevant segments from the data set with high precision. When many users actively use a data set, methods such as collaborative filtering exploit user behaviors for the retrieval process, and the multi-users preferences are

---

labels to the data set. However, when only few users access the data set, e.g., large personal audio logs collected over hearing devices, such high-level information is no longer available. Hence, the retrieval process has to operate on the level of audio content. For this purpose, we motivate a graph-based approach whilst executing the retrieval in the presence of relevance feedback. The graph connects similar data instances, and the connectivity may expose structures of the underlying manifold of the data. We show results on a publicly available music data set, where the graph based approach outperforms existing algorithms. An extension that supports feature re-weighting results in an additional performance gain.

## Zusammenfassung

Schwerpunkt der vorliegenden Arbeit ist das überwachte sequentielle Lernen mithilfe eines teilweise gelabelten Datensatzes. Das Lernen von einem Datensatz, in dem nur wenige Instanzen gelabelt sind, und in dem sonst vorwiegend ungelabelte Daten vorliegen, heisst “halb überwachtetes Lernen”. Das sequentielle Lernen nennt man “online Lernen”. Unsere Ziel ist das daher das “halb überwachtetes Lernen über die Zeit”. Die Motivation dieser Arbeit stammt aus der adaptiven Anpassung von Hörgeräten. Dieser Anpassungsprozess wird durch Benutzerfeedback ausgelöst, das allerdings nur sporadisch gegeben wird, woraus die unvollständig gelabelten Daten resultieren. Das halb überwachte Lernen ist sequentiell, da die Zeitreihendaten nicht vollständig speicherbar sind, sondern in Echtzeit bearbeitet werden müssen. In dieser Arbeit werden wir sowohl die Klassifikation von Audiosequenzen als auch die Suche relevanter Treffer in grossen Datenbanken untersuchen, die ähnlich zu einem Suchmuster sind.

Die Klassifikation wird von zwei Herangehensweisen betrachtet: In der ersten Vorgehensweise stellen wir eine vielseitig verwendbare Struktur vor, in der zwei frei wählbare bereits existierende Algorithmen – jeweils ein online Lernalgorithmus und ein halb überwachter Lernalgorithmus – kombiniert werden. Während des Lernprozesses bekommt der online Lernalgorithmus von dem halb überwachten Lernverfahren Vorschläge für die Labels der ungelabelten Instanzen, welche für das sequentielle Lernen genutzt werden können. Der resultierende zwei-Schritt-Algorithmus erbt die Vorteile beider Algorithmen. Der zweite Ansatz basiert auf der Erweiterung des “passive-aggressive” online Algorithmus für teilweise ungelabelte Daten. Der Passive-Aggressive online Algorithmus löst das online Lernen als ein konvexes Optimierungsproblem. In unserer Erweiterung werden die ungelabelten Daten als Regularisierungsterm benutzt. Wir zeigen, dass dieses neue Optimierungsproblem weiterhin konvex und analytisch zu berechnen ist. Die experimentiellen Ergebnisse dieser Arbeit bestätigen, dass das halbüberwachte online Lernen gerade dann von den ungelabelten Daten profitiert, wenn die zur Verfügung gestellten Labelinformationen nicht unabhängig verteilt sind.

Im letzten Teil betrachten wir die Suche nach Audiosegmenten in einer Menge unbeschrifteter Audiodaten. Anhand eines Suchmusters, z.B. Applaus, findet ein Algorithmus mit hoher Präzision ähnliche Audiosegmente. Mittels Feedback über die bisherigen Suchergebnisse passt der Algorithmus die Suche an und beeinflusst dadurch die zukünftigen Suchergebnisse. Unsere Herangehensweise stellt die Daten als ein Graph dar, welcher die Informationen bezüglich der Datenmanigfaltigkeiten beinhaltet, die dann bei der Suche ausgenutzt werden. In unseren Experimenten zeigen wir die höhere Präzision

dieser Methode verglichen mit herkömmlichen Methoden wie der Rocchio Algorithmus. Weitere Verbesserungen des Algorithmus werden durch eine Erweiterung um Merkmalgewichtung erzielt.