PRUNING SUBSEQUENCE SEARCH WITH ATTENTION-BASED EMBEDDING

Colin Raffel and Daniel P. W. Ellis

LabROSA
Columbia University
New York, NY

ABSTRACT

Finding the most similar sequence to a query in a database of sequences can be prohibitively expensive when the database size is large. Typically, the scalability of different approaches are dependent on various "pruning" techniques, which use heuristics to avoid expensive comparisons against a large subset of the database. We present an approximate pruning technique which involves embedding sequences in a Euclidean space. Sequences are embedded using a convolutional network with a form of attention which integrates over time, which is trained using matching and non-matching pairs of sequences. By using fixed-length embeddings, our pruning method effectively runs in constant time, making it many orders of magnitude faster when compared to full Dynamic Time Warping-based matching. We demonstrate the effectiveness of our approach on a large-scale musical score-to-audio recording retrieval task.

Index Terms— Sequence Retrieval, Attention-Based Models, Convolutional Networks, Dynamic Time Warping, Pruning

1. INTRODUCTION

Sequences[1] are a natural way of representing data in a wide range of fields, including multimedia, environmental sciences, natural language processing, and biology. The proliferation of sensors and decreasing cost of storage in recent years has resulted in the creation of very large databases of sequences. Given such a database, a fundamental task is to find the database entry which is most similar to a query. However, this task can be prohibitively expensive depending on both the method used to compare sequences and the size of the database.

An effective way to compare sequences is the Dynamic Time Warping (DTW) distance, first proposed in the context of speech utterances [1]. DTW first computes an optimal alignment of the two sequences of feature vectors being compared, and then reports their distance as the total distance among aligned feature vectors. The alignment step makes

---

[1]In this work, we refer to "signals" by the slightly more generic term "sequences", but the two can be used interchangeably.

DTW much more robust to phase or offset errors than typical distance metrics, such as the Euclidean distance [2]. DTW also naturally extends to the setting where subsequence matching is allowed. Using dynamic programming, the optimal DTW alignment can be found in time quadratic in the length of the sequences. Despite this optimization, naive application of DTW to nearest-neighbor retrieval of the most similar sequence in a database can be prohibitively expensive. Other dynamic programming-based "edit distance" metrics are used when appropriate, and are similar in both their effectiveness and inefficiency.

To mitigate this issue, a variety of "pruning methods" have been proposed. The idea behind these techniques is to use heuristics to avoid computing the full DTW alignment for a large proportion of the database. [2] gives a broad overview of different pruning methods, and shows that their application makes exact retrieval feasible in extremely large (e.g. trillions of sequences) databases. However, these methods rely on various assumptions, including that the query sequence is a subsequence of its correct match and the length of the alignment path is known a priori.

An alternative approach is to replace the DTW search with a surrogate problem which is faster to solve. For example, [3] proposes a technique which maps sequences to a fixed-length embedding, which avoids DTW calculation when matching the query to the database. The embedding is constructed by pre-computing the DTW distance between each sequence in the database and a small collection of "reference" sequences, which are chosen to optimize retrieval accuracy. Then, to match a query sequence to the database, its DTW distance is computed against the same collection of reference sequences, and the resulting vector of DTW distances is matched against the database of embeddings using a standard Euclidean distance-based nearest-neighbor search. The resulting algorithm is approximate, but provided state-of-the-art rsults both in terms of accuracy and speed. This technique relies on the same assumptions as the pruning methods described in [2]. In addition, it presents a trade-off between efficiency and accuracy where more reference sequences generally improve accuracy but require more full DTW calculations to be made, which can be prohibitively expensive for very long or very high-dimensional sequences.

Motivated by the above issues, we propose a learning-based system for producing sequence embeddings for approximate matching. Our approach is similar in spirit to [3], except that it is fully general, i.e. it does not rely on any assumptions about the alignment length or whether the query is a subsequence of its match. This is thanks to the fact that we utilize an attention-based neural network model which can adapt to any problem setting according to the training data provided. Our approach is also constant-time: By mapping sequences to a fixed-length embedding, comparing a query to each database entry is a single Euclidean distance calculation, which does not become less efficient for longer or higher-dimensional sequences.

In the following section, we motivate embedding and attention-based neural models based on recent work and propose the variant used in this work. In Section **??**, we evaluate the effectiveness of our proposed model on the task of matching musical transcriptions to audio recordings of their corresponding songs in a large database. Finally, we discuss possibilities for improvement in Section **??**.

2. REFERENCES

[1] Hiroaki Sakoe and Seibi Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no. 1, pp. 43–49, 1978.

[2] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 262–270.

[3] Panagiotis Papapetrou, Vassilis Athitsos, Michalis Potamias, George Kollios, and Dimitrios Gunopulos, "Embedding-based subsequence matching in time-series databases," *ACM Transactions on Database Systems (TODS)*, vol. 36, no. 3, pp. 17, 2011.