Statistical Learning of Knowledge from Sequential Data

Nick Harley
njharley@ai.vub.ac.be

December 16, 2021

ARTIFICIAL
INTELLIGENCE
RESEARCH GROUP

OUTLINE

- Introduction
- Modelling Method
- Application 1: Agent Behaviour
- Application 2: Music Cognition
- Future Work and Open challenges

OUTLINE

- **Introduction**
- Modelling Method
- Application 1: Agent Behaviour
- Application 2: Music Cognition
- Future Work and Open Challenges

INTRODUCTION

THE COMPUTATIONAL CREATIVITY LAB

Who We Are

- Prof. Geraint Wiggins
  Head of Group
  `geraint.wiggins@vub.be`
- **Nick Harley**
  Postdoc
  `nicholas.harley@vub.be`
- Steve Homer
  PhD Student
  `steven.thomas.homer@vub.be`
- Fernando Loro Velardo
  Developer
  `fernando.loro.velardo@vub.be`

Research Interests

- Computational Creativity:

> *leaves behind the problem solving paradigm, and moves forward to **problem-seeking** approaches.*

- Predictive Statistical Systems
- Learning and Reasoning Systems
- Cognitive Architectures: *The Information Dynamics of Thinking* (IDyOT: Wiggins, 2020)

MY RESEARCH INTERESTS

Representation and **learning** of hierarchical, cognitive knowledge structures.

Knowledge Representation

- Hierarchical knowledge structures.
- Formal ontology, logic and reasoning.
- Programming languages and type theory.

Knowledge Learning

- Statistical models of perception and cognition.
- Information theory
- Unsupervised learning of knowledge structures from sequence data.

OUTLINE

- Introduction
- **Modelling Method**
- Application 1: Agent Behaviour
- Application 2: Music Cognition
- Future Work and Open Challenges

METHOD

STATISTICAL MODELLING OF SEQUENTIAL DATA

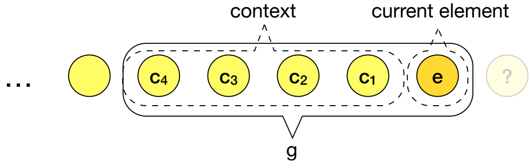
Objective Unsupervised learning of knowledge from statistical patterns in sequence data.

- Our method is based on *The Information Dynamics of Music* (IDyOM: Pearce, 2005).
- IDyOM uses variable-order markov models.
- IDyOM is one of the progenitors of IDyOT.

METHOD

N-GRAMS MODELS AND MAXIMUM LIKELIHOOD

N-gram models are the simplest way to model to a sequence of symbols.



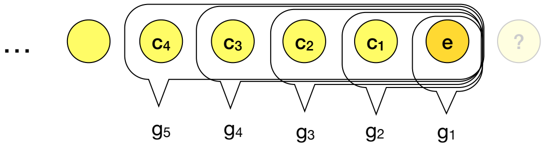
Maximum Likelihood:

$$ML(e, c) = \frac{count(e, c)}{\sum_{x \in T} count(x, c)}$$

METHOD

THE PREDICTION BY PARTIAL MATCH (PPM) ALGORITHM

Prediction by partial match (PPM: Cleary & Witten, 1984) combines n-gram predictions of different orders.



PPM:

$$p(e,c) = \alpha(e,c)+\gamma_4\cdot\alpha(e,c')+\gamma_3\cdot\alpha(e,c'')+\gamma_2\cdot\alpha(e,c''')+\gamma_1\cdot\alpha(e,[])+\gamma_0$$

EVALUATING PREDICTIVE MODELS USING INFORMATION

Information (Shannon, 1948) can be used to evaluate the predictive models.

- Information content quantifies unexpectedness.
- Entropy quantifies uncertainty.
- Mean information content quantifies model fit.

Information Content:

$$IC(e, c) = -log(p(e, c))$$

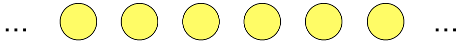
Entropy:

$$H(c) = \sum_{x \in T} p(x, c) \cdot IC(x, c)$$

METHOD

DEALING WITH MULTIDIMENSIONAL SEQUENCES

Multi-dimensional sequences are modelled using Multiple Viewpoint Systems (Conklin & Witten, 1995).



Element	**e1**	**e2**	**e3**	**e4**	**e5**	**e6**
Pitch	G	G	D	B	A	G
Onset	0	4	8	10	12	16
Duration	4	4	2	2	4	4

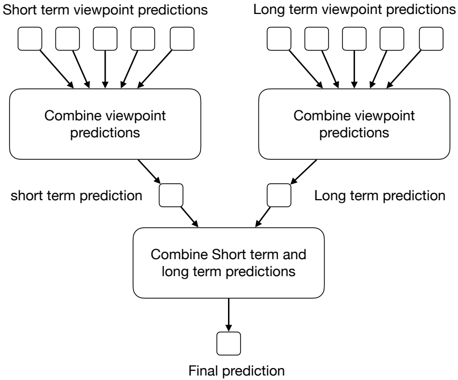
- IDyOM is a Multiple Viewpoint System.
- A viewpoint captures a feature of a sequence.
- Viewpoints can be basic, linked or derived.

METHOD

MODELLING ARCHITECTURE

Viewpoint predictions are combined to create long-term and short-term models.

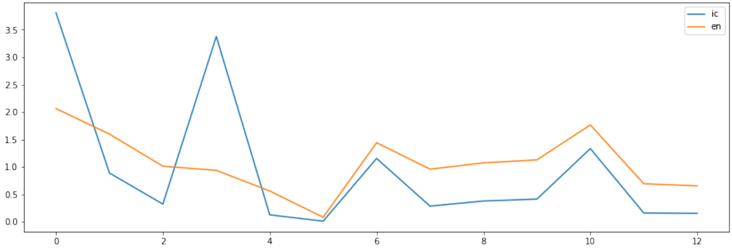
- A short-term model is trained incrementally on a sequence.
- A long-term model is pre-trained with example sequences.



METHOD

SEGMENTATION OF SEQUENCES

Information content and entropy profiles are used to locate the boundaries between statistically salient sub-sequences.



- Expectation plays a role in boundary perception and grouping (Pearce et al., 2010)

OUTLINE

- Introduction
- Modelling Method
- **Application 1: Agent Behaviour**
- Application 2: Music Cognition
- Future Work and Open Challenges

APPLICATION 1: AGENT BEHAVIOUR

OVERVIEW

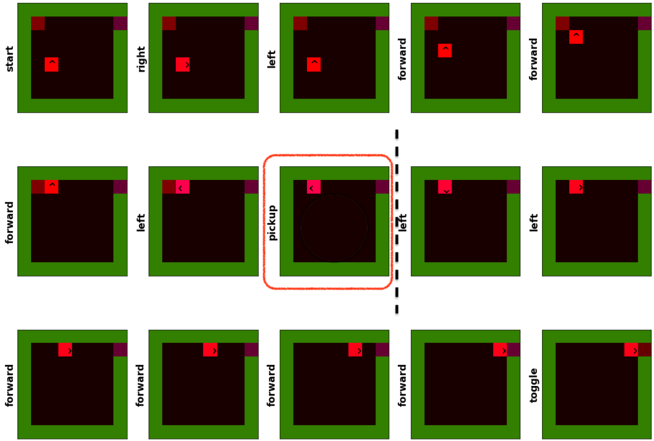
Objective Unsupervised learning of sub-task knowledge from agent behaviour sequence data.

- Partial observability, dynamic environments, hidden rewards and explainability are problems for RL models.
- Incorporating knowledge about sub-tasks can mitigate these problems.

APPLICATION 1: AGENT BEHAVIOUR

A SIMPLE TASK WITH TWO SUB-TASKS

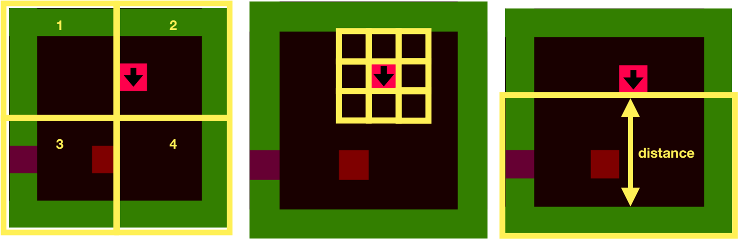
A Gridworld agent must open a door by first picking up a key and then unlocking the door.



APPLICATION 1: AGENT BEHAVIOUR

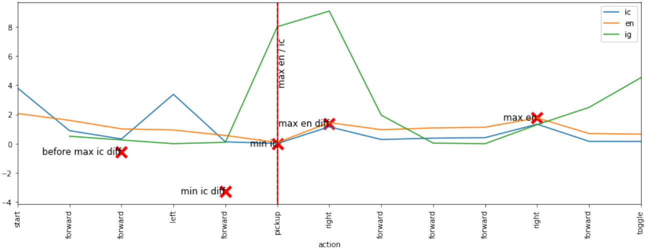
VIEWPOINT REPRESENTATION OF GRIDWORLD

Basic viewpoints capture the current state of the grid (from a birds-eye and a first-person perspective) as well as the last action performed.



APPLICATION 1: AGENT BEHAVIOUR

SUBTASK DETECTION



97% of boundaries correctly identified using minimum information content.

- First-person viewpoints perform best.
- Bi-gram models perform best.
- The short-term models don't help.
- Model selection by mean IC not optimum for segmentation.

APPLICATION 1: AGENT BEHAVIOUR

NEXT STEPS

- Study more complex tasks.
- Define a more general viewpoint representation.
- Perform automatic segmentation.
- Perform classification of segments.

OUTLINE

- Introduction
- Modelling Method
- Application 1: Agent Behaviour
- **Application 2: Music Cognition**
- Future Work and Open Challenges

APPLICATION 2: MUSIC

OVERVIEW

Objective Unsupervised learning of perceptual and cognitive representations of musicological structure from sequence data.

- Music is important and challenging from the perspective of computational creativity.
- Predictive models of music sequences might provide statistical insight into general perception and cognition.

APPLICATION 2: MUSIC

A CORPUS OF 1000 FOLK MELODIES

A corpus of 1000 folk melodies with phrase boundary annotations is used to train a predictive model.

APPLICATION 2: MUSIC

VIEWPOINT REPRESENTATION OF MUSICAL MELODIES

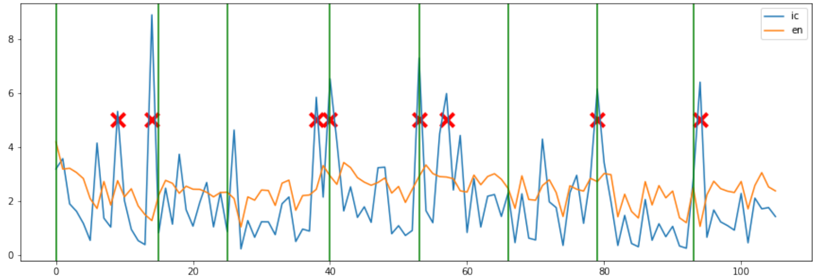
Viewpoints are use to construct an extremely rich representation of musical sequences.

- Basic viewpoints capture note attributes.
- More complex viewpoints capture musical relationships between notes.

APPLICATION 2: MUSIC

PHRASE BOUNDARY DETECTION

Musical phrase boundaries correspond closely with prominent features in the information profiles of melodies.



APPLICATION 2: MUSIC

NEXT STEPS

- Fine-grained segmentation of musical sequences.
- Automatic classification of segments.
- Higher-level predictive models of segment classes.

OUTLINE

- Introduction
- Modelling Method
- Application 1: Agent Behaviour
- Application 2: Music Cognition
- **Future Work and Open Challenges**

FUTURE WORK AND OPEN CHALLENGES

Future Work

- Develop composable and reusable tools.
- Integrate knowledge representation, learning and reasoning.
- Higher-level prediction.

Open Challenges

- What is the best way to generalise viewpoint representations?
- When can two segments be considered equal?

REFERENCES

- Wiggins, G. A. (2020). Creativity, information, and consciousness: The information dynamics of thinking. Physics of Life Reviews, 34–35, 1–39.
- Pearce, M. T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition.
- Cleary, J. & Witten, I. (1984). Data Compression Using Adaptive Coding and Partial String Matching. In IEEE Transactions on Communications, vol. 32, no. 4, pp. 396-402, April.
- Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423 and 623–656.
- Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. Journal of New Music Research, 24(1), 51–73.
- Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. Perception, 39(10), 1367–1391.

THANK YOU!

ARTIFICIAL INTELLIGENCE RESEARCH GROUP

Computational Creativity Lab