Grammar Based Feature Generation for Time-series Prediction

Anthony Mihirana de Silva & Farzad Noorian

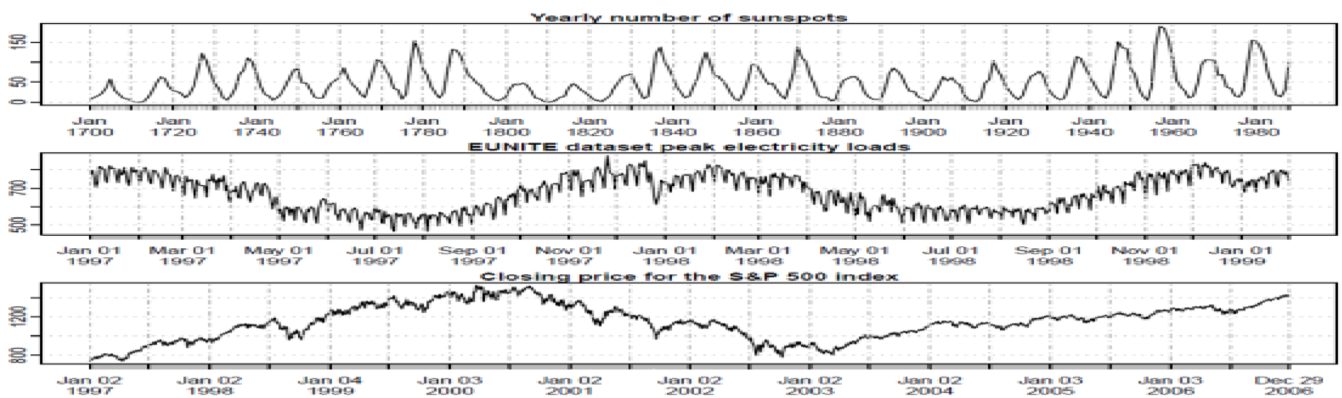
Computer Engineering Laboratory

School of Electrical and Information Engineering

THE UNIVERSITY OF SYDNEY

Time-series Prediction

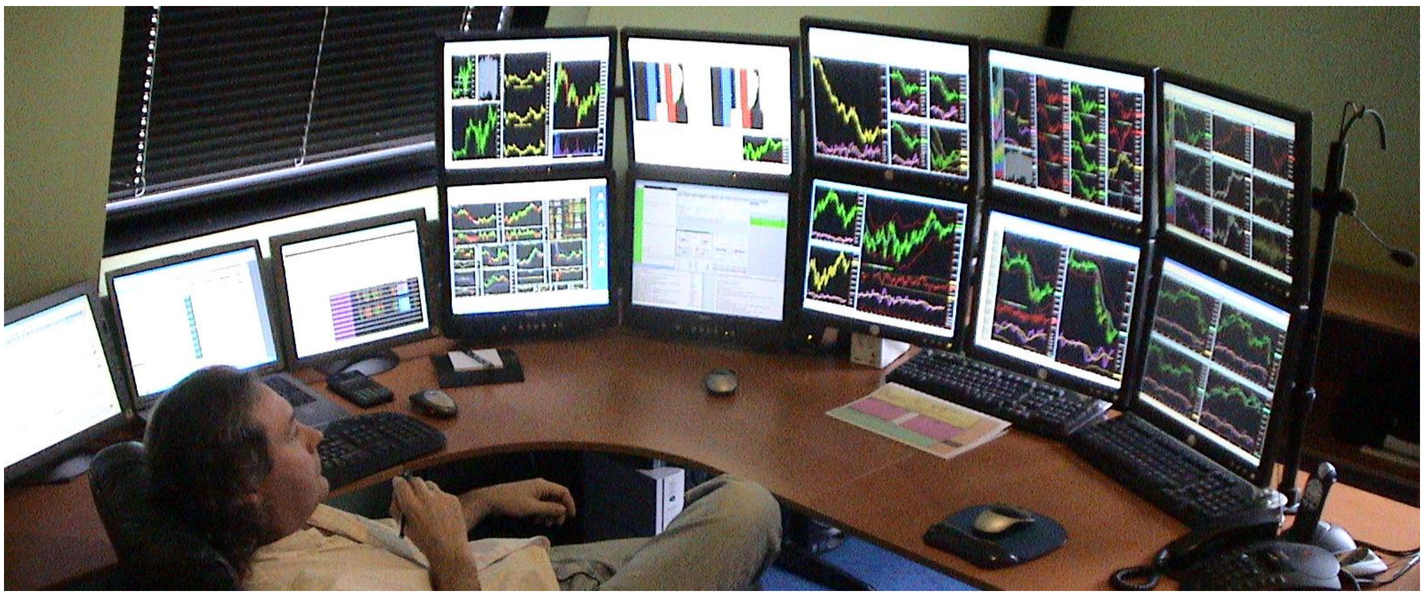
- Time-series: A sequence of vectors (or scalars) recorded at successive points in time.
- The learner is presented with the training samples $(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \ldots, (\boldsymbol{x_n}, y_n)$, where $\boldsymbol{x_i} \in \mathbb{R}^N$ is a vector of features.
- The learner's objective is to find structure in the training samples to infer a general function (a hypothesis) that can predict y_k for previously unseen $\boldsymbol{x_k}$.



Applications - Financial Time-series Prediction

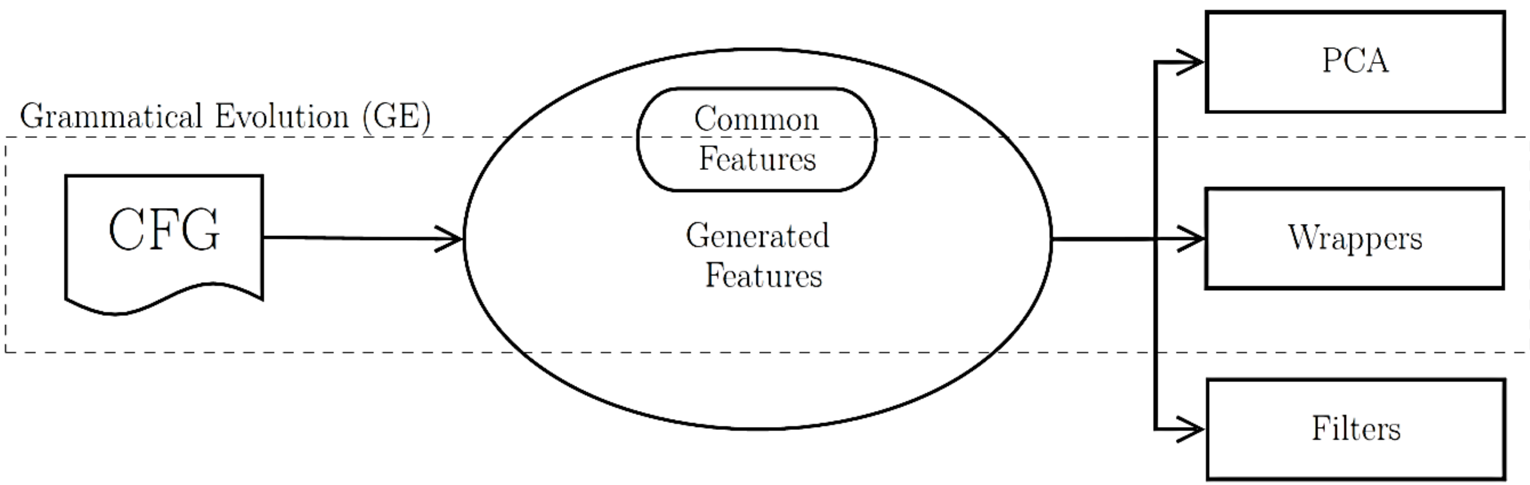
- Motive: Predict financial time-series
- Technical Indicators:
  - Used by "Technical Analysts" (A.K.A Chartists)
  - Identify patterns and market trends in financial markets

- Examples:
  - OHLC: Open/High/Low/Close
  - Moving Averages: $\mathrm{ema}(C_k, n), \mathrm{sma}(H_k, n)$
  - Momentum: $\frac{C_k - L^-_{k-n}}{H^+_{k-n} - L^-_{k-n}}, \frac{\mathrm{sma}(C_k, n_1) - \mathrm{sma}(C_k, n_2)}{\mathrm{sma}(C_k, n_1)}$
  - Volatility: $\mathrm{sma}(H_k - L_k, n) \pm 2\sigma$



Automatic Feature Generation - Why?

- Technical indicators have been widely used as input features to ML algorithms.
  - Technical indicators are a type of "Feature extractors".
- The number of technical indicators selected is not the same with some overlap between different works. However, the choice is generally ad-hoc.
- Can automatic feature generation using an expert defined framework produce better/competitive features?



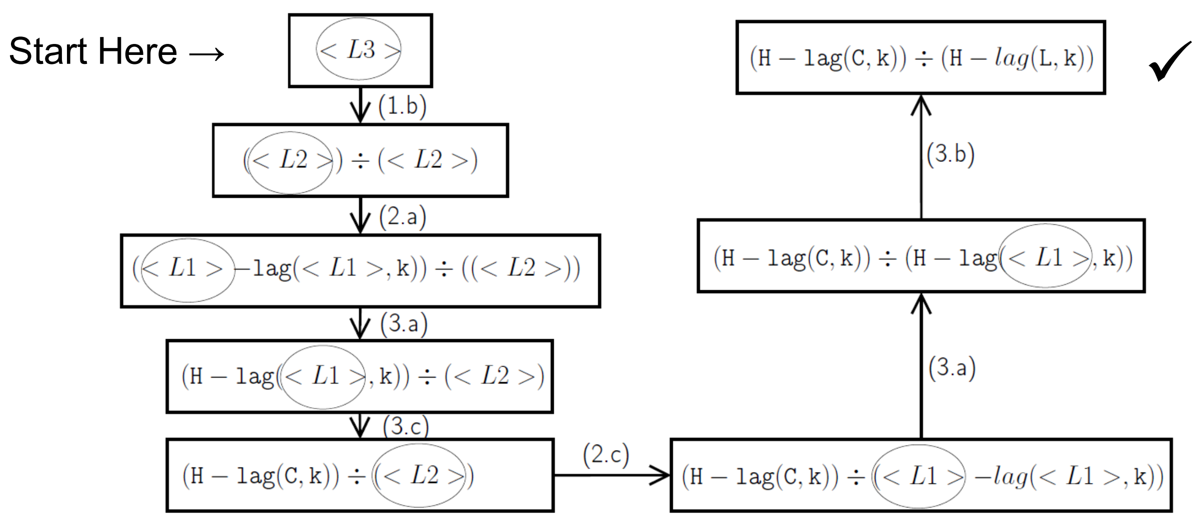
Generation of a Feature, e.g. ADO Indicator

Family 1

$\mathcal{N}$ = {L1, L2, L3}
$\mathcal{T}$ = {− , ÷, lag, sma, meandev, sum, H, L, C, M, n, k, N, (,) }
$\mathcal{S}$ = {L3}

Production rules : $\mathcal{R}$

⟨*L3*⟩	::=	((⟨*L2*⟩) ÷ (lag(⟨*L2*⟩, k)) | ((⟨*L2*⟩) ÷ ((⟨*L2*⟩))	(1.a), (1.b), (1.c)
	|	(((⟨*L2*⟩) − ((⟨*L2*⟩)) ÷ N | ⟨*L2*⟩	(1.d), (1.e)
⟨*L2*⟩	::=	⟨*L1*⟩ − lag(⟨*L1*⟩, k) | ⟨*L1*⟩ − sma(⟨*L1*⟩, n)	(2.a), (2.b)
	|	meandev(⟨*L1*⟩, n) | sum(⟨*L1*⟩, n) | ⟨*L1*⟩	(2.c), (2.d), (2.e)
⟨*L1*⟩	::=	H | L | C | M	(3.a), (3.b), (3.c), (3.d)



More Grammar Families

Family 3

$\mathcal{N} = \{\text{L1, L2, L3}\}$
$\mathcal{T} = \{- \ ,\ \div,\ \texttt{lag},\ \texttt{sma},\ \texttt{meandev},\ \texttt{sum},\ H_h,\ L_l,\ \texttt{C},\ \texttt{n},\ \texttt{k},\ (\ ,\ )\ \}$
$\mathcal{S} = \{\text{L3}\}$

Production rules : $\mathcal{R}$

$$
\begin{array}{lcll}
\langle L3 \rangle & ::= & (\langle L2 \rangle) \div (\langle L2 \rangle) \mid \texttt{sma}(\langle L2 \rangle, \texttt{n}) \mid \langle L2 \rangle & (1.a), (1.b), (1.c) \\
\langle L2 \rangle & ::= & \langle L1 \rangle - \texttt{lag}(\langle L1 \rangle, \texttt{k}) \mid \texttt{sma}(\langle L1 \rangle, \texttt{n}) & (2.a), (2.b) \\
 & \mid & \texttt{meandev}(\langle L1 \rangle, \texttt{n}) \mid \texttt{sum}(\langle L1 \rangle, \texttt{n}) \mid \langle L1 \rangle & (2.c), (2.d), (2.e) \\
\langle L1 \rangle & ::= & H^{+} \mid L^{-} \mid \texttt{C} & (3.a), (3.b), (3.c)
\end{array}
$$

Family 4

$\mathcal{N} = \{\text{L1, L2, L3, L4, L5}\}$
$\mathcal{T} = \{- \ ,\ \div,\ \texttt{lag},\ \texttt{ema},\ \texttt{sma},\ \texttt{meandev},\ \texttt{sum},\ \texttt{H},\ \texttt{L},\ \texttt{C},\ \texttt{n},\ \texttt{k},\ i^{+},\ i^{-},\ (\ ,\ )\ \}$
$\mathcal{S} = \{\text{L5}\}$

Production rules : $\mathcal{R}$

$$
\begin{array}{lcll}
\langle L5 \rangle & ::= & (\langle L3 \rangle \div \langle L4 \rangle) \mid (\langle L3 \rangle \div \texttt{N}) \mid \langle L4 \rangle & (1.a), (1.b), (1.c) \\
\langle L4 \rangle & ::= & \texttt{ema}(\langle L1 \rangle, \texttt{n}) \mid \texttt{sum}(\langle L1 \rangle, \texttt{n}) \mid \texttt{max}(\langle L1 \rangle, \texttt{n}) & (1.d), (1.e) \\
 & \mid & \texttt{min}(\langle L1 \rangle, \texttt{n}) \mid (\langle L1 \rangle) \div \texttt{N} \mid \langle L1 \rangle & (2.a), 2.b), (2.c) \\
\langle L3 \rangle & ::= & \langle L2 \rangle - \texttt{ema}(\langle L2 \rangle, \texttt{n}) \mid \texttt{ema}(\langle L2 \rangle, \texttt{n}) \mid \texttt{meandev}(\langle L2 \rangle, \texttt{n}) & (3.a), (3.b) \\
 & \mid & \texttt{sum}(\langle L2 \rangle, \texttt{n}) \mid \texttt{max}(\langle L2 \rangle, \texttt{n}) \mid \texttt{min}(\langle L2 \rangle, \texttt{n}) \mid \langle L1 \rangle & (3.c), (3.d), (3.e), (3.f) \\
\langle L2 \rangle & ::= & \texttt{H} \mid \texttt{L} \mid \texttt{C} & (4.a), (4.b), (4.c) \\
\langle L1 \rangle & ::= & i^{+} \mid i^{-} & (5.a), (5.b)
\end{array}
$$

Results

Feature Selection using mRMR + Integer GA

Method	FTSE	N225	NDX	HSI	SSMI
ARIMA	34.33	207.07	14.35	138.95	47.96
ETS	33.54	207.13	14.05	139.95	47.14
AR(1)	33.71	207.21	**13.98**	140.24	47.60
EMA	42.63	275.20	14.05	183.51	61.42
SVM (TIs)	40.27	203.99	15.15	136.70	46.26
SVM (Grammar)	**32.64**	**203.30**	14.37	**135.11**	**46.17**

Method	SSEC	TWII	AORD	GDAXI	GSPC
ARIMA	19.78	69.52	29.97	44.19	7.58
ETS	19.78	69.46	29.80	44.08	7.60
AR(1)	19.71	69.41	29.97	44.18	7.61
EMA	28.16	96.55	38.81	58.26	9.39
SVM (TIs)	18.03	**66.85**	29.23	**43.70**	7.54
SVM (Grammar)	**17.96**	67.13	**28.61**	44.49	**7.49**

Table: RMSE for test data for major stock indices using the ARIMA, ETS, AR(1), EMA($p = 5$) and SVM using technical indicators (TIs) and grammar features

Grammar Generated Features

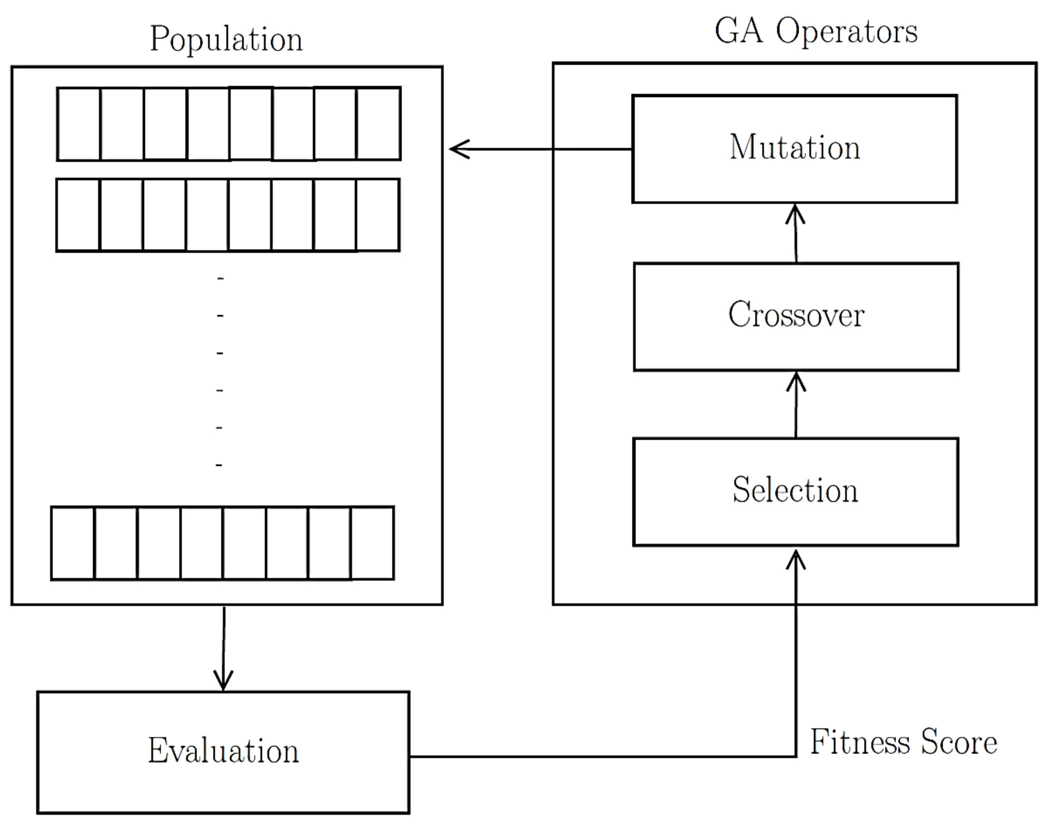
	Feature	Freq.
1	**Disparity**	32
2	C-(sd(lag(C, k), n))	24
3	C-(sd(lag(M, k), n))	23
4	C-(sd(lag(L, k), n))	22
5	H-(sd(lag(C, k), n))	21
6	L-(sd(lag(L, k), n))	21
7	sd(diff(H), n))/(sd(delt(L), n))	21
8	(C-(M-lag(C, k)))/n	20
9	(C-(M-lag(C, k)))/n	19
10	**Bias**	18
11	L-(sd(lag(H, k), n))	18
12	(M-(H-lag(H, k)))/n	18
13	(C-(M-lag(H, k)))/n	17
14	H-(sd(lag(L, k), n))	17
15	L-(sd(lag(M, k), n))	17
16	(L-(H-lag(L, k)))/n	17
17	(sd(diff(H), n))/(sd(delt(M), n))	17
18	**Lower Bollinger Band**	17
19	(C-(meandev(M, n)))/n	16
20	(H-(L-lag(H, k)))/n	16
21	M-(sd(lag(M, k), n))	16
22	M-(sma(ema(diff(L), n), n))	16
23	(C-(H-lag(C, k)))/n	15
24	C-(sd(diff(C), n))	15
25	C-(sma(ema(diff(L), n), n))	15
26	L-(sd(diff(C), n))	15
27	sma(L, n) - 2*(sd(L, n))	15
28	(C-(H-lag(H, k)))/n	14
29	(C-(L-lag(C, k)))/n	14
30	(C-(L-lag(H, k)))/n	14
31	(C-(L-sma(L, n)))/n	14
32	**CLV**	14

	Feature	Freq.
33	(lag(M, k))-(sd(lag(H, k), n))	14
34	M-(sma(ema(diff(M), n), n))	14
35	sma(L, n) + 2*(sd(L, n))	14
36	C-(H-lag(L, k)))/n	13
37	C-(H-sma(L, n)))/n	13
38	C-(sd(lag(H, k), n))	13
39	H-(sd(lag(M, k), n))	13
40	(M-(L-lag(H, k)))/n	13
41	(sd(diff(C), n))/(sd(delt(C), n))	13
42	(sd(diff(C), n))/(sd(delt(M), n))	13
43	(sd(diff(L), n))/(sd(delt(M), n))	13
44	sma(M, n) + 2*(sd(M, n))	13
45	C-(sma(diff(H), n))	12
46	C-(sma(ema(diff(M), n), n))	12
47	H-(sma(diff(H), n))	12
48	L-(sd(lag(C, k), n))	12
49	M-(sd(lag(L, k), n))	12
50	(L-(H-lag(C, k)))/n	12
51	(M-(L-lag(C, k)))/n	12
52	(sd(diff(M), n))/(sd(delt(L), n))	12
53	**Upper Bollinger Band**	12
54	**Aroon**	11
55	(C-(H-lag(M, k)))/n	11
56	(C-(M-lag(L, k)))/n	11
57	(H-(meandev(L, n)))/n	11
58	**Lagged closing price**	11
59	H-(sd(diff(L), n))	11
60	H-(sd(lag(H, k), n))	11
61	M-(sd(lag(C, k), n))	11
62	M-(sma(ema(diff(H), n), n))	11
63	(sum(L, n))/(max(i^+, n))	11
64	sma(H, n) + 2*(sd(H, n))	11

Table: Technical indicators and selected grammar feature frequency

Introduction to Evolutionary Algorithms

- Good in solving NP hard problems, e.g. Feature selection
- Many variations:
  - Genetic Algorithms
  - Genetic Programming
  - Differential Evolution
  - Many others …



Grammatical Evolution

› Phenotype to genotype mapping on a Grammar

11011110	01111101	01000111	00101100	10011010	00001011	10110101	11101001	Binary Chromosome
222	125	71	44	154	11	181	233	Integer Chromosome

$$
\begin{aligned}
\langle expr \rangle &::= (\langle expr \rangle)\langle op \rangle(\langle expr \rangle) \mid \langle var\text{-}op \rangle \\
\langle var\text{-}op \rangle &::= \texttt{lag}(\langle var \rangle, \texttt{k}) \mid \texttt{ema}(\langle var \rangle, \texttt{n}) \\
\langle op \rangle &::= \div \mid \texttt{-} \\
\langle var \rangle &::= \texttt{H} \mid \texttt{L} \mid \texttt{C} \mid \texttt{delt(H)}
\end{aligned}
$$

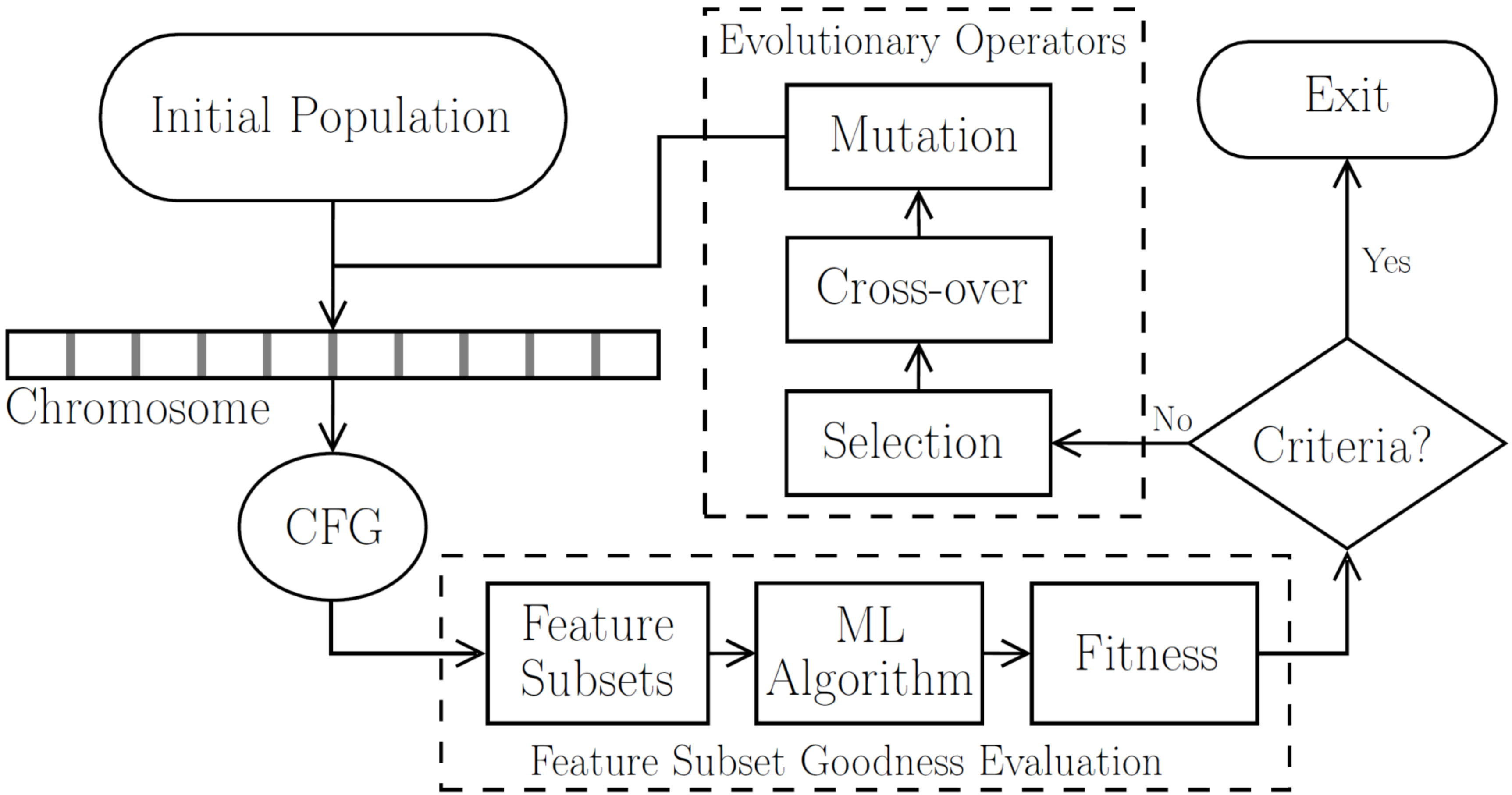
Momentum Grammar Family

<expr> ::= (<expr>)<op>(<expr>)	0
<expr> ::= <var-op>	1
<var-op> ::= lag(<var>, k)	0
<var-op> ::= ema(<var>, n)	1
<op> ::= ÷	0
<op> ::= -	1
<var> ::= H	0
<var> ::= L	1
<var> ::= C	2

Derivation Sequence
(Exapnding the left-most non-terminal)

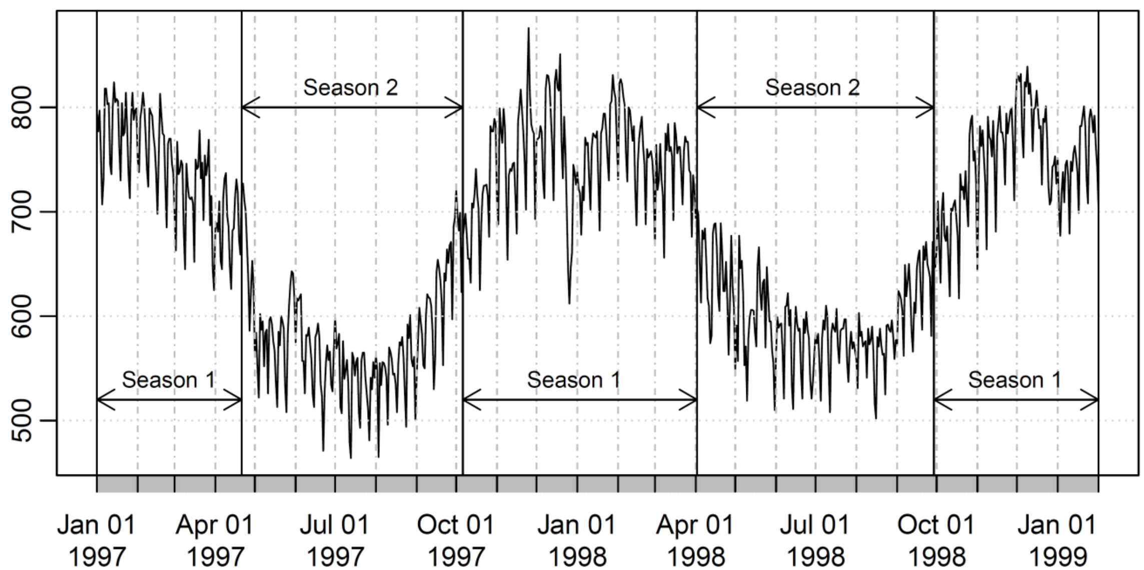
<expr>
$222\%2 = 0$
(<expr>)<op>(<expr>)
$125\%2 = 0$
(<var-op>)<op>(<expr>)
$71\%2 = 1$
(ema(<var>, n))<op>(<expr>)
$44\%3 = 2$
ema(C, n)<op>(<expr>)
$154\%2 = 0$
ema(C, n)-(<expr>)
$11\%2 = 1$
ema(C, n)-(<var-op>)
$181\%2 = 1$
ema(C, n)-(ema(<var>, n)
$233\%3 = 2$
ema(C, n)-(ema(C, n)

Feature Generation Using Grammatical Evolution



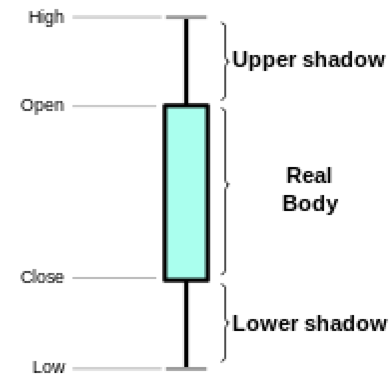
Applications in Electricity Load Prediction: E-UNITE

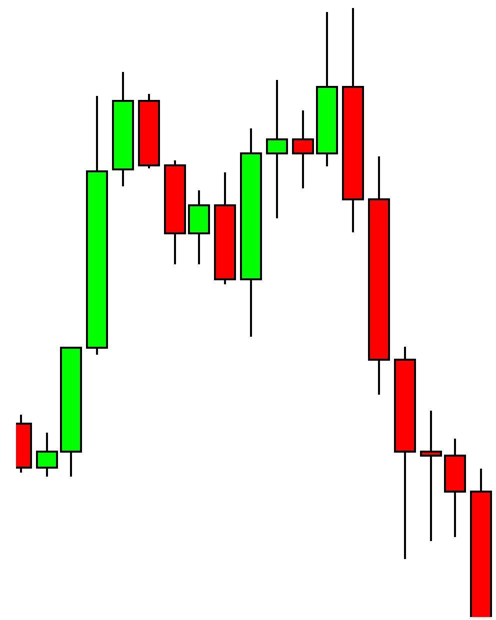
- E-UNITE dataset
- Original Competition: Given 2 years of 30 minutely data, predict a month's daily maximum
  - Previous attempts used complex learners but simple features.
  - Idea: Can feature generation help?



Applications in Electricity Load Prediction: Ideas

- Half-hourly load to daily peak value
- Similar to techniques used by chartists in stock market:
  - Open/Low/High/Close daily prices
- Use the grammar based on technical indicators
  - adding history windows and first differences
  - adding wavelets
- Not suitable for month-ahead





Applications in Electricity Load Forecasting: Results

- Results: Compared for 3 periods
- GE features improve prediction

Method and Features	Month-ahead MAPE %	Day-ahead MAPE %
Last Year's Data	3.43	3.43
ARIMA	3.50	2.62
ETS	3.33	2.23
Linear SVM with $H_i(\forall i \in \{k,\ldots,k-6\})$	2.66	2.00
Polynomial SVM with $H_i(\forall i \in \{k,\ldots,k-6\})$	2.52	2.03
Radial SVM with $H_i(\forall i \in \{k,\ldots,k-6\})$	2.19	**1.89**
Polynomial KRLS with $H_i(\forall i \in \{k,\ldots,k-6\})$	4.03	3.97
Radial KRLS with $H_i(\forall i \in \{k,\ldots,k-6\})$	**2.16**	1.92
Using radial KRLS with feature subsets		
$H_i(\forall i \in \{k,\ldots,k-6\})$ + Temp.	2.90	2.21
$H_i(\forall i \in \{k,\ldots,k-6\})$	**2.16**	1.92
D_k^1, D_k^2, D_k^3, S_k	2.27	**1.84**
$H_k, \Delta H_k, \Delta D_k^1, \Delta D_k^2, \Delta D_k^3, \Delta S_k$	2.36	1.88
$H_k, \Delta H_k, \Delta D_i^1, \Delta D_i^2, \Delta D_i^3, \Delta S_i(\forall i \in \{k-1,\ldots,k-6\})$	2.70	2.04
HLC grammar - best of each month	-	**1.48**
HLC grammar - average of each month	-	1.60
HLC grammar - worst of each month	-	1.68
Wavelet grammar - best of each month	**1.84**	1.81
Wavelet grammar - average of each month	2.08	1.96
Wavelet grammar - worst of each month	2.31	2.26

Conclusion

- Grammar based FS can deliver customized features for a given problem.
- Can be further enhanced using GE.
- Is a framework which can be automated and incorporate domain specific knowledge.
- Applied to financial indices and electricity load prediction with good results.

Publications

- **Full-length Conference Papers**
- Anthony Mihirana de Silva, Farzad Noorian, Richard I. A. Davis and Philip H. W. Leong, A Hybrid Feature Generation and Selection Algorithm for Electricity Load Prediction using Grammatical Evolution, In 12th IEEE International Conference on Machine Learning and Applications (ICMLA), page to appear, Dec. 2013.

- **Journal Papers Submitted**
- Anthony Mihirana de Silva, Richard I. A. Davis, Syed A. Pasha and Philip H. W. Leong, Forecasting Financial Time-series with Grammar Guided Feature Generation, Computational Intelligence, Submitted.