Evolutionary Learning Genetic Algorithm Strategy

Evolutionary Learning Genetic Algorithm Strategy

Genetic Algorithm

- A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The **algorithm** repeatedly modifies a population of individual solutions.
- Genetic algorithms can find near optimal solutions for NP complete type of problems(NP complete means that problem is so complex that the time required to solve it cannot be determined using a polynomial).This method uses natural selection “Survival of fittest” to solve optimization problems
- The main steps of genetic algorithm optimization are
    - The first step involves creating initial population of groups by allocating features as genomes in chromosome. A fitness function is then defined to evaluate each chromosome in solving a defined problem.
    - The next step of the algorithm will create new set of chromosomes using crossover and mutation which will be evaluated using the same fitness function.
    - We then delete a member of population that is less fit (with lesser fitness score) than the new chromosome and insert the new chromosome in the population
    - These steps will then be carried out till maximum number of iterations are achieved and the best chromosome will be returned

Motivation

- Genetic algorithms are a robust adaptive optimization technique based on a biological paradigm and are able to perform efficient search on poorly-defined spaces
- Genetic algorithms use acquired information to direct the search hence as a result there is less chance of getting trapped in the local optima. It can search hypotheses containing complex interacting parts and the optimization can be easily parallelized
- They can also be used for finding the best combination elements to make a good trading system or indicator
- Such an algorithm can be used in cases when there are multiple trade indicators and we need to find the best set of indicators that maximize a given fitness function like Sharpe ratio or winning %.
- Such algorithms are very useful in choosing the optimal set of indicators from a population of a much larger indicators

Evolutionary Learning Genetic Algorithm Strategy

Implementation and Input parameters

- In such a strategy we use a set of technical indicators to find the best group that optimizes a fitness function
- The technical indicators used in our strategy are
  - RSI: It is a momentum indicator that compares magnitude of recent gains to losses in an attempt to determine overbought/oversold signals.RSI produces a score that ranges from 0 to 100
  - Simple Moving Average crossover: Simple moving average crossovers (SMA crossover) between long term average and short term average can provide an indication if the momentum is moving in a particular direction. Generally a buy signal is generated when short term average crosses above the long term average and vice versa
  - William R% indicator: It is a momentum indicator measuring over bought/ over sold signals . It produces a score between 0 t0 -100 and determines entry and exit signal based on the computed score
- For each of these indicators a Long and short threshold is define as follows
  - RSI: Long signal when RSI < 30 and sell signal when RSI >70
  - SMA crossover: Short term and long term horizons are 12 and 26 respectively. Long signal when short term crosses above long term average and short signal when short term average crosses below the long term average
  - William R% indicator: Long signal when william R%>-50 and short signal when william R% < -50

Signal Generation

- After defining the input parameters(technical Indicators) we move ahead with finding the optimal set of technical indicators which maximize a given fitness function
- The fitness function used in our strategy is Sharpe Ratio
- The optimal set containing the technical indicators along with long short signals will have a bit string format shown below

Indicator1|connector| Indicator2|connector| Indicator3|Active

- In the above bit string format
  - **Indicators 1,2,3** are SMA crossover, RSI and WPR respectively, which take values of 1 or 0 for long or short signals
  - **Active** is a binary string of length 3 that contains 1 for indicators that are included and 0 for indicators that are not included
  - **Connector** is a binary string of length 2 that represents a Boolean operator defined as
    - 00 = AND
    - 01 = OR
    - 10 = XOR
- For example if the final output of genetic algorithm is of the form

0| 0 0 | 0 | 1 0 | 1 | 1 1 1 |

IND1 |AND |IND2|OR|IND3|IND1 IND2 IND3|

we observe that all bits in the active string are 1 hence all indicators are included in generating the trade signal , if SMA is short and RSI is short or William R% is long then a long signal is generated if not a short signal is generated

Evolutionary Learning Genetic Algorithm Strategy

Performance Analysis using Daily Prices

- We now go ahead and backtest the strategy using daily prices of NIFTY Index, the backtesting data includes open, high, low and close price from period 17-sep-2007 to 15-dec-2015
- The backtest data is split into two sets, set 1 for training the Genetic algorithm and set 2 for testing the out of sample performance
- The training and testing data is split in the ratio of 80:20 respectively, more specifically we use training data from period 17-sep-2007 to 9-Apr-14 and testing data from period 10-apr-14 to 15-dec-15
- The final output bit string i,e the optimal set generated by testing data is shown below

0	0	1	0	1	0	1	1	1	1
IND 1	OR		IND 2	XOR		IND 3	ACTIVE		

- The table below shows the performance of the strategy using daily prices(a comment in the performance)

APR	Annual Vol	APR/Vol	Max DD	Calmar
22.24%	15.15%	1.47	-9.13%	2.44

Return Profile of Daily Strategy

Evolutionary Learning Genetic Algorithm Strategy

Performance Analysis using Intraday 10 Minute prices

- We now apply the algorithm to intraday data at 1 minute and 5 minute intervals and analyze its out of sample performance
- As before the backtest data includes intraday 1 minute and 5 minute interval open, high, low and close prices. The methodology used for training and testing is a bit different from the one used for daily prices, in this strategy we split training and testing data in the ratio 90:10, more specifically
    - For intraday strategy using 1 min prices we use training data from period 9-sep-2015 to 10-Dec-15 and testing data from period 11-Dec-14 to 18-Dec-15
    - For intraday strategy using 5 min prices we use training data from period 8-Jun-2015 to 1-Dec-15 and testing data from period 2-Dec-14 to 18-Dec-15
- The reason behind using a larger data horizon for strategy using 5 min intraday data is to have adequate data points for training the algorithm
- We increase the allocation to training data since using intraday day data we have more data points as compared to daily prices and hence more data for training and testing the algorithm
- The final output bit string for 1 minute and 5 minute intraday data generated by testing data is shown below

1 Minute Interval	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**1**	**1**
	IND 1	**AND**		**IND 2**	**XOR**		**IND 3**	**ACTIVE**		
5 Minute Interval	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**1**	**1**
	IND 1	**AND**		**IND 2**	**XOR**		**IND 3**	**ACTIVE**		

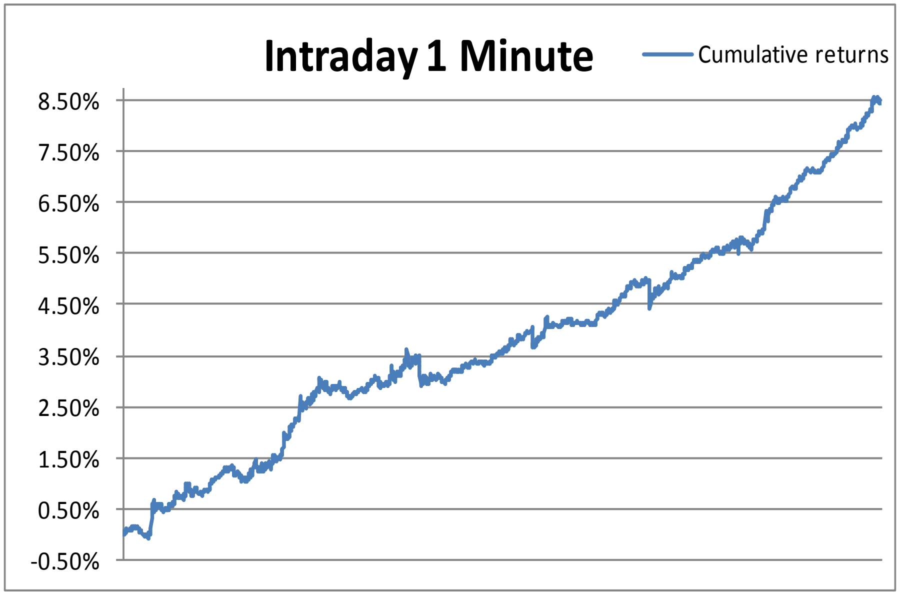
- Since both the strategies are subjected constant intraday rebalancing we are expected to incur significant transaction cost, from historical observation the average Bid-Ask spread observed in NIFTY is around 60 paisa(cents/pence).
- The table below shows the performance of both the strategies with and without the impact of transaction costs

Interval	Transaction cost	Return	APR	Annual Vol	APR/Vol
1 Min	No	8.86%	2584.37%	11.26%	229.56
1 Min	Yes	2.38%	148.79%	11.65%	12.77
5 Min	No	2.39%	55.46%	10.68%	5.19
5 Min	Yes	2.33%	55.33%	10.70%	5.17

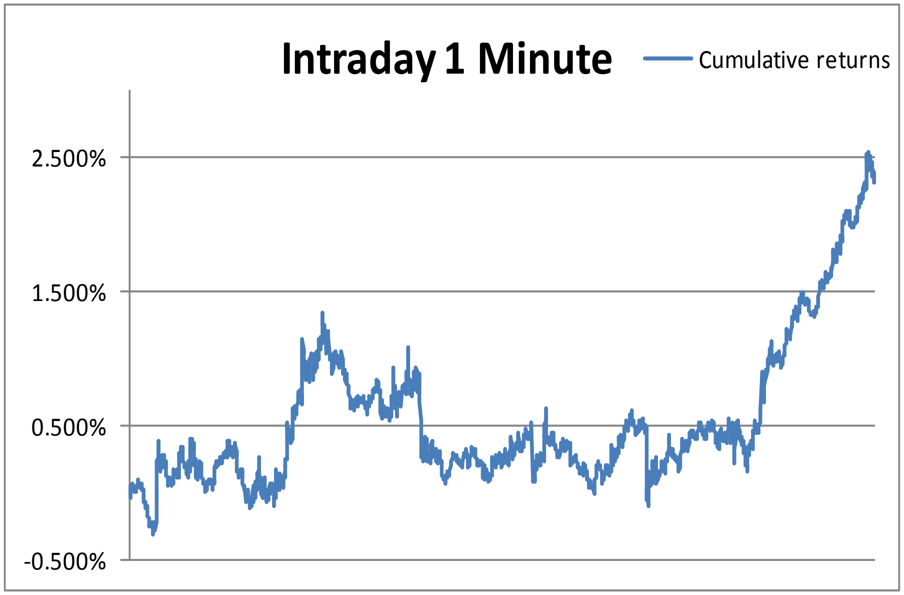
- From the table as expected we observe that impact of transaction cost on strategy using 1 min intraday data is more pronounced, as it will rebalance more frequently than strategy on 5 min intraday data

Evolutionary Learning Genetic Algorithm Strategy

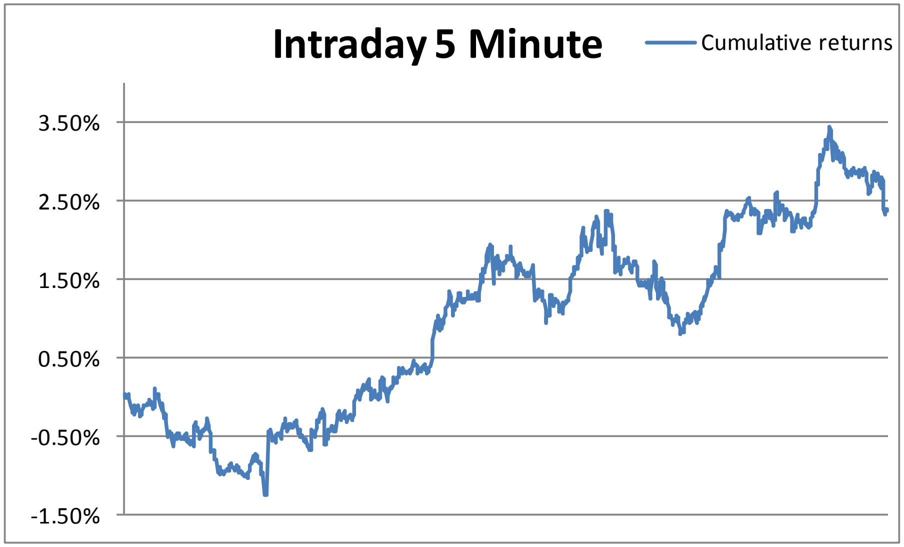
Return profile 1 minute interval without Transaction cost



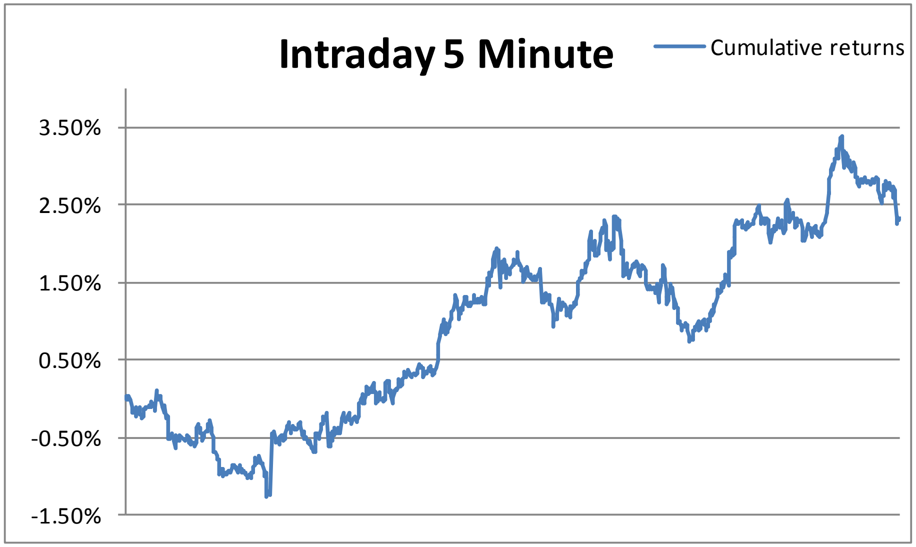
Return profile 1 minute interval with Transaction cost



Return profile 5 minute interval without Transaction cost



Return profile 5 minute interval with Transaction cost



Evolutionary Learning Genetic Algorithm Strategy

Future scope of strategy

- In this strategy we limited ourselves to just 3 indicators i.e RSI, SMA crossover and william R% respectively, which give information only about the state of momentum in the market. The scope of this strategy can be significantly expanded by using a larger and a more diverse set of indicators
- Ideally the choice of indicators should be such that it gives as much information about the state of the market as possible. Such an algorithm can be easily scaled up to include 20-30 indicators
- The set or type of features/indicators that can be added to make this strategy more effective are
  - **Order book information:** We can create features that indicate state of order book, the features or indicators that provide information about best bid/ask sizes, order book liquidity measures, the example of such features are show below
    - **VWAP Bid/Ask spread Indicator:** we can create a feature which indicates the order book imbalance using VWAP bid/ask spreads defined as $VWAP^{BID}$ spread=Trade price – $VWAP^{BID}$ and $VWAP^{ASK}$ spread= $VWAP^{ASK}$ - Trade price.

      So if as $VWAP^{BID}$ spread < $VWAP^{ASK}$ spread, then actual market price is lower than equilibrium price suggested by the order book hence one would expect the price to adjust upwards and downwards when $VWAP^{BID}$ spread < $VWAP^{ASK}$ spread
    - **Intraday Liquidity measures** : we can also use some intraday liquidity indicators like Rolls' spread, order ratio, quote slope etc.
  - **Short term forecasting outputs:** we can use some forecasting models which would themselves use intraday data to come up with long/ short signals. A few example of such models as shown below
    - Multifactor regression models:
    - Vector Auto-Regressive models:
  - **VIX index Features:** Since its historically observed that implied vols and underlying prices are correlated, we can develop features based on VIX index to generate long/short signals in the underlying i.e. NIFTY index
  - **Additional technical indicators :** In addition to technical indicators used in this strategy, additional technical indicators like ***Sentiment indicator*** , MACD, stochastic oscillators, ROC(rate of change), EMA crossover etc can also be included to expand the set of input features
- As observed the strategy performs relatively well in it's out of sample backtesting horizon providing high returns . Such a strategy has a very high potential to provide super normal returns by including a larger and a more diverse set of features, additionally using more features will also make this strategy more robust which could perform in multiple market regimes. However a monthly or fortnightly recalibration must be done to make the algorithm tuned to more recent market patterns