Candlestick encoding

Reindeer Flotilla

# Nomenclature

The principal components that we consider in a candlestick diagram are the candles. They look like the drawing below, where its main parts are named body and shadows.

Candlesticks are usually composed of the body (black / white or green / red), and an upper / lower shadows. The area between open and close is called the real **body**, price excursions above and below the real body are **shadows**. Shadows illustrate highest and lowest traded prices of a security during the time interval represented. Body illustrates the opening and closing trades.

The range between *open* and *close* the *OC* range.

The particular range between any *min* and *max* in an individual tick will be called .

To encode an OHLC tick and its corresponding candlestick, the following pieces of information are needed:

* range between the absolute minimum and maximum values found in the entire time series. If *min* *max*, the body size will be negative.
* range between min and max values of any particular tick.
* Direction and length of change wrt. previous candlestick.
* Body position
* Upper shadow size.
* Lower shadow size.

All these elements result in a three-components encoding: , body position and direction. The first three are encoded considering the previous element in the series. The result is 3-tuple that will encode candlestick dynamics. Size is relative to the previous elements, thus giving information about whether this candlestick is greater or smaller than previous one. Position is telling us how many shadows it has, whether the body is centered or shifted, and its relative size compared with the range. Finally, direction informs about whether this candlestick represents an increase or decrease of the different ranges with respect to the previous one.

## Normalization

We must consider what is the maximum range found in the time series, between the *min* and *max* of an individual tick. Considering that range, we will normalized the entire time series. We will call the range between *min* and *max* the range.

# High-Low range –Size

## Absolute or relative Encoding

To encode range size, it is possible to use **absolute** or **relative** encoding.

Absolute

determines what the current range size with respect to the maximum range found in the series () as a discrete value for the percentage of : 0, 10, 25, 50, 75, 100 or more.

Relative

encodes the body size as a percentage of the range, again, as a discrete value for the percentage of : 0, 10, 25, 50, 75, 100 or more.

Relative encoding may refer to the percentage of change in size of the current candlestick with respect to the previous one. That way, encoding provides contextualized information about the current *tick*, instead of simply informing about the individual characteristics of the element.

## Size encoding

Options for the size encoding are:

{P|N}{0..6}

The first character indicates greater or smaller range with respect to the previous one, and the range between 0 and 6 indicates whether that increase/decrease of size is 0%, 10%, 25%, 50%, 75%, 100% or more.

# Candlestick body position

## 0

### Two-shadowed

Let’s start by those covering different ranges of the *min*-*max* range (10%, 25%, 50%, 75% and 100%), and *open*-*close* range centered between them. By *centered* we can assume that it can be considered as with equal-length shadows at both sides of the body.

Now let’s consider those candlesticks that are not centered but shifted either upwards/downwards. Base assumption is that the range *open*-*close* is 50% of the range between *min*-*max*.

### One-shadow

Now, consider ranges between *open*-*close* that exceed the 50% limit, so they are on values around 75% of the *min*-*max* range.

### Zero-shadows

It is also possible that ranges matches the *OC* range. In that case we find zero shadows or complete match between the open-close interval and the min-max.

In that case, body position and size might be encoded with respect to the maximum range, but that will not add information about the current position in the trade. Alternatively, *ticks* with zero shadows will be encoded only by the body size. Shadows and position will be considered as zero.

## = 0

The case where is zero corresponds to an almost non-existent body. To correctly compute this case, it is proposed to consider null body when its *OC* range is less than 1-2% of the range.

## Encoding body

We have three different situations: Two shadows, One shadow or Zero shadow. We will use the first letter to encode number of shadows, and the second one for the position and size of the body.

{T|O|Z}{0..9}

A two-shadowed centered body covering 75% of the range will be encoded as T3, following the order in which the different options have been presented. A two-shadowed not centered body covering 25% of the upper part of the range should be encoded as: T6.

A one-shadow candlestick with body covering 10% of the lower range should be encoded as O3 (fourth element in the picture –where 0 should be the first one).

And finally, a zero shadows with the OC cross line in the lower minimum of the range is encoded as Z4 (4 to encode the 5th element in the illustration –where 0 should be the first one).

# Candlestick direction

In addition to the distribution and size of candlestick elements, we also have to encode the change of candlestick values with respect to its predecessor.

## Individual components

We have to consider the value evolution of the four components: O, H, L, C. Any of them can increase, decrease or remain steady. Any change in components value will be characterized by a difference with the reference value in by a given percentage over .

To encode any change in direction (positive or negative) and its length we can consider seven possibilities: 0, 10, 25, 50, 75, 100 or , for the new position or size.

To encode the change in position of a new candlestick we do need to consider the four points, and for each of them, the two different possibilities (increase or decrease –steady is coded as positive 0 change), and the grade of difference (one among the seven possible values). This encoding results therefore in possibilities.

## Range shift

The other possibility is to encode the direction and size of the ranges and as shift in a given direction. The first couple of options are that the range will increase its values (both) or decrease them.

But it can also increase or shrink its range without apparent shift upward or downward. In this case, the endpoints show a contradictory movement: one is increasing its value but the other one is decreasing it.

All the possibilities enumerated in this section lead to nine different combinations:

* shift upward, same size.
* shift upward, shrink.
* shift upward, expand.
* shift downward, same size.
* shift downward, shrink.
* shift downward, expand.
* shrink, no shift.
* expand, no shift.
* same size, no shift.

## Encoding movement

To encode the 9 possibilities, a two letter encoding is proposed. The first one determines size behavior (shrink, expand or no change), and the second one the direction (upward, downward and no shift).

{S|X|N}{U|D|N}

# Optimal Encoding for maximum distance between distinct elements

However, the total 45 possible candlestick codes need to be encoded in a way that maximizes the distance (hamming or levenhstein) between elements with big differences.