Linguistic summaries of time series

Paweł Gelar, Jakub Tarkowski

Warsaw 01.2023

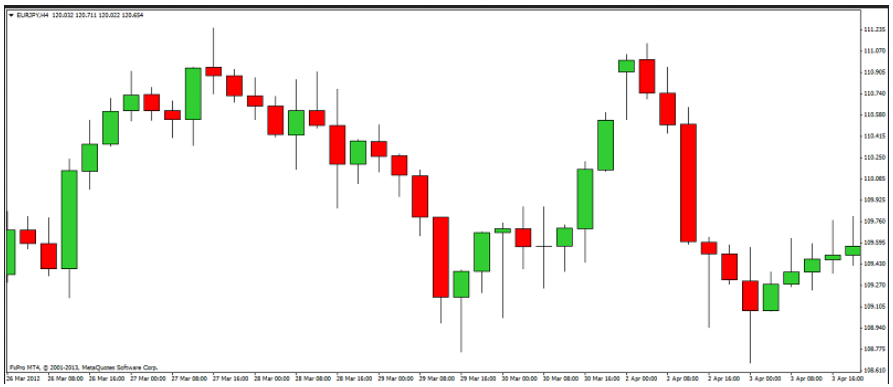
Contents

What are linguistic summaries of time series?

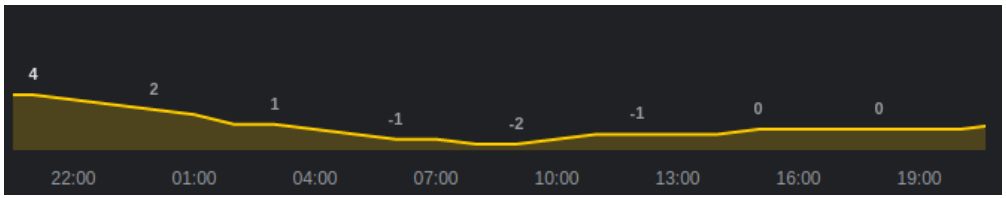
Time series

Examples

- Stock market



- Weather forecast



Linguistics summaries

Examples

- "The weather outside is nice."
- "Almost everyone scored high on the exam"
- "Among all short segments, most are slowly increasing"

Motivation

- Linguistic summaries are natural for humans.
- They are easier to grasp than raw data.
- Sometimes only text/audio communication channels are available
- Some people may have troubles with reading charts correctly.

Restrictions

1. Trends are linear.
2. Sentences are build using protoforms.

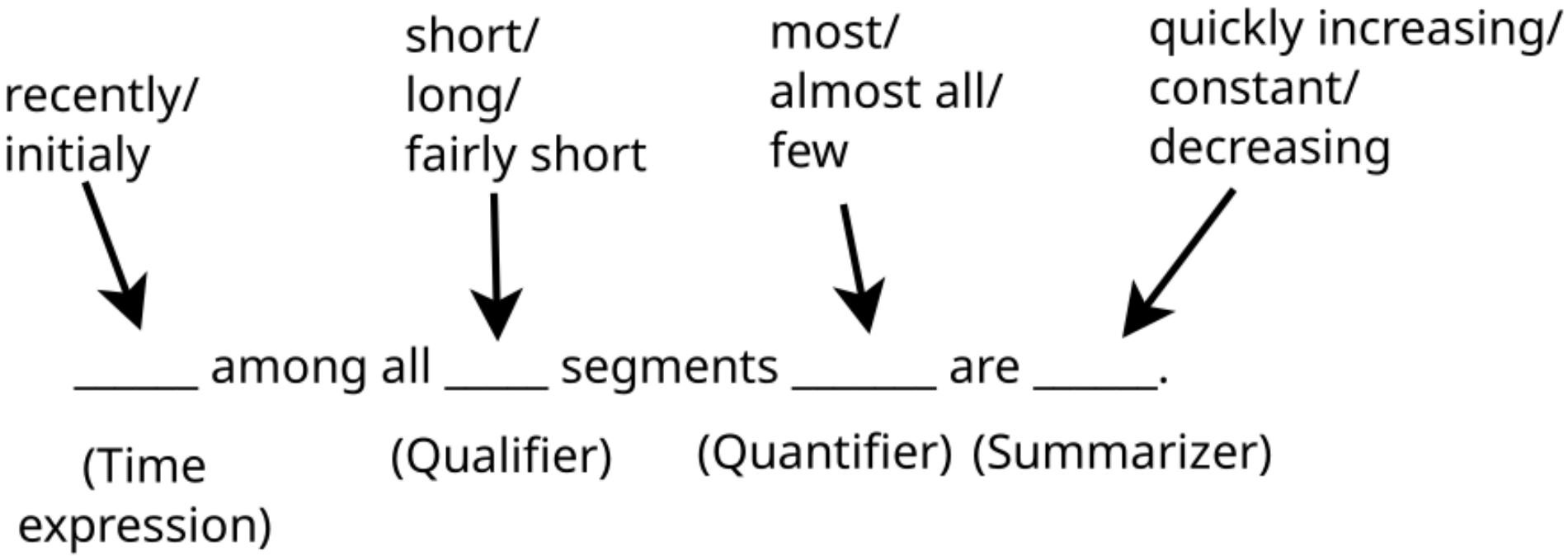


Figure: Protoform with examples of sentences

Creating time series summary step-by-step

Different approaches to segmenting time series

1. On-line (sliding window) algorithms
   - Broken line
   - **Cone based**
   - Linear regression
2. Global algorithms
   - top-down
   - bottom-up
3. Other
   - SWAB (Sliding Window and Bottom Up)
   - Evolutionary Algorithms

Cone based algorithm

Algorithm

1. Draw circles with radius ϵ (hyperparameter) aroud each point
2. Draw angles starting in the first point containing those circles until their intersection is empty
3. Start constructing new trend with the last point as starting point

Algorithm visualisation

Figure: Step 1

Algorithm visualisation

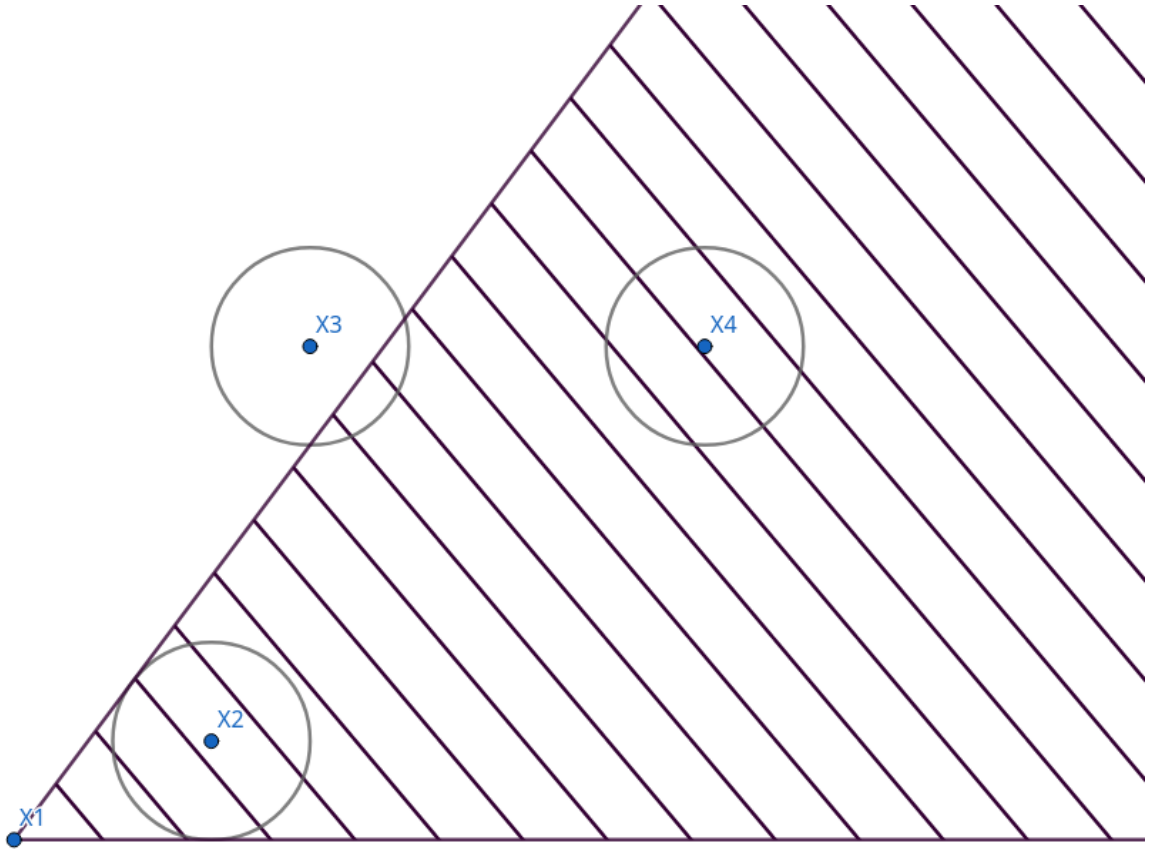


Figure: Step 2

Algorithm visualisation

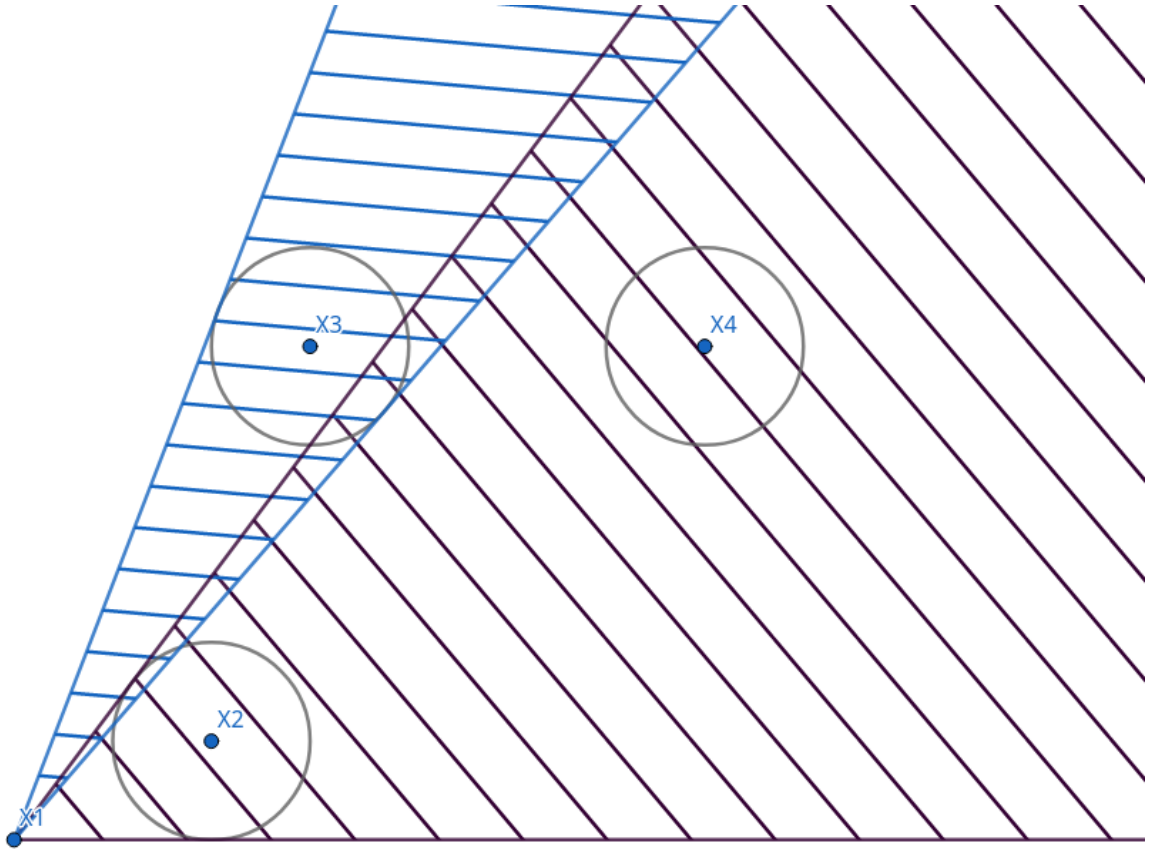


Figure: Step 3

Algorithm visualisation

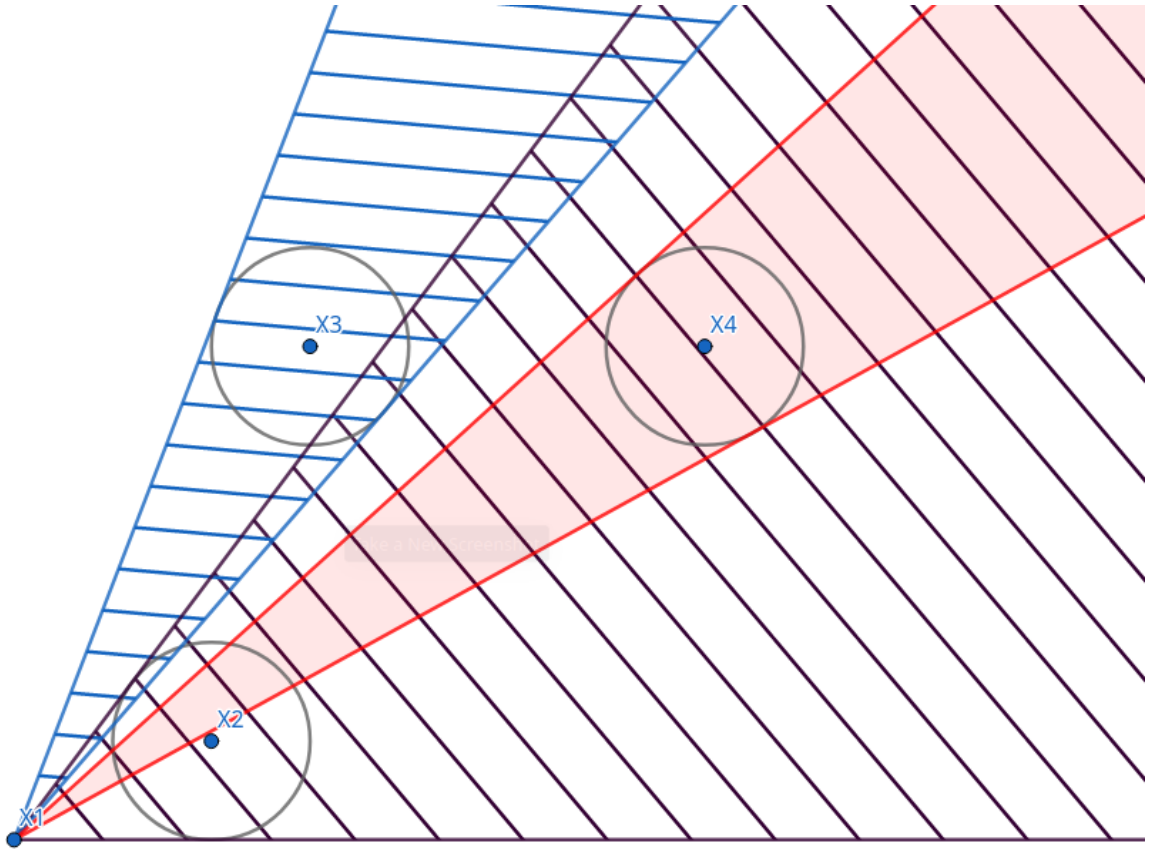


Figure: Step 4

Algorithm visualisation

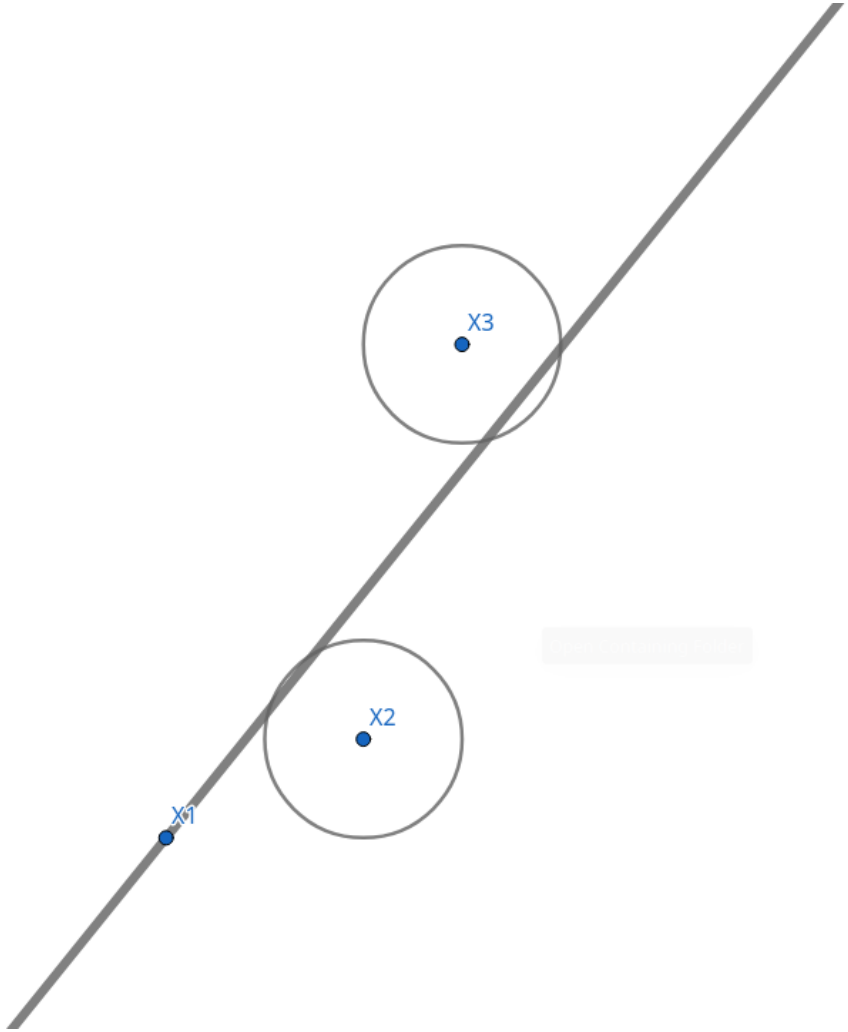


Figure: Step 5

Extracting features of segments

1. Dynamics of change - slope of the trend line
2. Duration - length of the segment
3. Variability
   - range
   - IQR
   - variance
   - standard deviation
   - MAD

All of the above can be easily represented as a fuzzy set.

Evaluation of the linguistic summaries

Available metrics:

1. **degree of truth (validity),**
2. **degree of imprecision,**
3. degree of specificity,
4. degree of fuzziness,
5. **degree of covering,**
6. **degree of focus,**
7. **degree of appropriateness,**
8. degree of informativeness
9. **length of the summary**

Validity

The most important metric of linguistic summaries.

Sentence with low degree of truth

Most of ravens are white.

Imprecision

It describes how many objects can be described by qualifier and summarizer separately (how imprecise they are)

Sentence of high imprecision

Young or old people usually are high or short.

Does not depend on the data to be summarized, but only on the form of a summary and the definition of linguistic values.

Focus, Covering (support)

Degree of focus depicts how common is given qualifier

Sentence with low focus

Among all women those studying civil engineering have blue eyes.

Degree of support depicts how common are objects satisfying both qualifier and summarizer.

Sentence with low support

Women with blue eyes study civil engineering.

Appropriateness

The degree of appropriateness indicates to which degree the obtained summary is surprising to us.

Sentence with low appropriateness

Among all MiNI students those with even index number have mostly good grades. [if all MiNI students have mostly good grades]

Summaries have low degree of appropriateness when qualifier and summarizer are independent.

Sentence with high appropriateness

Among all MiNI students those with even index number have mostly good grades. [if half of all MiNI students have mostly good grades]

Length

Different measures proposed, but the idea is pretty self-explanatory.

First sentence

Orange cats are long.

Second sentence

Among all orange and fast cats, almost all are long.

Evaluated summaries

Summary	truth	focus	inf.	appr.
Among all y, most are short	1.0000	1.0000	0.4500	0.0700
Among all low y, most are long	1.0000	0.3202	0.5000	0.0663
Among all moderate y, most are short	0.8922	0.3166	0.6067	0.0652
Among all decreasing y, most are low	0.8798	0.3578	0.4660	0.0506
Among all increasing y, most are low	0.8701	0.2966	0.5147	0.1056
Among all increasing y, most are short	0.8901	0.3726	0.4756	0.0110

Quality criteria

Example ranking of the criteria:

1. truth (0.4589)
2. informativeness (0.2398)
3. focus (0.1757)
4. appropriateness (0.1256)

Ranked summaries

Order	Summary
1	Among all y, most are short
2	Among all low y, most are long
3	Among all moderate y, most are short
4	Among all increasing y, most are short
5	Among all increasing y, most are low
6	Among all decreasing y, most are low

Thanks for your attention

Questions?

Literature

1. A. Wilbik, *Linguistic summaries of time series using fuzzy sets and their application for performance analysis of mutual funds*, Ph.D. Dissertation, Systems Research Institute, Polish Academy of Sciences, 2010
2. J. Kacprzyk, A. Wilbik, S. Zadrożny, *On Linguistic Summarization of NumericalTime Series Using Fuzzy Logic with LinguisticQuantifiers*, System Research Institute, Polish Academy of Sciences