Generating Comparative Explanations of Financial Time Series

Jacopo Fior(✉), Luca Cagliero, and Tommaso Calò

Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy
jacopo.fior@polito.it

Abstract. Private and professional investors can easily access large amounts of financial data describing the temporal evolution of the stock prices. Making appropriate decisions about financial activities often entails performing comparative studies to get an increased comprehension of the underlying assets. The aim of this work is to automatically generate summarized explanations of financial stock series based on the most established fundamental indicators. Unlike any previous summary protoform, the newly proposed time series explanations (i) are suited to comparative analyses, i.e., they express a relative strength of the summary claim about a given stock compared to a reference stock cluster, and (ii) are based on a time series embedding representation indicating the level of similarity between different stocks/stock groups in various periods. The preliminary results demonstrate the usefulness and applicability of the proposed approach.

Keywords: Time series explanation · Time series embeddings · Data summarization

1 Introduction

One of the most labour-intensive activities of financial investors is to explore market-related data, such as financial reports, stock price series and macroeconomic indicators [3,11]. To get an increased comprehension of the underlying assets investors are very interested in getting readable explanations of time-variant events. The present work focuses on generating explanations of stock price series in textual form. Formulating the resulting summaries in natural language allows human users to better understand the temporal evolution of the analyzed series and to effectively support decision-making.

Standard protoforms are the most popular way to summarize time series in textual form. They exemplify relevant patterns in databases using explainable summary templates [8]. For instance, "*The samples of Time Series T acquired in the last week are very similar to those observed in most of the previous 10 weeks*" is an example of protoform, where *last week* is the time window under consideration, whereas *very similar* and *most* are respectively denoted as protofom *quantifier* and *summarizer*.

S. Chiusano et al. (Eds.): ADBIS 2022, LNCS 13389, pp. 121–132, 2022.
https://doi.org/10.1007/978-3-031-15740-0_10

Recently proposed protoform-based approaches automatically generate protoforms from time series data by adopting clustering and fuzzy modelling [1,4]. The main drawbacks of state-of-the-art solutions are enumerated below:

- Standard protoforms are not suited to perform comparative analyses of time series data at multiple granularity levels (e.g., "*The samples of Time Series T acquired in the last week are very similar to those observed in the time series group G*").
- Since protoform quantifiers and summarizers are defined using static domain-specific rules, their values do not necessarily reflect the underlying data distribution.
- Protoforms are not tailored to the financial domain. Hence, they do not consider domain-specific aggregations (e.g., fundamental indicator levels, market sector and sentiment).

Paper Contribution. This work aims at generating *comparative explanations* of the financial stock price series by exploiting a self-supervised time series embedding representation. The key idea is to first encode the underlying stock series characteristics (e.g., price trend, seasonality, momentum, sentiment about the stock) into a unified vector space and then summarize the key differences between single stock vectors and the encoding of a stocks belonging to a reference group (e.g., the stocks of the same sector with highest operating profit). Quantifier and summarizer values are both dynamically defined based on a data-driven approach on top of the inferred latent space.

Running Example. The summary

Stock S is very similar to most of the most virtuous stocks of year 2020 for EBITDA indicator

compares the historical price series of a specific stock S with those of a group of correlated stocks clustered by means of an established fundamental indicator (EBITDA). Notice that the comparative term *very similar*, i.e., the summarizer, synthesizes the observed level of similarity between S and the reference group, whereas the quantifier *most* indicates the required level of similarity. *year 2015* indicates the reference time period in which the statement holds.

The self-supervised procedure of time series encoding is applied to both historical stock series and stock-related news data. It synthetizes the key information about a stock on a daily basis. The purpose is to inherently capture not only the observed stock price trends, seasonality, and momentum but also the underlying market movers (e.g., the sentiment of the main market actors).

To empirically evaluate the effectiveness and applicability of the proposed approach we carry out both intrinsic and extrinsic evaluations. Specifically, in the intrinsic evaluation the generated summaries are first shortlisted using established protoform-based metrics and then evaluated with the help of a domain expert. In the extrinsic evaluation, we backtest the reliability of the generated stock recommendations. The preliminary results achieved on the U.S. stock market confirm the applicability of the proposed approach to support stock trading activities.

The rest of the paper is organized as follows. Section 2 overviews the prior works. Section 3 describes the analyzed financial data. Sect. 4 presents the proposed method. Section 5 summarizes the main empirical outcomes, whereas Sect. 6 draws the conclusions of the presented work.

2 Literature Review

The generation of explainable summaries can be based on static domain-specific rules, statistic/probabilistic approaches, or neural methods [10]. A joint effort of the Deep Learning community has been devoted to relieving experts of the definition of static rules by leveraging data and algorithms. Albeit state-of-the-art probabilistic/statistical approaches and neural methods generate the text automatically, the quality and readability of the output summaries is not always guaranteed. For these reasons, most existing time series summarization techniques still partly rely on rule-based methods, which generate standard summary templates called protoforms [1,2,4,7]. For instance, [2,7] generate narratives of data that summarize the key series trends (e.g., increasing, decreasing), whereas [1] uses Evolutionary Genetic Algorithms to explore the set of candidates summary templates and pick those meeting specific (user-specified) constraints. More recently, [4] adopt sequence pattern mining and clustering techniques to support the generation of protoforms. This work presents an hybrid approach to time series summarization that combines the reliability of rule-based strategies with the flexibility of neural NLP methods. Specifically, it leverages a high-dimensional vector representation of the time series, generated by an ad hoc embedding models [13], to dynamically construct summaries providing comparative explanations. To the best of our knowledge, the use of neural network-based approach to define comparative summaries of financial time series has never been proposed so far.

3 Data Overview

To generate explainable summaries of financial data we analyze stock-related data under multiple aspects, i.e., the raw time series of historical prices and exchange volumes, the most established price trend and volatility indicators, the news sentiment, and the values of main fundamental indicators.

Time Series Data. We focus on the time series T_s of the daily closing prices of each stock s belonging to the Standard&Poor (S&P500) index. In our study we consider historical stock data spanning from 2007 to 2018[1].

[1] We crawled data from AlphaVantage (https://www.alphavantage.co/). In the considered time span historical data are available for 468 out of the 500 firms.

Price-Related Indicators. We consider the following technical indicators describing the momentum, trend, and volatility of the stock prices [12].

– Exponential Moving Average (EMA) with 5, 20, 50, and 200 periods.
– Moving average convergence divergence (MACD) with the following EMA combinations: (5, 20), (20, 50), and (50, 200).
– Relative Strength index (momentum oscillator) with cutoff thresholds 30% (over-sold) and 70% (over-bought).
– Aroon oscillator (trend descriptor).
– Accumulation/distribution indicator (price and volume divergence).

News Sentiment. We analyze the sentiment ss of the news related to each stock and compute an average per-day and per-stock sentiment score between -1 and 1. A positive score ($ss(s) >> 0$) indicates a positive sentiment of the market about the stock s, whereas a negative score ($ss(s) << 0$) provides a negative feedback. In our experiments, we collect English-written news on the S&P500 stocks in the period 2007–2018 from Reuters[2] and apply VADER to perform rule-based sentiment analysis [6]. We consider around 5253 news articles per stock. Notice the number of daily news per stock is rather variable, as most popular stocks are more likely to be cited.

Fundamental Indicators. They describe the economical and financial factors that mainly influence the stock and the underlying assets. In this study we consider the following established indicators: Earnings Before Interests Taxes Depreciation and Amortization (EBITDA), Return On Equity (ROE), Return On Assets (ROA), Research & Development investments (R&D), Net Income [5].

4 Financial Data Summarizer

A sketch of the proposed method for financial time series summarization is depicted in Fig. 1. The summarization pipeline consists of the following steps:

1. *Financial data encoding*, whose goal is to transform the raw time series and news data into a unified vector representation of the stocks encoding both price-related trends, momentum and seasonality, and market sentiment.
2. *Quantifier/summarizer evaluation*, which entails estimating for each stock and reference time period the values of the corresponding summarizers and quantifiers on top of the encoded stock representation.
3. *Protoform generation*, whose goal is to compose the explainable summaries of the financial time series and compute the corresponding quality indices.

[2] https://www.reuters.com/.

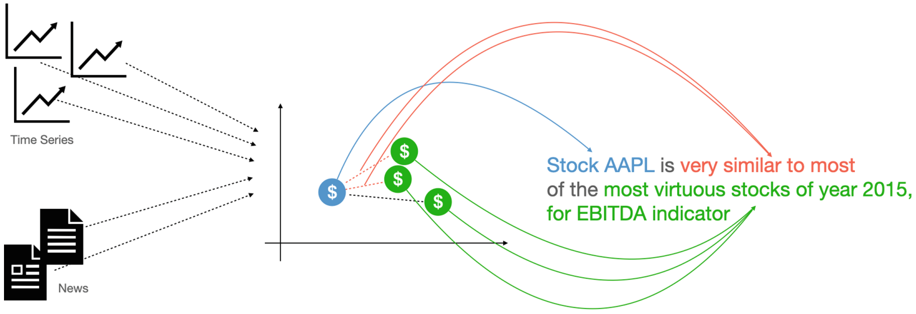

Fig. 1. Sketch of the time series summarization system.

Financial Data Encoding. This step entails encoding multimodal financial data into a unified, high-dimensional vector space. Each stock is represented by a vector in the latent space, which encodes both its most significant price-related features (i.e., trends, seasonalities, momentum) and the sentiment of the market extracted from the news articles.

Time series and news data are first transformed into a discrete sequence of symbols using a SAX representation [9] and then encoded the established Signal2Vec encoder [13][3]. In the SAX representation the daily samples of the series of stock prices, the technical indicators and news sentiment scores are mapped to a unique symbol to condense the daily multimodal information about each stock. Signal2Vec encodes discrete sequences of different time periods (e.g., yearly periods) annotated with the corresponding stock identifier. In such a way, sequences that refer to the same stock are used to describe the underlying behavior of the same stock.

Quantifier and Summarizer Estimation. Quantifiers and summarizers are the core elements of the comparative summaries. They express the level of adherence of the summary claim with the analyzed data.

We leverage the multimodal stock vector space trained at the first step to assign reliable quantifier/summarizer values. Specifically, let s_1, s_2 be the stocks under consideration for summary types *virtuos_stocks*, *year_to_stock*, and *virtuos_multivariate*. Let $v(s_1)$ and $v(s_2)$ be the corresponding vectors encoding the time series and news contents. The data-driven procedure instrumental for quantifier and summarizer estimation is described by the following procedure:

[3] In the experiments Signal2Vec is trained using the PV-DBOW architecture with vector size 100, 10 epochs, and a training window of 5 symbols.

Input: stock set S, fundamental indicator set I, reference time period T
for $\forall\ s \in S$ **do**
  for $\forall\ i \in I$ **do**
    $d(s_1, s_2) \leftarrow$ compute-distance($v(s_1)$, $v(s_2)$)[4]
    $v_s^i \leftarrow$ value of indicator i for s within the reference time period T
    $q_s^i \leftarrow$ quantile of stock s according to i, depending on v_s^i
    if $q_s^i == 1st$ **then**
      $R^i = R^i \cup \{s\}$ (set of reference stocks according to i)
      for $\forall\ s \in S$ **do**
        $R_s^i \leftarrow$ Nearest neighbors of stock s according to i calculated using set of distances d
      end for
    end if
  end for
end for
Output: $R_s^i\ \forall s \in S, i \in I$

For each fundamental indicator we first shortlist the top-ranked stocks (i.e., the stocks in the first quantile for the fundamental indicator. Then, for each stock in the vector space we compute the distance with each reference stock/group of stocks (e.g., the sector). Distances among vectors are used to quantify the similarity level with the reference group. Finally, quantifiers and summarizers in natural language are derived by uniformly discretizing the per-stock similarity scores.

Protoform Generation. We generate the five different types of comparative summaries reported in Table 1. Each summary is a sentence, called *protoform* [4], that provides a explanable comparison between time series data in natural language. Each protoform contains one or more fields denoting any of the following items:

- *Stock*: the name of the stock under consideration.
- *Sector*: the market sector under consideration.
- *Indicator*: the fundamental indicator under consideration.
- *Quantifier*: A word or phrase that specifies how often the summarizer is true.
- *Summarizer*: Word or phrase denoting a level of match between the compared items.
- *Time window*: A time window of interest for the given protoform.

Table 2 reports the possible values taken by each field and the summaries in which they appear. A more detailed description of the proposed protoforms is given below.

Given a fundamental indicator as reference metric of stock virtuosity, the summary type named *virtuous_stocks* compares a single stock with the most virtuous stocks, whereas the specular type *sectors* compares market sectors instead

[4] In the experiments we adopt the cosine distance in compliance with [13].

of single stocks. The type *virtuous multivariate* specifies a percentage of reference stocks with the given level of similarity and also allows the inclusion of multiple indicators for the definition of virtuous stocks.

Summaries *years_to_stock* and *year_to_years* perform time-based comparisons between time series. Specifically, the former type indicates for how many years one stock have been similar to all years of another one. The latter type compares a single year of the one stock with all the years of the other one defining a level of similarity and again how many years correspond to it.

Table 1. Proposed protoforms.

Summary type	Protoform template
virtuous_stocks	Stock [stock] is [summarizer] to most of the most virtuous stocks, for [indicator]
sectors	Sector [sector] is [summarizer] to most of the most virtuous stocks of [sector] sector, for [indicator]
years_to_stock	In [quantifier] years the stock [stock_1] has been [summarizer] to the stock [stock_2]
year_to_years	In year [period] the stock [stock_1] has been [summarizer] to [quantifier] years of the stock [stock_2]
virtuous_multivariate	Stock [stock] is [summarizer] to [quantifier] of the most virtuous stocks, for [indicator_1]..[indicator_n]

Table 2. Fields of the protoforms in Table 1.

Field	Values	Summaries
[stock]	Stock ticker choosen for the comparison	virtuous_stocks, years_to_stock, year_to_years, virtuous_multivariate
[summarizer]	very similar, fairly similar, not similar	all
[indicator]	EBITDA, ROE, ROA, R&D, Net Income	virtuous_stocks, sectors, virtuous_multivariate
[sectors]	Market sector (e.g. Energy, Healthcare, ...)	sectors
[quantifier]	none, few, many, all	years_to_stock, year_to_years
[period]	reference year	year_to_years
[quantifier]	percentage of reference stocks with given level of similarity	virtuous_multivariate

5 Experiments

Hardware and Code. We run the experiments on a hexa-core 2.67 GHz Intel Xeon with 32 GB of RAM, running Ubuntu Linux 18.04.4 LTS. The framework is written in Python and is available for research purposes upon request to the authors.

Execution Times. Table 3 summarizes the execution times taken by each phase of the time series summarization process. The computational time required to generate the time-dependent summaries (i.e., years_to_stock and years_to_years) is roughly one order of magnitude higher than all the other ones. The reason is that their generation entails partitioning stock series data into multiple time periods and recompute the indicator ranks separately for each reference period.

Table 3. Execution times per summary type.

Task	Execution time (avg ± std)
Time series and news data transformation	240 s ± 3 s
Sentiment Analysis	324 s ± 5 s
Multimodal data encoding	594 s ± 15 s
"virtuous_stocks" summary generation	245 s ± 5 s
"sectors" summary generation	220 s ± 3 s
"years_to_stock" summary generation	2793 s ± 23 s
"year_to_years" summary generation	5000 s ± 500 s
"virtuous_multivariate" summary generation	260 s ± 4 s

5.1 Intrinsic Evaluation

We characterize the generated summaries using a set of reference quality metrics first introduced in [4]. Metric values are normalized between zero and one. A brief description of the used metrics is given below.

- *Degree of truth (T1)*: it quantifies the truth of the quantifier-summarizer pair expressed by the summary. It is valid only for summary types year_to_years and years_to_stock.
- *Degree of Imprecision (T2)*: it measures the precision of the summary with respect to the whole data collection.
- *Degree of covering (T3)*: it indicates the percentage of data instances that are covered by the summary statement.
- *Degree of Appropriateness (T4)*: it quantifies the gap between the observed summarizers' values and the expected ones. This metric is valid only for the virtuous_multivariate type.

Table 4 reports some representative summary examples belonging to different type and the corresponding quality metric values. The generated summaries can be ranked by decreasing coverage and precision to shortlist the most reliable stock explanations. For example, the summaries of type *sectors* allow end-users to compare the market sector *Energy* with *Utilities* and *Industrial*, respectively. The Degree of covering (T3) indicates that the reported *sectors* summaries are supported by roughly half of the covered data instances. According to the Degree of Imprecision (T2), their precision is almost maximal (99%) in both cases.

Table 4. Examples of generated summaries.

Summary type	Summary	T1	T2	T3	T4
virtuous_stocks	Stock DU is not similar to most of the most virtuous stocks, for the ROA indicator		1	0.99	
virtuous_stocks	Stock AAPL is very similar to most of the most virtuous stocks, for the ROA indicator		0.99	0.41	
sectors	Most of the stocks of Energy sector, has been not similar to most of to the most virtuous stocks of Utilities sector for the ROA indicator		0.99	0.47	
sectors	Most of the stocks of Energy sector, has been very similar to most of to the most virtuous stocks of Industrial sector for the R&D indicator		0.99	0.65	
years_to_stock	In few years the stock AAPL has been fairly similar to the stock SYF	0.83	0.90	0.17	
years_to_stock	In most years the stock AAPL has been very similar to the stock ANSS	0.77	0.74	0.42	
year_to_years	In year 2015 the stock HPE has been very similar to few years of the stock JEF	0.71	0.71	0.14	
year_to_years	In year 2015 the stock HPE has been fairly similar to most years of the stock FDX	0.93	0.75	0.33	
virtuous_multivariate	Stock FITB is fairly similar to 22% of the most virtuous stocks, for the ROE and the Net Income indicator		0.93	0.22	0.24
virtuous_multivariate	Stock FITB is fairly similar to 32% of the most virtuous stocks, for the ROE and the ROA indicator		0.9	0.32	0.17

5.2 Extrinsic Summary Validation

We validate the usability of the information provided by per-stock summaries via extrinsic evaluation. Specifically, Table 5 reports two summary examples of type *Sectors* that compare the performance of the *Industrials* sector with that of the *Communication Services* and to the *Materials* sectors, respectively. In Fig. 2 we show the corresponding temporal price variations. The summaries are coherent with the observed price series trends: the *Industrials* sector is highly similar to the most virtuous *Communication Services* stocks, whereas is weakly similar to those of the *Materials* sector.

Table 5. Sectors-type summary examples.

Summary	T1	T2	T3	T4
Most of the stocks of Industrials sector, has been very similar to most of to the most virtuous stocks of Communication Services sector for the ROE indicator		0.99	0.44	
Most of the stocks of Industrials sector, has been not similar to most of to the most virtuous stocks of Materials sector for the ROE indicator.		0,99	0,41	

Table 6 reports two summaries of type *years_to_stock* that compare the performance of the Apple stock with that of the Vertex Pharmaceuticals and CME Group stocks. According to the generated summaries, the price movements of the stocks AAPL are expected to be more similar to those of stock VRTX than those of CME. The expected result is confirmed by historical time series depicted in Fig. 3 (see, for example, years 2014–2016).

Table 6. years_to_stock summary examples.

Summary	T1	T2	T3	T4
In most years the stock AAPL has been very similar to the stock VRTX	0,77	0,74	0,42	
In most years the stock AAPL has been not similar to the stock CME	0.77	0.74	0.42	

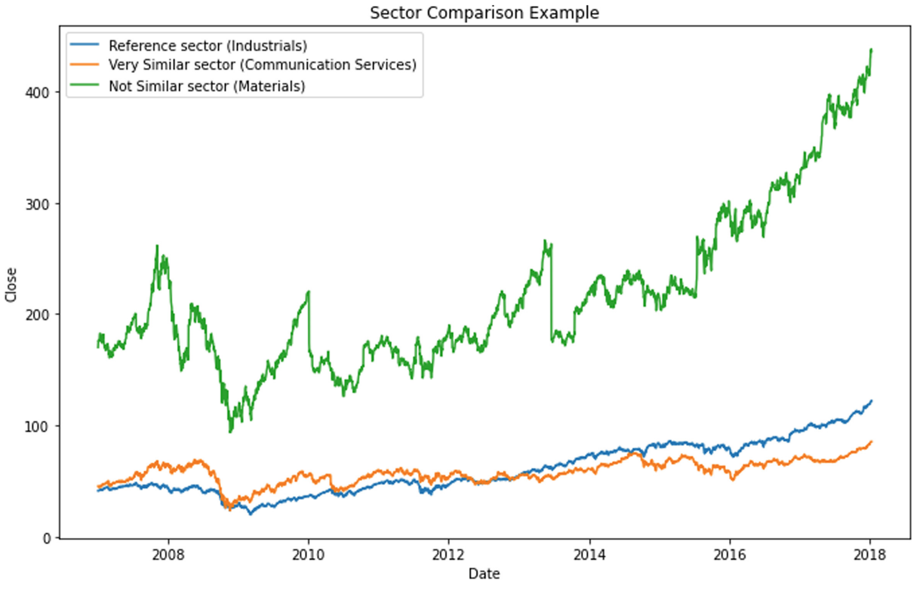

Fig. 2. Comparison between the Energy Sector and the most virtuous stocks for Industrial and Utilities Sectors.

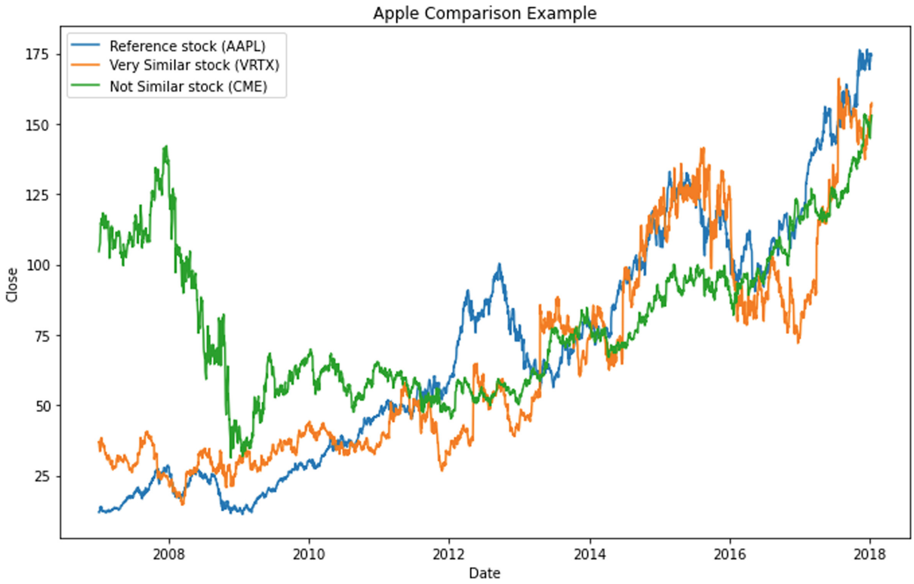

Fig. 3. Comparison between APPL, VRTX, and CME stocks.

6 Conclusions and Future Works

The paper proposed a new approach to generate explainable summaries of financial time series in textual form. The key idea is to represent the key information about the stocks into a unified latent space, among which price-related time series data and news sentiment scores. By leveraging the vector representation to get reliable stock and stock group similarities we are able to automatically estimate the quantifiers and summarizers needed to generate the protoforms.

The preliminary results show that the provided summary examples (1) achieve satisfactory quality levels according to the metrics defined in [4], (2) are coherent with the expectation, and (3) can be exploited by domain experts to support decision-making.

We plan to extend the empirical validation by designing and test a dedicated mobile application through which private and professional investors can access and evaluate the generated summaries and the corresponding evaluation metrics. We will collect subjective user feedbacks with the twofold aim at improving the robustness of the empirical validation and exploiting the relevance feedback in order to selectively filter the generated summaries.

Acknowledgements. The research leading to these results has been partly supported by the SmartData@PoliTO center for Big Data and Machine Learning technologies.

References

1. Catillo-Ortega, R.M., Marín, N., Sánchez, D.: A fuzzy approach to the linguistic summarization of time series. J. Multiple-Valued Logic Soft Comput. **17** (2011)
2. Gatt, A., et al.: From data to text in the neonatal intensive care unit: using NLG technology for decision support and information management. AI Commun. **22**(3), 153–186 (2009)
3. Goyal, K., Kumar, S.: Financial literacy: a systematic review and bibliometric analysis. Int. J. Consum. Stud. **45**(1), 80–105 (2021). https://doi.org/10.1111/ijcs.12605. https://onlinelibrary.wiley.com/doi/abs/10.1111/ijcs.12605
4. Harris, J.J., Chen, C.H., Zaki, M.J.: A framework for generating summaries from temporal personal health data. ACM Trans. Comput. Healthcare **2**(3), 1–43 (2021). https://doi.org/10.1145/3448672
5. Huang, Y., Capretz, L.F., Ho, D.: Machine learning for stock prediction based on fundamental analysis. In: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–10 (2021). https://doi.org/10.1109/SSCI50451.2021.9660134
6. Hutto, C., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, pp. 216–225, May 2014. https://ojs.aaai.org/index.php/ICWSM/article/view/14550
7. Kacprzyk, J., Wilbik, A., Zadrozny, S.: An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation. Int. J. Intell. Syst. **25**(5), 411–439 (2010). https://doi.org/10.1002/int.20405
8. Kacprzyk, J., Zadrony, S.: Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. Inf. Sci. **173**(4), 281–304 (2005)
9. Keogh, E., Lin, J., Fu, A.: HOT SAX: efficiently finding the most unusual time series subsequence. In: Proceedings of the Fifth IEEE International Conference on Data Mining, ICDM 2005, pp. 226–233. IEEE Computer Society, USA (2005). https://doi.org/10.1109/ICDM.2005.79
10. van der Lee, C., Krahmer, E., Wubben, S.: Automated learning of templates for data-to-text generation: comparing rule-based, statistical and neural methods. In: Proceedings of the 11th International Conference on Natural Language Generation, pp. 35–45. Association for Computational Linguistics, Tilburg University, The Netherlands, November 2018. https://doi.org/10.18653/v1/W18-6504
11. Liapis, C.M., Karanikola, A., Kotsiantis, S.: A multi-method survey on the use of sentiment analysis in multivariate financial time series forecasting. Entropy **23**(12), 1603 (2021). https://doi.org/10.3390/e23121603. www.mdpi.com/1099-4300/23/12/1603
12. Murphy, J.J.: Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. Penguin, New York (1999)
13. Nalmpantis, C., Vrakas, D.: Signal2Vec: time series embedding representation. In: Macintyre, J., Iliadis, L., Maglogiannis, I., Jayne, C. (eds.) EANN 2019. CCIS, vol. 1000, pp. 80–90. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20257-6_7