European Journal of Interdisciplinary Research and Development
Volume-13 Mar. - 2023
Website: www.ejird.journalspark.org ISSN (E): 2720-5746

TEXT MINING

Israil Tojimamatov
Teacher

Todjiyeva Dilfuza Abdusalomovna
Haydarova Ozodaxon Adhamjon qizi
Kamoliddinova Odinaxon Abdukarim qizi
Mamasoliyeva Dilnoza Baxtiyor qizi
Faculty of Mathematics and Informatics of Fergana State University
field of applied mathematics and informatics

Annotation
It is a science about the generalized language of informatics, which appeared on the basis of the methods of machine learning and the rules of statistics . Text mining (also called text analytics) is the use of natural language processing (NLP) to analyze free (unstructured) text in documents and databases or transform it into normalized, structured data suitable for machine control. artificial intelligence (AI) technology. learning (ML) algorithms.In our independent work, we provide an introduction to these technologies and highlight some of the features that contribute to an efficient solution.

Keywords: Text Analytics, Textmining, Deep Analytics, Descriptive Analytics, Predictive Analytics, Prescriptive Analytics, Diagnostic Analytics, NLP, Business Intelligence, Sentiment Analysis.

What is text mining ? Text mining, widely used in knowledge- based organizations, is the process of examining large sets of documents to discover new information or help answer specific research questions. identifies remaining facts, relationships, and assertions. Once extracted, this information is converted into a structured form that can be further analyzed or presented directly using clustered HTML tables, mind maps, charts, etc. Text mining uses different methodologies for text processing, one of the most important is **Natural Language Processing (NLP)** . Structured data created by text mining can be integrated into databases, data warehouses or business intelligence dashboards and descriptive, can be used for prescriptive or predictive analytics. **What is Natural Language Processing (NLP)?** Natural language understanding helps machines "read" text (or other input such as speech) by simulating a human's ability to understand a *natural language such as English, Spanish, or Chinese.* Natural language processing includes both natural language understanding and natural language generation, which mimics the human ability to produce natural language text, such as summarizing information or engaging in dialogue. As a technology, natural language processing is the latest o Developed over the years, products such as Siri, Alexa, and Google Voice Search use NLP to understand and respond to user requests. Advanced text mining applications are also being developed in various fields such as medical research, risk

management, customer service, insurance (fraud detection), and contextual advertising. developed.

Deep analysis requires both critical thinking and creative thinking. It often involves the use of various methods and tools, such as data visualization, statistical analysis, and qualitative research. It's a time-consuming process, but it can be very rewarding. It leads to discoveries, greater understanding, and improved decision-making.

Deep analysis is a necessary tool for anyone who wants to understand the world around them and what the market wants or needs. What is deep analysis? In-depth analysis is the process of studying a problem or issue in detail. This often involves looking at data or information in a new way to find hidden patterns or uncover previously unknown relationships. Deep analysis can be used to solve complex problems or understand complex systems. It is an important tool for businesses and organizations to make the right decisions and improve their performance. In-depth analysis is also essential for academic research and writing. This allows scientists to study a topic in depth and develop new insights or perspectives on it.

The most popular approach to in-depth analysis is qualitative and quantitative. Some people prefer a more quantitative approach, while others prefer a more qualitative approach. Qualitative research methods include focus groups, online interviews, etc. Quantitative research, on the other hand, is conducted through online surveys with various sampling methods. Tips for In-Depth Analysis In general, in-depth analysis involves four main steps: data collection, data analysis, interpretation of results, and communication of findings. It's also important to do your due diligence on the panel company you choose. Also, market research and in-depth analysis go hand in hand. • Data collection: This is the first step for in-depth analysis. This can be done through observation, experience or surveys. This process involves looking for patterns and relationships within the data. • Interpretation of the result: This step requires critical thinking and creative thinking to complete the data. Finally, the results of the in-depth analysis should be communicated to others. This can be done through writing, presentation or discussion. Deep analysis is a complex process, but it is a powerful tool for understanding the world around us. Market research is the process of gathering information about the needs and wants of consumers. This information can be used to help businesses make better decisions about product development, marketing and sales. Market research can be conducted in a variety of ways, including surveys, interviews, focus groups, and observations. Why in-depth analysis is important Choosing the right method for each situation is important because different methods can provide different information. In-depth analysis is an essential tool for any business looking to succeed in today's competitive market. It helps businesses understand consumer behavior and how it has changed in response to the current situation. In addition, market research can identify new opportunities that may arise as a result of the pandemic. For example, there may be an increase in demand for certain products or services that businesses can handle. Market research can also help businesses understand the potential impact of the pandemic on their industries and make the necessary changes to their operations. In short, market research is an essential tool for businesses to weather the Covid pandemic and emerge successfully on the other side. At QuestionPro, we can help you with both – providing the best survey software and the right audience to answer them. You name it and we have it.

B2B, B2C and niche we have it all. When it comes to data collection, we have various layers of security in place to weed out respondents who say "Yes", ignorers, duplicate IP addresses, etc. before even entering the actual survey.

There are four main methods of text analysis:

1. **Descriptive analysis** is the most common method, it is "what happened?" answers the question and analyzes real-time incoming and historical data. The main goal is to determine the reasons and patterns of success or failure in a particular area and to use this information effectively.
2. **Predictive analysis** - based on available data, helps to predict the most likely development of events. For this , it uses ready-made templates based on any objects or events that have a similar set of properties .
3. **Prescriptive analytics** is the next level compared to predictive analytics. With the help of Big Data and modern technologies, it is possible to identify problem points in business or other activities and calculate how to avoid them in the future.
4. **Diagnostic analytics** – uses data to analyze the causes of an incident. It helps to identify anomalies and random connections between events and actions.

As data becomes the new source of wealth, big data solutions are expected to play an increasingly important role in the growth of companies. Companies have always had access to data , but the ability to access and process data to obtain meaningful information is limited. "Big data" solutions allow companies to overcome these challenges . " Big data" analysis forms the basis of companies, helps manage, process, optimize large data sets in real time and improves decision-making ability. In addition, the main purpose of big data and business analytics is to help companies better understand their customers and improve their marketing campaigns.

While some industries have made great strides in using big data, others have yet to focus on big data enough. By 2025, the big data analytics market in the banking sector may grow to $ 62.10 billion. According to Big Data in Banking statistics , the infrastructure of the global banking sector already includes big data analytics. For example, as of 2013 , 64 percent of the global financial sector has already incorporated big data into its infrastructure, and in 2015, the market size reached $12 billion. By 2019, the big data banking transaction analytics market was worth $29.87 billion. The banking big data analytics market is expected to grow at a CAGR of 12.97 percent between 2020 and 2025. The Big Data market is valued at $198.08 billion in 2020 and will grow to $274.3 billion by 2022, with a CAGR of 13.2% over the next 5 years. It is expected to grow to $103 billion by 2027, and the software segment will account for 45 percent. According to BARC , organizations are benefiting from "big data". Specifically, the ability to make strategic decisions increased by 69 percent, the ability to gain more control over operational processes by 54 percent, the ability to better understand consumers by 52 percent, and the ability to reduce costs by 47 percent. Companies taking advantage of big data have reported an average 8% increase in revenue and 10% reduction in costs. Forbes predicts that by 2025, more than 150 zettabytes or 150 trillion gigabytes of data will be analyzed in real time . According to Forbes , more than 95 percent of companies need help managing multiple sets of unstructured data, and 40 percent of companies say they need to deal with big data.

According to StrategyMRC, the Hadoop and big data market is currently valued at $138.9 billion and is expected to grow to $30 billion by 2022 at a CAGR of 28.5 percent. . According to Wikibon, the volume of big data, analytics, and application database solutions will grow at a CAGR of 6 percent over the decade, from $6.4 billion in 2017 to 2027. it is expected to grow to 12 billion dollars. Demand for open source platforms in the big data ecosystem, such as Hadoop, Kafka, Spark, and TensorFlow, may decrease due to its direct application to artificial intelligence, machine learning, deep learning, or data science. But the hybrid deployment of data analytics platforms such as Hadoop, NoSQL, in-memory, streaming and many other databases is increasing the market share of data scale and data mining solutions. Today, in the market of "Big Data" technologies of the USA, Amazon.com, Inc., IBM, Microsoft, Oracle, Dell Technologies, Cisco Systems, Inc., Cloudera, Inc., Salesforce.com, Inc., Teradata, Tableau Software, Hewlett- Packard, Prolifics, Inc., Xplenty, Clairvoyant, Teradata, EquBot Inc., Dell Technologies, Vmware, Inc. Databrix Corporations and SAS Institute, Germany's SAP SE and Software AG, Japan's Fujitsu Limited and CMIC Co., Ltd., India's Infosys Limited, China's Huawei Technologies Co., Ltd. and Deloitte Touche Tohmatsu Limited of Great Britain are leading companies. The share of countries in the global big data and business analytics market in 2021: USA – 51%; Japan - 5.7 %; China - 5.5% Great Britain - 5.1%; Germany – 4.4%; others - 28.3%. The future of "big data" requires analysts. As the volume of big data continues to grow, so does the need for trained data analysts to dive into it and extract actionable insights. Big data analytics offers incredible opportunities to create change in industries such as finance, government, and healthcare, as well as save lives by preventing fraud, allocating resources in the event of a natural disaster, or improving healthcare. helps to change.

Natural Language Processing (NLP)

This type of data analysis requires Natural Language Processing (NLP), a form of machine learning algorithm. NLP is seen as the next big thing in data analytics, providing the ability to use big data to gain insights using innovative methods to produce useful insights into market trends – current or forecast.

Although NLP research has been going on for several decades, the field has only made significant progress in the last 3 years. Machine learning methodologies using NLP are now widely deployed across enterprises through big data partner consulting.

NLP uses statistics and machine learning to analyze linguistics and semantics to learn patterns that appear in text entries in big data and extract important entities and relationships in the context of what customers want to say in their posts. In fact, instead of focusing on a single word or string of words, NLP analyzes sentences comprehensively for purpose. The most common methodologies used in NLP are automatic generalization, disambiguation, part-of-speech marking, relation extraction and object extraction, and most importantly, natural language understanding and recognition.

How can NLP help you make better use of information in unstructured big data?

In every field - medical, legal, pharmaceutical, sports, education, etc., let's say, large amounts of data are archived every day in the form of documents, customer data, sales data, etc. This

data is mostly text and therefore NLP. It is essential to get effective results from the analysis - whether it is predictive, real-time or historical.

NLP can help in the following areas.

Interactions

Siri on iOS is a great example of NLP in the interactive realm. Online banking and retail self-service tools also use NLP, as do automatic translation applications. The interaction management experience using NLP has evolved to the levels where traditional customer support calls can be effectively handled and resolved through the introduction of artificial intelligence.

Business intelligence

Tracking a particular social media label may require the analyst to enter all possible "hashtags" and keywords that cover the topic. NLP can perform search operations on queries entered in natural language, thereby covering all possible scenarios and reducing statistical errors in determining how many people are talking about a topic.

Sentiment analysis

Brands can now collect data through direct channels, from social media conversations to formal customer feedback. NLP is a specific product or service in the target market segments

- can create a definitive picture of acceptance across demographic and/or geographic segments. It is estimated that by 2020, all big data analysis provided by every Big Data Solution will be done using NLP, as the data volume will exceed 44 trillion gigabytes worldwide, so for Big Data Analytics The scope of NLP continues to expand. Instead of being just a buzzword and a "maybe" technology, big data is now one of the most important strategic imperatives for enterprises and startups. Big Data – Information such as customer online shopping behavior, social media activity and internal data log such as point of sale information is continuously collected from various channels. Real-time logging data is continuously logged and stored. in the cloud. Large enterprises can have more than a petabyte of data.

Because data is taken from its source in an "as is" state, it is largely unstructured, while it is constantly growing. The format of this big data is basically text shared on social media or customer feedback about products. In other words, big data is a string of words in natural language that the average person might use in real-life conversations.

NLP for translation

One of the most important and valuable applications of natural language processing is translations. Although languages around the world originate from a few language families, they have spread to countless countries and cultures. English is different in the US, UK, Asia and Europe; Spain, Central and South America have their own versions of Spanish.

These and other linguistic issues have been a major obstacle for machine translation for years. It should be noted that machine translation is completely different from sentiment analysis and document classification.

This challenging task needs a model that predicts word sequences rather than labels. Machine Learning also underlines all the talk and excitement about Deep Learning, as it has changed the game for analyzing sequential data.

You can check out more about RNN and advanced translation results here.

There is also a need for more effective articulation and rewording. Let's say you need an automatic text summarization model and you want to summarize the text while preserving its complete and original meaning. The attention mechanism (included as a module in the latest solutions) proves valuable in these tasks, as it can understand whole blocks of texts by looking for meaning in only certain segments of the texts.

In addition, so far there is an answer to questions that is close to Hollywood-level artificial intelligence. This requires the algorithm model not only to understand the question, but also to fully understand the context and related texts, as well as where to look for answers.

Deep Learning represents different types of data in vectors. Thus, algorithm models targeting different domains can be easily created. Thus, "visual question answering" was born. This is a very easy task for a child: to answer a question about an image.

However, think of this in the context and capabilities of a machine that has yet to understand words and images. Deep models were the first to achieve significant results in this task without human supervision.

References

1. Devi Silen , Arno Meisman , Mohamed Ali . Basic Data Science and Big Data. Python i nauka o dann yx. SPb.: Peter, 2017. -336 p. (Series "Biblioteka programmista").
2. Henrik, Richards Joseph, Feverolf Mark, Mashinnoe obuchenie, SPb.: Peter, 2017. -336 p. (Series "Biblioteka programmista").
3. V. E. Tumanov. Proektirovanie khranilish dann y x dlya pilojeni system delovay osvedomlennosti (Business Intelligence System). izd. INTUIT. 2016g.
4. P. Flax. The science and art of building algorithms, which explore knowledge from the world. Published by Cambridge University Press, 2012.