UDC 331.214: 338.27

Hosseini Seyed Iman,
Research scholar,
Graduate School of Economics and Management,
Ural Federal University
named after the first President of Russia B.N. Yeltsin
Ekaterinburg, Russian Federation

Tarasyev A.,
Research scholar,
Graduate School of Economics and Management,
Ural Federal University
named after the first President of Russia B.N. Yeltsin
Ekaterinburg, Russian Federation

MACHINE LEARNING METHODS IN INDIVIDUAL MIGRATION BEHAVIOR[31]

Abstract:
Machine learning is described as "a field of computer science that gives a machine the ability to learn". In fact, machine learning is considered as a sub branch of Artificial Intelligence(AI). In recent years the rise of big data and cloud computing gives AI expert and specifically machine learning expert to dive deeply in data and extract knowledge from it by using machine learning algorithms. In this paper we try to introduce the basic concepts of machine learning algorithms including supervised learning, unsupervised learning and reinforcement learning and its usage in different applications. We describe specifically how to use machine learning in migration process modeling and focus on an approach for migration description, that is based on one of machine learning methods, the decision tree algorithm. We apply this method for the description of the economic behavior of an individual in the question of continuing his work in Russia based on the panel data and the data from the sociological survey. The accuracy of our estimation using decision tree

31 Research was supported by the grant of RFHS № 15-02-00072 "Modelling educational trajectories of individuals and monitoring the higher education system as a basis for decision-making to improve its competitiveness".

is 67 percent for this specific task. All in all, the main objective of this paper is to introduce the important aspects of machine learning and its usages in the state-of-the-art technologies.

Keywords:
artificial intelligence, machine learning, supervised learning, unsupervised learning, reinforcement learning, migration process, labor migration, economic behavior

Introduction

Machine learning is a field of computer science that gives machines the ability to learn without being explicitly programmed. Machine learning is an application that is fed by data as the input and gives up knowledge as the output. By this definition we can imply that it plays an important role in a world which is comprised of extremely large structured and unstructured data (so called Big Data). If we look at the world, we are living in, a little bit deeper we can see that data almost touches every aspect of our lives, from the way we transact commerce on the web, to how we measure our fitness and safety, to the way doctors treat our illnesses, to economic decisions that affect entire nations. Scientific fields are transitioning from data-poor to data-rich and across industries, science, and government-methods for making decisions are becoming more data-driven as large amounts of data are being harvested and stored. The first challenge anybody finds when starting to understand how to build intelligent machines is how to mimic human behavior in many ways or, to put it even more appropriately, how to do things even better and more efficiently than humans. In order to build an intelligent machine, we need to be familiar with some tasks which I think are closely integrated with each other.

Machine learning methods

Statistics, the science, which deals with learning from data and measuring, controlling, and communicating uncertainty, is the most mature of the data sciences. Over the last two centuries, and particularly the last 30 years with the ability to do large-scale computing, this discipline has been an essential part of the social, natural, biomedical, and physical sciences, engineering, and business analytics, among others. Statistical thinking not only helps make scientific discoveries, but it quantifies the reliability,

reproducibility and general uncertainty associated with these discoveries. Because one can easily be fooled by complicated biases and patterns arising by chance, and because statistics has matured around making discoveries from data, statistical thinking will be integral to Big Data challenges [3].

Tom Mitchell in his paper The Discipline of Machine Learning defined the fundamental question as "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?" He further explains, "To be more precise, we say that a machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system reliably improves its performance P at task T, following experience E." He also illustrates that, the definition of computer science is "How can we build machines that solve problems, and which problems are inherently tractable/intractable?" whereas Statistics is "What can be inferred from data plus a set of modeling assumptions, with what reliability?" Having asked these questions, we can clearly see the difference between machine learning and statistics. Another line of thought, in the paper "Statistical Modeling: The Two Cultures" by Leo Breiman in 2001 [4], argued that statisticians rely too heavily on data modeling, and that machine learning techniques are instead focusing on the predictive accuracy of models.

John McCarthy, a well-known American computer scientist in the field of *Artificial Intelligence* (AI) in his paper *What is Artificial Intelligence defines* "AI is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable" [5]. He also makes a list of different branches of AI including logical AI, search AI, pattern recognition, representation, inference, common sense knowledge and reasoning, learning from experience, planning, epistemology, ontology, heuristics, genetic programming. It should be pointed out that machine learning is also a subset of AI.

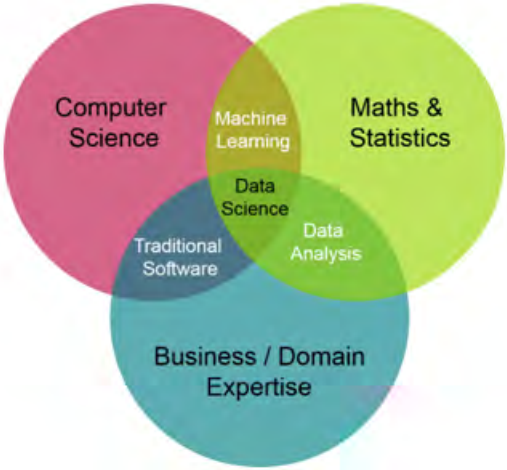

Figure 1: Data science toolbox

In *Data Mining Concepts and Techniques 2nd edition published by Jiawei Han and Micheline Kamber* book [6], data mining is "extracting or mining knowledge from large amounts of data." By considering this definition, we can obviously see that there is a strong correlation between data mining and machine learning. We can see that machine learning is used at the heart of data mining.

Data Science is a big umbrella that brought everything that we need in order to extract data and show insight from data. Data science employs many fields including: mathematics, statistics, machine learning, business, communication science and computer science. Figure 1 can be a sample of a toolbox that is being prepared by data science for us.

Different Types of Machine Learning Algorithms

Supervised Learning

This algorithm consists of dependent variable (so called target variable) and a set of independent variables (so called predictors). The algorithm uses independent variables to generate a function in order to predict the dependent variable.

Common supervised machine learning included: [7]

- Predictive analysis based on regression or categorical classification
- Spam detection
- Pattern detection
- Natural Language Processing
- Sentiment analysis
- Automatic image classification
- Automatic sequence processing (for example: music or speech)

Unsupervised Learning

In this type of machine learning algorithm, we do not have any dependent variables. In fact, we use it for clustering our data in different categories.

Common unsupervised applications include: [7]

- Object segmentation (for example: users, products, movies, songs and so on)
- Similarity detection
- Automatic Labeling

Reinforcement Learning

In this algorithm, the agent is exposed into an environment and it learns from the environment. According to its behavior it gets a feedback which is usually called reward in case it is positive and if it is negative it gets a penalty and it is useful to know that a specific action performed in an environment is positive or negative.

There is a very informative description of reinforcement-learning model in the paper Reinforcement Learning: A survey published on Journal of Artificial Intelligence Research in 1996 [8], "In the standard reinforcement-learning model, an agent is connected to its environment via perception and action. On each step of interaction, the agent receives as input, i, some indication of the current state, s, of the environment; the agent then chooses an action, a, to generate as output. The action changes the state of the environment, and the value of this state transition is communicated to the agent through a scalar reinforcement signal, r. The agent's behavior, B, should choose actions that tend to increase the long-run sum of values of the reinforcement signal. It can learn to do this over time by

systematic trial and error, guided by a wide variety of algorithms".

Figure 2 shows the different types of machine learning.

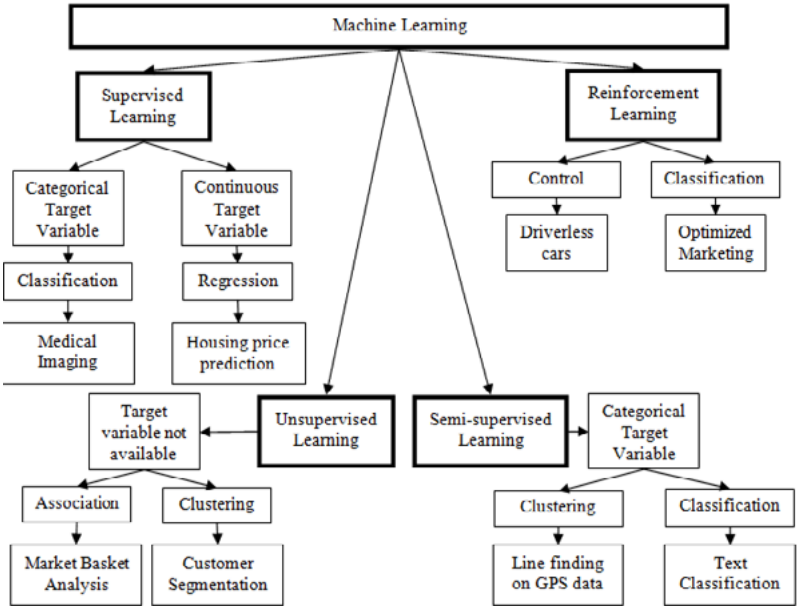

Figure 2: Machine Learning types

Top Machine Learning Use Cases

Bernard Marr wrote an article in Forbes(an American business magazine) in 2016 about the top AI and Machin machine learning Use case that are briefly summarized below.

In 2014, Kaspersky Lab said it had detected 325,000 new malware files every day. But institutional intelligent company Deep Instinct says that only between 2% to 10% of the files change. Machine learning can look for pattern and report anomalies that could predict Data Security breaches [9].

It's a huge opportunity for businessmen if they are capable of predicting the stock market and Financial Trading. Many prestigious trading firms use machine learning to predict and execute trades at high speed and high volume [9].

Computer-aided diagnosis (CAD) can help radiologists find early-stage problems. One study used computer assisted diagnosis (CAD) when to review the early mammography scans of women who later developed breast cancer, and the computer spotted 52% of the cancers as much as a year before the women were officially diagnosed. Additionally, machine learning can be used to understand risk factors for disease in large populations [9].

The more you know about your customers, the better you can serve them. That is the foundation behind Marketing Personalization. Companies can personalize which emails a customer receives, which direct mailings or coupons, which offers they see, which products show up as "recommended" and so on, all designed to lead the consumer more reliably towards a sale [9].

Machine learning is getting better and better at spotting potential cases of Fraud Detection across many different fields. PayPal, for example, is using machine learning to fight money laundering. The company has tools that compare millions of transactions and can precisely distinguish between legitimate and fraudulent transactions between buyers and sellers [9].

The best example of utilizing machine learning is this area is Amazon. When you want to buy a book, Amazon automatically choose other books that is closely related to the books you want to buy. YouTube also has the same strategy when you watch a video clip.

Perhaps the most famous use of machine learning, Google and its competitors are constantly improving what the online search engine understands.

Machine translation using Natural Language Processing (NLP) helps us conquer language barriers that we often encounter by translating.

Automatic summarization for summarizing the meaning of documents and information, it also used to understand the emotional meanings inside the information.

Sentiment Analysis is used to identity how positive or negative people think or feel about different products or services.

Text classification helps us to classify text. An example of text classification is spam filtering in email.

Machine Learning in Migration Process Modelling

It is necessary to highlight, that socio-economic factors causing and determining the migration processes, are usually interconnected. According to the new theory of migration it is also necessary to take into account the earnings level of migrant's environment when describing the decision on migration [10]. Among the various factors, the wage difference between sending and receiving countries is the main economic factor. The neoclassical economic theory of migration is provided with migration models in which immigrant workers respond to the main economical factors, such as cross-region difference in salaries levels, migration costs and the labour market situation in the countries of origin and destination [11]. The standard theory of equilibrium which states that the uncontrollable inflow of foreign labour to a country reduces the relative wage of local workers because of the raising level of competition between migrants. The human capital model takes into account such migration factors as a fully rational decision on migration depending on the wage gap or the fact of perfect competition on the local labour market. Migration networks help disseminate information about the conditions of life and the labour market situation in different countries and help in finding employment for involved into these networks migrant workers. As a result, the migratory redistribution helps in regulating wages and stabilize the global labour market in two ways [12]: directly, through the reduction of supply of the labour market of the country with excessive labour resources and the increase of supply in the country lacking labour resources, and indirectly through international money transfers made by migrants [13].

One of the goals of our research was to estimate migrants behavior in the hosting country and to analyze their decisions on continuing their work as a migrant employee. In this paper we use an approach for migration description, that is based on machine learning methods, in particular the decision tree and the k-nearest neighbour methods. In the age of modern technology, there is a large amount of structured and unstructured amount of data. Machine learning evolved as a subfield of Artificial Intelligence that involved self-driving algorithms that derive knowledge from data in order to make prediction. Instead of requiring humans to manually derive rules and build models from analyzing large amounts of data machine learning offers a more efficient way for capturing the knowledge in data to gradually improve the performance of predictive models and make data-driven decisions. This is to do with what type of response variable we have in the training data. In our work we used an approach for migration description, that is based on machine learning methods, in particular

the decision tree and the k-nearest neighbour methods. The best accuracy of estimation during our work was observed when using the decision tree method (Table 1). According to our approach 8 percent of migrants wish to stay for work in Russia for additional three months, 35 percent for an additional year and 57 percent wish to stay for a time period from one to three years, which can be explained by a high attractiveness of well developed labour markets for migrants.

Table 1. Comparison of different prediction models

Prediction Model Used	Application	Evaluation Measure
K-Nearest Neighbour	Immigration	Accuracy = 58.13%
Decision Tree	Immigration	Accuracy = 67.44%

The wage gap between Russia and CIS countries will diminish, while the wages in the Moscow region will go downward. This tendency however may call for attention to the negative effects on wages and employment due to immigration, where the arguments are that immigrants may compete in labour markets with native-born workers, thus displacing native-born workers, or causing real wages for native-born workers to be pushed down. These preliminary results need deeper analysis to ascertain the migration impact on both the sending and hosting economies, including dimensions of wage, unemployment, and labour productivity [14].

Conclusion

The complexity of migration processes comes from the interaction of economic, demographic, and other social factors, which co-determine the size, age structure, and skill level of migration flows from sending to receiving countries. The analytical tools are needed to simulate the dynamics of the migration processes in a holistic context.

It is necessary to highlight that the methods of machine learning are very good fitting when describing the variants of system's development based on a big statistic dataset. When modelling the overall migration system for a country it is necessary to combine the machine learning, the agent based models and the dynamic models of migration based on the game theory assumptions. According to our results the main share of migrants prefers to stay for work in the country of destination so long, as possible, which is explained by growing migration costs and the bigger possibility of finding a job when staying inside of the developed labor market.

REFERENCES

- Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. № 11. 1967. 601-617.
- Ramasubramanian Karthik, Singh Abhishek. Machine Learning Using R: A Comprehensive Guide to Machine Learning. Apress. 566 p.
- Cynthia Rudin, David Dunson, Rafael Irizarry et all. Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society. A report of a Working Group of the American Statistical Association. (July 2, 2014) // http://www.amstat.org/asa/files/pdfs/POL-BigDataStatisticsJune2014.pdf
- Leo Breiman. Statistical Modeling: The Two Cultures. Statistical Science 2001, Vol. 16, No. 3, 199–231.
- John McCarthy. Defending AI research : a collection of essays and reviews. CSLI lecture notes: no. 49. Center for the Study of Language and Information, 1996. distributed by Cambridge University Press.
- Jiawei Han and Micheline Kamber. Data Mining Concepts and Techniques, Second Edition. Morgan Kaufmann Publishers, San Francisco, CA 94111, USA, 2011, 743 p.
- Bonaccorso Giuseppe. Machine Learning Algorithms. Packt Publishing Ltd. Electronic book. ISBN-10:1785889621. July 24, 2017. Birmingham B3 2PB, UK. 510 p.
- Leslie Pack Kaelbling, Michael L. Littman, Andrew W. Moore. Reinforcement Learning: A Survey. Journal of Artifcial Intelligence Research, №4, 1996, 237-285
- Bernard Marr. The Top 10 AI And Machine Learning Use Cases Everyone Should Know About // https://www.forbes.com/sites/bernardmarr/2016/09/30/what-are-the-top-10-use-cases-for-machine-learning-and-ai/#495d815b94c9
- Harris J.R., Todaro M.P. (1970). Migration, Unemployment and Development: A Two- Sector Analysis. The American Economics Review, Vol.60. Is.l. 126-142.
- Hicks J.R. (1963). The Theory of Wages. / London: Macmillan. 1932, 2nd ed.
- Friedberg, R.M., Hunt, J. (1995). The impact of immigrants on host country wages, employment and growth. Journal of Economic Perspectives, Vol.9, 23–44.
- J. Bijak, D. Kupiszewska, M. Kupiszewski, K. Saczuk, A. Kicinger (2007), Population and labour force projections for 27 European countries, 2002-2052: impact of international migration on population ageing. European Journal of Population, 23(1): 1-31.
- Dustmann, C., and I. Preston. (2011). Estimating the Effect of Immigration on Wages. Northface Migration, Discussion Paper, No. 26