

APPLICATION OF IBM WATSON FOR THE BEST SPECIALITY OPTION PREDICTION FOR PROSPECTIVE STUDENTS

*Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus*

Victor Kozub, Alexander Alexandrov

Ivan Piletski – associate professor

This paper investigates a problem of choosing a speciality by prospective students. Every year, it's a difficult problem for school graduates to choose a speciality that's the most suitable for them. Furthermore, there is a diversity of specialities and faculties at one university to choose from. Often when choosing a speciality, prospective students don't take into account their own personal characteristics. We've developed an application in IBM Bluemix cloud that predicts the best speciality option for prospective students of our university. The application uses IBM Watson service to get psychological characteristics of a person and then to predict a proper speciality. In this paper, we describe the IBM Bluemix cloud platform as a service, tools and philosophy of it by the example of the developed application. Obtained results are planned to be extended to satisfy business needs and to search for a staff needed for a particular project at a company.

The object recognition task can be solved using different methods. It depends on structures and algorithms that are used to store data. Simple example is when data is stored in a relational database: we can answer if a new object is known by simply accessing the database. Similarly, when objects are stored as elements of a set with known properties, enumeration method can be used to find whether an object is within the set or not. When it is impossible to compare a new object with a known one, new key characteristics of the object are picked. Then, these key characteristics are used by deep search algorithms to make a prediction with high probability rate. For example, car properties can be predicted by using a photo of the car.

But the task of making a personality portrait of a particular person is much more difficult than tasks above. That is because every person is different, and its personality portrait can be examined through its behavior and actions as well as its writings. It is practically impossible to create a database of personality portraits for entire human population in advance. Thus, methodology of incremental development of such databases for small groups of people needs to be developed. This task is especially important when applying for a job (e.g., pilot) or when choosing a speciality (e.g., mathematician) or a university. In this case, a personality portraits for a certain job position, speciality, university need to be created. Remember that it is more difficult to create an organization personality portrait than a human personality portrait. When a person needs to choose a speciality, more informed choice decision can be done by comparing these portraits.

In the last few years cognitive systems have had a boost in development. Now it is possible to use cognitive computing to solve problems of this kind.

There were psychological research in USA that have proven connection between one's writing style, word order, emotional tone and personality traits [1]. A prediction about one's predisposition to certain types of activities can be made based on particular personality traits. The research results were practically applied to a development of cognitive systems. Cognitive system can simulate person's perception system that was established in his mind during a process of development a character, education, learning, observation and reflection of the world.

Cognitive modelling allows us to find the best management decisions or case scenarios based on concepts that describe the evolving situation both qualitatively and quantitatively. Interference of different factors is also taken into account. Cognitive modelling can be done within companies for a purpose of supporting executives and managers who make decisions concerning future development of a company.

IBM Watson is one of the first cognitive systems in the world. It has many functions and also can understand and work with natural language. It uses large amounts of global unstructured data (Big Data). IBM Watson is a technology platform that uses natural language processing and machine learning to reveal insights from large amounts of unstructured data [2]. Today IBM Watson is available within a new cloud platform called IBM Bluemix that has introduced last year.

IBM Bluemix is an open-standards cloud platform for building, running, and managing applications. With Bluemix, developers can focus on building excellent user experiences with flexible compute options, choice of DevOps tooling, and a powerful set of IBM and third-party APIs and services [3].

There is a service available in Bluemix that allows developer to build personality portraits. It is called Personality Insights. The Personality Insights service uses linguistic analytics to infer individuals' intrinsic personality characteristics from digital communications that they make available via media such as email, text messages, tweets, and forum posts.

Personality Insights analytics are developed based on the psychology of language in combination with data analytics algorithms. The characteristics are described in terms of three models [1]: Big Five, Needs, Values.

One of the main goals of this work is to explore possibilities provided by IBM Watson and its application in everyday life. We have decided to create an application to help school graduates choose a speciality. Using Personality Insights service, the application gets personality profiles of a prospective student and specialities in our university and then compares them. Furthermore, it is possible to modify the application and use it in different IT companies when looking for candidates to a specific job.

The majority of school graduates don't know what they want to do in life because they unaware about how their interests and hobbies can be applied in life and what universities can offer to them. There are different psychological

tests that they can take in order to identify predisposition for a particular job. But it is possible to cheat on these tests, aware or unaware of it. Cognitive modelling allows building the most accurate personality portrait based on just writings. This approach takes into account word order, emotional tone and other characteristics of the text to make a prediction about person who wrote it.

We have developed an application called 'Choice of a speciality at a university', or simply 'Speciality helper'. It works as follows:

- User loads his writing (or provides his Twitter-ID);
- Application connects to IBM Watson services and gets user's personality profile;
- Obtained user profile is being compared to personality profiles of specialities at our university in order to find the best match.

The result is a prediction about the most suitable speciality to study. When comparing very different profiles it is possible to inform a user to consider other universities.

The method of obtaining specialities personality profiles requires clarification. There was no significant difference between specialities because of the texts we have used to get their profiles. Due to analysis of Personality Insights service, we have discovered two significant characteristics that affect the results:

- The text that is sent to the service should be written by user himself;
- The text should have some emotional tone, express the user's position and his opinion regarding the object or phenomenon described in it.

Now, when we know what kind of text we need, we have struggled with another problem: how to obtain speciality personal profile because speciality is not a person. Then we came to the assumption that the text about particular speciality should be written by graduates of this speciality. Speciality is a collective image of its graduates. We assume that there is minimum number of reviews about speciality that we can summarize in order to get its personality profile.

Thus, the main problem we are working on now is collecting and summarizing specialities reviews in order to create sufficient personality portraits. That is, we are developing optimal methodology of creating collective personality profiles.

Now we are working on methods and application that can be applied in other areas of human life. E.g., it is possible to create personality portrait of a desired candidate for a certain job position in company using collective personality portrait of successful employees of this particular position.

The application under development is based on an IT company template and will be used to upgrade the Talent Constructor application [4]. Talent Constructor is a packaged software product for effective management of organizational capacity.

In IT company template, the structure of a company is represented by a tree view, which starting node contains information about company at all, and child nodes contain information about departments. The edges show the structural subordination of the departments. Each node has associated competences: general, professional, leadership. The following entities form the basis of this approach. Employee (personality portrait, professional characteristics) – information is extracted from projects, emails, social networks. Candidate (requirements for a specific project) – information is extracted from projects, emails, social networks. Customer – information is extracted from documentation, emails, social networks, activity history. Project (completed projects, their characteristics, possible projects). Vacancy (requirements for a specific department).

The application compares a concrete entity profile with an abstract profile required by organization. Some reports are being formed.

The most effective administrative decision can be made using the results of comparison. Thus, company executives can manage the company more effectively.

As you can see, general solution of the problem can be effectively applied in different businesses. The comparison of personality profiles allows us to make a more informed decision, whether we are choosing a speciality to study or a company to work for.

The methodology and technology of creating a personality portrait of a person or speciality has been developed. This methodology has been tested using an application that helps prospective students choose a speciality. The application is written in Python using Django framework. It uses IBM Watson Personality Insights service available in IBM Bluemix cloud platform. Personality Insights service accepts text which should have at least three thousand words and returns so called 'Big Five' personality model. When we have person's personality portrait and speciality profile the comparison can be made. In developed application we use Euclidean distance to find suitable speciality.

References:

1. IBM Watson Developer Cloud (No date) *The science behind the Personality Insights service* [Online] IBM Watson Developer Cloud. Available: <http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/doc/personality-insights/science.shtml> [Accessed 21 March 2016]
2. IBM Watson (No date) *What is Watson* [Online] IBM Watson. Available: <http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html> [Accessed 21 March 2016]
3. I.Piletski et al. Virtual IT-environment for Big Data and VCL research: BIG DATA and Predictive Analytics. Conference proceedings, pp. 21-32. Minsk, BSUIR (2015). / И.И. Пилецкий и др. Виртуальная ИТ среда БГУИР для исследования Big Data и VCL, стр. 21-32, BIG DATA and Predictive Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий : сборник материалов междунар. науч.-практ. конф. / редкол. : М. П. Батура [и др.]. – Минск : БГУИР, 2015. – 220 с.
4. IBA Group (No date) *Talent Constructor* [Online] IBA Group. Available: <http://talentconstructor.by/> [Accessed 21 March 2016]