Knowledge discovery for business process optimization at Enterprise 2.0

Knowledge discovery für Geschftsprozessoptimierung in Unternehmen 2.0 (Enterprise 2.0)

Konstantina Plamenova Dimitrova
Master Student at Technical University - Sofia, FDIBA
e-mail: konstantina.dimitrova@fdiba.tu-sofia.bg

Abstract — The goal of this paper is to explore the possibilities of deploying Knowledge Discovery into the Enterprise 2.0 for business process optimization and the corresponding benefits. For the purpose, an exemplary theoretical implementation of Knowledge Discovery Process using Decision Trees and Excel Macros is presented.

Zusammenfassung — Das Ziel dieses wissenschaftlichen Artikel ist Erforshung der Implementierung von Knowledge Discovery in dem Unternehmen 2.0 zum Zweck Geschäftsprozessoptimierung und davon resultierenden Gewinn. Zum Veranschaulichen ist eine theoretische Lösung von Knowledge Discovery Implementierung mittels Entscheidungsbaum (Decision Tree) und Excel Macros dargestellt.

I. Prerequisites and current environment

In the era of the Zetta-byte the trend of exponentially growing data is no longer a myth. In the past few years the business investing even more in knowledge and data management in order to process the already gathered data and preparing for the vast amounts of data that are yet to come [1, 2]. But efficient data storage methods are not enough. Support-oriented services, for example, independently from their focus, accumulate an enormous amount of unstructured data like call and chat logs, tickets, emails, photos that is significantly more challenging to process than the structured. Accumulating data without processing it and putting the gathered knowledge into practice can result in losses for the Enterprise 2.0 that vary in form.

The development of the Web and the Internet, two fundamental conditions, have not only lead to exponential data growth but also presented the opportunity for inventions, like tools for improving our working routines and techniques. The term Web 2.0, which was coined by OReilly [2], represents elaborated combination of technologies like AJAX, SaaS, tagging, blogging as an example of collective knowledge and the collaborative web, the social networks. They are used on daily bases in every aspect of a work process for communicating information in the professional sphere and outside, managing the information flow from every entry point via cloud-based enterprise platforms, harvesting collaborative knowledge from the employees and easing the information and data flow within the organization. This is what defines an Enterprise 2.0 an organization that takes advantage of the Web 2.0 technologies and uses them to fuel up the companys growth and improvement in the constantly changing and challenging business environment [3, 4]. Although the Web 2.0 technologies that are used in the Enterprise 2.0 like Yammer or other forms of blogs for collaborative knowledge exchange, facilitate the working processes in the organization, they also produce additional valuable data as a byproduct that needs further processing. For example, a logged ticket about a service issue together with the accompanying email or call log are stored in databases, but the communication between the employees for providing a resolution is saved as well. Thus, one single event can result in the induction of newly generated data of unstructured type and of a great importance to the Enterprise 2.0.

II. Knowledge Discovery

Processing the stored data is challenging task for every Enterprise 2.0. The difficulty of accomplishing this task is even higher when the data to be processed is unstructured. A support-oriented service with two communication channels for customer direct calls and emails, for example, can handle hundreds even thousands of tickets per day. What is more important than the current issue described in the ticker is whether it is a one-time event or a part of an ongrowing trend. This could be determined using Knowledge Discovery techniques.

The aim of Knowledge Discovery is to reveal new, novel and useful information hidden in the unstructured data and not apparent upon first sight. This information can be discovered by deploying Knowledge Discovery techniques into the best practices of the Enterprise 2.0. the so-called Knowledge Discovery Process (KDP) can be implemented into the business processes and by its definition is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [5]. Although the process emphasizes on databases as the main

source of data, it also can be applied to the nondatabase source of data. The process consists of many steps, one of which is Data Mining, and each attempts to complete a particular discovery task. In order to put the KDP into practice, a Knowledge Discovery Model is realized. There are three types of KDPM academic, industrial and hybrid. The latter represents a compilation of the first two it provides more general, research-oriented description of the steps like in the academic model, while at the same time focuses on the understanding of objectives and requirements from the business perspective, like the industrial model. The models have iterative nature and introduce the Data Mining step instead of the modelling one. The six-step hybrid model by Cios et al. [6] is based on the CRISP-DM model and its steps are:

1. Understanding of the problem domain it involves working closely with the domain experts in order to gain better insight of the problem and thus easily defining it and determining the project goals, identifying the key people and learning about current solutions to this problem;

2. Understanding of the data firstly, a collection of sample data is gathered, checked for completeness, redundancy, missing values, the plausibility of attribute values, etc.;

3. Preparation of the data it is defined by sampling, running correlation and significance tests, data cleaning;

4. Data Mining data mining methods are used by the data miners to derive patterns and thus knowledge from the processed data;

5. Evaluation of the discovered knowledge in this step the goal is the understanding the results, checking whether the discovered knowledge is novel and interesting, interpretation of the results by domain experts;

6. Use of the discovered knowledge.

The KDPM puts emphasis on the iterative aspect of the process drawing from the experience of users of previous models. Several explicit feedback loops are identified in this model, e.g. [6]: (1)from the preparation of the data to the understanding of the data in order to guide the choice of specific data processing algorithms additional or more specific information about the data is demanded; (2)from data mining to the understanding of the data poor understanding of the data causes this loop, as it leads also to the incorrect selection of a data mining method and therefore to failure, etc.

The hybrid model with its characteristics is the most suitable KDPM for deployment into the business process of the Enterprise 2.0, as it reflects the need of the business while providing a sufficiently detailed technical description with respect to data analysis. Therefore, the fielding of a hybrid model into the business process can facilitate the processing of the fast amount of generated data within an Enterprise 2.0. A good example would be an enterprise, which besides its many services and products also provides support services via a variety of channels. Such example will be reviewed closed later in this paper.

III. Data Mining using Classification Trees

Data Mining is the fourth step of the KDPM [6]. It is the step of processing the produced data in order to discover data patterns and the deriving from them knowledge. There are four basic techniques for Data Mining Regression, Association Rule Discovery, Classification and Clustering [7]. All of them can be implemented to serve the needs of support-oriented services, depending on the expected end results. This paper focuses on the usage of Classification Trees and provides an example for classifying the generated information from support-centers for the purpose of further use in the Enterprise 2.0 for product development, improvement of the provided services, decision making etc.

Decision Trees (DT), a.k.a. Classification Trees, is one of the most commonly used data mining technique, because of its model, which is easy to understand [8]. It has a flow-chart-like structure, where each internal node denotes a test on an attribute, each branch represents the outcome of this test and leaf nodes that represent classes or class distribution [9]. DTs most common task is to build models for prediction of a class of an object based on attributes. An object can be a companys customer, a patient, email, transactions, even single character. The attributes used in DT are such that describe the object. For example, in textual data like an email, the combination of the characters *failed part* or *not syncing*, can be used as attributes and refer to an object of a class. These attribute that can be used to classify the data, based on if the particular combination of characters can be found or not in the textual data. Based on the attributes, the class of an object can be either positive or negative, depending on, if the object fits certain criteria or not [10]. For better classification results of unstructured raw data, a combination of attributes can be used to describe a class. Thus, improving the classification algorithms implemented in the KDP for business process optimization in the Enterprise 2.0.

IV. Example

When fielding a new process, technology even a workflow, there are plenty of factors that should be taken into consideration. For example the very structure and functionality, the employee training, the time for introduction, security etc. Regarding the last one security of the employees as well of the information, no compromises could be tolerated. The security measures at the Enterprise 2.0 can be kept while implementing and fielding a KDP for the purpose of process optimization as it could be realized by using already available software tools of the working environment. The simplified example is for a customer support team of an Enterprise 2.0, which uses could-based platform for documentation of the support service and standard set of office work tools include at least the Microsoft Office Packet as well as other specifically designed tools to fits the need of the work process. The provided service is technical support for cars. The aim of the team to process the gathered information so that new, valid, useful and novel knowledge is discovered and a corresponding report about the discoveries is issued. As the example is simplified, it is assumed that there is only one channel for communication with the owners of the products/ customers via email.

The induced data by the communication flow is stored in a cloud-based platform like Salesforce e.g. The theoretical and practical implementation can be carried out by a single individual, who is skilled in programming. The solution is realized with Macros for Microsoft Excel using VBA (Visual Basic for Application) and it consists of the following:

1. Microsoft Excel documents serving as dictionaries for the DT algorithm;
2. VBA Macros where the classification algorithms (decision trees in particular) is implemented.

The integrated KDP in the Enterprise 2.0 is based on decision tree algorithm. For the purpose, VBA macro in Excel is implemented in order to semi-automate the processing of the generated data. It sorts out data to predefined classes based on pre-set attributes in the decision three on which the solution is. The DT algorithm is described in Fig. 1 below. Firstly the algorithm checks, if any replacement parts regarding the product are marked as used in the report that is pulled out of the could enterprise platform. If it is so, this means that the car had a hardware defect that has been repaired. Then the algorithm continues to further process the data in order to determine how was the hardware issue fixed by sending a replacement part to the customer (CSR) or by sending a technician to replace the broken part (CE onsite). In order to do this, the algorithm relies on a dictionary, where all parts that are currently on the market and suitable for the model are enlisted. So the algorithm determines, if the part is in the dictionaries and use this check to spit the classes into a total of four classes of hardware issued, to which a case can be assigned to: CSR for known part, CSR for unknown part, CE onsite for known part, CE onsite for unknown part. Other possible failures of the product are software ones. In this case, no parts have been replaced in the car, so the data are processed further using pattern matching to determine the issue in the filed case against the warranty. For the pattern matching, a dictionary is required for reference. The second Excel dictionary consists of patterns combinations defining the most commonly occurring issued. It will be used as a database of references to the different detailed representation of the attributes. Examples for patterns variations are *sync*, *Synchronization*, *lost* + *signal* and *c* + *mobile* + *fail*. By using such dictionaries, the quality of the classification can be increased in at least two different ways. Either the dictionaries are developed further by adding new possible combination the dictionary, or/ and adding different threshold values for registering a hit (classification of an object). Thus, the simple VBA implementation can guarantee the aspect of continuous improvement in the business process of the Enterprise 2.0, which is basic feature of the business process management [11, 12].

By fielding knowledge discovery into the business process can significantly reduce the average handling time for the tasks and improve the quality. The increased quality of the service is assured by unifying the process flow and introducing standards and basic rules, which can further secure the execution and the quality of the results. Thus, the end product is not only more coherent, but also the all employees are encouraged to engage in the improvement of the business processes in the Enterprise 2.0.

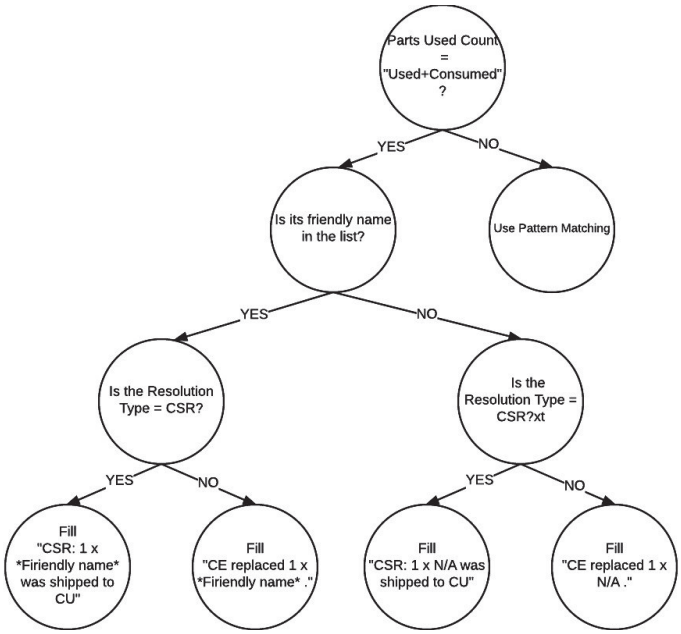

Fig. 1. Decision tree of the logic implemented in the solution

```
while Case != 0 DO
      IF Spare_Part == unused
            SearchIn(Dict_1)
            IF Spare_Part is in the Dict_1
                  IF CE_onsite==1
                        Classify(Case,Class_1)
                  ELSE
                        Classify(Case,Class_2)
            ELSE
                  IF CE_onsite==1
                        Classify(Case,Class_3)
                  ELSE
                        Classify(Case,Class_4)
      ELSE
            Use_Pattern_Matching()
            Classify(Case,Class_X)
```

Fig. 2. Exemplary implementation of the solution with pseudo code

V. Conclusion

Optimizing a business process by deploying Knowledge Discovery can result in different benefits for the Enterprise 2.0. It can help process the exponentially growing volumes of data, so that they can be transformed into valid, novel, potentially useful knowledge. Using simple automation solutions like shown in the example can not only significantly reduce the average time for processing unstructured raw data at the support center, the quality of the service, the business process, but also stimulate the employees to engage in the development process, to seek innovation and perfection.

References

[1] CISCO. (2017, June) The Zettabyte Era: Trends and Analysis. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html

[2] T. O'Reilly. (2009, September) What Is Web 2.0. Design Patterns and Business Models for

the Next Generation of Software. [Online]. Available: http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1

[3] AIIM. (2009) Collaboration and Enterprise 2.0. Work-meets-play or the future of business? [Online]. Available: http://www.aiim.org/PDFDocuments/36789.pdf

[4] A. McAfee. (2006, Spring) Enterprise 2.0: The Dawn of Emergent Collaboration. [Online]. Available: http://www.wikiservice.at/upload/ChristopheDucamp/McAfeeEntrepriseDeux.pdf

[5] F. Gullo. (2015) From Patterns in Data to Knowledge Discovery: What Data Mining Can Do, Vol. 4, No. 1. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187538921500036X

[6] K. Cios, W. Pedtycz, R. Swiniaerski, and L. Kurgan, *Data Mining. A Knowledge Discovery Approach*. Germany: Springer, 2007.

[7] M. Brown. (2012, December) Data mining techniques. [Online]. Available: https://www.ibm.com/developerworks/library/ba-data-mining-techniques/

[8] ZenTut. (2017) Data Mining Techniques. [Online]. Available: http://www.zentut.com/data-mining/data-mining-techniques/

[9] A. Ashari, A. M. Tjoa, and I. Paryudi, "Performance Comparison between Naive Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 4, no. 11, pp. 33 – 39, 2013.

[10] P. Ozer, "Data Mining Algorithms for Classification," January 2008, bSc Thesis Artificial Intelligence.

[11] J. Xiao. (2017) life cycle of business process management. [Online]. Available: http://bigthink.com/articles/life-cycle-of-business-process-management

[12] S. Grier. (2008-2017) The 7 Stages of Business Process Management. [Online]. Available: http://itmanagersinbox.com/846/the-7-stages-of-business-process-management/