A New Way to Classify Physical Effects for Ontology Instantiation

Pei Zhang[1](✉), Denis Cavallucci[1,3], and Cecilia Zanni-Merk[2]

[1] LGéCo (Design Engineering Laboratory), 67000 Strasbourg, France
pei.zhang@insa-strasbourg.fr
[2] INSA Rouen Normandie, LITIS, Normastic (FR CNRS 3638), Rouen, France
[3] INSA de Strasbourg, 24 boulevard de la Victoire, Strasbourg, France

Abstract. As one of the most important knowledge sources of TRIZ, the collection of Physical effects is currently searched in a very basic way. In order to enhance its use, different proposals have been brought to the community to classify effects into different categories. Among them, a rule-based approach classified the collection of physical effects into four categories and built a set of rules to facilitate its direct use. However, this approach is not robust enough due to the lack of instances of physical effects. In this paper, we propose a new approach to classify physical effects in order instantiate the existing ontology. In addition, preliminary results are presented to demonstrate the feasibility of the approach. The results brought us evidences that we facilitate the direct access to the collection of physical effects.

Keywords: TRIZ · Inventive design theory · Wikipedia · Physical effects

1 Introduction

According to Altshuller [1] the use of the scientific effects is one of the most important approaches that facilitates solution finding. For the reason that this collection of knowledge contains natural phenomena previously unused in the engineering domain. The use of this collection of effects often give rise to simple and reliable designs. In order to facilitate its use, the collection of physical effects provides the mapping between the technical functions and the available technical laws.

However, the technical functions and the available technical laws while the physical effects are a collection of physical phenomenon. They are at different levels of abstraction, thus making the access of the collection of physical effects from technical functions quite difficult. In order to address this problem, existing researches have proposed different ways to classify this collection of knowledge in order to ease its usage. These classifications are either based on the physical parameters that describe the effects or based on the categories defined by the constructed domain models such as ontologies. However, the existing classifications do not support the direct use of the collection of the physical effects, as a consequence, it requires knowledge both in engineering and physical domain.

Tackling at this drawback, existing research have proposed a physical effects ontology to facilitate its direct access by classifying the physical effects into two

Published by Springer Nature Switzerland AG 2019
R. Benmoussa et al. (Eds.): TFC 2019, IFIP AICT 572, pp. 70–78, 2019.
https://doi.org/10.1007/978-3-030-32497-1_7

categories (substance effects and field effects) [2]. However, the knowledge base is not complete thus limits its use. Therefore, in this paper, we focus on instantiating the existing physical effects ontology based on machine learning to support the construction and the population of the physical effect knowledge base, with the aim of support decision making by reusing existing ontology.

The remainder of this paper is organized as follows. In chapter two, we give a literature review about the existing classifications of physical effects and the need of ontology instantiation. In chapter three, we present the proposed method. In chapter four, we present the preliminary result to validate the proposed method. Finally, in chapter five, we conclude this paper by discussion and conclusion.

2 Literature Review

The classical way to classify the physical effects is by their functions, which is knownas the pointers to scientific-engineering effects (as it is depicted in Fig. 1). The pointers classified the effects by different technical functions they perform. For example, the technical function *Change shape* can be achieved by the use of the effects like *Curie point, Evaporation, Ferromagnetism, crystallization etc.*

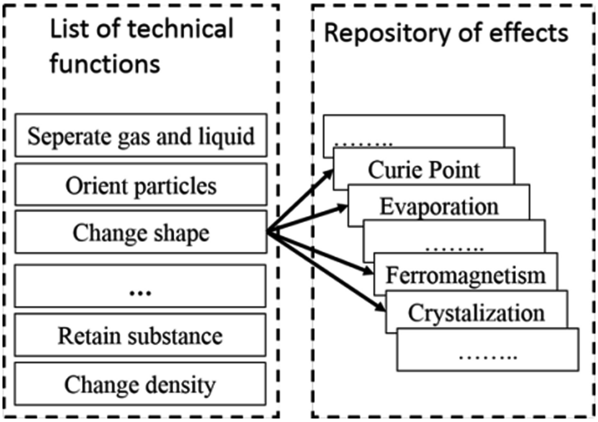

Fig. 1. Pointers to scientific-engineering effects

In this way, the effects can be accessed by searching for the technical functions. However, the use of the scientific-engineering depends on the user's experience in the engineering domain to map between conceptual (technical functions) and actual (scientific-engineering effects) solutions. Therefore, this type of classification is useful for experienced users but making its use for novice users very difficult.

Apart from the pointers to scientific-engineering effects, the classification of effects based on the construction of a domain ontology is getting more and more attention. This is because the construction of an ontology not only facilitates the description of the modeling domain, but also enables the knowledge induction by the establishment of rules. Combined together, the construction of ontology enables the automatic inference of needed knowledge without the intervention of human.

An ontology is a formal and explicit specification of a shared conceptualization. It is composed of classes, instances and their binary relations that are used to express knowledge about the domain of interest. In an ontology, classes are abstract collection of objects. The instances are concrete objects of its class. Relations are links between pairs of classes, pairs of instances or between classes and instances. Rules are another form of expressing knowledge in the domain of interest [2]. They are used to reflect the notion of consequence and are in the form of IF-THEN-constructs. In this way, rules are able to express complex statements of different types. Based on the construction of rules, techniques of automated reasoning allow a computer system to draw conclusions from the existing ontology.

The work of [3] organized the effects by a chain of design elements at different abstraction level. It creates the causal relationship between the functions and structures by physical variables. It is based on the fact that most of the physical laws include variables and constants. And the application of these physical laws depends on such descriptors. The work in [4] constructed an ontology that classifies the physical effects based on the language descriptors. In this ontology, the classes are input, output and object. The relations for representing the interactions between the classes are: Cause Action, EffectAction and ActionObject. In this way, different effects can be classified and represented by different text descriptors that are extracted from natural language texts. Other direction of this research has conducted by authors in [5] as well, who proposed a dynamic approach that does not rely on the classification in advance.

The work in [6] developed the physical effects ontology that classified the physical effects into two categories: Sub_PE and Field_PE (Fig. 2) in order to facilitate its use with the Inventive Standards. The physical effects ontology classified the physical effects based on the substances and fields it concerns. Along with it, this work also constructed the rules based on the Inventive Standards. These notions are particularly interesting because they enable the user to access directly to the needed effects by reasoning on rules once they have obtained the substance-field model of the his/her problem. Therefore, the novice user can access to the needed effects through the substance-field model without understanding what function that is performed by which effects. However, even though the physical effects ontology has been built, it should be instantiated largely in order to provide enough knowledge for the users. Therefore, we should consider a method to instantiate the physical effects ontology to facilitate its reuse.

Ontology instantiation is the process of building the knowledge base. It consists of adding new instances of concepts and relations into an existing ontology. This process usually starts after the conceptual model of ontology is built [7].

The construction of the knowledge base makes it possible to perform reasoning tasks with the aim of assisting decision making. However, to construct a knowledge base is not an easy task because it is based on the capture of categorized knowledge. Therefore, it is often done manually by domain experts.

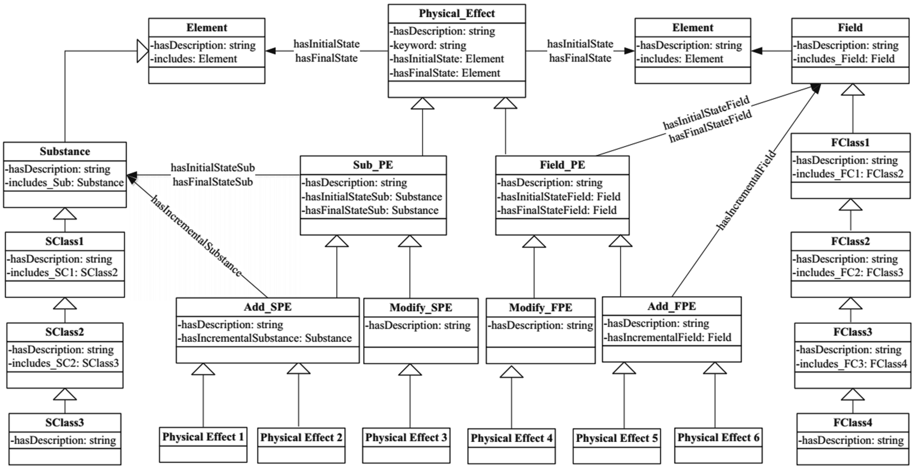

Fig. 2. The physical effects ontology (from [5])

In order to solve this problem, we adopt machine learning to automatically classify the physical effects in order to instantiate the physical effect ontology and in this way, constructing the physical effects knowledge base.

3 Methodology

In this chapter, we propose a new approach based on machine learning which enables the instantiation of physical effects knowledge base. The proposed approach is implemented to classify the collection of physical effects as two classes (Sub_PE and Field_PE). The proposed methodology is composed of three steps as it is illustrated in Fig. 3:

- Step1: data collection
- Step2: feature extraction
- Step3: classification and ontology instantiation

The first step is to collect the data needed. In here, the data should be in the machine interpretable form in order to apply the classification techniques later. However, there are two main difficulties. One is that the physical effects are in the form of natural language and we need to find a way to represent them in a computer process able way. The other is to find an appropriate text similarity measure in order to perform the classification task.

We address these problems by using Wikipedia[1] as a knowledge base to create a graph for each effect based on the Wikipedia category network. In this way, each

[1] https://en.wikipedia.org/wiki/Main_Page.

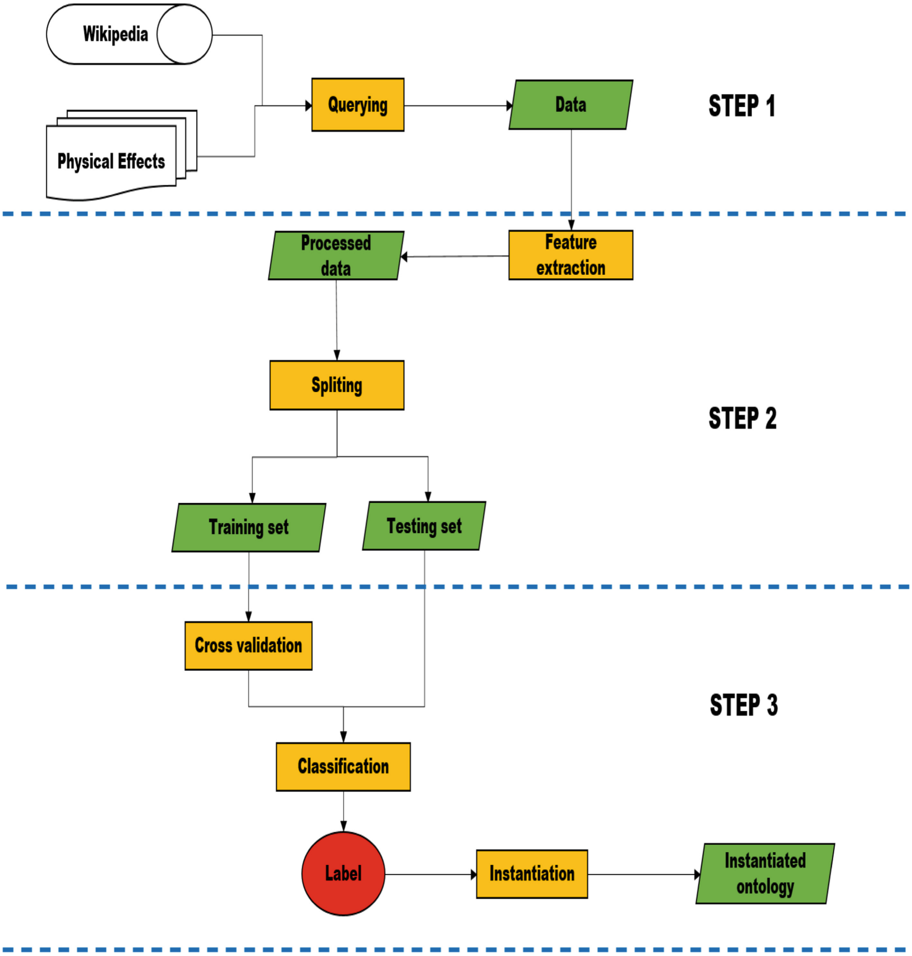

Fig. 3. The proposed methodology

physical effects will be represented as a hierarchical graph composed of a set of its related categories.

The second step is feature extraction. Based on the characteristics of the obtained graphs, which are not suitable for applying classification algorithm, we have to find a way to extract features from them. Feature extraction consists in transforming arbitrary data, such as text or images, into numerical features usable for machine learning tasks (e.g. classification task). In order to extract features from the obtained graphs, we are inspired by the vector-space model [8] and represent each effect by all the categories associated with it. We assign the value of each feature by the distance between the effect and each category on the shortest path from the effect to the category. In this way, we obtained a multi-dimensional vector for each effect, where each category related to this effect on the retrieved graph corresponds to an axis. Such that the values along the axes for the effect correspond to the distance between this category and the effect on the shortest path linking them.

In this way, we can construct a large feature matrix where each row corresponds to an effect and each column corresponds to a category that is retrieved from the Wikipedia network. However, the resulting feature matrix is too large which will increase the training time. Therefore, we have to reduce the dimensions of the matrix by preserving the most significant features to shorten the training time for the predictive model. To do so, the Principal Component Analysis (PCA) is applied to achieve this goal.

Once the feature extraction is done, we can apply the classification algorithm to classify the data. To do so, we have to train a classifier. A classifier is a function that maps an unclassified piece of data to a class by applying an induction algorithm, which builds a classifier from a given dataset [9]. In order to train the classifier, we have to apply a proper classification algorithm [10] on the training set obtained from the previous step. One of such method is the k-Nearest Neighbor algorithm (kNN) [10]. The advantages of applying the kNN classification algorithm is obvious: There is no existing model to classify the physical effects but a collection of correctly classified effect instances that is labelled by the domain experts. kNN method assumes that the observations which are close together will have the same classification, making it possible to classify the given effects based on the effects with a label. Once the classifier is obtained, we can instantiate the physical effect ontology based on the obtained label.

4 Preliminary Result

In our experiment, we try to validate the proposed approach to classify the physical effect Boiling. We take 11 physical effects to conduct our experiment. Among them, we use 10 as the training set and 1 as the testing set. The aim is to classify the effect boiling into one of the two categories.

Firstly, we have to obtain the graph of each effect. It is obtained by retrieving the Wikipedia category network. Therefore, for each effect, we obtain a category graph by querying Wikipedia. In Fig. 4, an excerpt of the obtained graph of boiling is presented. Once the graph of each effect is obtained, we eliminate the irrelevant categories. For example, at the first level of the category graph of boiling, there are two irrelevant categories. They are All_Articles_needing_additional_references and All_Articles_needing_additional_refences_from_June_2017. In Fig. 4, these irrelevant categories presented by the red nodes. In our approach, these categories are deleted in step 1.

The next step is to apply the feature extraction method to transform the obtained graph into a multi-dimensional vector. To achieve this goal, we assign the shortest distance between a category and the effect on the shortest path between them. For example, the value of feature Phase_transitions is 1 because the distance between the category Phase_transitions and the effect Boiling on the shortest path is one. In addition, a special case is that there is no path between the effect and a category, for example, the category Electromagnetic_radiation, Electrodynamics and Radiation. In this case, we assume that the distance is 11. This is because in our graphs, the distance from an effect to the Contents category (root node) ranges from 3–10, therefore, 11 means that the value is infinity.

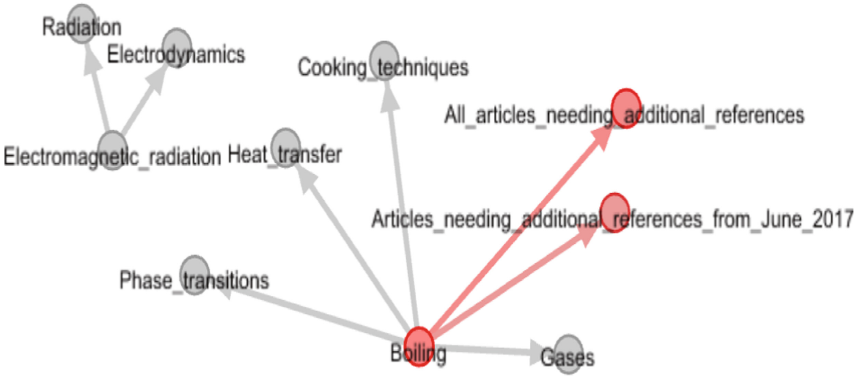
Radiation
Electrodynamics
Cooking_techniques
All_articles_needing_additional_references
Electromagnetic_radiation
Heat_transfer
Articles_needing_additional_references_from_June_2017
Phase_transitions
Boiling
Gases

Fig. 4. An excerpt of the effect Boiling

Once the vector of each effect is obtained, we construct a feature matrix with 11 rows and 89 columns where each row corresponds to an effect and each column corresponds to a category that is retrieved from the Wikipedia network (as it is presented in Fig. 5). With the obtained matrix, PCA method is employed for feature selection. We fit the PCA preserving 95% of components and obtained a new matrix with 11 rows and 5 columns, where each row corresponds to an effect and each column corresponds to the assigned k principle components after the feature selection.

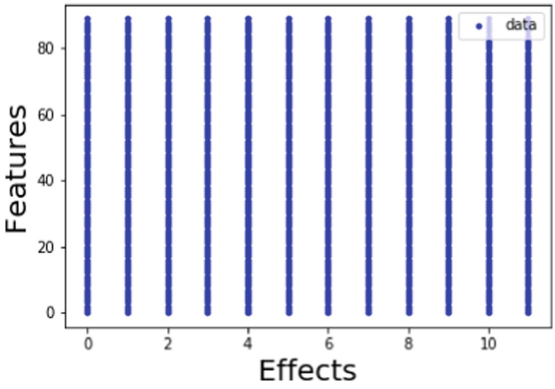

Fig. 5. Visualization of the feature matrix

Then, the obtained training set is used as the input for classification task by applying the kNN method. And finally the 10-fold cross validation is applied to determine the value of k. To do so, we divide the training set into 10 subsets of equal size and repeat the 10-fold cross validation 10 times, where each time one subset is assigned as the Testing_set_cv and the rest are assigned as the Training_set_cv. Therefore, each time the cross validation is performed, we can evaluate its performance by calculating its accuracy rate. The accuracy rate is obtained by dividing the sum of

TP and TN by the sum of P and N, where TP is the number of true positive samples, TN is the true negative samples, P is the number of positive samples and N is the number of positive samples. In this way, the mean accuracy rate is obtained by calculating the sum of the 10 accuracy rates and divide it by 10 as it is depicted by Eq. (1).

$$\text{ACC\_mean} = \frac{\sum_{x=1}^{10} (TP + TN)/(P + N)}{10} \tag{1}$$

The 10-fold cross validation is used to determine the value of k, which is the number of the nearest neighbors. We varied the value of k from 1 to 5 and the ACC_mean of k = 1, k = 2, k = 3 and k = 5 are respectively presented in Table 1. From this result, we can observe that k = 5 is a better choice than the others since it yields a better accuracy rate than the other values of k.

Table 1. Experiment result

Category	K = 1	K = 2	K = 3	K = 5
Substance_effect	0.5	0.5	0.873	0.955
Field_effect	0.5	0.7	0.7	0.873

5 Discussion and Conclusion

In this paper, we proposed a new approach to classify the physical effects based on machine learning. More specifically, it is based on the use of Wikipedia and kNN method. The proposed method can be applied to instantiate the physical effects database automatically which contributes the reusability of domain ontologies and the independence from domain experts.

The preliminary result showed that we have successfully applied the proposed method to classify the physical effects, and achieved the mean accuracy of 0.995 when k = 5. This encouraging result enables further directions of research:

- To largely populate the physical effects into two classes by a bigger data set;
- To populate the relations of the physical effects ontology with more refined classifications.

However, the proposed method relies on labelled data which is sometimes difficult to obtain, therefore, there is a need to find some ready to use dataset in order to instantiate the physical effects ontology with more individuals. Moreover, it is also interesting to test other machine learning techniques on a larger dataset in order in increase the precision of the classification result, such as naive bayes classifiers [11], support vector machines [12], radial basis function (RBF) networks [13] etc.

References

1. Altshuller, G., Shulyak, L.: And Suddenly the Inventor Appeared: TRIZ, the Theory of Inventive Problem Solving. Technical Innovation Center, Inc. (1996)
2. Grimm, S., Hitzler, P., Abecker, A.: Knowledge Representation and Ontologies Logic, Ontologies and Semantic Web Languages (2007)
3. Rihtaršič, J., Žavbi, R., Duhovnik, J.: Application of wirk elements for the synthesis of alternative conceptual solutions. Res. Eng. Des. **23**(3), 219–234 (2012)
4. Korobkin, D., Fomenkov, S., Kamaev, V., Fomenkova, M.: Multi-agent model of ontology-based extraction of physical effects descriptions from natural language text. In: Information Technologies in Science, Management, Social Sphere and Medicine (2016)
5. Russo, D., Montecchi, T., Caputi, A.: Tech-finder: a dynamic pointer to effects. In: 16th International TRIZ Future Conference, vol. 1, no. 3, pp. 79–87 (2018)
6. Yan, W., Zanni-Merk, C., Rousselot, F., Cavallucci, D., Collet, P.: A heuristic method of using the pointers to physical effects in su-field analysis. Procedia Eng. **131**, 539–550 (2015)
7. Makki, J., Alquier, A.-M., Prince, V.: Semi automatic ontology instantiation in the domain of risk management. In: Shi, Z., Mercier-Laurent, E., Leake, D. (eds.) IIP 2008. ITIFIP, vol. 288, pp. 254–265. Springer, Boston, MA (2008). https://doi.org/10.1007/978-0-387-87685-6_30
8. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM **18**(11), 613–620 (1975)
9. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI **14**, 1137–1145 (1995)
10. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13**(1), 21–27 (1967)
11. Rish, I.: An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3. no. 22 (2001)
12. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0026683
13. Howlett, R.J., Jain, L.C.: Radial basis function networks 2: new advances in design, vol. 67. Physica (2013)