AGILE Association of Geographic Information Laboratories in Europe

Comparing supervised learning algorithms for Spatial Nominal Entity recognition

Amine Medad[1], Mauro Gaio[1], Ludovic Moncla[2]
Sébastien Mustière[3], and Yannick Le Nir[4]

[1] Pau University (UPPA), LMAP, UMR 5142 Pau, France,
{im.medad,mauro.gaio}@univ-pau.fr
[2] INSA Lyon, LIRIS UMR 5205, Lyon, France,
ludovic.moncla@insa-lyon.fr
[3] Univ. Paris-Est, LASTIG IGN, Saint-Mandé, France,
sebastien.mustiere@ign.fr
[4] EISTI, Pau, France, yl@eisti.eu

Abstract. Discourse may contain both named and nominal entities. Most common nouns or nominal mentions in natural language do not have a single, simple meaning but rather a number of related meanings. This form of ambiguity led to the development of a task in natural language processing known as Word Sense Disambiguation. Recognition and categorisation of named and nominal entities is an essential step for Word Sense Disambiguation methods. Up to now, named entity recognition and categorisation systems mainly focused on the annotation, categorisation and identification of named entities. This paper focuses on the annotation and the identification of spatial nominal entities. We explore the combination of Transfer Learning principle and supervised learning algorithms, in order to build a system to detect spatial nominal entities. For this purpose, different supervised learning algorithms are evaluated with three different context sizes on two manually annotated datasets built from Wikipedia articles and hiking description texts. The studied algorithms have been selected for one or more of their specific properties potentially useful in solving our problem. The results of the first phase of experiments reveal that the selected algorithms have similar performances in terms of ability to detect spatial nominal entities. The study also confirms the importance of the size of the window to describe the context, when word-embedding principle is used to represent the semantics of each word.

Keywords: Geographic Information Retrieval, Natural Language Processing, Nominal Entity Recognition

1 Introduction

A critical aspect of polysemy (i.e., the ability to have multiple meanings) is that the different meanings of a word can be conceptually closely related but in very distant semantic categories. In most cases, the contextual support of evocation makes it possible to retain the appropriate meaning. For example, consider the word '*church*' used to refer an organisation sense as in sentence (1). In this sentence (1), it allows to personify an affirmation, versus a building sense, as in sentence (2) here used as a spatial reference point.

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

(1) 'Depending on the tradition, these organisations may connect local churches to larger **church** traditions.'[5]

(2) 'Continue straight on towards the **church** then turn right to reach the park.'[6]

In geographic information retrieval approaches involving Natural Language Processing (NLP) methods, the task of Word Sense Disambiguation (WSD) is defined by Vanetik and Litvak [1] as a set of methods that automatically assign the appropriate meaning to a polysemous word. Lesk [2] used the context (short phrase containing the ambiguous word) to look for partial matching with the definitions in dictionaries (glossaries) of the ambiguous word and its context words in order to disambiguate the word sense. Lesk's method aims to disambiguate the sense of any word of the vocabulary, which depends on words' definitions in dictionaries that are often short and do not provide enough context. In this paper, we also consider the local context of words to disambiguate them, with a focus on a more specific case of spatial entities.

The recognition of entities is an important task in NLP and according to Vicente [3], named entity is defined as a concept used to designate a mono-referential discursive element, which coincides with a proper name definition and follows specific syntactic patterns. The 'reference' refers to the link between the linguistic expression and a single element of the world (i.e., the referent). There may also be situations where, due to a specific context, a common noun or a nominal mention could be used as a co-reference to a named entity or may have the ability to refer to a unique object as presented in the example (2).This will be referred to as a nominal entity.

As part of the Choucas project [7], we are interested in the identification of the two types of entities referring to places (i.e., named and nominal). This identification must be carried out on the basis of hikes descriptions in the form of unstructured text [4]. With regards to named entities, Gaio and Moncla [5] have proposed the concept of *Extended Named Entity (ENE)*. They argue that a named entity can be composed of a proper name and a descriptive expansion. The descriptive expansion is made of common nouns that can change the default type of the object referenced by the proper name on its own, (e.g: *'maire de Gavarnie'*: Gavarnie (populate place), mayor of Gavarnie (social function)). They also define the concept of Extended Spatial Named Entity (ESNE) as an ENE that designates a specific spatial object (e.g: *'Boulevard du Général Charles de Gaulle'*: Charles de Gaulle (person), Général Charles de Gaulle (social function), Général Charles de Gaulle Boulevard (location)).

Nominal entity detection can be viewed as an extension of the Named Entity Recognition (NER) task. Therefore, the problems described for named entities must also be considered for nominal entities. As shown in examples (1 and 2) the same noun can be used to refer to a spatial object or else, which leads to ambiguities. Therefore, the recognition and disambiguation of spatial nominal entities in a corpus of an unstructured text compose a challenging problem to which we propose a solution. In this

[5] Source: Wikipedia

[6] Source: hiking description (https://www.visorando.com)

[7] The CHOUCAS project (http://choucas.ign.fr) is a French interdisciplinary research project aiming to respond to a need expressed by the high mountain gendarmerie platoon to help localising victims in mountain area. When they describe their position by referring to surrounding elements like in "I can see Mont Blanc and I am close to a lake"

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

paper, we first introduce the concept of spatial nominal entity (SNoE) and we propose an approach for their recognition from unstructured texts. For this purpose, we trained several supervised learning algorithms to study their ability to detect whether a nominal entity identified in the sentence is used as a reference to a spatial object or not.

The remainder of this paper is structured as follows. In section 2 we present an overview of tasks and methods from NLP domain related to our work such as: named and nominal entity recognition and categorisation, word-embedding and transfert learning. Section 3 is dedicated to the definition of the concept of SNoE and provides methodological details of our approach for SNoE recognition. In section 4 we describe the dataset we have generated and manually tagged. This dataset is used to train and evaluate the studied algorithms mainly to demonstrate the feasibility of our approach; we also expose and discuss the experimental results of these algorithms. Finally, section 5 concludes this paper.

2 Related work

NER is considered as the key task in the field of WSD. NER implementations are based on a wide variety of methods and their role is to recognise named entities in a sentence and classify them in various classes (e.g. Name of Location, Person, Organisation, Quantity, Time, Percentage etc.). Despite the many implementations available, there is a great need to develop methods to refine the capabilities of NER, since existing tools have a limited scope. In particular, the vast majority, if not all of these tools are not able to recognise nominal entities (without proper nouns). Whatever these limitations, it appears in the literature that NER has been addressed by both machine learning and knowledge-based approaches.

Learning methods are based on labelled learning data sets, usually by a human who does not need to be an expert in linguistics such as in [6,7,8,9,10]. Knowledge-based approaches use hand crafted syntactic and semantic rules developed by linguistic experts. They involve morpho-syntactic structures and specific resources (e.g., lexicons, gazetteers) [11,12,13,14]. In [15,16,17] both type of approaches have been combined to build hybrid methods where the input features of the machine learning algorithms are provided by knowledge-based systems. Once entities are recognised, the considered categories may vary. For example, some NERs have the 'acronym' category while others do not but can categorise dates, etc. However, the 'location' category is always present.

For instance, the well-known Stanford NER [18] is based on a features extraction and Conditional Random Fields (CRF), the system categorises named entities in three classes ('person', 'organisation' and 'location').

NER systems dedicated to location are known as: 'geoparsers'. In general, geoparsers proceed two sub process: geotagging and geocoding. The geotagging (i.e. recognition) consists in marking in texts all segments containing a named entity referring to a place (i.e. place name. The geocoding (i.e. resolution) assigns a single couple of geographical coordinates to the previously identified (in geotagging step) named entity. Karimzadeh and al. [19] have proposed a geoparser called Geotex. This geotagging system is a web-based geotagger where; users have a choice within a list of 6 publicly

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

available NER systems (Standford NER [8], ANNIE [9], Illinois NER [10], MITIE [11], Apache OpenNLP [12], LingPipe [13]). The Edinburgh Parser [20] is a major geoparser whose; the geotagging task is performed by a multi-rule based geotagger. Moncla et al. [21] have proposed a system called Perdido[14], consisting in a rule-based method implemented with a cascade of transducers for the generic recognition of ESNE structures. The resolution is done within specific corpora composed exclusively of textual descriptions of pedestrian movements.

As stated in the introduction, words used to construct nominal entities are polysemous and the context is the main available information for identifying the used meaning. A solution that currently seems to be very promising is the Word-Embeddings (WEs). WEs are continuous space language models built using Neural Networks (NN). The main idea behind WEs is to project a set of words of a vocabulary of size N_v into a continuous vector space of a lower dimension N_d (knowing that $N_d << N_v$). As a result, each word of the vocabulary is represented as a real-valued vector in a low-dimensional space and words with similar representations appear in similar contexts. WEs can be learned in an unsupervised way to capture distributional similarities between words of the vocabulary, and be fine-tuned in a supervised context. Several works such as [22,23,24,25,26,27] have used NN to learn distributed representations for words. These approaches differ in the type of the model and the data used to train the model.

The principle of producing WEs through neural networks was first introduced by Bengio [28]. Recently, Bojanowski et al. [29] have proposed FastText, a WE method that takes into consideration the internal structure of words by including character sequences in the learning process of word representation, which has proved to be of a great impact when working with morphological rich languages such as French or Finnish. WEs has opened a new direction for many NLP tasks based on NN such as question answering [30,31], sentiment analysis [32,33,34], relation extraction and classification [35,36], NER [8] and mention detection [37].

In our context of implementing a WSD process, geoparsing and geotagging named entities and their spatial-based context is fundamental but not sufficient. Therefore, it is essential to apply the same kind of processing to nominal entities. To the best of our knowledge, none of the actual state-of-the-art works attempt to identify SNoE, at least for French language.

In the absence of a French annotated corpus of nominal entities, our methodology is based on the principle of transfer learning (TL). According to the proceedings of the NIPS-95 workshop entitled 'Learning to Learn' [38], TL was primarily motivated by: "the need for lifelong machine-learning methods that retain and reuse previously learned knowledge". Moreover, the information Processing Technology Office (IPTO) of the Defense Advanced Research Projects Agency (DARPA) published

[8] http://nlp.stanford.edu/software/CRF-NER.shtml
[9] https://gate.ac.uk/sale/tao/splitch6.html#chap:annie.
[10] http://cogcomp.cs.illinois.edu/page/demo_view/ner
[11] https://github.com/mit-nip/MITIE
[12] https://opennlp.apache.org
[13] http://alias-i.com/lingpipe/demos/tutorial/ne/read-me.html
[14] http://erig.univ-pau.fr/PERDIDO/demonstration/

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

a Broad Agency Announcement N° 05-29 in 2005 [15] where they define TL as "the ability of a system to recognise and apply knowledge and skills learned in previous tasks to novel tasks". Following the principles of TL, we propose to use the FastText[16] pre-trained WE model as input of different supervised learning algorithms. Then, we compare the obtained results with two manually labelled datasets.

3 Automatic identification of SNoE

3.1 Concept and definition

The SNoE is defined as a nominal phrase that refers to a physical object which is usually involved in a spatial-based context. SNoE may be a common noun composed of a single token (village, hut, church) or composed of several tokens (boundary marker, tourist office, transformer substation). The concept of SNoE derives from the concept of nominal entity that was defined in the Entity Discovery and Linking task[17] as "A nominal mention consists of a common noun which refers to an entity in place of a name" and is classed into 5 different types ('person', 'location', 'organisation', 'facilities', 'geo-political entity'). Hence, a SNoE is composed of at least one common noun (i.e. the pivot) involved in a spatial-based context (e.g. [...] reach the **summit**.) in which there is no proper name present, because otherwise the noun 'summit' is a component of an ESNE (e.g. [...] climbing to the **summit of Mont Blanc**).

In our definition, the concept of SNoE covers:

- Physical static entities that have fixed geographical coordinates, such as 3a.
- Spatial objects with the property of being able to be in motion, as shown in example 3b.
- A group of physical objects forming a unique spatial reference point, such as 3c.

Consequently, this concept does not cover:

- Nominal phrase involving a common noun, which may refer to a physical object, but associated with a proper name, such as 4a.
- Nominal phrase only used for its ability to evoke the object (abstract or physical) as a concept without a spatial reference, such as 4b.
- A spatial reference to a virtual object without physical existence, such as an ephemeral entity which exists only in a specific moment of a narrative, as illustrated in example 4c

(3) a. Continuer la descente en **sous-bois** pour rejoindre le lac.
*'Continue downhill in the **undergrowth** to reach the lake.'*

b. Prendre le sentier qui passe sous le **téléphérique**.
'*Take the path that goes under the **cable car**.*'

[15] http://logic.stanford.edu/tl/TransferLearningPIP.pdf
[16] https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md
[17] https://tac.nist.gov/2016/KBP/guidelines/TAC_KBP_2016_EDL_Guidelines_V1.1.pdf

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

c. Le sentier grimpe au-dessus du hameau avec un passage dans les **rochers**.
*'The path climbs through **rocks**, above the hamlet .'*

(4) a. À partir des **chalets de l'Échet**, faire demi-tour et rejoindre le carrefour précédent.
*'From the **chalets de l'Échet**, make a U-turn and walk back to the previous crossroads.'*

b. Le **chalet** est un bâtiment rural des régions de montagne, dont le bois est le constituant essentiel.
*'The **chalet** is a rural building of mountain regions, essentially built of wood'*

c. Pour la **descente**, revenir sur ses pas pour une bonne centaine de mètres de dénivelé pour rejoindre le carrefour de montée.
*'For the **descent**, retrace your steps for a good hundred meters of drop to reach the crossroads of the ascent.'*

Furthermore, as mentioned previously, words may be polysemic, they may have a different meaning depending on different syntactic or semantic contexts. In the scope of our problem, we distinguish two main categories covering three different senses for a specified word:

1. The word is used to identify a physical object used as a landmark, such as example 2.
2. The word is used to identify a non physical object, an abstract entity with no physical borders, it could be a reference to an organisation such as 1 or the word is used to identify a physical object which is not used as a landmark, as in the following example 5a.

(5) a. [...] **les églises** à l'époque gothique sont parfois revêtues d'une épaisse couche d'enduit ou de mortier [...]
'[...] ***churches** in the Gothic period are sometimes covered with a thick layer of plaster or mortar* [...]'

3.2 Methodology

In order to detect SNoE, we are considering the development of a system based on a supervised machine learning approach. As shown on figure 1 the process-chain is divided in two main phases: the pre-processing phase and the learning phase.

Pre-processing The pre-processing phase is an input preparation step for the learning phase. Three tasks are performed: 1) establishing a lexicon containing a varied list of terms that can constitute the pivot 2) context setting and 3) semantic representation of words.

Although it is recognised that the left context is generally more important in French than the right context, there are cases where the right-hand context is useful to improve discrimination. We extract n-grams from sentences because both the right and the left

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

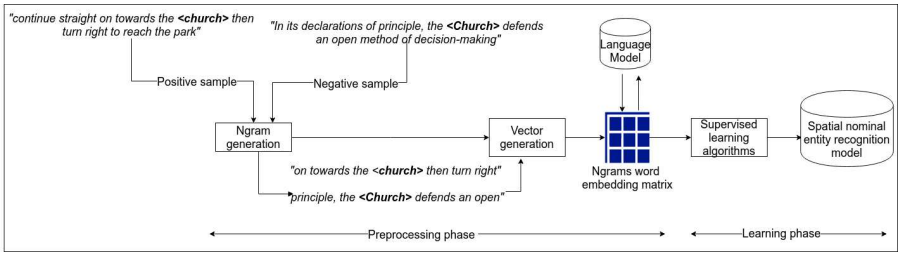


Fig. 1. Supervised machine learning approach for spatial nominal entity recognition

context are useful and important to determine whether or not the pivot in the sequence of n-grams considered is used as a spatial entity. In order to obtain the n-grams from a given corpus we have constructed a lexicon of terms that can refer to spatial entities. For building this lexicon, we propose to manually extract a set of words used as SNoE from a set of French hiking description texts, such as *lac*, *pont*, *église*, *avenue*, and *office du toursime* (respectively lake, bridge, church, avenue, tourist office). This lexicon is then used to extract n-grams from a sentence (see example 6) while the n-grams represents the context of the pivot. The extracted sentences are then manually annotated in order to build the different datasets used for training, testing and validation. Our hypothesis is that the principle of the n-grams (with the size of n yet to be defined) associated with the principle of TL are sufficient for the different algorithms under study to achieve a reasonably good rate of expected decision.

(6) The sentence: [...] *'the path climbs through rocks, above the lake to reach the country road'* [...]
The pivot word: *'lake'*
The 7-grams context: *' , above the lake to reach the'*

Once the n-grams extracted, the next step of the prepossessing phase is to vectorize the inputs. In accordance with the principles of TL, each word x_i of the n-grams N is transformed into a vector e_i of dimensionality d_e by looking it up in the WE table of a pre-trained FastText. As a result, the original n-grams can be now viewed as a matrix X of size $n * d_e$:

$$X = [e_1, e_2, ..., e_n] \tag{1}$$

Notice that sometimes the pivot could be at the beginning or at the end of a sentence and not enough words can be found before or after the pivot. For these cases the best alternative is to pad using a *White noise*. The concept of white noise in WE could be related to the concept of neutral vector [39,40]. Unfortunately the concept of neutral vector does not exist in WE. However, we solve this issue by randomly extracting words from a French corpus.

Supervised Learning algorithms As shown in Figure 1, during the learning phase the matrix X is fed into the input layer of a supervised machine learning algorithm.

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

As the experiments were designed, the algorithm must make a dichotomous choice in order to decide whether the pivot word of the input matrix represent a spatial phrase or a non-spatial phrase. We have used two types of machine learning models: classical machine learning (ML) and deep neural networks (DNN). Five different algorithms were selected (two ML and three DNN) based on some of their characteristics that we considered potentially relevant to our problem and described below. Each ML algorithm is fine tuned on a training dataset, then the best model is chosen following the empirical results on a testing dataset (see Section 4.1).

For classical ML algorithms we choose to evaluate the performances of Support Vector Machine (SVM) and Random Forest (RF) algorithms for the task of SNoE recognition. These two algorithms have been commonly used for NLP and information retrieval tasks such as text classification [41,42,43].

Support vector machine (SVM), is a vector space based machine learning method proposed by Cortes and Vapnik [44] where the goal is to find a decision boundary between two classes that is maximally far from any point in the training data (possibly discounting some points as outliers or noise). The SVM algorithms have been used in text classification task [43,42,41].

SVM models are known to scale well with high dimensional data with a good capacity of generalization and with a limited risk of over-fitting. Additionally, SVM is efficient when the number of input dimensions is greater than the number of samples. Thus, SVM appears as a good choice for our study and experiments using WEs. Random Forest (RF), is a supervised machine learning algorithm introduced by Breiman et al. [45] that has been widely used for classification and regression tasks. The principle behind RF is to create a forest with n number of decision trees. Then by a sampling process based on the bootstrapping principle [46] the algorithm created n subsets of the learning dataset and each tree is trained on one of these subsets. In order to classify an example, a 'tree voting' operation is conducted (i.e., where a tree predicts a class). Each vote is recorded and the forest chooses the class with the highest number of votes. In general, the greater the number of trees in the forest is, the stronger the prediction and the higher the accuracy are. The purpose of choosing the RF algorithms is motivated by the fact that the results of a trained RF models could be more interpretable than other complex models such as neural networks.

Deep learning models have recently led to significant and rapid progress in several NLP tasks such as: NER, Relationship extraction and question answering. We have experimented three DNN models: Multilayer Perceptron with Auto-Encoder, a Multilayer Perceptron with Principal Component Analysis, and a Gated Recurrent Unit.

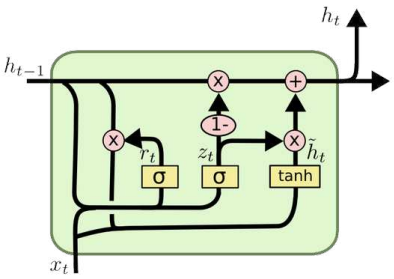
Multilayer perceptron with an auto-encoder (MLP+AE), is a pipeline composed of an auto encoder (encoding layer, decoding layer) and a deep multilayer perceptron (MLP). The main idea behind using autoencoder (AE) is dimensionality reduction. We have made the hypothesis that as *FastText* is a model of pretrained vectors (300 dimensions) on *Wikipedia* and *Common Crawl*, it provides a generic representation of words. As a result, similar words (such as the plurals) have independent embeddings. In that way a vector representation of a word contains a lot of redundant information. What if we could take out the redundancy and express the same information in a fraction of the numbers (compression)? An AE can be used for that purpose. The AE receives the

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

input matrix X representing a sentence and learns to encode it into a less dimension representation X'. The AE starts out by compressing the data into a lower-dimensional representation z (encoding step), and then converts it back to a reconstruction of the original input (decoding step). With the convergence of the AE, the representation z is a compressed version of the data but still encodes the same quantity of the information. The encoded representation matrix X' is fed into the deep MLP which performs the prediction task.

Multilayer Perceptron with a Principal Component Analysis (MLP+PCA) is a pipeline with a Principal Component Analysis (PCA) and deep MLP. As AE, the PCA is a method for data compression. The basic idea of PCA is to reduce the dimensionality of inputs by transforming elements of the input vector e to a new set of variables known as the principal components *PCs*. The *PCs* are a linear combination of the original variables, the *PCs* are orthogonal i.e., the correlation between any pair of variables is 0. The obtained vector is an eigenvector and represents the feature vector which is fed into the deep MLP. Both MLP+PCA and MLP+AE models uses a dimensionality reduction of the inputs, the hypothesis behind dimensionality reduction is to learn on a discriminating information. Indeed, the vector representation of the inputs using WEs provides a set of all the possible semantic spectrum for a given word. However, our context could be seen as a language of specialty with a specific terminology and therefore we assume that we just need a subset of possible semantics, so a subset of components of the vector representation.

Gated Recurrent Unit (GRU), was introduced by Cho et al. [47] and is an improvement of the standard recurrent neuron network (RNN) to solve the vanishing gradient problems that comes with RNN by bringing up the concepts of update gate and reset gate. As shown in Figure 2 the update gate z_t helps the model to determine how much of the past information (from previous time steps) needs to be passed along the future, while the reset gate r_t is used to decide how much of the past information to forget. In others terms, z_t and r_t are two vectors that decide which information should be passed to the output, therefore the GRU can be trained to keep information for long term and remove information that is irrelevant to the prediction.



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Fig. 2. The gated recurrent unit (source: [47]).

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

GRU is design as a solution for short-term memory such as LSTM (Long Short Term Memory) [48]. While LSTM has three gates (input, output, and forget gates), GRU uses only two (reset and update gates). The GRU network is less complex than LSTM and is trained faster. This makes the GRU less complex than LSTM and so GRU models are trained faster than LSTM. In addition, the GRU unit controls the flow of information like the LSTM unit, but without having to use a memory unit. It just exposes the full hidden content without any control. According to Yin et al. [49] GRU has shown better performance on certain smaller data-sets for text classification tasks in NLP. For these reasons we have decided to evaluate the performances of the GRU model in our case study. An advantage of GRU being in the fact that it takes into account the sequentially of the input, which has a great impact as we work on text classification where the words order in a sentence is an important information.

4 Experiments and evaluations

This section describes the experimental study to demonstrate the feasibility of our approach, i.e., examine the ability of each algorithm to detect SNoEs. Thus, we have conducted a series of experiments based on two manually annotated gold standard datasets (cf. 4.1). The validation dataset is dedicated to evaluate the categorisation performances of each system on the SNoEs task. Pivots used to compose the samples of the validation dataset, are the same ones that were used for learning (training dataset) but in different contexts.

The second dataset (emergence dataset) is used to study the emergence performance of the different systems. Emergence assessment allows us to measure the ability to properly classify samples holding new pivots that do not have a sample in the training data set. As our study is a classification task we used the evaluation metrics: Precision (P), Recall (R), Accuracy (ACC) and F1_Score (F1) which is a combination of both recall and precision.

4.1 Datasets

As previously mentioned, there is no standard French dataset available to train and evaluate nominal entities recognition algorithms. Therefore, we have decided to build our own dataset. This implies two steps: 1) building the lexicon, 2) extracting and annotating a set of sentences containing at least a word from the lexicon.

As explained in section 3.2, the lexicon is built by manually extracting all the words used as SNoEs from a set of 14 French hiking description texts. As a result, 141 words have been extracted and constitute the elements of the lexicon called *Aléa*. Starting from this lexicon, we have extracted a corpus of sentences containing at least one lexical entry from two different sources:

1. *Wikipedia* articles, using the OpeanSearch API [18].

[18] https://fr.wikipedia.org/wiki/OpenSearch

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

2. Textual hiking descriptions from two hiking and outdoor sharing community websites (*Visorando*[19] and *Camptocamp*[20]) using a custom web crawler.

A corpus of 78,785 sentences were extracted from different web sources, 25,821 sentences were extracted from both *Visorando* and *Camptocamp*, where 52,964 sentences were extracted from *Wikipedia*. A total of 956 sentences were randomly selected from the corpus then manually labelled and distributed as shown in Table 1. As illustrated in the introduction and explained in Section 3.1, an example is annotated as positive only if the pivot designates a SNoE, otherwise it is labelled as negative.

Table 1. Distribution of sentences manually labelled in our dataset

	C1			C2
Datasets	Train	Test	Validation	Emergence
Positives examples	289	112	112	51
Negatives examples	279	82	82	42
Total	568	194	194	93

We dedicated 568 samples for the training dataset that is used to adjust the model parameters (weights and biases in the case of Neural Network), 194 samples for the test dataset to fine-tune the hyper-parameters of the trained models. Finally, 194 samples for the validation dataset that is used once a model is fully trained (using the train and test datasets) to evaluate competing models. Each sample is a sentence annotated according to the pivot meaning, the sample is annotated as SNoE if the pivot is used in its spatial meaning in the context of the sample, and is classified as non-SNoE otherwise. This dataset is called 'C1'.

In order to study the ability of different algorithms to detect new SNoE, we have extracted a set of sentences using a new lexicon of 15 new pivots extracted from the Geonto ontology [50]. These pivots do not correspond to any lexical entry of *Aléa*, therefore, no sentences containing any of these new pivots are present in the training, testing and validation datasets. A set of 93 sentences containing new pivots was then manually labelled according SNoE or non-SNoE.

This dataset allows us to evaluate the ability of the systems to detect new pivots that have not been seen before (during training). We have named this task as the emergence of new pivots and we identify this dataset as 'C2'.

All datasets (Datasets 'C1' and 'C2') supporting this publication are available in a github repository.[21]

4.2 Resources

According to the TL principle, all the experiments below use a pre-trained WEs. The WEs set has a dimension of $d_e = 300$ and was produced by a *Fastext* [51] trained on

[19] https://www.visorando.com
[20] https://www.camptocamp.org
[21] https://github.com/ANRChoucas/Spatial-Nominal-Entity-Recognition/tree/master/data

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

Common Crawl and *Wikipedia* using the CBOW method. This WEs set is publicly available[22]. Furthermore, we have used the implementation of the deep learning algorithms (GRU, MLP+AE, MLP+ACP) provided within the python library Keras[23]. For the classical machine learning algorithms (SVM, RF) we have used the implementations provided by the python library Scikit-learn[24]. All the trained models supporting this publication are available in a github repository.[25]

4.3 Evaluation results

We have conducted a series of experiments on two datasets in order to evaluate the absolute categorisation performances of each system (using the dataset 'C1') and study the emergence of new pivots (using the dataset 'C2'). We have evaluated the performance of each machine learning model (GRU, MLP+ACP, MLP+AE, SVM, RF) with three different context sizes (1 gram, 5 grams, 7 grams), which results in 15 systems.

Categorisation performances We have conducted an experiment based on the validation dataset (from 'C1') in order to compare the performances of each algorithm, table 2 shows the performances of the 15 models. A general observation is that the results of almost all the tested models have better results when the value of n increases. This is consistent with the hypothesis that the context holds important information about the spatial semantics of a SNoE. An exception was found for models based on the MLP+ACP architecture (systems: 13, 14, 15) as there was a slight decrease in performance from 1 gram to 7 grams.

More precisely, we notice that both MLP+AE 7 grams and GRU 7 grams slightly outperforms the other algorithms. The MLP+AE-7grams had obtained an accuracy of 79,38% and a F1-score of 83,19%, while GRU-7grams obtained a closed result of 78,35% and 80,9% for accuracy and F1-score respectively. As the differences are rather small, this observation requires to be confirmed by a larger scale experiment.

It can already be said that the chosen approach is viable. In particular, this makes it possible to consider the use of neural network algorithms despite the fact that only a small corpus is available.

Emergence performance In order to study the emergence performance of the algorithms on the SNoE recognition task, we evaluate them using the 'C2' dataset. Table 3 shows the evaluation results of the emergence capacity of each system. As a reminder, the emergence capacity makes it possible to measure the ability of a system to recognise expressions (ngrams) whose pivot is used with a spatial meaning and is not part of the *Aléa* lexicon that helped to build both learning and validation datasets.

The same observation on the validation results can be made on the emergence results. The increase of the context size improves the global classification performance

[22] https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md
[23] https://keras.io
[24] https://scikit-learn.org/stable/
[25] https://github.com/ANRChoucas/Spatial-Nominal-Entity-Recognition

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

Table 2. Evaluation results on the validation dataset

System	*Algorithm*	*Ngrams*	*TP*	*TN*	*FP*	*FN*	*ACC*	*P*	*R*	*F*1
1	SVM	1g	81	56	26	31	70,62%	75,70%	72,32%	73,97%
2		5g	85	63	19	27	76,28%	81,37%	75,80%	78,70%
3		7g	87	63	19	25	77,31%	82,07%	77,67%	79,81%
4	RF	1g	79	56	26	33	69,58%	75,24%	70,53%	72,81%
5		5g	77	65	17	35	73,19%	81,91%	68,75%	74,75%
6		7g	82	64	18	30	75,25%	82,00%	73,21%	77,35%
7	GRU	1g	78	53	29	34	67,52%	72,90%	69,64%	71,23%
8		5g	84	63	19	28	75,77%	81,53%	75,00%	78,13%
9		7g	89	63	19	23	78,35%	82,41%	79,46%	80,90%
10	MLP+AE	1g	83	52	90	29	69,58%	73,45%	74,10%	73,77%
11		5g	84	60	22	28	74,22%	79,24%	75,00%	77,00%
12		7g	99	55	27	13	79,38%	78,57%	88,39%	83,19%
13	MLP+PCA	1g	77	52	30	35	66,49%	71,96%	68,75%	70,31%
14		5g	78	55	27	34	68,56%	74,28%	69,64%	71,88%
15		7g	68	51	31	44	61,34%	68,68%	60,71%	64,45%

TP: True Positives, *TN*: True Negatives, *FP*: False Positives, *FN*: False Negatives, *ACC*: Accuracy, *P*: Precision, *R*: Recall, *F*1: F1_Score.

Table 3. Evaluation results on the emergence dataset

System	*Model*	*ngrams*	*TP*	*TN*	*FP*	*FN*	*ACC*	*P*	*R*	*F*1
1	SVM	1g	37	15	27	14	55,90 %	57,81 %	72,55 %	64,35 %
2		5g	36	27	15	15	67,70 %	70,58 %	70,58 %	70,58 %
3		7g	40	28	14	11	73,12 %	74,07 %	78,43 %	76,19 %
4	RF	1g	48	3	39	3	54,83 %	55,17 %	94,12 %	69,56 %
5		5g	43	20	22	8	67,70 %	66,15 %	84,30 %	74,14 %
6		7g	46	22	20	5	73,11 %	69,69 %	90,19 %	78,63 %
7	GRU	1g	44	6	36	7	53,76 %	55,00 %	86,27 %	67,17 %
8		5g	44	26	19	7	72,04 %	69,84 %	86,27 %	77,19 %
9		7g	47	26	16	4	82,45 %	74,60 %	92,15 %	82,46 %
10	MLP+AE	1g	44	7	35	7	54,80 %	55,69 %	86,27 %	67,69 %
11		5g	44	27	15	7	76,00 %	74,57 %	86,27 %	80,00 %
12		7g	48	23	19	3	76,00 %	71,60 %	91,10 %	81,30 %
13	MLP+PCA	1g	38	12	30	13	54,37 %	55,88 %	74,50 %	63,80 %
14		5g	36	23	19	15	63,44 %	65,45 %	70,58 %	67,92 %
15		7g	38	23	19	13	65,59 %	66,66 %	74,51 %	70,37 %

TP: True Positives, *TN*: True Negatives, *FP*: False Positives, *FN*: False Negatives, *ACC*: Accuracy, *P*: Precision, *R*: Recall, *F*1: F1_Score.

of each algorithm. The GRU 7 grams system obtained the best results with 82,45%, 82,46% for both accuracy and F1-score respectively, which outperforms all the others algorithms regarding the accuracy score. Nevertheless, none of the tested algorithms differs with regard to the F1 score.

It can be assumed that one way to improve the performance of most of the studied algorithms is to increase the size of the context. Another possibility that appears very

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

promising would be to use contextualised WE models such as those produced by the BERT model proposed by Devlin et al.[52].

5 Conclusion

This paper presents a methodology comparing five supervised machine learning algorithms for the automatic identification of SNoE from raw texts. The approach uses a pre-trained WEs model as input according to the TL principle. The WEs used as input data for these algorithms, come from the FastText model pre-trained on a huge corpus of generic texts in French. The FastText model was chosen because it produced better results, compared to other equivalent WEs models, on so-called morphological rich languages such as French. The experimental results demonstrate: 1) the feasibility of our approach for the SNoE recognition task, 2) the importance of the context on this kind of task. Thanks to the use of the principle of transfer learning we have been able to show that it is possible to test methodological and algorithmic choices by relying on small corpora. Nevertheless, in order to obtain better performances, the size of our corpus seems insufficient. As a result, an extension of our dataset is already being developed. Given new models of WEs that seem to exceed the performance of the models we have used, we also plan to reproduce the same type of study using this time TL principle from a BERT model pre-trained on a French corpus like the one proposed by Le et al. [53].

According to the obtained results, none of the presented algorithms significantly outperforms, however, regarding the properties of each models presented in section 3.2 the GRU system seems to have a greater potential when working with the whole sentence. For this reason we are interested to invest more in this track. As future work, we aim to study the ability of the GRU to improve the performances on the SNoE recognition task, in particular by providing the whole sentence as input of the system (not only the n-grams) and thus fully use the ability of the GRU model to take into account the sequence aspect of the data in the input. Considering the entire of our WSD problematic it is necessary to be able to distinguish between sentences where the pivot is used to describe a static spatial situation and those where it is used to describe a motion (an itinerary). Then we will also work to categorise the context of SNoE in order to detect spatial relationships and different categories of verbs (e.g., displacement, description, perception) involved in the context.

References

1. Vanetik, N., Litvak, M.: Multilingual Text Analysis: Challenges, Models, And Approaches. World Scientific (2019) 2
2. Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In: Proceedings of the 5th annual international conference on Systems documentation. pp. 24–26. Citeseer (1986) 2
3. Vicente, M.R.: La glose comme outil de désambiguïsation référentielle des noms propres purs. Corela. Cognition, représentation, langage (HS-2) (2005) 2

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

4. Olteanu-Raimond, A.M., Davoine, P.A., Gaio, M., Gouardéres, E., Van Damme, M.D., Villanova-Oliver, M., Brasebin, M., Domingues, C., Duchêne, C., Favre, O., Mustière, S., Devin, F., Le Nir, Y., Moncla, L., Bouveret, S., Genoud, P., Gensel, J., Ziebelin, D.: Projet CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. In: Spatial Analysis and GEOmatics 2017 (Sageo 20017). INSA de rouen, Rouen, France (Nov 2017), ¡hal-01649156¿ 2
5. Gaio, M., Moncla, L.: Extended named entity recognition using finite-state transducers: An application to place names. In: The Ninth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2017) (2017) 2
6. Wang, J., Liu, Z.: A novel arithmetic of named entity identification. In: 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery. vol. 4, pp. 457–461. IEEE (2008) 3
7. Florian, R., Ittycheriah, A., Jing, H., Zhang, T.: Named entity recognition through classifier combination. In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. pp. 168–171. Association for Computational Linguistics (2003) 3
8. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360 (2016) 3, 4
9. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. Journal of Machine Learning Research 12, 2493–2537 (Aug 2011) 3
10. Zhou, J., Xu, W.: End-to-end learning of semantic role labeling using recurrent neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). vol. 1, pp. 1127–1137 (2015) 3
11. Friburger, N., Maurel, D.: Finite-state transducer cascades to extract named entities in texts. Theoretical Computer Science 313(1), 93–104 (2004) 3
12. Chan, S.K., Lam, W., Yu, X.: A cascaded approach to biomedical named entity recognition using a unified model. In: Seventh IEEE International Conference on Data Mining (ICDM 2007). pp. 93–102. IEEE (2007) 3
13. Maurel, D., Friburger, N., Antoine, J.Y., Eshkol, I., Nouvel, D.: Cascades de transducteurs autour de la reconnaissance des entités nommées. Traitement automatique des langues 52(1), 69–96 (2011) 3
14. Moncla, L., Renteria-Agualimpia, W., Nogueras-Iso, J., Gaio, M.: Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. In: Proceedings of the 22nd acm sigspatial international conference on advances in geographic information systems. pp. 183–192. ACM (2014) 3
15. Béchet, F., Sagot, B., Stern, R.: Coopération de méthodes statistiques et symboliques pour l'adaptation non-supervisée d'un système d'étiquetage en entités nommées. In: TALN'2011-Traitement Automatique des Langues Naturelles (2011) 3
16. Srihari, R.: A hybrid approach for named entity and sub-type tagging. In: Sixth Applied Natural Language Processing Conference (2000) 3
17. Nouvel, D., Antoine, J.Y., Friburger, N., Soulet, A.: Coupling knowledge-based and data-driven systems for named entity recognition. In: Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data. pp. 69–77. Association for Computational Linguistics (2012) 3
18. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd annual meeting on association for computational linguistics. pp. 363–370. Association for Computational Linguistics (2005) 3

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

19. Karimzadeh, M., Pezanowski, S., MacEachren, A.M., Wallgrün, J.O.: Geotxt: A scalable geoparsing system for unstructured text geolocation. Transactions in GIS 23(1), 118–136 (2019) 3
20. Grover, C., Tobin, R., Byrne, K., Woollard, M., Reid, J., Dunn, S., Ball, J.: Use of the edinburgh geoparser for georeferencing digitized historical collections. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368(1925), 3875–3889 (2010) 4
21. Moncla, L., Gaio, M., Nogueras-Iso, J., Mustière, S.: Reconstruction of itineraries from annotated text with an informed spanning tree algorithm. International Journal of Geographical Information Science 30(6), 1137–1160 (2016) 4
22. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. Journal of machine learning research 3(Feb), 1137–1155 (2003) 4
23. Mnih, A., Hinton, G.: Three new graphical models for statistical language modelling. In: Proceedings of the 24th international conference on Machine learning. pp. 641–648. ACM (2007) 4
24. Collobert, R., Weston, J.: A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th international conference on Machine learning. pp. 160–167. ACM (2008) 4
25. Mnih, A., Hinton, G.E.: A scalable hierarchical distributed language model. In: Advances in neural information processing systems. pp. 1081–1088 (2009) 4
26. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: Proceedings of the 48th annual meeting of the association for computational linguistics. pp. 384–394. Association for Computational Linguistics (2010) 4
27. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013) 4
28. Bengio, Y.: New distributed probabilistic language models. Dept. IRO, Université de Montréal, Montréal, QC, Canada, Tech. Rep 1215 (2002) 4
29. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5, 135–146 (2017) 4
30. Liu, X., Shen, Y., Duh, K., Gao, J.: Stochastic answer networks for machine reading comprehension. arXiv preprint arXiv:1712.03556 (2017) 4
31. Clark, C., Gardner, M.: Simple and effective multi-paragraph reading comprehension. arXiv preprint arXiv:1710.10723 (2017) 4
32. Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., Chatzisavvas, K.C.: Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications 69, 214–224 (2017) 4
33. Severyn, A., Moschitti, A.: Unitn: Training deep convolutional neural network for twitter sentiment classification. In: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). pp. 464–469 (2015) 4
34. Rezaeinia, S.M., Ghodsi, A., Rahmani, R.: Improving the accuracy of pre-trained word embeddings for sentiment analysis. arXiv preprint arXiv:1711.08609 (2017) 4
35. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 207–212 (2016) 4
36. Gábor, K., Buscaldi, D., Schumann, A.K., QasemiZadeh, B., Zargayouna, H., Charnois, T.: Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 679–688 (2018) 4

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

37. Sanh, V., Wolf, T., Ruder, S.: A hierarchical multi-task approach for learning embeddings from semantic tasks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 6949–6956 (2019) 4
38. Caruana, R., Silver, D., Baxter, J., Mitchell, T., Pratt, L., Thrun, S.: Learning to learn: knowledge consolidation and transfer in inductive systems. In: Workshop held at NIPS-95, Vail, CO, see http://www.cs.cmu.edu/afs/cs/project/cnbc/nips/NIPS95/Workshops.html (1995) 4
39. Connor, R.J., Mosimann, J.E.: Concepts of independence for proportions with a generalization of the dirichlet distribution. Journal of the American Statistical Association 64(325), 194–206 (1969) 7
40. Fessler, J.A.: On transformations of random vectors. Commun. Signal Process. Lab., Dept. Elect. Eng. Comput. Sci., Univ. Michigan, Ann Arbor, MI, http://www. eecs. umich. edu/˜ fessler (1998) 7
41. Wang, Z.Q., Sun, X., Zhang, D.X., Li, X.: An optimal svm-based text classification algorithm. In: 2006 International Conference on Machine Learning and Cybernetics. pp. 1378–1381. IEEE (2006) 8
42. Sun, A., Lim, E.P., Liu, Y.: On strategies for imbalanced text classification using svm: A comparative study. Decision Support Systems 48(1), 191–201 (2009) 8
43. Goudjil, M., Koudil, M., Bedda, M., Ghoggali, N.: A novel active learning method using svm for text classification. International Journal of Automation and Computing 15(3), 290–298 (2018) 8
44. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning 20(3), 273–297 (1995) 8
45. Breiman, L., et al.: Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science 16(3), 199–231 (2001) 8
46. Efron, B.: Bootstrap methods: another look at the jackknife. In: Breakthroughs in statistics, pp. 569–593. Springer (1992) 8
47. Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014) 9
48. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), 1735–1780 (1997) 10
49. Yin, W., Kann, K., Yu, M., Schütze, H.: Comparative study of cnn and rnn for natural language processing. arXiv preprint arXiv:1702.01923 (2017) 10
50. Mustière, S., Abadie, N., Aussenac-Gilles, N., Bessagnet, M.N., Kamel, M., Kergosien, E., Reynaud, C., Safar, B.: Géonto: Enrichissement d'une taxonomie de concepts topographiques (2009) 11
51. Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T.: Learning word vectors for 157 languages. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018) (2018) 11
52. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). pp. 4171–4186 (2019) 14
53. Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., Schwab, D.: Flaubert: Unsupervised language model pre-training for french (2019) 14

Supplemental Material : Hyper-parameters Settings

Here we show the hyper-parameters for each model and the range tried for the hyper-parameters in parentheses. For GRU models, hyper-parameters include the number of

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.

GRU units (5,10,100,1000,100), GRU units activation function (hyperbolic tangent), recurrent activation function (hyperbolic tangent), dropout (0.0, 0.3, 0.5, 0.8, 0.9), recurrent dropout (0.0, 0.3, 0.5, 0.8, 0.9), dense activation function (hyperbolic tangent, sigmoid), the number of epochs for training (500, 1000, 2000), the optimiser (adam) with learning rate (0.001). For the RF the hyper-parameters include the number of trees in the forest (50, 60, 70, 80, 100, 200, 300, 500), the maximum depth of the tree (1, 3, 6, 12, 15, 20, 22, 25, 27, 29, 32, 34, 36, 38, 40, 43, 46, 48, 50, 60, 65, 70, 75, 80), the function to measure the quality of a split (Gini impurity, Entropy). For the SVM the hyper-parameters include the kernel type (Polynomial, Linear, Sigmoid, Radial Basis Function), regularisation parameter (1e-3, 1e-2, 1e-1, 0.5, 1, 10, 100), the kernel coefficient gamma (1e-3, 1e-2, 1e-1, 1, 10, 100, 1000, scale). For MLP+PCA hyper-parameters include activation function for each layer (Exponential Linear Unit, Rectified Linear Unit, Softplus), the output layer activation function (sigmoid), dropout (0.0, 0.3, 0.5), PCA information, the optimiser (adam) with learning rate (0.001). For MLP+AE the hyper-parameters are the same as in the MLP+AE except for the dropout (0.0, 0.5, 0.9), and the dimension of the encoding layer (500).

AGILE: GIScience Series, 1, 2020.
Full paper Proceedings of the 23rd AGILE Conference on Geographic Information Science, 2020.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras
This contribution underwent peer review based on a full paper submission.
https://doi.org/10.5194/agile-giss-1-15-2020 | © Authors 2020. CC BY 4.0 License.