Learning Bayesian Networks from the Knowledge of a Never-Ending Learning System

Rogers R. Avelar de Carvalho and Edimilson B. dos Santos

***Abstract*—Never-Ending Learning (NEL) is a new paradigm of machine learning in which computer systems can learn continuously and incrementally. The first never-ending learning system reported in literature is named NELL (Never-Ending Language Learning). NELL has built a Knowledge Base (KB) containing many types of knowledge. In this paper, we propose to learn Bayesian networks from the knowledge stored in the NELL's KB aiming to assist in the development of expert systems in the future. In addition, Bayesian networks have shown to be promising to solve the problem of representing semantic relations and extending the NELL's initial ontology. In the initial experiments, we have built a dataset on the domain of diseases from relations existing in NELL and applied two learning algorithms of Bayesian networks named K2 and DMBC. The empirical results have shown that this proposal is promising.**

***Index Terms*—Bayesian networks, never-ending learning, machine learning.**

I. INTRODUCTION

Never-ending learning (NEL) is an alternative learning paradigm in which computer systems learn many types of knowledge such as humans [1]. The central idea underlying the proposal of this new paradigm of machine learning is that computer systems can learn continuously and incrementally. The maturation and dissemination of this paradigm can have a major impact on the learning ability of computer systems and modify the way they are currently constructed.

NELL (Never-Ending Language Learner) has been the first never-ending learning system reported in literature [2]. This system has been running 24hrs/day since January 2010, extracting information from web text to populate and extend its own Knowledge Base (KB). This base is represented by an ontology-based structure characterized by categories, (e.g., person, sportsTeam, fruit, emotion), relations (e.g., PlaysOnTeam(athlete,sportsTeam)) and their instances (e.g. “Obama” for “person” category; “Neymar plays soccer” for “AthletePlaysSport(athlete, sport)” relation). To learn these items, the NELL system uses different components (or modules), such as CPL [1], CSEAL [3], Prophet [4], OntExt [5] and Conversing Learning [6], with the goal of having the ability of “self-supervision” and “self-reflection” of your KB.

Although a NEL system is not something new in the machine learning community [7] [8], developing methodologies to help both, extending and populating such KB, and improving their coverage, is still a challenge. Some works [5] [9] [10] [11] [12] [13] [14] [15] have presented proposals to help in such tasks. These proposals have obtained promising results which shown the feasibility of building a NEL system based on the methodology proposed in [2]. However, there is still no well-defined methodology for the development of this type of system.

In this sense, we propose to explore the knowledge stored by NELL and apply the formalism of Bayesian networks [16] to make inferences and to identify new semantic relationships that can be inserted in its KB. Thus, Bayesian networks can assist to populate the NELL's KB and, since NELL has built a KB containing many types of knowledge, they can also be used to assist in the development of expert systems in the future. For this, we present in this paper:

- a methodology to built datasets from semantic relations stored in the NELL's KB. Such as a case study, we built a dataset on the domain of diseases to show that is possible to draw information about some specific domain from NELL's KB;
- the application of two learning algorithms of Bayesian networks named K2 [17] and DMBC [18]. These algorithms have important characteristics which allow to induce consistent and representative Bayesian networks. For example, K2 efficiently induces unrestricted Bayesian networks from a suitable attribute ordering. Whereas DMBC can provide good classification accuracies and better probability estimates than other algorithms, like Naive Bayes and TAN, while being more computationally efficient than K2;
- an analysis on the graphs of induced Bayesian networks to verify if the relationships between nodes (which represent the variables/attributes) correspond to the semantic relations extracted from the NELL's KB used to create the dataset. Other relationships can be suggested as new semantic relations for NELL and new facts can be discovered through inference algorithms applied to Bayesian networks.

The remainder of this paper is organized as follows. Section II brings an overview about the NELL system. In section III, Bayesian networks and the K2 and DMBC algorithms are presented. Section IV explains the proposed methodology and describes how K2 and DMBC have been applied to NELL's KB to induce Bayesian networks which represent the domain of diseases. Besides, the experiments and analysis of results are presented. Finally, Section V brings the concluding remarks and points out some future work.

Edimilson B. dos Santos is with Department of Computer Science, Federal University of Sao Joao del-Rei, Brazil, e-mail:edimilson.santos@ufsj.edu.br.
Rogers R. Avelar de Carvalho are with Department of Computer Science, Federal University of Sao Joao del-Rei, Brazil, e-mail:rogers@ufsj.edu.br

II. NEVER ENDING LANGUAGE LEARNER

Never-Ending Language Learner (NELL) is a computer system able to learn continuously and incrementally about diverse knowledge domains by reading the Web. NELL was reported by Carlson et al. [2] and it has been running non-stop since January 2010, each day extracting more beliefs from the Web and removing old incorrect beliefs. Thus, it retrains itself to improve its competence and builds its Knowledge Base (KB) which has approximately 120 million interconnected beliefs, along with millions of learned phrasings, morphological features, and Web page structures [1].

The NELL system receives as input an initial ontology defining hundreds of categories (e.g., University, City) and binary relations that hold between members of these categories (e.g., UniversityInCity(x, y)). Also, some instances for each category and each relation are provided (e.g. "Carnegie Mellon University" for University; "Carnegie Mellon University is in Pittsburgh" for UniversityInCity(x, y)). Then, NELL reads more beliefs from the Web and learns to read better than the previous day, running 24hrs/day, forever, and each day.

To carry out the learning, there is a software architecture for NELL, depicted in Fig. 1, which includes various learning and inference modules. NELL's growing KB serves as a shared blackboard through which its various reading and inference modules interact. The modules are explained as follows:

- Coupled Pattern Learning (CPL): performs learning from textual patterns of web pages.
- Set Expander for Any Language (SEAL): performs web queries and extracts knowledge from HTML standards.
- Coupled Morphological Classifier (CMC): is a morphological classifier that examines the results obtained through CPL and SEAL, aiming to find morphological patterns.
- Learned Embeddings (LE): learns vector embeddings of the noun phrases.
- OpenEval: uses somewhat different context features from the above CPL system, and uses real time Web search to collect the distribution of text contexts found around the noun phrase.
- Path Ranking Algorithm (PRA): infers new beliefs from the analysis of paths taken by the learned fact. This path refers to the links between the relationships that extracted a fact.
- Never Ending Image Learner (NEIL): analyzes visual images associated with the noun phrase, when the noun phrase is given to an image search engine.
- OntExt: extends the NELL's ontology by inventing new relational predicates. OntExt searches for new relations by considering every pair of categories in NELL's current ontology.

III. BAYESIAN NETWORKS

Bayesian Networks are graphical representations of multivariate joint probability distributions and have been widely used to model and represent uncertainty [19]. They are described by directed acyclic graphs (DAG) – or graphical structure – in which the nodes represent the attributes and the arcs represent probabilistic dependencies between connected nodes (attributes). Also, there is a set of conditional probability table (CPT) – or numerical parameters.

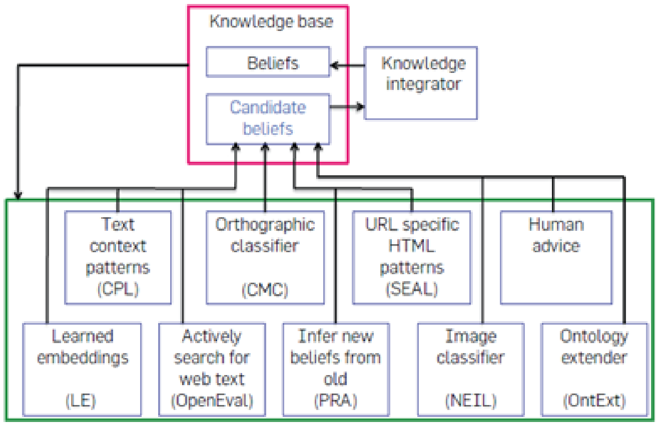

Fig. 1. NELL's software architecture (Reproduced from [1]).

Computational methods for learning Bayesian networks may be seen as a manner to identify a probabilistic model which describes the dependence (and independence) among attributes from a given domain [17]. Once the Bayesian network is defined, knowledge represented on it can be drawn through a process called inference.

Among several Bayesian network learning algorithms available in the literature, this work applies two specific methods: K2 and DMBC. These algorithms have important characteristics for the induction of Bayesian networks and the solution of the problem of identifying semantic relations in NELL's knowledge base, as it is explained in Subsection III-A and Subsection III-B, respectively.

A. K2 Algorithm

The K2 algorithm [17] constructs a Bayesian network from data using a heuristic search. It expects as input a complete dataset D and a attribute ordering (AO). Given D and AO, K2 searches for the Bayesian network structure that best represents D. The algorithm is broadly used due to its good performance in terms of computational complexity (time) as well as its good results when an adequate AO is provided [20].

The attribute preorder assumption is a critical issue since it aims to reduce the number of possible structures to be learned. In this sense, K2 uses an ordered list (containing all the attributes including the class), which asserts that a given attribute A can only be parent of attributes that follow it in the list. The first attribute in the list has no parent.

K2 uses a greedy search in its attempt to find the best structure. Given a dataset D with M objects and a AO, the algorithm begins as if every node has no parent. Then, starting with the second attribute in AO list (the first one has no parents in the network structure), its eligible parents are evaluated and those that maximize the probability of the whole structure are added. When the addition of any single parent does not increase the probability, the algorithm stops adding parents to the current node. The process is repeated for all attributes in order to get the best possible structure. The K2 metric for evaluating the parent assignment to each attribute is defined by (1):

$$g(i,\pi_i)=\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{k=1}^{r_i}N_{ijk}! \tag{1}$$

where each attribute $x_i(i=1,...n)$ has r_i possible values $(v_{i_1},v_{i_2},...,v_{i_{r_i}})$ and has as its set of parents π_i. The number of instantiations of π_i is represented by q_i. N_{ijk} is the number of objects in D in which x_i has value v_{ik} and π_i is instantiated as w_{ij} (which represents the *j-th* instantiation relative to D of π_i. Finally, $N_{ij}=\Sigma N_{ijk}$ for $k=1,...,r_i$.

B. DMBC Algorithm

DMBC (Dynamic Markov Blanket Classifier) [18] was developed from K2 algorithm [17] to learn Bayesian networks designed especially for classification problems. The idea of inducing classifiers using the K2 score is aimed at reducing the bias introduced by the independence assumptions embedded in the Naive Bayes classifier, thus improving probability estimates while maintaining good classification rates and, also, allowing the identification of important relationships among the most relevant attributes. In addition, DMBC tends to reduce the number of attributes in the induced network.

The DMBC algorithm is basically a version of K2 designed to explore the Class' Markov Blanket, which can lead to faster Bayesian network structure induction and more accurate estimates of class probabilities. According to [21], for each attribute X in the Bayesian network, the set of all parents of X, children of X, and parents of the children of X, is a Markov Blanket of X.

The DMBC main idea is excluding from possible structures those having attributes out of the dynamically created Class' Markov Blanket. Considering only attributes present in Class' Markov Blanket, the usually high computational time needed to learn an unrestricted Bayesian network can be reduced. Thus, DMBC tends to generate a probabilistic graphical model focused on the class attribute and discarding attributes that are not inside the Markov boundary defined by the Markov Blanket.

IV. APPLYING K2 AND DMBC TO NELL'S KB

This paper proposes to apply learning algorithms of Bayesian networks to NELL's KB aiming to represent stored knowledge about some domain through induced Bayesian networks. The Bayesian networks can be used to assist in the development of expert systems for decision making and to assist extending and populating the NELL's KB.

To assess this proposal, we extracted information about the domain of diseases from NELL's KB to build a dataset used in the training of two learning algorithms of Bayesian networks: K2 and DMBC. These algorithms have been chosen for the initial experiments due to their individual characteristics. K2 induces unrestricted Bayesian networks and performs well in terms of computational complexity (time) as well as obtains good results when an adequate attribute ordering is supplied [20]. Thus, it can be used for identifying relationships among attributes from given domain. DMBC runs quickly and induces a simpler model that represents relationships among attributes of interest in a classification task.

A. Creation of the Dataset from the NELL's KB

Initially, a dataset representing the domain of diseases has been built from information contained in the NELL's KB for the training of the K2 and DMBC algorithms. The NELL's KB has information about many types of knowledge. We would like to assess whether this knowledge can be used to train K2 and DMBC and to induce Bayesian networks capable of assisting to solve problems. In this case, we intend to assess the knowledge acquired by NELL related to medicine to assist in the diagnostic of diseases.

The NELL's complete KB is available at http://rtw.ml.cmu.edu/rtw/resources. The NELL's KB was initially pre-processed. The pre-processing procedure allowed building the dataset having only the data from NELL's semantic relations related to diseases.

To apply the pre-processing step, we have used the tab-separated-value file with every belief in the KB, one per line. The file has thirteen columns, of which only the first three were used: Entity, Relation and Value (forming a kind of SPO, Subject-Predicate-Object, triple). The column Relation presents relations existing between two categories stored in Entity and Value. Each line of this file contains one category or relation instance that NELL believes to be true. Nominally, each belief is an (Entity, Relation, Value) tripple; instances of relations have the form: "cancer", "DiseaseMayBeCausedByDrug", "alcohol". The steps to build the dataset are explained as follows.

Step 1: Initially, some relations from NELL's KB related to diseases have been chosen to create the dataset used in this work. For this, the Pandas library in Python has been applied to search instances that had the word "disease", both in the column Entity and in the column Value. From these instances, relation names about diseases were searched in the Relation column and 16 relations were found.

Step 2: Each relation with the data of the columns "Entity" and "Value" was stored in a file of the type ".csv".

Step 3: After, all files ".csv" referring to the relations were concatenated in a single file ".csv". The file resulting from the concatenation has 16 columns, each referring to a relation, and the column Entity with the names of the diseases, which was renamed to disease, totaling 17 columns.

Step 4: The name of each column referring to the relations was changed, forming the attributes of the dataset (16 attributes plus the "disease" class), as can be seen in Table I.

Step 5: The disease domain relations found in the NELL knowledge base file did not bring up data on symptoms. Since an attribute about disease symptoms is important for the representation of this domain, it was decided to search for it on the NELL project website, where there is a page with already validated data on categories and relations. Thus, the "symptoms" attribute was included in the dataset.

Step 6: After obtaining the attributes and instances from the relations, some steps of data transformation have still

TABLE I

SEMANTIC RELATIONS PRESENTS IN NELL'S KB USED TO PROVIDE ATTRIBUTES AND THEIR VALUES TO CREATE THE DATASET. THE "RELATIONS" COLUMN BRINGS THE NAME OF THE RELATIONSHIP CHOSEN IN THE NELL KNOWLEDGE BASE. THE "ATTRIBUTES" COLUMN DISPLAYS THE NAME CHOSEN FOR THE ATTRIBUTE CORRESPONDING TO THE RELATIONSHIP.

Relations	**Attributes**
AgriculturalProductContainChemical	chemprod
BacteriaIsTheCausativeAgentOfPhysiologicalCondition	bacteria
BacteriaCanCauseIllnessFromFood	food
BeverageContainsProtein	protein
BodypartWithinBodypart	partbody
DrugHasSideEffect	effdrug
DrugWorkedOnByAgent	wagent
InverseOfFoodDecreasesTheRiskOfDisease	fooddec
InverseOfFoodCanCauseDisease	foodcan
InverseOfBeverageCanSoftenDisease	bevsoft
InverseOfDiseaseCausesPhysiologicalCondition	phycond
InverseOfEmotionAssocietedWithDisease	emotion
InverseOfAnimalDevelopDisease	animal
PhysiologicalConditionPossiblyTreatedByDrug	drug
PhysiologicalConditionStudiedByPerson	scientist
SideEffectCausedByDrug	sidedrug

been necessary. For this, we have used the Weka software [22] which provides many filters to alter the dataset in some way. First, the missing values have been filled using the filter named "ReplaceMissingValues". This filter replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data. Many attributes of the dataset have missing values because the NELL system has not yet been able to learn a lot about them. Next, attributes having many distinct values are changed to have only the most repeatable values. Thus, the filter called "MergeInfrequentNominalValues" has been used. This filter merges all values of the specified nominal attributes that are insufficiently frequent.

The dataset built on the disease domain from NELL data is composed of 15,534 instances (or records) and 18 attributes: 17 predictor attributes (including the new attribute "symptoms") and the class attribute "disease".

B. Initial Experiments and Analysis of Results

For the experiments, K2 and DMBC were performed using the previously described dataset (Subsection IV-A) for the training. The Bayesian network structures induced by these algorithms were analyzed by observing the relationship among attributes through edges presents in graph. Next, inference algorithms were applied to Bayesian network for evaluating the represented knowledge and their capacity of prediction.

1) Results obtained by DMBC: The DMBC algorithm induces a Bayesian classifier and thus induces a simpler structure (with fewer edges and fewer attributes), identifying only the relationships that directly influence the class attribute. Due to this, however, the algorithm has a reduced computational cost. It needs two input parameters beyond the dataset: i) the class attribute, defined here like being "disease"; and ii) the maximum number of parents for each node in the Bayesian network structure, which was defined here as 10 (the goal was to let DMBC find as many parents as possible for each node). Fig. 2 presents the Bayesian network induced by DMBC from created dataset.

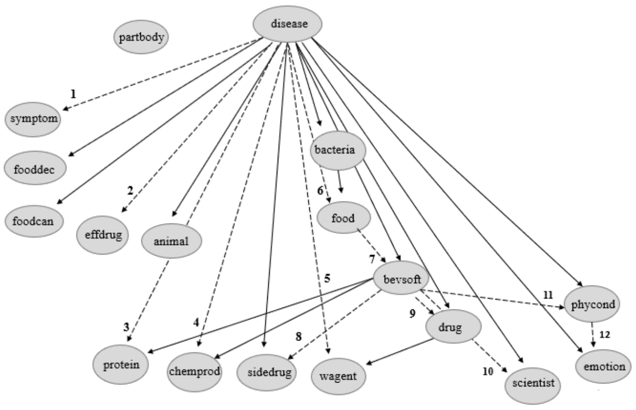

Fig. 2. Bayesian network induced by DMBC using the dataset related to disease created from the NELL's KB. The dashed and numbered edges suggest new semantic relations for NELL, while the solid-line edges represent semantic relationships existing in the base of NELL.

The edges of the Bayesian network structure in Fig. 2 represent relationships among the attributes and can be understood like semantic relations for the NELL's KB. Observing the edges in structure, it is possible to note that DMBC can especially learn the semantic relations that involve the class attribute (in this case, disease).

Some edges in this strucure represent semantic relations existents in NELL's KB, which were used to create the dataset (see Table I). It is possible to note that DMBC was able to represent 14 of 16 used semantic relations. These relationships are represented in the structure by solid-line edges. This demonstrates that DMBC may be able to identify existing relationships in the NELL' KB. For instance, the edge between disease and drug represents the semantic relation "PhysiologicalConditionPossiblyTreatedByDrug", which exists in NELL's KB (as can be seen in Table I). Only two relationships existing in the NELL's KB (and used to build the dataset) were not identified by the DMBC: the "BodypartWithinBodypart" and the "DrugHasSideEffect". Note, in Fig. 2, that no edges were formed between the attributes "disease" and "partbody" and "effdrug" and "drug". There are relations, such as these, with few values learned by NELL and thus they generate attributes having many missing values for the dataset. Then, not all semantic relations can be identified by DMBC if there is no proper treatment of these missing values.

The Bayesian network structure in Fig 2 presents others edges in addition to those that represent the existing semantic relationships. These edges are dashed and numbered in the structure and suggest new relationships found by DMBC, in a total of 12. For instance, the edge between "symptom" and "disease" represents a new relation which does not exist in NELL' KB. Therefore, an analysis about the possible new relations can be done by experts to verify the possibility of expanding the NELL's initial ontology. This result demonstrates that DMBC has potential to find out new relations and extend the initial ontology.

2) Results obtained by K2: Unlike the DMBC algorithm, K2 induces unrestricted Bayesian networks (not only for classification problems). Therefore, it is able to find more relationships among attributes. However, these relationships are influenced by the attribute ordering, as K2 induces the Bayesian network structure considering that only the attributes to the left of a given attribute can be its parents in the structure. Thus, for the experiments here, in addition to the number of parents (set to 10), seventeen different attribute orderings were also defined looking for the optimal Bayesian network (or closer to the optimal), which best fits the dataset and it is able to represent the most correct relationships. Ten orderings were randomly defined and the other seven through attribute selection methods, available in the Weka software (*GainRatioAttributeEval, InfoGainAttributeEval, OneRAtributeEval, ReliefFAttributeEval, SymmetricalUncertAttributeEval, ClassifierAttributeEval* and *CorrelationAttributeEval*). Attribute selection methods evaluate the importance of attributes in relation to the class and thus may generate good orderings for K2.

The K2 algorithm was executed with each of the 17 orderings, inducing 17 different Bayesian networks, which were compared by the value of the function g (1) - it expresses the quality of the structures of the Bayesian networks. Table II displays the values of the function g obtained by K2 for each induced Bayesian network.

Fig. 3 shows the best Bayesian network ($g = -420175, 65$) induced by K2 using the dataset related to disease created from the NELL's KB and the attribute ordering generated from *GainRatioAttributeEval* method. When by observing the structure of the network (see solid-line edges), it can be seen that K2 was able to represent 14 semantic relations existing in the NELL's KB (of the 16 used to create the dataset - see Table I). Only two relations were not identified by K2: "BacteriaCanCauseIllnessFromFood" and "BodypartWithinBodypart". K2 also identified 15 new relationships that are represented in the structure of Fig. 3 by dashed and numbered edges. For comparison, the Bayesian network induced by K2, which has the worst value of g ($-450113, 36$, with Random_3), identified 11 existing semantic relations and found 12 new relationships. This result demonstrates that the K2 algorithm, receiving a suitable attribute ordering, has potential to represent the semantic relations existing in the NELL's KB and suggest new relations which would assist NELL to learn more and better.

C. Results of Bayesian Inference

After the induction of Bayesian networks by the DMBC and K2 algorithms, a Bayesian inference algorithm (based on junction tree) was applied to evaluate the knowledge stored in the networks. Thus, the capacity of Bayesian networks to discover new facts that can be inserted in the NELL's KB was evaluated, in addition to existing data in the base. The inferences were performed using the Nética software [23].

TABLE II

VALUES OF THE FUNCTION G FOR THE BAYESIAN NETWORKS INDUCED BY K2, RECEIVING DIFFERENT ORDERING OF ATTRIBUTES AT EACH EXECUTION. THE VALUES OF G ARE SORTED IN DESCENDING ORDER. SEVENTEEN ATTRIBUTE ORDERING WERE GENERATED TO RUN K2: 7 ORDERINGS OBTAINED FROM ATTRIBUTE SELECTION METHODS AND 10 ORDERINGS OBTAINED RANDOMLY.

Sorting	**G function**
GainRatioAttributeEval	-420175.65
ReliefFAttributeEval	-421574.86
ClassifierAttributeEval	-422893.57
Random_5	-425531.13
SymmetricalUncertAttributeEval	-426327.62
InfoGain	-427503.38
OneRAttributeEval	-427503.38
Random_10	-430480.82
Random_8	-430986.36
Random_9	-431215.24
Random_1	-436026.49
Random_7	-436157.14
Correlation AttributeEval	-436866.89
Random_6	-437769.39
Random_2	-439651.81
Random_4	-440927.64
Random_3	-450113.36

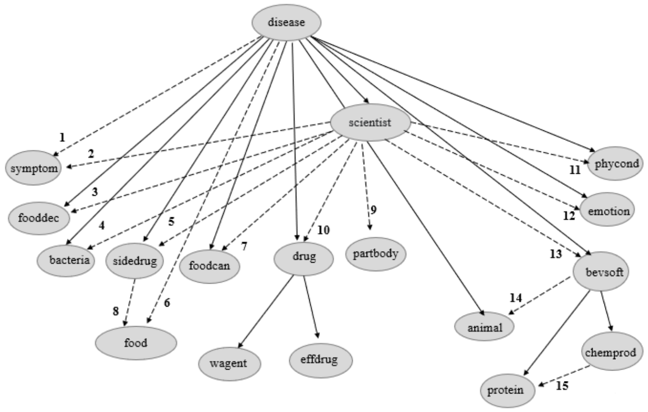

Fig. 3. Bayesian network induced by K2 using the dataset related to disease created from NELL's KB. The dashed and numbered edges suggest new semantic relations for NELL, while the solid-line edges represent semantic relationships existing in the base of NELL.

The inference experiments performed in this work evaluated two aspects: whether the values inferred in Bayesian networks were correct according to the NELL's KB, and whether the inference brought new values that do not exist in the NELL's KB for the relationships between nodes. For this, 20 diseases were analyzed. As input attribute, we used only "disease" and as output attributes: sidedrug, drug, fooddec, foodcan, bevsoft, symptoms.

Table III displays the inference results obtained from Bayesian networks induced by K2 and DMBC. In this table, the inferred values for the output attributes already existed in the NELL's KB. For the 20 instances shown, both Bayesian

TABLE III
INFERRED (AND EXISTING IN NELL) VALUES FROM BAYESIAN NETWORKS INDUCED BY K2 AND DMBC. THE % VALUES REPRESENT THE PROBABILITY THAT THE PREDICTION OF THE ALGORITHMS IS CORRECT.

Input	Output	K2 (%)	DMBC (%)
Depression	sidedrug: Accutane	17.2	15.2
Bronchitis	sidedrug: Singulair	59.5	40.7
Heart_attack	sidedrug: Tylenol	6.67	5.56
Bladder_cancer	sidedrug: Actos	54.7	34.5
Anxiety	sidedrug: Remeron	86.8	79.7
Diabetes	drug: Avandia	14.5	8.39
Depression	drug: Ssris	3.99	3.66
Infection	drug: Levaquin	80.5	62.8
Allergy	drug: Claritin	34.7	29.6
Bipolar_disorder	drug: Seroquel	52.1	40.1
Breast_cancer	fooddec: mushrooms	4.19	4.70
Hypertension	fooddec: vegetables	5.29	6.15
Obesity	fooddec: vegetables	6.76	7.49
Cancer	fooddec: broccoli	5.66	1.13
Diabetes	fooddec: coffee	31.1	28.9
Depression	foodcan: sugar	41.2	45.2
Inflammation	foodcan: sugar	7.35	7.82
Cancer	foodcan: meats	2.60	2.68
Heart_disease	foodcan: junk_food	7.38	8.36
Blood_pressure	foodcan: chocolate	0.98	1.08

TABLE IV
INFERRED (AND NOT EXISTING IN NELL FOR THESE ATTRIBUTES) VALUES FROM BAYESIAN NETWORKS INDUCED BY K2 AND DMBC. THE % VALUES REPRESENT THE PROBABILITY THAT THE PREDICTION OF THE ALGORITHMS IS CORRECT.

Input	Output	K2 (%)	DMBC (%)
Alzheimer	bevsoft: Pomegranate	36.9	32.0
Cardio_problems	bevsoft: Cranberry	88.8	67.4
Heart_attack	bevsoft: Red wine	90.6	74.9
Atherosclerosis	bevsoft: Orange juice	82.6	59.4
Cancer	bevsoft: Pomegranate	49.3	46.1
Depression	sidedrug: Cipro	16.1	14.4
Fatigue	sidedrug: Neurontim	20.1	14.0
Asthma	sidedrug: Singulair	42.4	36.5
Colitis	sidedrug: Aldara	48.1	32.3
Digestive_disorder	sidedrug: Celexa	37.3	19.0
Insomnia	foodcan: drink	5.44	5.69
Anxiety_disease	foodcan: coffee	21.6	22.2
Chronic_pain	foodcan: junk_food	80.7	84.1
Schizophrenia	foodcan: glucose	70.9	75.7
Celulitis	foodcan: coffee	4.90	5.47
Asthma	symptoms: cough	5.39	6.55
Anxiety	symptoms: insomnia	8.56	8.84
Crohn	symptoms: abdo_pain	7.36	1.03
Fibromyalgia	symptoms: fatigue	3.36	3.83
Pneumonia	symptoms: fever	4.77	5.20

networks induced by K2 and DMBC were able to correctly infer the value of the output attributes. This demonstrates that induced Bayesian networks are able to adequately represent this knowledge learned by NELL.

Table IV shows interesting inference results for the output attributes: bevsoft, sidedrug, foodcan and symptoms. From the values of the "disease" input attribute, values that did not exist in the NELL's KB were inferred for these output attributes. Although the inferred values have not yet been validated by an expert, they make sense for the attributes. Based on this, it is believed that Bayesian networks induced by K2 and DMBC are able to use the knowledge already learned for one category to learn another. Thus, it is possible to identify new facts that can be inserted into the NELL's KB.

The K2 and DMBC algorithms were also compared to two other well-known classifiers, Naive Bayes (NB) and Decision Tree (DT), to assess the inference of the class attribute "disease" and thus verify the knowledge of the dataset. All algorithms were trained and tested using the hold-out training strategy, separating the dataset into 70% for training and 30% for testing. The classification results (accuracy and recall) obtained by the algorithms are displayed in Table V.

The DT obtained the best accuracy (94.8%) and recall (94.8%) values. The Bayesian networks induced by K2 and DMBC have also obtained good values, with results very close to the DT. In fact, the results obtained by all algorithms are very close and superior to 90%, which indicates that the dataset constructed from NELL's base has a very good knowledge. Therefore, the inference results of other attributes obtained from the Bayesian networks induced with this dataset can be reliable.

TABLE V
CLASSIFICATION RESULTS OBTAINED BY DECISION TREE (DT), NAIVE BAYES (NB), K2 AND DMBC.

Metric	DT (%)	NB (%)	K2 (%)	DMBC (%)
accuracy	**94.8**	92.4	94.4	94.3
recall	**94.8**	92.5	94.3	93.6

V. CONCLUSIONS AND FUTURE WORKS

This paper presents a study on the application of Bayesian networks to the knowledge base (KB) of a never-ending learning system, known as NELL. The main objective is to show the possibility of using the knowledge acquired by the NELL system to induce Bayesian networks and assist in the development of an expert system in the future. In addition, we have evaluated the ability of Bayesian networks to represent semantic relations between attributes as well as their capacity of inference.

The NELL system's KB allowed to extract information about the domain of diseases that enabled the induction of Bayesian networks using the learning algorithms named K2 and DMBC. The results of the experiments showed that both algorithms were able to induce suitable Bayesian networks to represent the semantic relations between the attributes.

The Bayesian network structures induced by the K2 and DMBC represented 87.5% of the semantic relations extracted from NELL's KB, which suggests that the new relationships identified may be safely used to extend the initial ontology of NELL. This reinforces the hypothesis that Bayesian networks are able to identify relationships between attributes that represent existing semantic relations and suggest new relations for NELL. Furthermore, these Bayesian networks induced by both algorithms provide good inference results, being also able to deduce new facts from the existing knowledge in NELL, improving its learning ability.

Based on these results, we intend to improve the methodology through another very interesting line of investigation: by exploring relational (and first order) approaches [24] to Bayesian networks, such as the approaches proposed in [25] [26].

ACKNOWLEDGMENT

Authors acknowledge the Brazilian research agency CNPq for their financial support.

REFERENCES

[1] T. M. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohammad, N. Nakashole, E. A. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling, "Never-ending learning," *Commun. ACM*, vol. 61, p. 103–115, Apr. 2018.

[2] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. H. Jr., and T. Mitchell, "Toward an architecture for never-ending language learning," in *Proceedings of the Conference on Artificial Intelligence (AAAI)*, pp. 1306–1313, AAAI Press, 2010.

[3] A. F. Mansano *et al.*, *Aprendizado de máquina multivisão aplicado à análise de correferência em um sistema de aprendizado sem fim*. dissertation, Universidade Federal de São Carlos: UFSCar, 2018.

[4] A. P. Appel and E. R. Hruschka Junior, "Prophet – a link-predictor to learn new rules on nell," in *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 917–924, 2011.

[5] P. H. Barchi, *Expansão de ontologia através de leitura de máquina continua*. dissertation, Universidade Federal de São Carlos: UFSCar, 2014.

[6] T. Mohamed, E. Hruschka, and T. Mitchell, "Discovering relations between noun categories," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1447–1455, 2011.

[7] S. Thrun and T. M. Mitchell, "Lifelong robot learning," in *The Biology and Technology of Intelligent Autonomous Agents* (L. Steels, ed.), (Berlin, Heidelberg), pp. 165–196, Springer Berlin Heidelberg, 1995.

[8] S. Thrun and L. Pratt, eds., *Learning to Learn*. USA: Kluwer Academic Publishers, 1998.

[9] R. G. L. Miani, S. D. d. S. Pedro, and E. R. Hruschla, "Association rules to help populating a never-ending growing knowledge base," in *Ibero-American Conference on Artificial Intelligence*, pp. 169–181, Springer, 2014.

[10] R. G. L. Miani and E. H. Junior, "Exploring association rules in a large growing knowledge base," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 7, pp. 106–114, 2015.

[11] S. Verma and E. R. H. Jr., "Coupled bayesian sets algorithm for semi-supervised learning and information extraction," in *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part II* (P. A. Flach, T. D. Bie, and N. Cristianini, eds.), vol. 7524 of *Lecture Notes in Computer Science*, pp. 307–322, Springer, 2012.

[12] M. C. Duarte, *Leitura da web em português em ambiente de aprendizado sem-fim*. dissertation, Universidade Federal de São Carlos: UFSCar, 2016.

[13] D. T. Wijaya and T. Mitchell, "Mapping verbs in different languages to knowledge base relations using web text as interlingua," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 818–827, 2016.

[14] E. B. dos Santos, M. L. Fernandes, E. R. H. Jr., and M. C. Duarte, "Bayesian networks for identifying semantic relations in a never-ending learning system," in *Intelligent Systems Design and Applications - 16th International Conference on Intelligent Systems Design and Applications (ISDA 2016) held in Porto, Portugal, December 16-18, 2016* (A. M. Madureira, A. Abraham, D. Gamboa, and P. Novais, eds.), vol. 557 of *Advances in Intelligent Systems and Computing*, pp. 279–288, Springer, 2016.

[15] E. B. dos Santos, I. H. P. Rodrigues, E. R. H. Jr., and L. R. Bruno, "Bayesian networks for inference and discovery of semantic relations in a never-ending learning system," in *8th Brazilian Conference on Intelligent Systems, BRACIS 2019, Salvador, Brazil, October 15-18, 2019*, pp. 687–692, IEEE, 2019.

[16] J. Pearl, "Probabilistic reasoning in intelligent systems: networks of plausible inference—judea pearl," 1988.

[17] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Mach. Learn.*, vol. 9, pp. 309–347, 1992.

[18] E. B. dos Santos, E. R. H. Jr., E. R. Hruschka, and N. F. F. Ebecken, "Bayesian network classifiers: Beyond classification accuracy," *Intell. Data Anal.*, vol. 15, no. 3, pp. 279–298, 2011.

[19] J. Pearl, *Probabilistic reasoning in intelligent systems - networks of plausible inference*. Morgan Kaufmann series in representation and reasoning, Morgan Kaufmann, 1989.

[20] W. H. Hsu, "Genetic wrappers for feature selection in decision tree induction and variable ordering in bayesian network structure learning," *Information Sciences*, vol. 163, no. 1, pp. 103–122, 2004. Soft Computing Data Mining.

[21] R. E. Neapolitan, *Learning Bayesian Networks*. Prentice Hall, 2003.

[22] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems, Amsterdam: Morgan Kaufmann, 3 ed., 2011.

[23] "https://www.norsys.com/download.html,"

[24] D. Poole, "First-order probabilistic inference," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, (San Francisco, CA, USA), p. 985–991, Morgan Kaufmann Publishers Inc., 2003.

[25] A. Kimmig, L. Mihalkova, and L. Getoor, "Lifted graphical models: a survey," *Machine Learning*, pp. 1–45, 2014.

[26] O. Schulte, Z. Qian, A. E. Kirkpatrick, X. Yin, and Y. Sun, "Fast learning of relational dependency networks," *Mach. Learn.*, vol. 103, p. 377–406, jun 2016.

Rogers R. Avelar de Carvalho Master from the Federal University of São João del-Rei in Computer Science (2021). He has 15 years experience in the technology market. Acting as Systems Analyst, Data Administrator, BI Analyst and Software Engineer. Experience as a teacher in technical, technological and higher education.

Edimilson B. dos Santos Graduated in Computer Science from the Federal University of Lavras/MG (2003), Master in Computer Science from the Federal University of São Carlos/SP (2007) and Ph.D. in High Performance Computational Systems from the Federal University of Rio de Janeiro (PEC/COPPE/UFRJ) (2011). He is currently professor at the Department of Computer Science at the Federal University of São João del Rei/MG (UFSJ). He has experience in Computer Science, with emphasis on Machine Learning, Data Mining and working mainly on the following topics: machine learning, never-ending learning, Bayesian models, evolutionary algorithms and intelligent systems.