

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Rogers Ricardo de Avelar Carvalho

**Redes Bayesianas aplicadas à Base de
Conhecimento de um Sistema de Aprendizado
Sem Fim**

São João del-Rei

2021

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Rogers Ricardo de Avelar Carvalho

**Redes Bayesianas aplicadas à Base de Conhecimento de
um Sistema de Aprendizado Sem Fim**

Dissertação apresentada como requisito para
obtenção do título de mestre no Curso de
Mestrado do Programa de Pós Graduação em
Ciência da Computação da UFSJ.

Orientador: Prof. Dr. Edimilson Batista dos Santos

Universidade Federal de São João del-Rei – UFSJ

Mestrado em Ciência da Computação

São João del-Rei

2021

Rogers Ricardo de Avelar Carvalho

Redes Bayesianas aplicadas à Base de Conhecimento de um Sistema de Aprendizado Sem Fim

Dissertação apresentada como requisito para obtenção do título de mestre no Curso de Mestrado do Programa de Pós Graduação em Ciência da Computação da UFSJ.

Trabalho aprovado. São João del-Rei, 17 de dezembro de 2021 :

Prof. Dr. Edimilson Batista dos Santos
Orientador

Prof. Dr. Diego Roberto C. Dias
Convidado 1

Profa. Dra. Lúcia Helena de Magalhães
Convidado 2

São João del-Rei
2021

"O estudo é chave da oportunidade".

Gordon B. Hinckley

À meus pais e familiares.

Agradecimentos

Inicialmente agradeço ao Prof. Dr. Edimilson Batista dos Santos a oportunidade de tê-lo como meu orientador, por todo apoio, dedicação, paciência e incentivo na condução do projeto desenvolvido. Serei eternamente grato.

Gostaria também de agradecer aos professores com os quais fiz as disciplinas do curso e com quem pude aprender muito: Leonardo Rocha, Flávio Luiz Schiavoni, Vinícius Vieira, Daniel Madeira. E a todo incentivo dado pelos professores do IF SUDESTE MG - campus São João del-Rei: Teresinha Magalhães, Lúcia Magalhães, Priscila Pereira, Gracinha Alves, Ivete Sara de Almeida, Alexandre Silva de Almeida, Fernando Machado da Rocha, Icaro Machado Prado, Aldair Fernandes da Silva.

O apoio da minha família foi fundamental para conseguir superar todos os desafios vividos ao longo do curso. Compartilho essa vitória com cada um deles, reconhecendo a grandeza de meu Pai Celestial que sempre esteve ao meu lado me abençoando com força, proteção e direção.

Faço uma dedicatória especial aos meus pais que sempre me incentivaram e apoiaram, contribuindo de todas as formas possíveis para o meu sucesso. Dedico também esse trabalho de forma especial ao meu querido avô Aristoteles Ferreira da Silva, que sempre me incentivou a buscar conhecimento.

Resumo

O primeiro sistema de aprendizado sem fim descrito na literatura é chamado NELL (*Never-Ending Language Learning*). O objetivo principal do sistema NELL é aprender a ler a web, cada dia melhor, para armazenar o conhecimento adquirido em uma base de conhecimento crescente e sem fim. Para ajudar o sistema NELL nesta tarefa de aprendizagem, neste projeto, é proposta a aplicação de Redes Bayesianas para realizar inferências na base de conhecimento do NELL, bem como para identificar novas relações semânticas que possam ser inseridas na base de conhecimento. Um conjunto de dados foi construído a partir de informações das relações semânticas existentes na base de conhecimento do sistema NELL. Os resultados apresentados mostram que as Redes Bayesianas induzidas pelos algoritmos de aprendizado DMBC e K2 podem representar relações existentes e sugerir novas relações para estender a ontologia inicial do NELL. As inferências realizadas também indicam que as Redes Bayesianas induzidas por ambos algoritmos são capazes de descobrir novas informações para serem inseridas na base de conhecimento do NELL. Os algoritmos K2 e DMBC foram comparados aos classificadores Naive Bayes e Árvore de Decisão com o objetivo de avaliar a inferência da variável classe.

Palavras-chaves: Aprendizado de máquina, Aprendizado sem-fim, Redes Bayesianas.

Abstract

The first never-ending learning system described in the literature is called NELL (Never-Ending Language Learning). The main objective of the NELL system is to learn to read the web, getting better each day, in order to store the acquired knowledge in a growing and never-ending knowledge base. To help the NELL system in this learning task, this project proposes the application of Bayesian Networks to make inferences in the NELL knowledge base, as well as to identify new semantic relations that can be inserted in the knowledge base. A dataset was built from information on the semantic relations existing in the knowledge base of the NELL system. The results found show that Bayesian Networks induced by the DMBC and K2 learning algorithms can represent existing relationships and suggest new relationships to extend the initial ontology of NELL. The inferences made also indicate that the Bayesian Networks induced by both algorithms are capable of discovering new information to be inserted in the NELL knowledge base. The K2 and DMBC algorithms were compared to the Naive Bayes and Decision Tree classifiers in order to evaluate the inference of the class variable.

Key-words: Machine Learning, Never-Ending Learning, Bayesian Networks

Lista de ilustrações

Figura 1 – Processo de classificação -[Adaptado((REZENDE, 2003))]	17
Figura 2 – Subconjunto da ontologia do NELL [Adaptado de Duarte (2016)]	19
Figura 3 – Arquitetura do sistema NELL (MITCHELL et al., 2018)	20
Figura 4 – Exemplo de uma rede Bayesiana. A rede, à esquerda, representa um conjunto de suposições de independência condicional. Cada nó é associado a uma tabela de probabilidade condicional, que especifica a distribuição condicional para as variáveis, dados seus pais imediatos no grafo. A tabela de probabilidade condicional para o nó <i>Fogueira</i> é mostrada à direita, onde <i>Fogueira</i> é abreviado como <i>C</i> , <i>Tempestade</i> como <i>S</i> e <i>Turista</i> como <i>B</i> . [Adaptado de (MITCHELL, 1997)]	22
Figura 5 – Exemplo de ordenação de variáveis para o K2. Neste exemplo, a lista <i>X</i> traz as variáveis <i>A, B, C</i> e <i>D</i> de um conjunto de dados. O K2 segue a ordem da esquerda para a direita na busca de pais para cada variável. Como a variável <i>A</i> é a primeira da lista, ela será o nó raiz e não terá pais. A partir de <i>B</i> todas as variáveis à esquerda são pais candidatos. Então, <i>A</i> pode ser pai de <i>B</i> , <i>A</i> e <i>B</i> podem ser pais de <i>C</i> e <i>A, B</i> e <i>C</i> podem ser pais de <i>D</i> .	25
Figura 6 – Pseudocódigo do algoritmo K2 [Adaptado de (COOPER; HERSKOVITS, 1992)].	26
Figura 7 – Os nós em destaque representam a cobertura de <i>Markov</i> do nó <i>X</i> .	27
Figura 8 – Exemplo de uma estrutura induzida pelo algoritmo DMBC a partir de um conjunto de dados. Nesta figura, é simulado o aprendizado de estrutura pelo DMBC e pelo K2 para destacar o número de variáveis que podem ser eliminadas e, portanto, as estruturas que não são candidatas. É possível notar que DMBC incluiu apenas um subconjunto de variáveis do problema (nós hachurados) na estrutura, aquelas que fazem parte da cobertura de <i>Markov</i> do nó classe. Se a estrutura fosse induzida pelo K2, os nós não hachurados também seriam incorporados.	28
Figura 9 – Pseudocódigo do algoritmo DMBC [Adaptado de (SANTOS, 2011)].	29
Figura 10 – Estrutura do classificador Naive Bayes.	30

Figura 11	– Estrutura da rede Bayesiana induzida pelo algoritmo DMBC a partir do conjunto de dados sobre doenças construído com dados do NELL. As arestas contínuas representam as relações semânticas existentes na base do NELL e utilizadas na construção do conjunto de dados. As arestas tracejadas e numeradas representam novas relações sugeridas para expandir a ontologia inicial do NELL.	41
Figura 12	– Estrutura da melhor rede Bayesiana induzida pelo algoritmo K2 a partir do conjunto de dados sobre doenças construído com dados do NELL. As arestas contínuas representam as relações semânticas existentes na base do NELL e utilizadas na construção do conjunto de dados. As arestas tracejadas e numeradas representam novas relações sugeridas para expandir a ontologia inicial do NELL.	45
Figura 13	– Estrutura da pior rede Bayesiana induzida pelo algoritmo K2 a partir do conjunto de dados sobre doenças construído com dados do NELL. As arestas contínuas representam as relações semânticas existentes na base do NELL e utilizadas na construção do conjunto de dados. As arestas tracejadas e numeradas representam novas relações sugeridas para expandir a ontologia inicial do NELL.	46
Figura 14	– Exemplo de inferência realizada no software Nética. Neste exemplo, o valor <i>heart_disease</i> (doença do coração) da variável <i>disease</i> foi selecionado e, depois da inferência, é observado o valor <i>red wine</i> , inferido para a variável <i>bevsoft</i> , como uma bebida que ameniza a doença.	47

Lista de tabelas

Tabela 1	– Exemplos de dados de relações armazenados na base de conhecimento do NELL.	35
Tabela 2	– Relações selecionadas a partir da base de conhecimento do NELL e usadas para a construção do conjunto de dados.	36
Tabela 3	– Exemplo de instâncias (registros) para a relação "inverseoffooddecreasestheriskofdisease".	36
Tabela 4	– Definição dos nomes dos atributos da base, de acordo com as relações utilizadas.	37
Tabela 5	– Informações sobre os atributos do conjunto de dados construído.	39
Tabela 6	– Novas relações identificadas pelo DMBC.	42
Tabela 7	– Descrição dos métodos de seleção de atributos.	43
Tabela 8	– Valores da função g para as Redes Bayesianas induzidas pelo K2 com diferentes ordenações de variáveis. Os valores de g estão ordenados de forma decrescente.	44
Tabela 9	– Novas relações identificadas pelo K2.	45
Tabela 10	– Valores inferidos (e presentes no NELL) a partir das Redes Bayesianas induzidas por K2 e DMBC.	48
Tabela 11	– Valores inferidos (e não presentes no NELL) a partir das Redes Bayesianas induzidas por K2 e DMBC.	49
Tabela 12	– Comparação dos resultados de inferência para a variável classe <i>disease</i>	50

Sumário

1	Introdução	12
1.1	Contexto inicial	12
1.2	Justificativas	13
1.3	Objetivos gerais e específicos	14
1.4	Organização do texto	15
2	Referencial Teórico	16
2.1	Aprendizado de máquina	16
2.2	Pré-processamento	17
2.3	<i>Never-Ending Language Learning</i> (NELL)	18
2.4	Redes Bayesianas	21
2.4.1	Definição	21
2.4.2	Inferência em Redes Bayesianas	22
2.4.3	Aprendizado de Redes Bayesianas	23
2.4.3.1	Algoritmo K2	24
2.4.4	Aprendizado de classificadores Bayesianos	26
2.4.4.1	Algoritmo DMBC	26
2.4.4.2	Algoritmo Naive Bayes	29
3	Trabalhos Relacionados	31
4	Construção de um conjunto de Dados a partir do NELL	34
4.1	Etapas de construção do conjunto de dados	34
4.2	Informações sobre o conjunto de dados	39
5	Redes Bayesianas a partir do NELL	40
5.1	Rede Bayesiana induzida por DMBC	40
5.2	Redes Bayesianas induzidas por K2	42
5.3	Resultados de inferência Bayesiana	46
5.3.1	Inferência da variável classe	50
6	Conclusão e Trabalhos Futuros	51
6.1	Trabalhos futuros	52
6.2	Contribuições do projeto	52
	Referências	53

1 Introdução

1.1 Contexto inicial

O aprendizado de máquina (AM) (MITCHELL, 1997) é uma área de pesquisa que vem recebendo bastante atenção dentro da Inteligência Artificial e Ciência da Computação em geral. Este fato contribui para um grande avanço e progresso nos resultados obtidos por métodos e algoritmos desta área de pesquisa nas últimas décadas (KUROSE et al., 2006).

Os algoritmos tradicionais de AM aprendem de forma diferente dos seres humanos. Estes algoritmos aprendem a induzir uma função (ou hipótese) a partir de um conjunto de dados e, assim, geram um modelo capaz de fazer previsões (FACELI et al., 2017).

Desta forma, os métodos de AM são usados para encontrar regularidades, padrões ou conceitos num conjunto de dados (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). Mesmo assim, não existem, hoje em dia, ainda muitos sistemas computacionais capazes de aprender de maneira cumulativa para sempre. É mais importante, sistemas que se utilizem do conhecimento adquirido ontem para melhorar sua habilidade de aprendizado hoje, num processo contínuo e sem fim.

Inspiradas no aprendizado humano, novas pesquisas têm contribuído para o desenvolvimento de um novo paradigma de AM chamado de aprendizado sem fim. O primeiro sistema de aprendizado sem fim relatado na literatura é chamado NELL (*Never-Ending Language Learning*) (CARLSON et al., 2010). NELL é um sistema computacional que executa 24 horas por dia, 7 dias por semana, aprendendo, de maneira contínua e incremental, por meio de informações de textos da web, extraíndo conhecimento sobre diferentes itens e armazenando-o na ontologia¹.

A proposta do sistema NELL também sugere que, assim como os seres humanos, estes sistemas de aprendizado sem fim devem ser capazes de se utilizar de conhecimento adquirido a cada dia para, autonomamente, aprenderem cada vez melhor.

A base de conhecimento do sistema NELL é representada por uma estrutura baseada em ontologia e caracterizada por categorias, relações e suas instâncias. As categorias são os tipos de conhecimento, por exemplo, “pessoa”, “equipeEsportiva”, “fruta”, “emoção”, etc. Já as relações são os relacionamentos entre as categorias, por exemplo, “atletaJogaParaEquipeEsportiva(atleta, equipeEsportiva)”, “musicoTocaInstrumento(músico, instrumento)”, etc. Algumas das categorias e relações da ontologia são marcadas como itens

¹ ontologia é um modelo que representa um conjunto de conceitos a partir de um domínio

que devem ser “populados”. “Popular” um item significa, neste contexto, extrair conhecimento sobre este item, a partir da Web, e armazenar na ontologia. Para isto, o sistema NELL utiliza diferentes componentes (ou sub-sistemas), como CPL (MITCHELL et al., 2018), SEAL (MANSANO et al., 2018), Prophet (APPEL; JUNIOR, 2011), OntExt (BARCHI, 2014) e Conversing Learning (MOHAMED; HRUSCHKA; MITCHELL, 2011), com o objetivo de ter a habilidade de “autossupervisão” e “autorreflexão” de sua base de conhecimento. Apesar de terem sido obtidos resultados promissores na construção do NELL, não há ainda uma metodologia bem definida para o desenvolvimento deste tipo de sistema.

NELL tem mostrado a viabilidade da construção de um sistema de aprendizado sem fim com base na metodologia proposta em (CARLSON et al., 2010). No entanto, ainda não é capaz de apresentar todas as características essenciais para que possa aprender continuamente de maneira autônoma, autossupervisionada e possa expandir sua ontologia inicial automaticamente sempre que necessário. Assim, é proposto neste trabalho utilizar o formalismo das Redes Bayesianas (veja Seção 2.4) para realizar inferências na base de conhecimento do NELL, bem como para identificar novas relações semânticas que possam ser inseridas na base de conhecimento (expandindo assim a ontologia inicial do NELL).

1.2 Justificativas

Existem diferentes abordagens para o aprendizado automático, as quais proporcionam métodos ou técnicas que podem se adaptar a diferentes problemas representados por conjuntos de dados. Entre estas abordagens, é possível destacar o formalismo de Redes Bayesianas, muito utilizado em problemas que envolvem incerteza.

Há situações bem específicas em que a construção de um modelo gráfico de Redes Bayesianas é possível com a ajuda de um especialista, porém nem sempre é fácil, devido a existência de centenas de variáveis. Logo, torna-se necessário o uso de métodos de aprendizado automático para que a estrutura da rede seja aprendida (ou induzida), e assim, o modelo gráfico seja construído.

Algoritmos de aprendizado de Redes Bayesianas podem ser usados como uma ferramenta para encontrar relações entre variáveis. Então, estes algoritmos podem descobrir novas relações semânticas e, desta forma, ajudar a estender a ontologia inicial do NELL. Além disso, algoritmos de inferência podem ser aplicados às Redes Bayesianas induzidas a partir da base de conhecimento do NELL, com o objetivo de descobrir novos “fatos” (ou instâncias) que podem ser inseridos na base do NELL.

1.3 Objetivos gerais e específicos

O objetivo principal deste projeto é propor o uso de Redes Bayesianas para ajudar a estender e popular a base de conhecimento do NELL. Mais especificamente, busca-se a ampliação da capacidade de inferência do NELL em sua própria base de conhecimento para permitir generalizações e aprendizado automático com base no que já foi aprendido. Com isso, pode-se obter a melhoria do desempenho do sistema a partir do conhecimento previamente adquirido (característica que descreve o processo de aprender a aprender cada vez melhor).

Como as Redes Bayesianas são modelos probabilísticos que, além de permitirem inferência, também podem ser utilizadas para a descoberta de relações entre as variáveis, os métodos e algoritmos aplicados para inferência (na base de conhecimento do NELL) podem também ser utilizados para a extensão automática da ontologia do NELL (por meio da inserção de novas relações semânticas). Assim, têm-se como objetivos específicos:

- Apresentar uma metodologia para a construção de conjuntos de dados a partir do conhecimento armazenado na base de conhecimento do sistema NELL para induzir Redes Bayesianas. Neste caso, é proposto utilizar as informações adquiridas a partir das relações semânticas existentes na ontologia no NELL. Como um estudo de caso, neste trabalho, foi construído, a partir desta metodologia, um conjunto de dados que representa o domínio de doenças;
- Utilizar Redes Bayesianas para representar o conhecimento adquirido pelo sistema NELL e permitir inferência (com base nesta representação). Para tanto, foram aplicados algoritmos de aprendizado de Redes Bayesianas e de inferência no conjunto de dados construído. Entre os algoritmos de aprendizado, destacam-se os algoritmos DMBC (SANTOS *et al.*, 2011) e K2 (COOPER; HERSKOVITS, 1992) por suas características específicas.
- Utilizar Redes Bayesianas para a descoberta de novas relações semânticas e novos fatos a serem inseridas na base de conhecimento do NELL (expandindo assim a ontologia inicial). Após a indução das Redes Bayesianas com o conjunto de dados criado a partir do conhecimento do NELL, as estruturas gráficas (grafos) são analisadas para verificar se os relacionamentos entre nós (que representam as variáveis/atributos) correspondem às relações semânticas extraídas da base de conhecimento do NELL para criar o conjunto de dados. Outros relacionamentos podem ser sugeridos como novas relações semânticas para o NELL e novos fatos podem ser descobertos através de algoritmos de inferência aplicados às Redes Bayesianas.

1.4 Organização do texto

Esta dissertação é organizada como segue. No Capítulo 2, será apresentado o referencial teórico sobre os conceitos utilizados neste trabalho. Inicialmente, na Seção 2.1, são abordados conceitos básicos de aprendizado de máquina, seguidos de uma pequena introdução sobre pré-processamento de dados na Seção 2.2. A Seção 2.3 apresenta o sistema NELL e explica seu funcionamento. Na Seção 2.4, são exibidos conceitos e definições de Redes Bayesianas, explicando brevemente sobre a inferência Bayesiana e o problema do aprendizado de estruturas.

O Capítulo 3 traz os trabalhos relacionados, que propõem formas diferentes de aprender novas relações e novos fatos para contribuir com o aprendizado do NELL. Também é apresentado um trabalho que propõe a criação de uma instância do NELL em português.

No Capítulo 4, é apresentada uma proposta para a criação de conjuntos de dados a partir das relações semânticas existentes na base de conhecimento do NELL. A partir desta proposta, foi criado, para este trabalho, um conjunto de dados com informações de um domínio de conhecimento específico, neste caso, sobre doenças. O objetivo foi verificar se o conhecimento do NELL sobre algum domínio específico pode ser usado também para a construção de um sistema especialista com base em Redes Bayesianas.

O Capítulo 5 apresenta os resultados obtidos pelos algoritmos de aprendizado de Redes Bayesianas, tendo como entrada o conjunto de dados criado. As redes induzidas são exibidas e comparadas, destacando a identificação de relacionamentos entre as variáveis que formam as relações semânticas. Além disso, resultados de inferência das redes também são mostrados.

Por fim, no Capítulo 6, são apresentadas as conclusões sobre os resultados obtidos e apontados alguns trabalhos futuros.

2 Referencial Teórico

2.1 Aprendizado de máquina

O Aprendizado de Máquina (AM) é uma área de Inteligência Artificial que desenvolve técnicas computacionais para construir sistemas capazes de adquirir conhecimento de forma automática (REZENDE, 2003). É possível distinguir três formas de adquirir conhecimento por meio das técnicas de AM: aprendizado supervisionado, não-supervisionado e por reforço.

O aprendizado supervisionado se caracteriza pela existência de um supervisor externo que conhece a saída (rótulo) para cada conjunto de valores de entrada. Isso permite ao supervisor avaliar a capacidade da hipótese induzida de prever o valor de saída para novos exemplos de entrada (FACELI et al., 2017). No aprendizado não-supervisionado, não se faz uso do atributo de saída; os algoritmos aprendem com dados de teste que não foram rotulados, buscando formar agrupamentos ou *clusters*. Já o aprendizado por reforço consiste em treinar modelos de AM para tomar uma sequência de decisões.

Os algoritmos do paradigma de aprendizado supervisionado podem ser aplicados em tarefas de classificação e regressão. Um algoritmo de classificação tem por tarefa classificar itens com valores discretos, previamente rotulados por um supervisor. Já os algoritmos de regressão prevêem valores contínuos ou reais.

Os algoritmos de classificação, assim como de regressão, tem por objetivo aprender um modelo ou hipótese a partir de um subconjunto de dados de treinamento que é capaz de relacionar os valores dos atributos de entrada ao valor de seu atributo alvo. Após terminar a fase de treinamento do conjunto de dados, obtém-se um modelo capaz de prever corretamente o valor de novos exemplos (REZENDE, 2003). Neste projeto, é adotado o paradigma de aprendizado supervisionado, utilizando algoritmos de classificação.

Na Figura 1, é demonstrado o processo de construção de um classificador. Inicialmente, os dados brutos devem ser preparados (em um conjunto de exemplos para que possam ser processados). Suponha um conjunto de dados de pacientes de um hospital. Cada dado (registro ou objeto) corresponde a um paciente. Esses dados passam por um pré-processamento gerando um conjunto de exemplos no formato da tabela abaixo. O nome, idade, sintomas de um paciente (X_1, X_2, \dots, X_m) são chamados de atributos de entrada ou previsores e um dos atributos, por exemplo resultados clínicos (Y), é escolhido para ser o atributo alvo ou saída (classe). A aplicação de um algoritmo de AM, no caso Árvore de Decisão, sob esse conjunto de exemplos vai buscar por uma hipótese na forma de regra ou conjunto de regras que possa definir qual doença o paciente possui, de acordo

com os atributos de entrada. Logo, esse conjunto de regras corresponde a um modelo ou classificador, no caso, uma Árvore de Decisão.

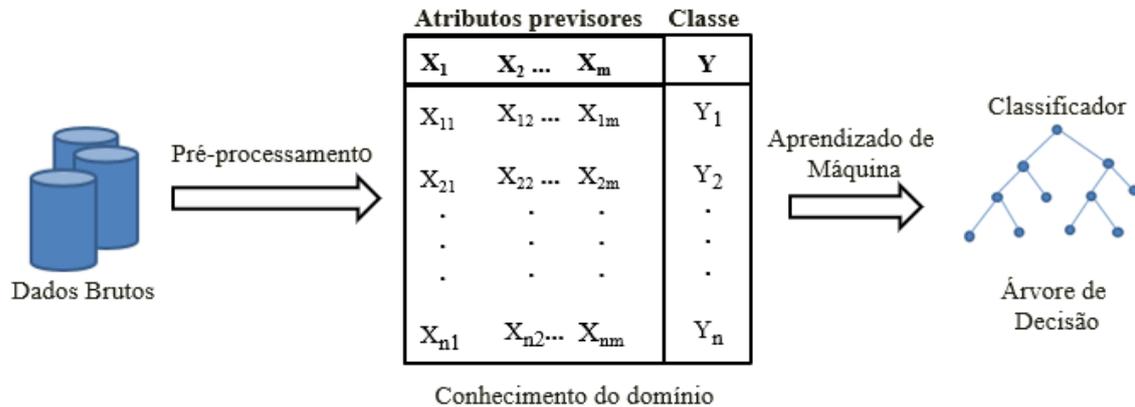


Figura 1 – Processo de classificação -[Adaptado((REZENDE, 2003))]

2.2 Pré-processamento

As técnicas de pré-processamento de dados são utilizadas para melhorar a qualidade dos dados, eliminando ou minimizando problemas existentes, tornando-os mais adequados para a aplicação de determinado algoritmo de AM (FACELI et al., 2017). As principais operações de pré-processamento de dados são:

1. Seleção de dados: compreende retirar atributos redundantes ou irrelevantes.
2. Transformação de dados: essa etapa visa codificar os dados de forma a atender às necessidades específicas dos algoritmos de AM. Existem algoritmos que trabalham apenas com valores numéricos e outros apenas com valores categóricos, exigindo a necessidade de transformação dos dados.
3. Limpeza dos dados: É comum que os dados sobre os quais se deseja extrair algum conhecimento estejam incompletos (ausência de informação), ruidosos (dados errados ou divergentes do esperado) (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Algumas técnicas podem ser utilizadas para resolver esses problemas, quando encontrados em uma base de dados. Segundo a literatura, quando se trata de valores incompletos ou ausentes, pode-se utilizar as seguintes técnicas: criar para o atributo com valor ausente um novo valor, indicando que o atributo possuía um valor desconhecido. Também, é possível calcular a média (atributos numéricos), moda ou mediana (atributos

categoricos) dos valores conhecidos para os valores ausentes do atributo. Outra variação dessa abordagem é preencher o atributo faltante com o valor mais frequente (FACELI et al., 2017).

A exclusão de casos e a correção de erros são também alguns métodos utilizados para a limpeza de inconsistências. O primeiro método consiste em excluir os registros que possuem pelo menos uma inconsistência e a identificação dos casos pode ser feito por meio de consultas. Já o método de correção de erros consiste em substituir valores errôneos identificados no conjunto de dados de forma manual ou utilizando comandos de atualização de dados existentes (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

2.3 *Never-Ending Language Learning* (NELL)

Considerando que um dos mais importantes métodos pelos quais o ser humano aprende é pela leitura, muitas iniciativas, bem como projetos de pesquisa, têm dedicado suas investigações a fim de construir máquinas capazes de aprender a partir de leitura (CLARK et al., 2007).

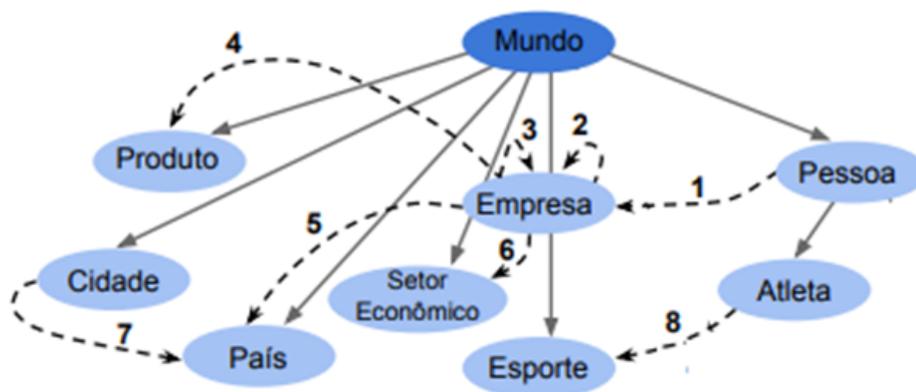
O aprendizado sem fim sugere um agente inteligente capaz de reduzir a dificuldade de resolver problemas usando conhecimento adquirido. Dessa forma, o processo de aprendizagem sem fim acontece de maneira incremental, ou seja, cada conceito aprendido pode ser usado para auxiliar o aprendizado posterior. Este processo tem sido analisado, implementado e aplicado em diferentes trabalhos nas áreas de robótica e extração de conhecimento a partir de grandes fontes como a web. O projeto NELL é um exemplo da implementação de aprendizado sem fim.

NELL foi o primeiro sistema de aprendizado sem fim descrito na literatura, ativo desde de janeiro de 2010, e que funciona 24 horas por dia, 7 dias por semana, extraíndo informações de texto da web (CARLSON et al., 2010). A NELL tem como objetivos alimentar a sua base de conhecimento (*Knowledge Base - KB*) por meio da coleta diária de fatos a partir da web, aprender cada vez mais com esses fatos e alcançar uma alta confiança no aprendizado para que possa garantir a continuidade do aprendizado (DUARTE, 2016).

A NELL é representada por uma base ontológica estruturada e caracterizada por categorias, relações e suas instâncias (MIANI; JUNIOR, 2015). As categorias são, por exemplo, os tipos de conhecimento: pessoa, organização, universidade, emoção, músico, etc. Já as relações são os relacionamentos entre as categorias, por exemplo, *TrabalhaPara* (pessoa, organização), *musicoTocaInstrumento* (músico, instrumento), etc.

Para que a NELL inicie o aprendizado, a ontologia também deve possuir exemplos de instâncias para cada uma das categorias, por exemplo, pessoa (João), fruta (laranja). E da mesma forma, deve possuir exemplos de cada uma das relações, por exemplo, atleta-

JogaParaEquipeEsportiva (Zico, Flamengo), musicoTocaInstrumento (Eric Clapton, guitarra), etc. A Figura 2 apresenta exemplos de categorias e relações do sistema NELL.



Relações:

- | | |
|---|---|
| 1- Chefia (Pessoa, Empresa) | 5- PossuiEscritorio (Empresa, País) |
| 2- VendePara (Empresa, Empresa) | 6- CompetênciaEm (Empresa, SetorEconômico) |
| 3- ParceiraDe (Empresa, Empresa) | 7- LocalizaEm (Cidade, País) |
| 4- Fabrica (Empresa, Produto) | 8- Joga (Atleta, Esporte) |

Figura 2 – Subconjunto da ontologia do NELL [Adaptado de Duarte (2016)]

As instâncias de categoria e de relações adicionadas à base de conhecimento são particionadas em fatos e crenças candidatos. Os componentes do subsistema podem ler a partir da KB e consultar outros recursos externos (por exemplo, corpora de texto ou a Internet) e, em seguida, propor novos fatos candidatos. Componentes fornecem uma probabilidade para cada candidato proposto e um resumo da evidência da fonte que o suporta.

A arquitetura de software da NELL inclui uma KB que age como um quadro-negro através do qual vários módulos de aprendizagem e inferência se comunicam. Esses módulos mapeiam os métodos de aprendizagem (CPL, CMC, SEAL, OpenEval, PRA e NEIL) para os diferentes tipos de funções (MITCHELL et al., 2018). Na Figura 3, é apresentada a arquitetura de software do NELL.

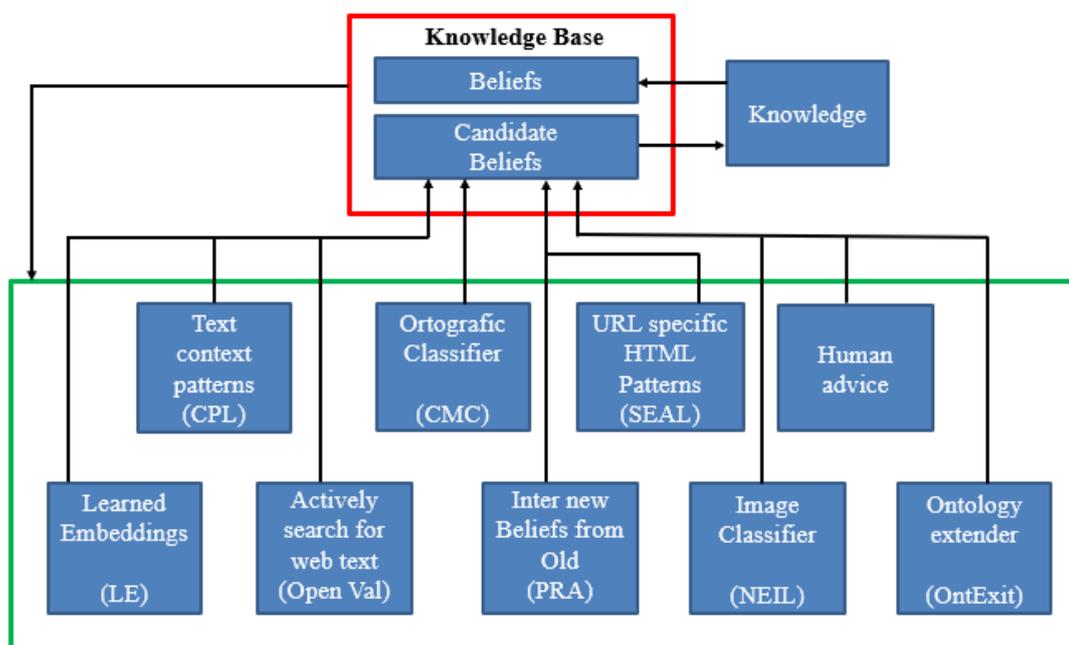


Figura 3 – Arquitetura do sistema NELL (MITCHELL et al., 2018)

A seguir, são descritos os módulos de aprendizagem:

1. CPL - Coupled Pattern Learning: é um subsistema que realiza o aprendizado extraíndo padrões de um corpus a partir de páginas web.(MITCHELL et al., 2018).
2. CMC - Coupled Morphological Classifier: é um classificador morfológico que examina os resultados obtidos através do CPL e do SEAL, procurando por padrões morfológicos (DUARTE, 2016).
3. SEAL - Sistema Expander for Any Language: extrator semi-estruturado que vasculha a Internet com conjuntos de crenças de cada categoria ou relação, buscando extrair novas instâncias (MANSANO et al., 2018).
4. OpenEval: o sistema OpenEval usa diferentes recursos de contexto do sistema CPL, pesquisando na Web em tempo real para coletar informações(MITCHELL et al., 2018).
5. NEIL - Never Ending Image Learner: mecanismo de busca de imagens, aplicando-se a apenas determinado subconjunto de categorias de ontologia do NELL(por exemplo, não para categorias não visuais como gênero musical) (MITCHELL et al., 2018).
6. PRA - Path Ranking Algorithm: é um método proposto que visa melhorar a cobertura da base de conhecimento executando inferência diretamente sobre o gráfico KB (GARDNER et al., 2013).

7. OntExt - Ontology extender: é um subsistema componente que possui um papel importante na expansão da ontologia do NELL. Ele objetiva expandir continuamente a KB, identificando e adicionando novas relações. Ele utiliza matrizes de co-ocorrência para estruturar os valores normalizados de co-ocorrência entre as frases verbais para cada par de categorias, e dessa forma identifica padrões de contexto que interligam as categorias (BARCHI, 2014).

2.4 Redes Bayesianas

2.4.1 Definição

Redes Bayesianas são redes causais ou modelos gráficos probabilísticos e representam uma distribuição de probabilidade sobre um conjunto de variáveis (MITCHELL, 1997). Em síntese, as Redes Bayesianas utilizam a teoria da probabilidade para modelar as incertezas de um determinado ambiente (SILVA, 2014).

O formalismo de Redes Bayesianas foi desenvolvido no início dos anos 80 para facilitar a tarefa de predição em sistemas de inteligência artificial. As Redes Bayesianas passaram a ser adotadas em situações onde existem incertezas, apoiando-se na teoria da probabilidade, e têm sido utilizadas em diversas aplicações, principalmente em diagnósticos médicos (MARQUES; DUTRA, 2002).

Uma rede Bayesiana é composta por dois componentes: i) um grafo direcionado acíclico (GDA), ou estrutura gráfica, onde os nós representam as variáveis aleatórias de um determinado problema. Os nós são conectados por arcos que definem relações de dependência entre os nós. Se houver um arco a partir de um nó X para um nó Y , diz-se que X é um dos pais de Y ; ii) para cada variável, há uma tabela de probabilidade condicional (ou parâmetros numéricos) que descreve a distribuição de probabilidade da variável de acordo com seus pais.

Numa rede Bayesiana, a ausência de alguns arcos no GDA (conhecido como componente qualitativo) representa a existência de relações de independência condicional entre as variáveis e a presença deles pode representar a existência de relações de dependência direta. A tabela de probabilidades (conhecida como componente quantitativo) é uma coleção de medidas de probabilidades condicionais e são atualizadas com o uso do Teorema de Bayes, com base em uma nova informação amostral. A distribuição de probabilidade conjunta sobre um conjunto de variáveis ($X_1 \dots X_n$) pode ser calculada pela fórmula 2.1:

$$P(X_1 \dots X_n) = \prod_{i=1}^n P(X_i | Pais(X_i)) \quad (2.1)$$

onde $Pais(X_i)$ denota o conjunto de pais de X_i na rede. Note que os valores de $P(X_i | Pais(X_i))$ são os valores armazenados na tabela de probabilidade condicional associada com o nó

X_i (MITCHELL, 1997).

A Figura 4 apresenta uma rede Bayesiana que representa a distribuição de probabilidade conjunta sobre as variáveis booleanas *Tempestade*, *Raio*, *Trovão*, *Queimada*, *Fogueira* e *Turista*. O lado direito da figura mostra a tabela de probabilidade condicional associada a variável *Fogueira*. O conjunto de tabelas de probabilidade condicional para todas as variáveis, junto com o conjunto de suposições de independência condicional descritas pela rede, descreve a distribuição de probabilidade conjunta completa para a rede.

Uma característica atrativa de Redes Bayesianas é o fato de que elas permitem um jeito interessante de representar conhecimento causal como, por exemplo, o fato de que *Raio* causa *Trovão*. Na terminologia de independência condicional, isto é expresso pela declaração de que *Trovão* é condicionalmente independente de outras variáveis na rede, dado o valor de *Raio*. Esta suposição de independência condicional é demonstrada pelos arcos na rede Bayesiana da Figura 4.

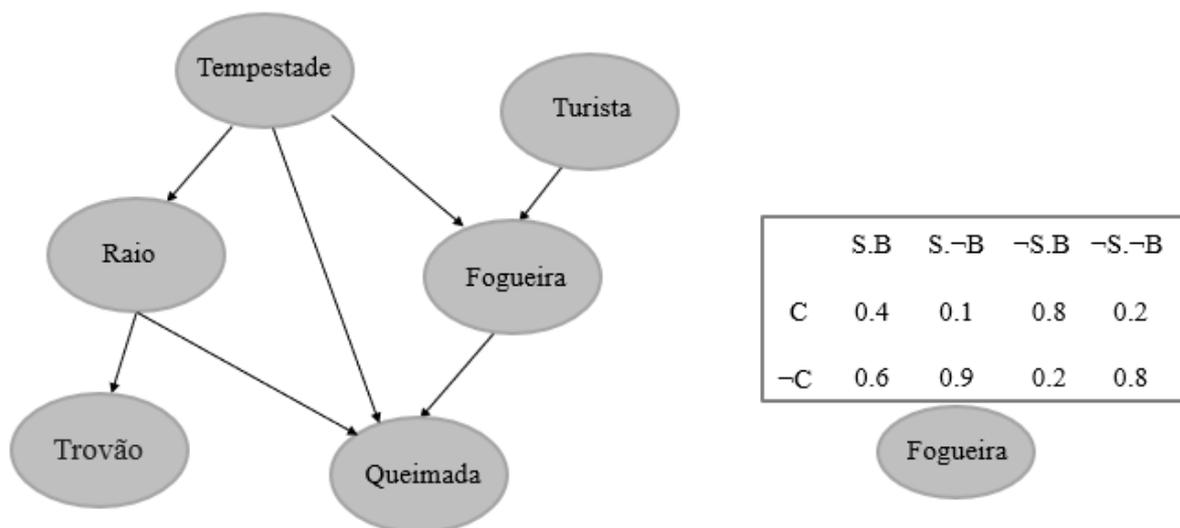


Figura 4 – Exemplo de uma rede Bayesiana. A rede, à esquerda, representa um conjunto de suposições de independência condicional. Cada nó é associado a uma tabela de probabilidade condicional, que especifica a distribuição condicional para as variáveis, dados seus pais imediatos no grafo. A tabela de probabilidade condicional para o nó *Fogueira* é mostrada à direita, onde *Fogueira* é abreviado como *C*, *Tempestade* como *S* e *Turista* como *B*. [Adaptado de (MITCHELL, 1997)]

2.4.2 Inferência em Redes Bayesianas

Após a construção (ou aprendizado) de uma rede Bayesiana, é possível extrair conhecimentos por meio dela, utilizando um processo chamado inferência. A inferência Bayesiana calcula a distribuição de probabilidade condicional de um conjunto de variáveis, de acordo com um conjunto de evidências (SANTOS, 2011).

A probabilidade condicional de eventos de interesse são calculados a partir de probabilidades conhecidas, usando o Teorema de Bayes. A inferência pode ser realizada simplesmente para adquirir conhecimento da rede, como saber quais os sintomas de uma doença ou para consultar determinada situação. Por exemplo, uma inferência possível seria: dado que um paciente possui a doença diabetes e utiliza determinada droga no tratamento, seria importante conhecer os possíveis efeitos colaterais. A partir das informações fornecidas à rede, é gerada uma propagação dos fatos ou evidências por toda a rede, havendo uma atualização do conhecimento, permitindo consultar uma variável alvo. A propagação dos fatos é responsável pela atualização do conhecimento a partir de evidências.

As Redes Bayesianas, além de permitirem a consulta a partir de fatos fornecidos à rede, também proporcionam a tomada de decisões baseada em probabilidades, ajudam a definir quais evidências devem ser observadas a fim de se obter informações úteis, dão visibilidade aos aspectos do modelo que possuem maior impacto sob as variáveis de consulta, e ainda, possibilitam explicar os resultados de uma inferência probabilística ao usuário (MARQUES; DUTRA, 2002).

2.4.3 Aprendizado de Redes Bayesianas

Problemas podem ser modelados a partir do uso de Redes Bayesianas. Porém, construir uma rede Bayesiana não é uma tarefa fácil. Ela precisa representar o domínio em questão, identificando corretamente as dependências e independências das variáveis e a direção da causalidade. Uma das dificuldades encontradas nesse processo é a falta de informação, muitas vezes, pela ineficiência em lidar com dados estatísticos. Logo, surgiu a necessidade de desenvolver metodologias que pudessem construir Redes Bayesianas de forma automática.

A aprendizagem Bayesiana é o mecanismo utilizado para a construção de uma rede Bayesiana a partir de um domínio e uma base de dados. O processo de aprendizado em Redes Bayesianas deve identificar a estrutura da rede e induzir as distribuições de probabilidades da rede. Este processo pode ser dividido em: aprendizado da estrutura ou grafo e aprendizado dos parâmetros numéricos ou probabilidades condicionais (CASTILLO; GUTIERREZ; HADI, 2012).

O aprendizado da estrutura identifica as variáveis, suas relações e a direção da causalidade. Este processo de aprendizado busca construir a estrutura gráfica da rede (ou grafo), a partir de uma base de dados, sendo considerado um problema difícil, principalmente devido ao fato de que o número de possíveis estruturas para um dado problema cresce exponencialmente de acordo com o número de variáveis (SANTOS, 2011). Tendo a estrutura e os dados, o aprendizado dos parâmetros numéricos foca em obter as probabilidades condicionais para cada variável, as quais identificam a “força” dos relacionamentos

entre elas.

A estimação da estrutura de uma rede Bayesiana, também conhecida na literatura como aprendizado de estrutura, pode ser obtida a partir de duas classes principais de métodos: a primeira baseada em uma busca heurística e a segunda baseada no conceito de independência condicional entre atributos. Assim, algoritmos são requeridos para ambos tipos de estimação (NETO; CYMBALISTA, 2006). Sendo que ainda existem implementações de versões híbridas que combinam as duas categorias citadas acima.

Um bom exemplo de algoritmo de aprendizado de Redes Bayesianas com base em busca heurística é o algoritmo K2, proposto por (COOPER; HERSKOVITS, 1992), o qual é apresentado na Subseção 2.4.3.1.

2.4.3.1 Algoritmo K2

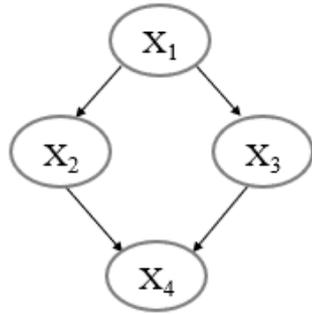
K2 é um algoritmo de busca heurística desenvolvido para o aprendizado de estruturas de Redes Bayesianas. Ele recebe, como entrada, um conjunto de dados e uma ordenação das variáveis e produz, como saída, a estrutura da rede (SANTOS, 2007). De acordo com (VERMA; PEARL, 1990) e (HSU, 2004), K2 tem bom desempenho, em termos de complexidade computacional e resultados precisos, quando uma ordenação de variáveis adequada é fornecida.

A ordenação de variáveis permite reduzir o espaço de busca por possíveis estruturas, visto que o número de estruturas candidatas cresce exponencialmente de acordo com o número de variáveis (a Figura 5 apresenta um exemplo de ordenação para o K2).

Uma busca por todas as estruturas pode tornar o processo inviável computacionalmente para diversos domínios (SANTOS, 2007). Assim, K2 usa uma lista ordenada (contendo todas as variáveis) e considera que somente variáveis à esquerda de uma determinada variável X_i podem ser consideradas pais de X_i . A primeira variável na lista será um nó raiz na estrutura da rede e, portanto, não tem pais.

O algoritmo K2 inicia sua execução supondo que todos os nós não têm pais. Então, a partir da segunda variável da lista ordenada (já que a primeira é um nó raiz), os possíveis pais para cada variável (ou nó na rede) são testados de forma incremental e, aqueles que maximizam a probabilidade da estrutura estar de acordo com a base de dados são adicionados à rede. Se a adição de uma variável ao conjunto de pais não maximiza esta probabilidade, o algoritmo encerra a busca de pais para a variável atual e passa a buscar pais para a próxima variável da lista. Este processo é repetido para todas as variáveis até que a melhor estrutura possível seja encontrada (SANTOS, 2011).

$$Y = \{X_1, X_2, X_3, X_4\}$$



$$Z = \{X_4, X_3, X_2, X_1\}$$

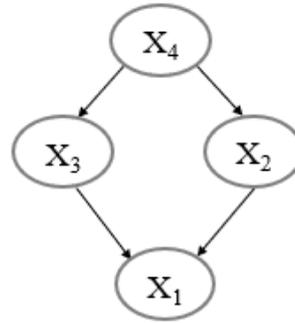


Figura 5 – Exemplo de ordenação de variáveis para o K2. Neste exemplo, a lista X traz as variáveis A, B, C e D de um conjunto de dados. O K2 segue a ordem da esquerda para a direita na busca de pais para cada variável. Como a variável A é a primeira da lista, ela será o nó raiz e não terá pais. A partir de B todas as variáveis à esquerda são pais candidatos. Então, A pode ser pai de B , A e B podem ser pais de C e A, B e C podem ser pais de D .

É importante salientar que K2 usa um método guloso para encontrar a melhor estrutura. As estruturas candidatas são avaliadas de acordo com um critério de pontuação (ou métrica), que testa cada conjunto de pais possíveis, para cada nó, de acordo com a seguinte equação 2.2:

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (2.2)$$

onde cada variável $x_i (i = 1, \dots, n)$ tem r_i possíveis valores $(v_{i_1}, v_{i_2}, \dots, v_{i_{r_i}})$. Cada variável x_i tem um conjunto de pais π_i e q_i é o número de instanciações de π_i . N_{ijk} é o número de objetos no conjunto de dados D , onde x_i tem o valor v_{ik} e π_i é instanciado como w_{ij} , o qual representa a j -ésima instanciação relativa a D de π_i . Finalmente, $N_{ij} = \sum N_{ijk}$.

A Figura 6 apresenta o pseudocódigo do algoritmo K2. Após construir a estrutura da rede Bayesiana, as probabilidades condicionais já podem ser determinadas usando, por exemplo, uma estimação Bayesiana da probabilidade da estrutura da rede (SANTOS, 2011).

```

Algoritmo K2:
{Entrada: Um conjunto de n nós, uma ordenação dos nós, limite superior u sobre o
número de pais que um nó pode ter e uma base de dados D, contendo m casos.}
{Saída: Para cada nó, a identificação dos pais do nó.}
1. for i := 1 to n do
2.    $\pi_i := \{\}$ ;
3.   Pold := g(i,  $\pi_i$ ); {Equação (2.2).}
4.   OKToProceed := true
5.   while OKToProceed and  $|\pi_i| < u$  do
6.     let z be the node in Pred(xi) -  $\pi_i$  that maximizes g(i,  $\pi_i \cup \{z\}$ );
7.     Pnew := g(i,  $\pi_i \cup \{z\}$ );
8.     if Pnew > Pold then
9.       Pold := Pnew;
10.       $\pi_i := \pi_i \cup \{z\}$ ;
11.    else OKToProceed := false;
12.  end {while};
13.  write ('Node:', xi, 'Parents of this node:',  $\pi_i$ )
14. end {for};
15. end {K2};

```

Figura 6 – Pseudocódigo do algoritmo K2 [Adaptado de (COOPER; HERSKOVITS, 1992)].

As Redes Bayesianas induzidas por K2 podem ser aplicadas também para tarefas de classificação. Neste caso, a rede Bayesiana é utilizada com o fim específico de inferir o comportamento de uma única variável X_i , considerada então como uma variável classe. Quando este processo é executado, diz-se que se construiu um classificador Bayesiano irrestrito. Contudo, a indução de classificadores baseada em Redes Bayesianas, por sua vez, pode gerar custos computacionais indesejáveis. Desta forma, têm sido desenvolvidos algoritmos específicos para construção de classificadores Bayesianos que distinguem a variável classe desde o início do processo de indução a partir dos dados. A Subseção 2.4.4 apresenta dois algoritmos de indução de classificadores Bayesianos: DMBC, desenvolvido a partir do K2, e o Naive Bayes.

2.4.4 Aprendizado de classificadores Bayesianos

2.4.4.1 Algoritmo DMBC

O algoritmo DMBC (*Dynamic Markov Blanket Classifier*) foi desenvolvido com base no algoritmo K2. É um algoritmo de aprendizado de Redes Bayesianas desenvolvido especificamente para induzir classificadores Bayesianos (SANTOS, 2011).

O DMBC foi desenvolvido usando a função g do K2. A indução de classificadores usando a função g do K2 pode melhorar as estimativas de probabilidade, contribui para boas taxas de classificação e também permite a identificação de relações entre as variáveis mais importantes. (SANTOS, 2011).

O algoritmo DMBC explora a cobertura de *Markov* da variável classe (CMB) a fim de aprender a estrutura da rede de forma mais rápida e produzir estimativas mais exatas com relação às probabilidades da classe. A ideia é excluir, das possíveis estruturas (GDA), aquelas que têm variáveis que não pertencem ao CMB (SANTOS, 2011).

A cobertura de *Markov* (ou *Markov Blanket*) de uma variável X , $CM(X)$, define um conjunto de nós da rede Bayesiana que tem influência sobre X . Diz-se, então, que X é condicionalmente independente de todas as outras variáveis, dado $CM(X)$. Em resumo, $CM(X)$ é formada pela união do conjunto de pais de X , os filhos de X e os pais dos filhos de X , como é mostrado na Figura 7.

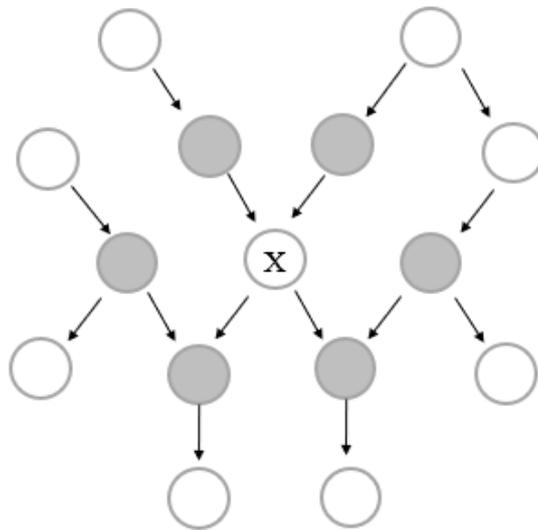


Figura 7 – Os nós em destaque representam a cobertura de *Markov* do nó X .

Utilizando a cobertura de *Markov* do nó classe, o algoritmo DMBC pode excluir variáveis irrelevantes para a classificação, buscando por uma estrutura que mapeia as relações entre a classe e as outras variáveis de um conjunto de dados. Desta forma, ele tende a reduzir o número de variáveis da rede gerada. O custo computacional, então, é reduzido quando comparado ao algoritmo K2, otimizando o aprendizado da estrutura do classificador Bayesiano. Na Figura 8, é apresentado um exemplo de estrutura induzida pelo DMBC.

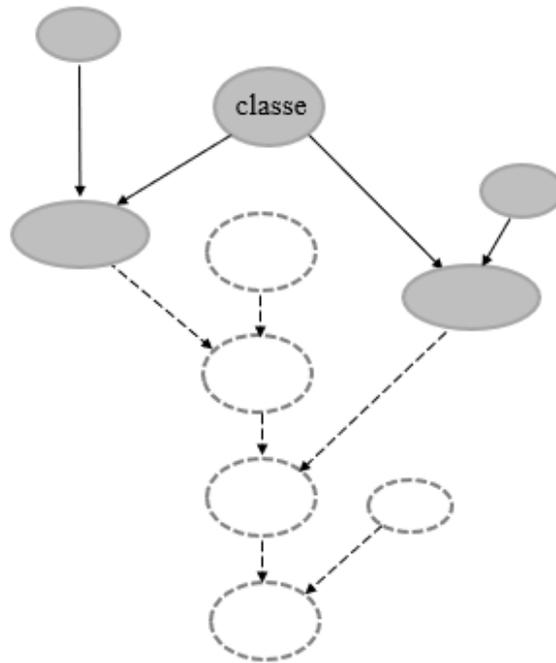


Figura 8 – Exemplo de uma estrutura induzida pelo algoritmo DMBC a partir de um conjunto de dados. Nesta figura, é simulado o aprendizado de estrutura pelo DMBC e pelo K2 para destacar o número de variáveis que podem ser eliminadas e, portanto, as estruturas que não são candidatas. É possível notar que DMBC incluiu apenas um subconjunto de variáveis do problema (nós hachurados) na estrutura, aquelas que fazem parte da cobertura de *Markov* do nó classe. Se a estrutura fosse induzida pelo K2, os nós não hachurados também seriam incorporados.

A Figura 9 apresenta o DMBC em uma forma algorítmica. Ao compará-lo com o algoritmo K2 (6), é possível notar que ele se diferencia nas linhas 10, 11, 15 e 16. Nas linhas 10 e 11, é verificado se o nó atual (da lista ordenada) está presente no CMB. Esta é uma restrição do DMBC para que todas as variáveis que pertençam ao CMB não sejam testadas e não sejam incluídas como pais de outras variáveis, reduzindo o número de possíveis estruturas a serem investigadas (SANTOS, 2011). Já as linhas 15 e 16, verificam se o nó atual é filho da classe; caso não seja, a busca por outros pais para este nó é interrompida (SANTOS, 2011).

```

Algoritmo DMBC
{Entrada: conjunto de nós  $n$ , uma ordenação dos nós (tendo a classe como primeiro
nó), um limite  $u$  sobre o número de pais que um nó pode ter e uma base de dados  $D$ 
contendo  $m$  casos.}
{Saída: Para cada nó, a identificação dos pais do nó.}
1. CMB := {}; / Cobertura de Markov da Classe /
2. for  $i := 1$  to  $n$  do
3.    $\pi_i := \{\}$ ;
4.    $P_{old} := g(i, \pi_i)$ ; / equação (2.2) /
5.   OKToProceed := true;
6.   while OKToProceed and  $\pi_i < u$  do
7.     let  $z$  be the node in  $Pred(x_i) - \{\pi_i \cup CMB\}$  that
           maximizes  $g(i, \pi_i \cup \{z\})$ ;
8.      $P_{new} = g(i, \pi_i \cup \{z\})$ ;
9.     if  $P_{new} > P_{old}$  then
10.      if ( $z$  is parent of any class child) or
           ( $z$  is a class child) then
11.        CMB := CMB  $\cup$   $\{z\}$ ;
12.         $P_{old} := P_{new}$ ;
13.         $\pi_i := \pi_i \cup \{z\}$ ;
14.      end {if};
15.      if (not_class_child( $i$ ) and class  $\notin \pi_i$ ) then
16.        break_while();
17.      else OKToProceed := false;
18.    end {while};
19.    write('Node:',  $x_i$ , 'Parents of this node:',  $\pi_i$ )
20.  end {for};
21. end {DMBC};

```

Figura 9 – Pseudocódigo do algoritmo DMBC [Adaptado de (SANTOS, 2011)].

2.4.4.2 Algoritmo Naive Bayes

O classificador Naive Bayes (Bayesiano Ingênuo) baseia-se no Teorema de Bayes e está relacionado ao cálculo de probabilidade condicional (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Ele se baseia numa forte suposição de que todos os atributos previsores são independentes entre si, dado o valor do atributo classe; por isso, o termo “ingênuo” em “classificador Bayesiano Ingênuo”.

Naive Bayes pode ser visto como uma rede Bayesiana particular onde qualquer variável tem seus nós correspondentes conectados apenas ao nó representado pela variável classe.

A Figura 10 ilustra a estrutura do classificador Naive Bayes. Nesta estrutura, a classe C influencia diretamente os atributos $A_i (i = 1, \dots, n)$, mas nenhum atributo A_i exerce influência sobre o outro; são independentes entre si, dado a classe C .

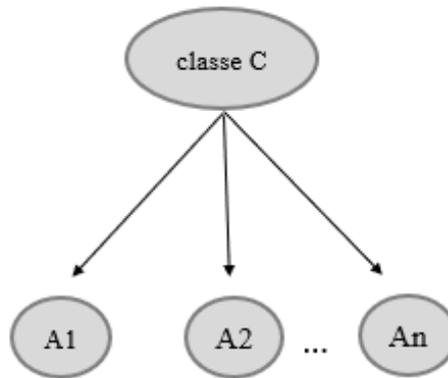


Figura 10 – Estrutura do classificador Naive Bayes.

Como classificador probabilístico, Naive Bayes calcula a probabilidade de uma classe c_j ($c_j \in C$), dados os atributos A_i , através da equação 2.3:

$$C_{NB} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_i P(A_i | c_j) \quad (2.3)$$

onde $P(A_i | c_j)$ é a probabilidade condicional dos atributos A_i , dado o valor c_j , e $P(c_j)$ é a probabilidade a priori de c_j . O resultado (C_{NB}) é a classe com o maior valor de probabilidade.

Naive Bayes tem fornecido bons resultados de classificação em vários domínios (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997). No entanto, suas estimativas de probabilidades não são realísticas e seu desempenho de classificação pode ser melhorado. Além disso, o modelo Naive Bayes induzido não pode capturar as relações reais entre as variáveis (SANTOS, 2011).

3 Trabalhos Relacionados

A busca por um sistema de aprendizado sem fim não é algo novo na comunidade de inteligência artificial. Entretanto, não há ainda uma metodologia bem definida para o desenvolvimento deste tipo de sistema, e nem tampouco um sistema capaz de apresentar todas as características essenciais para que possa aprender continuamente de maneira autônoma, autossupervisionada e possa expandir sua ontologia inicial automaticamente sempre que necessário.

([MITCHELL et al., 2018](#)) apresenta um estudo de caso do sistema NELL e, por meio de uma análise empírica, comprova que o NELL aprende com sucesso, usando sua competência para construir uma base de conhecimento de crenças sobre o mundo. De acordo com os experimentos, é sugerido quatro recursos de projeto: i) para atingir o aprendizado semi-supervisionado bem-sucedido, é preciso acoplar o treinamento de muitas tarefas de aprendizagem diferentes; ii) permitir que o agente aprenda restrições de acoplamento adicionais; iii) aprender novas representações que cobrem fenômenos relevantes além da representação inicial; iv) organizar um conjunto de tarefas de aprendizagem iniciando com um currículo fácil e tornando-o cada vez mais difícil. Estes recursos contribuem para o sucesso do aprendizado e pode ser adotado para qualquer sistema de aprendizagem sem fim. Porém foi constatado que o NELL também possui limitações. Por exemplo, o NELL apresenta uma baixa competência de monitorar seu próprio desempenho e progresso.

Em ([BARCHI, 2014](#)), é apresentado o subsistema newOntExt, como sendo uma abordagem atualizada com novos recursos para extensão da estrutura da base de conhecimento do NELL. Desde 2011, OntExt vinha sendo utilizado para a criação de novas relações e trabalhando para extensão da ontologia da base de conhecimento, e ainda, serviu de base para a nova abordagem do newOntExt. Este fornece uma nova implementação com novas características, recursos e estratégias para expandir a base de conhecimento do NELL, além de validar e nomear relações encontradas pelo Prophet ([APPEL; JUNIOR, 2011](#)), outro subsistema componente do NELL. Com os experimentos realizados, foi observado que o newOntExt (quando comparado com o OntExt que não gerou nenhuma relação correta para os mesmos subconjuntos de categorias) demonstrou de forma empírica que poderia trazer ganhos para o sistema de aprendizado sem-fim, encontrando relações corretamente. Constatou-se também que o subsistema newOntExt (assim como com o OntExt) é sensível a ruídos, logo o newOntExt passa a ser utilizado também como um alerta para correções na ontologia, colaborando com a autossupervisão e autorreflexão do sistema NELL.

(WIJAYA; MITCHELL, 2016) apresentam um método capaz de construir automaticamente um recurso verbal para qualquer linguagem (multilíngue), dado uma base de conhecimento e um conjunto de textos em um idioma específico (corpus de texto). A ideia seria mapear automaticamente frases verbais para relações da base de conhecimento, uma espécie de interlíngua. Este recurso seria muito útil para extrair fatos do texto para as bases de conhecimento, ajudando na integração de conhecimento em diferentes bases de conhecimento e idiomas, além de tarefas como leitura automática. Os autores demonstraram em seus experimentos que o recurso verbal é eficaz. Dado um corpus de texto e uma base de conhecimento em qualquer idioma, ele pode produzir um mapeamento das frases verbais da linguagem para as relações da base de conhecimento. Quando o recurso foi aplicado a uma base de conhecimento do NELL em português e um corpus de texto em português, o mesmo método construiu automaticamente um recurso de verbo em português capaz de extrair instâncias de relação do texto em português. Da mesma forma, quando aplicado a uma base de conhecimento do NELL em inglês e um corpus de texto em inglês, o mesmo método foi construído automaticamente e um recurso de verbo em inglês foi eficaz para extrair instâncias de relação do texto em inglês.

(MIANI; JUNIOR, 2015) propõem o uso de regra de associação (ou associação generalizada) para aumentar a base de conhecimento do NELL e as relações entre as categorias da base de conhecimento. Porém, algoritmos de regras de associação geram muitas regras, sendo uma tarefa árdua avaliar cada uma dessas regras. Logo, foi proposto e desenvolvido uma estrutura para facilitar o processo de avaliação das regras extraídas, podendo um conjunto de itens óbvios e regras de associação, diminuindo assim, a quantidade de regras descobertas. Também foi desenvolvido um novo parâmetro, denominado MSC (Cálculo de Suporte Modificado). Essa medida contribuiu para gerar regras significativas. Os autores, através dos experimentos realizados, comprovaram que, por medida MSC e regras de associação, foram descobertas regras importantes que ajudaram a preencher a base de conhecimento do NELL com instâncias. Além disso, houve um aumento das relações da base de conhecimento, usando regras de associação generalizadas.

O autor (DUARTE, 2016) propõem a criação de uma nova instância do NELL em português, uma vez que o sistema executa continuamente desde 2010 aprendendo em inglês. Então, foi apresentado a criação de uma base preprocessada (*all-pairs-data*), sendo este o primeiro passo para o sistema aprender e, posteriormente, a criação de uma abordagem híbrida para a resolução de correferências com o objetivo de melhorar a confiabilidade no aprendizado. O NELL em português, tendo um tamanho muito inferior em relação a base de conhecimento do NELL em inglês, não pode ser equiparado ao mesmo, pois a base de conhecimento não permitiu uma evolução do sistema por muito tempo. Porém, através dos experimentos realizados e evidências empíricas, o NELL desenvolvido e proposto em português obteve bons resultados.

Todos estes trabalhos mostram a viabilidade da construção de um sistema de aprendizado sem fim com base na metodologia proposta em (CARLSON et al., 2010). Mas, apesar de serem muito promissores, ainda não trouxeram contribuições definitivas em alguns pontos importantes do novo paradigma de aprendizado sem fim. Assim, é na busca por tais contribuições que este projeto se insere. Mais especificamente, pretende-se utilizar o formalismo das Redes Bayesianas para realizar inferências na base de conhecimento do NELL, bem como para identificar novas relações semânticas que possam ser inseridas na base de conhecimento.

4 Construção de um conjunto de Dados a partir do NELL

A base de conhecimento do NELL contém registros sobre vários domínios de conhecimento. Para este projeto, decidiu-se extrair dados apenas do domínio de doenças com o objetivo de avaliar um conhecimento específico obtido pelo NELL que pudesse contribuir para a construção de um sistema especialista futuramente. Para isto, algumas etapas de pré-processamento foram aplicadas sobre a base de conhecimento, utilizando a linguagem de programação Python e a biblioteca Pandas.

4.1 Etapas de construção do conjunto de dados

A base de conhecimento do sistema NELL está disponibilizada no site do projeto¹ e contém várias categorias e relações entre categorias. A base disponibilizada, com milhares de instâncias (2.766.048 na data atual), contém 13 colunas de dados sobre cada instância aprendida. Porém, apenas as três primeiras colunas foram utilizadas para a criação do conjunto de dados. Estas colunas apresentam dados relativos às relações entre categorias. Visto que um dos interesses deste trabalho é representar relações através de Redes Bayesianas, um conjunto de dados foi criado a partir de dados das relações existentes na base de conhecimento do NELL.

Os dados das relações estão armazenados em apenas três colunas da base de conhecimento do NELL: "Entity", "Relation" e "Value". Estas colunas armazenam os seguintes dados: coluna "Entity", armazena nomes de categorias; coluna "Relation", contém nomes das relações entre as categorias; e coluna "Value", contém valores aprendidos por NELL por meio destas relações. A Tabela 1 apresenta alguns exemplos de dados de relações. Por exemplo, a relação "inversefoodcausedisease" (na coluna "Relation") permitiu ao NELL aprender que *stress* (na coluna "Entity") é uma doença (*disease*) que pode ser causada pelo alimento (*food*) *sugar* (coluna "Value"). (Os dados originais apresentam os caracteres "concept:", que não são de interesse, e foram removidos nesta etapa de aquisição).

¹ <http://rtw.ml.cmu.edu/rtw/resources>

Entity	Relation	Value
Stress	Inverseoffoodcancausedisease	sugar
Diabetes	physiologicalconditionpossiblytreatedbydrug	actos
Depression	emotionassociatedwithdisease	anguish
Cancer	inverseoffooddecreasestheriskofdisease	pomegranate
Tuberculosis	inverseofanimaldevelopdisease	child

Tabela 1 – Exemplos de dados de relações armazenados na base de conhecimento do NELL.

Para a construção de conjuntos de dados a partir das relações da base de conhecimento do NELL, foram definidas as seguintes etapas:

Etapa 1:

A primeira etapa de construção do conjunto de dados consiste em obter as relações e seus dados a partir da base de conhecimento do NELL. Para isto, foram executados de forma automática alguns filtros utilizando comandos da linguagem Python (Biblioteca Pandas). O primeiro filtro selecionou todos os registros que tinham a palavra *disease*, tanto na coluna "Entity", quanto na coluna "Value". Desta forma, foi possível selecionar dados relacionados diretamente ao domínio de doenças. Foram retornados 9.397 registros. A partir desses registros retornados, aplicou-se outro filtro para selecionar apenas a coluna "Relation" e todos os nomes das relações de forma agrupada (sem repetir os nomes das relações). O resultado foi 61 relações obtidas. Essas relações foram analisadas para verificar se pertenciam ao domínio de doenças e partir delas foram selecionadas 10 relações para compor a base de dados. No entanto, apenas 10 relações não seriam suficientes para construir um conjunto de dados. Então, foi executado outro filtro a fim de obter mais relações. Desta vez, selecionou-se todos os registros que não tinham a palavra *disease* nas colunas "Entity" e/ou "Value" (inverso do primeiro filtro). O objetivo era encontrar relações que pertenciam indiretamente ao contexto de doenças. O resultado foi 635.497 registros. Posteriormente, foi executado outro filtro selecionando apenas a coluna "Relation" e todos os respectivos nomes das relações de forma agrupada existentes nesse conjunto de 635.497 registros. Foram retornadas 826 relações. Foi feita uma análise nessas relações e, das 826, apenas 6 relações estavam de acordo com o domínio proposto. Totalizando assim 16 relações, as quais são apresentadas na Tabela 2, com uma breve descrição sobre cada uma.

N	Relações	Descrição
01	agriculturalproductcontainchemical	química que compõe o produto
02	animaldevelopdisease	animal que desenvolve a doença
03	bodypartwithinbodypart	parte do corpo afetado pela doença
04	beveragecontainsprotein	proteína que a bebida contém
05	beveragecansoftendisease	bebida que ameniza a doença
06	bacteriaisthecausativeagentof physiologicalcondition	bactéria que causa doença
07	bacteriacancauseillnessfromfood	alimento onde bacteria pode se alojar
08	drugworkedonbyagent	fabricante da medicação ou droga
09	drughasideeffect	efeito causado pela droga que trata
10	diseasecausesphysiologicalcondition	condição fisiologica gerada pela doença
11	inverseofemotionassociatedwithdisease	emoção associada a doença
12	physiologicalconditionpossiblytreated bydrug	droga que trata a doença
13	inverseoffooddecreasestheriskofdisease	alimento que decresce o risco da doença
14	inverseoffoodcancausedisease	alimento que pode causar doença
15	physiologicalconditionstudedbyperson	cientista que estuda a doença
16	sideeffectcausedbydrug	doença que a droga causa

Tabela 2 – Relações selecionadas a partir da base de conhecimento do NELL e usadas para a construção do conjunto de dados.

Etapa 2:

A seguir, a coluna "Relation" de cada relação foi excluída e a coluna "Value" recebeu o nome da relação. Este processo foi realizado para cada relação.

Entity	Value
	Inverseoffooddecreasestheriskofdisease
heart_disease	blueberries
heart_disease	omega_3
heart_disease	seafood
heart_disease	tomatoes
heart_disease	oats
heart_disease	apples

Tabela 3 – Exemplo de instâncias (registros) para a relação "inverseoffooddecreasestheriskofdisease".

Etapa 3:

Cada relação com os dados das colunas "Entity" e "Value" foi armazenada em um arquivo do tipo '.csv' (uma tabela para cada relação e seus dados, como pode ser visto na

Tabela 3 para a relação “inverseoffooddecreasestheriskofdisease”).

Etapa 4:

Depois, todos os arquivos (".csv") referentes às relações foram concatenados num único arquivo do tipo ".csv". O arquivo resultante da concatenação apresentou 16 colunas, cada uma referente a uma relação, e mais a coluna "Entity" com os nomes das doenças, que foi renomeada para *disease*, totalizando 17 colunas.

Etapa 5:

O nome de cada coluna referente às relações foram alterados, formando os atributos do conjunto de dados (16 atributos e mais a classe *disease*), segundo a Tabela 4.

N	Relações	Atributos
1	Agriculturalproductcontainchemical	chemprod
2	Bacteriaisthecausativeagentofphysiologicalcondition	bacteria
3	Bacteriacancauseillnessfromfood	food
4	Beveragecontainsprotein	protein
5	Bodypartwithinbodypart	partbody
6	Drughassideeffect	effdrug
7	Drugworkedonbyagent	wagent
8	Inverseoffooddecreasestheriskofdisease	fooddec
9	Inverseoffoodcancausedisease	foodcan
10	Inverseofbeveragecansoftendisease	bevsoft
11	Inverseofdiseasecausesphysiologicalcondition	phycond
12	Inverseofemotionassociatedwithdisease	emotion
13	Inverseofanimaldevelopdisease	animal
14	physiologicalconditionpossiblytreatedbydrug	drug
15	Physiologicalconditionstudedbyperson	scientist
16	Sideeffectcausedbydrug	sidedrug

Tabela 4 – Definição dos nomes dos atributos da base, de acordo com as relações utilizadas.

Etapa 6:

O próximo passo foi acrescentar o atributo *symptomns* (sintomas) ao arquivo ".csv". As relações sobre o domínio de doenças encontradas no arquivo da base de conhecimento do NELL e utilizadas para a construção do conjunto de dados não trouxeram dados sobre sintomas. Visto que um atributo sobre os sintomas de doenças é importante para a representação deste domínio, decidiu-se incluí-lo no conjunto de dados como *symptomns* (sintomas), totalizando 18 atributos. No próprio site do projeto do NELL, existe uma

página com dados já validados sobre categorias e relações. Ao selecionar a opção "categories", pode-se pesquisar pelo nome da doença acompanhado da palavra "Symptoms"(por exemplo, digita-se "Diabetes Symptoms"para saber os sintomas da doença diabetes e a pesquisa retorna um link para acesso aos dados sobre os sintomas de diabetes).

Etapa 7:

Ao analisar o arquivo .csv, utilizando o software Weka, foram encontrados valores redundantes e ruidosos na base de dados, porém num grau pequeno. Esse problema foi resolvido por meio do método de exclusão de casos, basicamente para registros redundantes e alguns registros que estavam fora do domínio.

Etapa 8:

Vários atributos, no entanto, apresentaram valores ausentes após a coleta dos dados e transformação das relações. Muitas das relações semânticas existentes na base do NELL possuem poucas instâncias (valores). Desta forma, muitos atributos no conjunto de dados construído obtiveram poucas instâncias e assim tinham muitos valores ausentes. Os valores ausentes é um dos problemas mais comuns da base de conhecimento do NELL, pois o sistema não consegue aprender muito sobre algumas categorias e relações. Neste caso, as Redes Bayesianas podem ser usadas como uma ferramenta de inferência para tentar prever valores de variáveis que o NELL não foi capaz de aprender a partir da web. Na prática, nem todas as doenças possuíam todas as relações definidas para a construção do conjunto de dados, gerando assim registros com valores incompletos. Para resolver este problema, foi aplicada a técnica do cálculo da mediana, considerando todo o conjunto dos dados e permitindo substituir os valores ausentes pelo valor da mediana. O processo foi feito de forma automática, utilizando o método de atualização do atributo com dado ausente, através de comandos do Python.

Etapa 9:

Após a execução de todas as etapas anteriores, o conjunto de dados passa a apresentar 5.178 registros. Após uma análise dos dados, observou-se que havia determinados atributos com muitos valores diversificados e com pouca repetição. Por exemplo, o atributo *drug* possui diversos medicamentos que tratam a doença (*disease*) "diabetes", havendo pouca repetição de um mesmo medicamento. Por isso, resolveu-se triplicar a base, passando de 5.178 registros para 15.534 com o objetivo de gerar mais repetições para estes valores pouco frequentes, permitindo que os algoritmos de aprendizado de Redes Bayesianas pudessem encontrar padrões nos dados.

4.2 Informações sobre o conjunto de dados

O conjunto de dados construído sobre o domínio de doenças a partir de dados do NELL é composto por 15.534 instâncias (ou registros) e 18 atributos: 17 atributos previsores (incluindo o novo atributo *symptoms*) e o atributo classe *disease*. Todos os atributos têm valores categóricos, incluindo a classe, que têm 126 valores não muito desbalanceados. A Tabela 5 traz mais informações sobre os atributos e seus valores.

Atributos	Descrição	Quantidade
bacteria	bactéria que causa doença	137
food	alimento que a bactéria infecta	9
fooddec	decrece o risco de doença	148
bevsoft	bebida que suaviza a doença	15
protein	proteína que a bebida contém	19
chemprod	componente do alimento	68
foodcan	alimento que pode causar a doença	120
sidedrug	droga ou medicação que pode causar doença	89
drug	droga que trata doença	220
effdrug	efeito colateral da droga	149
wagent	fabricante da medicação	171
phycond	condição fisiologica causada pela doença	24
emotion	emoção associada a doença	77
animal	animal que desenvolve a doença	70
partbody	parte do corpo onde a doença atua	156
scientist	estudioso da doença	28
symptoms	Sintomas das doenças	425
disease	doenças	126

Tabela 5 – Informações sobre os atributos do conjunto de dados construído.

5 Redes Bayesianas a partir do NELL

Neste trabalho, é proposto utilizar as Redes Bayesianas para representar o conhecimento armazenado pelo sistema NELL e assim contribuir com seu aprendizado. Para isto, foi apresentada uma metodologia de construção de conjuntos de dados a partir das relações armazenadas na base de conhecimento do NELL. Os experimentos realizados aqui envolvem o conjunto de dados criado (apresentado no Capítulo 4) e algoritmos de aprendizado e inferência de Redes Bayesianas, buscando analisar dois objetivos:

- Primeiro, aplicar os algoritmos de aprendizado de Redes Bayesianas sob a base de dados construída utilizando o conhecimento do NELL de forma a representar relações semânticas e realizar inferências com precisão. Desta forma, as Redes Bayesianas induzidas podem identificar novas relações e inferir novos valores para serem inseridos na base de conhecimento do NELL. Para isso, foram aplicados dois algoritmos de aprendizado de Redes Bayesianas: i) K2, o qual induz Redes Bayesianas irrestritas e pode identificar mais relacionamentos entre as variáveis, e ii) DMBC, o qual induz classificadores Bayesianos com uma estrutura de rede mais simples, entretanto, exige menos esforço computacional;
- Segundo, pretende-se mostrar que o NELL pode fornecer conhecimento sobre um domínio específico para aprender Redes Bayesianas e, desta forma, auxiliar no desenvolvimento de sistemas especialistas. Para avaliar o conhecimento das Redes Bayesianas induzidas, são realizadas inferências sobre algumas variáveis para descobrir novos valores (ou fatos). Além disso, é realizada inferência sobre a variável classe para avaliar a taxa de acerto.

5.1 Rede Bayesiana induzida por DMBC

O algoritmo DMBC induz um classificador Bayesiano e assim constrói uma estrutura mais simples (com menos arestas e menos atributos), identificando apenas os relacionamentos que influenciam diretamente o atributo classe. Devido a isso, no entanto, o algoritmo tem um custo computacional reduzido. Ele precisa de dois parâmetros de entrada, além do conjunto de dados: i) a variável classe, definida aqui como sendo a variável *disease* e ii) o número máximo de pais para cada variável na estrutura da rede Bayesiana, o qual foi definido aqui como 10 (após testes realizados). A Figura 11 apresenta a rede Bayesiana induzida pelo DMBC a partir do conjunto de dados sobre doenças.

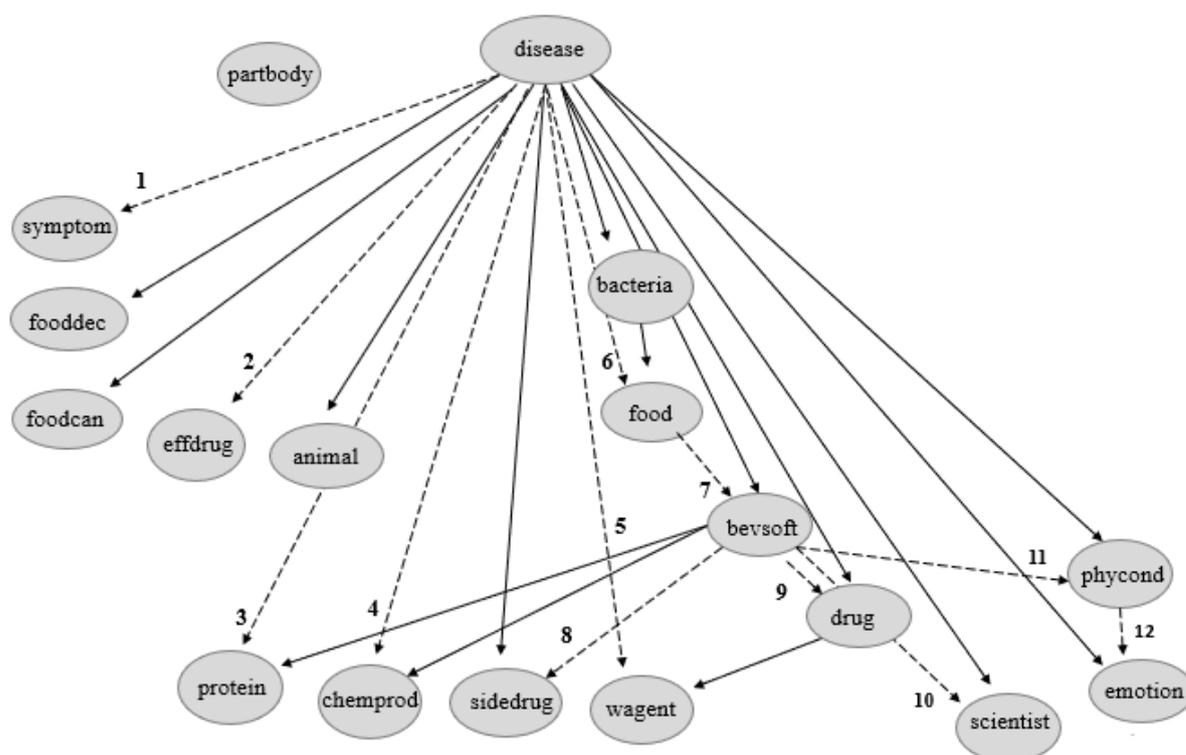


Figura 11 – Estrutura da rede Bayesiana induzida pelo algoritmo DMBC a partir do conjunto de dados sobre doenças construído com dados do NELL. As arestas contínuas representam as relações semânticas existentes na base do NELL e utilizadas na construção do conjunto de dados. As arestas tracejadas e numeradas representam novas relações sugeridas para expandir a ontologia inicial do NELL.

As arestas da estrutura da rede Bayesiana da Figura 11 representam relacionamentos entre os atributos (variáveis) e podem ser entendidas como representação de relações semânticas para a base de conhecimento do NELL. Observando as arestas na estrutura, é possível notar que o DMBC pode aprender especialmente as relações semânticas que envolvem o atributo classe (neste caso, *disease*).

Neste experimento, o DMBC foi capaz de representar 14 relações semânticas existentes na base do NELL (das 16 usadas para criar o conjunto de dados - veja a Tabela 2). Por exemplo, a aresta entre *disease* e *drug* representa a relação semântica “physiological-conditionpossiblytreatedbydrug”, a qual existe na base do NELL. Apenas duas relações existentes na base do NELL (e usadas para a construção do conjunto de dados) não foram identificadas pelo DMBC: uma é a “bodypartwithinbodypart” e a outra é a “drughasseideffect”. Nota-se, na Figura 11, que não foram formadas arestas entre os atributos *disease* e *partbody* e *effdrug* e *drug*. Há relações, como estas, com poucos valores aprendidos pelo NELL e assim elas geram atributos com muitos valores ausentes para o conjunto de dados. Então, nós acreditamos que nem todas as relações semânticas podem ser identificadas por DMBC se não houver um tratamento adequado destes valores ausentes. Esta estrutura demonstra que o DMBC é capaz de representar relações existentes a partir de seus dados.

Na estrutura da rede Bayesiana da Figura 11, também é possível observar que há outras arestas além daquelas que representam as relações semânticas existentes. Estas arestas estão tracejadas e numeradas na estrutura e sugerem novas relações encontradas por DMBC. São 12 novas relações. Por exemplo, a aresta entre *symptom* e *disease* sugere uma nova relação que não existe na base de conhecimento do NELL. Portanto, uma análise sobre as possíveis novas relações pode ser feita por especialistas para verificar a possibilidade de expandir a ontologia inicial do NELL. A Tabela 6 exibe as novas relações identificadas pelas arestas da estrutura da rede Bayesiana.

Arestas	Novas relações
01	disease -> symptom
02	disease -> effedrug
03	disease -> protein
04	disease -> chemprod
05	disease -> wagent
06	disease -> food
07	food -> bevsoft
08	bevsoft -> sidedrug
09	bevsoft -> drug
10	bevsoft -> scientist
11	bevsoft -> phycond
12	phycond -> emotion

Tabela 6 – Novas relações identificadas pelo DMBC.

5.2 Redes Bayesianas induzidas por K2

Diferente do algoritmo DMBC, K2 induz Redes Bayesianas irrestritas (não apenas para classificação). Por isso, é capaz de encontrar mais relacionamentos entre as variáveis. No entanto, estes relacionamentos são influenciados pela ordenação de variáveis, pois K2 induz a estrutura da rede Bayesiana considerando que apenas as variáveis à esquerda de uma determinada variável podem ser seus pais. Então, para os experimentos aqui, além do número de pais (definido como 10), também foram definidas diferentes ordenações em busca da rede Bayesiana ótima (ou mais próxima da ótima), que se adaptasse melhor à base de dados e conseguisse representar os relacionamentos mais corretos.

Foram geradas 17 ordenações de variáveis para a execução do K2. Dez ordenações foram definidas de forma aleatória e as outras sete através de métodos de seleção de atributos, fornecidos pelo software *Weka*. Os métodos de seleção de atributos avaliam a importância das variáveis em relação à classe e assim podem gerar boas ordenações para o K2. Estes métodos estão descritos na Tabela 7.

Métodos de seleção de atributos	Descrição
GainRatioAttributeEval	Avalia o valor de um atributo medindo a taxa de ganho em relação à classe.
InfoGainAttributeEval	Avalia o valor de um atributo medindo o ganho de informação em relação à classe.
OneRAttributeEval	Avalia o valor de um atributo usando o classificador OneR.
ReliefFAttributeEval	Avalia o valor de um atributo mostrando repetidamente uma instância e considerando o valor do atributo fornecido para a instância mais próxima da mesma classe diferente.
SymmetricalUncertAttributeEval	Avalia o valor de um atributo medindo a incerteza simétrica em relação à classe.
ClassifierAttributeEval	Avalia o valor de um atributo usando um classificador específico pelo usuário.
CorrelationAttributeEval	Avalia o valor de um atributo medindo a correlação entre ele e a classe.

Tabela 7 – Descrição dos métodos de seleção de atributos.

O algoritmo K2 foi executado com cada uma das 17 ordenações, induzindo 17 Redes Bayesianas diferentes, que foram comparadas pelo valor da função g (métrica aplicada pelo K2 para avaliar a qualidade da rede Bayesiana). A Tabela 8 exibe os valores da função g obtidos pelo K2 para cada rede Bayesiana induzida. A métrica *GainRation* apresenta o maior valor obtido e a "ordenação3" (ordenação aleatória) o menor valor.

Tipos de ordenação	Valor da função g
GainRation	-420175,65
ReliefFAttributeEval	-421574,86
ClassifierAttributeEval	-422893,57
OrdenaçãoAleatoria-5	-425531,13
SymmetricalUncert AttributeEval	-426327,62
InfoGain	-427503,38
OneRAttributeEval	-427503,38
OrdenaçãoAleatoria-10	-430480,82
OrdenaçãoAleatoria-8	-430986,36
OrdenaçãoAleatoria-9	-431215,24
OrdenaçãoAleatoria-1	-436026,49
OrdenaçãoAleatoria-7	-436157,14
Correlation AttributeEval	-436866,89
OrdenaçãoAleatoria-6	-437769,39
OrdenaçãoAleatoria-2	-439651,81
OrdenaçãoAleatoria-4	-440927,64
OrdenaçãoAleatoria-3	-450113,36

Tabela 8 – Valores da função g para as Redes Bayesianas induzidas pelo K2 com diferentes ordenações de variáveis. Os valores de g estão ordenados de forma decrescente.

A melhor rede Bayesiana ($valor\ g = -420175,65$) foi induzida a partir da ordenação de variáveis definida pelo método *GainRation*. A estrutura desta rede é exibida na Figura 12.

Ao observar a estrutura da rede nota-se que o K2 foi capaz de representar 14 relações semânticas existentes na base do NELL (das 16 usadas para criar o conjunto de dados - veja a Tabela 2). Estas relações estão representadas por arestas de linha contínua. Apenas duas relações não foram identificadas pelo K2: "bacteriacancauseillnessfromfood", "bodypartwithinbodypart". Ao observar a Figura 12, pode-se notar que não se formou uma aresta entre os seguintes atributos: *bacteria* e *food* (relation: bacteriacancauseillnessfromfood), e também, *disease* e *partbody* (relation: bodypartwithinbodypart).

O K2 também identificou 15 novas relações que estão representadas na estrutura da Figura 12 por linhas tracejadas e numeradas. Por exemplo, na linha tracejada de número 3, os atributos *scientist* e *fooddec* sugerem uma nova relação que não existe na base de conhecimento do NELL. Da mesma forma ocorre com os outros atributos interligados por linhas tracejadas. As novas relações são resumidas na Tabela 9.

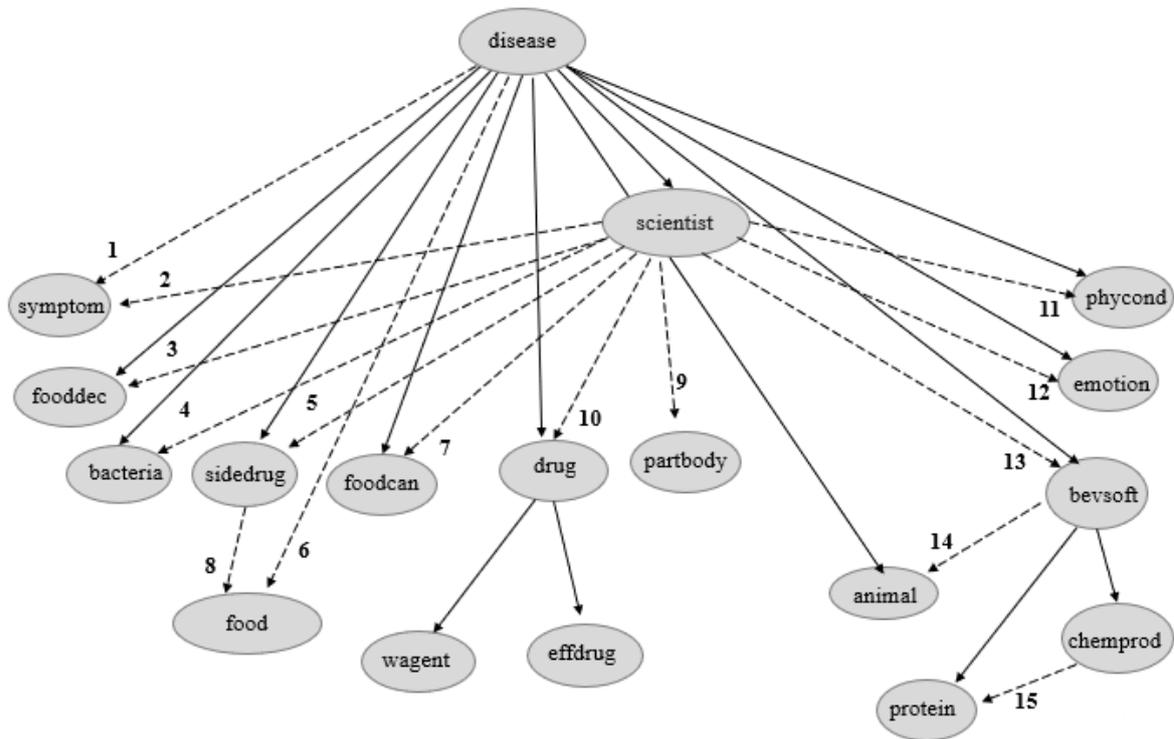


Figura 12 – Estrutura da melhor rede Bayesiana induzida pelo algoritmo K2 a partir do conjunto de dados sobre doenças construído com dados do NELL. As arestas contínuas representam as relações semânticas existentes na base do NELL e utilizadas na construção do conjunto de dados. As arestas tracejadas e numeradas representam novas relações sugeridas para expandir a ontologia inicial do NELL.

Arestas	Novas relações
1	disease -> symptom
2	scientist -> symptom
3	scientist -> fooddec
4	scientist -> bacteria
5	scientist -> sidedrug
6	disease -> food
7	scientist -> foodcan
8	sidedrug -> food
9	scientist -> partbody
10	scientist -> drug
11	scientist -> phycond
12	scientist -> emotion
13	scientist -> bevsoft
14	bevsoft -> animal
15	chemprod -> protein

Tabela 9 – Novas relações identificadas pelo K2.

Para efeito de comparação, a Figura 13 apresenta a estrutura da pior rede Bayesiana induzida pelo K2 (*valor $g = -450113.36$*), usando uma ordenação aleatória (veja tabela 8). Analisando as arestas da estrutura, é possível concluir que apenas 11 relações existentes na base de conhecimento do NELL foram representadas (do total de 16 usadas para construção da base de dados), enquanto a estrutura da melhor rede representou 14. Quanto ao número de novas relações, esta rede pior conseguiu encontrar 12, contra 15 da melhor. Isto demonstra como a ordenação de variáveis realmente influencia o resultado do algoritmo K2.

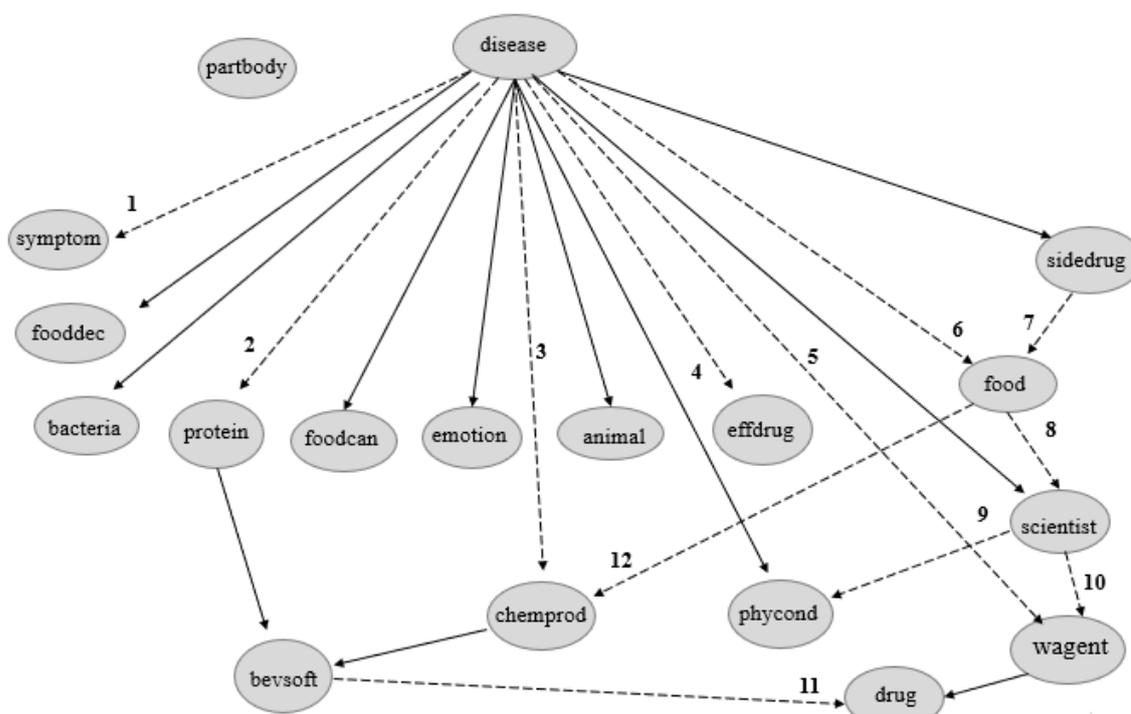


Figura 13 – Estrutura da pior rede Bayesiana induzida pelo algoritmo K2 a partir do conjunto de dados sobre doenças construído com dados do NELL. As arestas contínuas representam as relações semânticas existentes na base do NELL e utilizadas na construção do conjunto de dados. As arestas tracejadas e numeradas representam novas relações sugeridas para expandir a ontologia inicial do NELL.

5.3 Resultados de inferência Bayesiana

Após a indução das Redes Bayesianas pelos algoritmos DMBC e K2, foi aplicado um algoritmo de inferência Bayesiana (com base em *junction tree*) para avaliar o conhecimento armazenado nas redes. Foi avaliada a capacidade das Redes Bayesianas em descobrir novos fatos que possam ser inseridos na base de conhecimento do NELL, além de dados já existentes na base. As inferências foram realizadas através do *software* Nética¹.

¹ Nética é um *software* que usa as Redes Bayesianas para executar vários tipos de inferências. O *software* está disponível em <https://www.norsys.com/download.html>.

Os experimentos iniciais de inferência realizados neste trabalho avaliaram dois aspectos: se os valores inferidos nas Redes Bayesianas estavam corretos de acordo com a base de conhecimento do NELL, e se a inferência trouxe novos valores ou relações que não existem na base. Para isso, foram analisadas 20 doenças. Como variável de entrada foi usado apenas *disease* e como variáveis de saída: *sidedrug*, *drug*, *fooddec*, *foodcan*, *bevsoft*, *symptoms*.

Os arquivos (.xml) que armazenam as rede Bayesianas induzidas pelo DMBC e K2 (melhor rede) foram importados para o Nética, permitindo a visualização das estruturas gráficas, facilitando assim a escolha dos valores das variáveis de entrada e a análise do resultado das variáveis de saída (valor do atributo a ser inferido).

A Figura 14 apresenta um exemplo de inferência, na qual pode ser observado o valor inferido (*red wine*) da variável *bevsoft* a partir da escolha do valor *heart_disease* (doença do coração) para a variável *disease*. Neste exemplo, o interesse maior foi descobrir que tipo de bebida pode amenizar a ocorrência de uma doença.

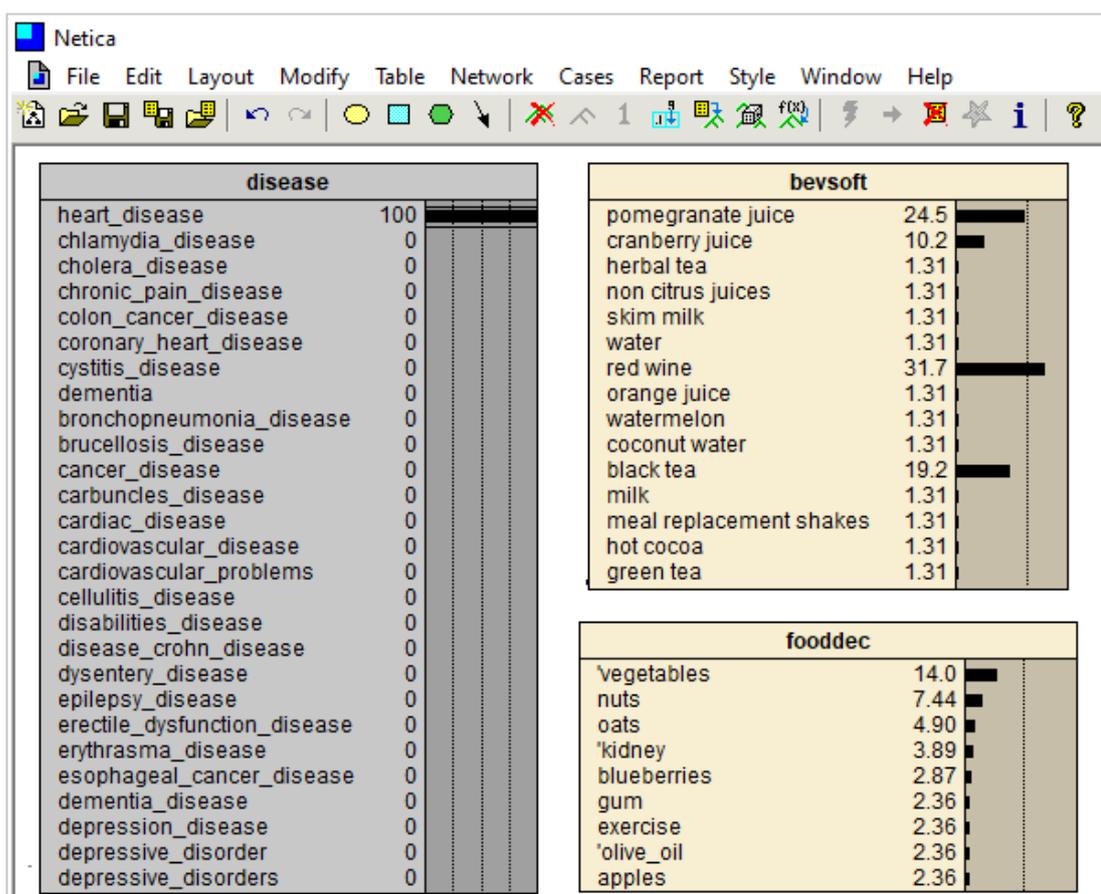


Figura 14 – Exemplo de inferência realizada no software Nética. Neste exemplo, o valor *heart_disease* (doença do coração) da variável *disease* foi selecionado e, depois da inferência, é observado o valor *red wine*, inferido para a variável *bevsoft*, como uma bebida que ameniza a doença.

A Tabela 10 exhibe os resultados de inferência obtidos a partir das Redes Bayesia-

nas induzidas pelo K2 e pelo DMBC. Nessa tabela, os valores inferidos para as variáveis de saída já existiam na base de conhecimento do NELL. Como variável de entrada, foi usado apenas valores da variável *disease* e como variáveis de saída: *sidedrug*, *drug*, *fooddec*, *foodcan*. Por exemplo, na linha 03, a doença *Heart_attack* (ataque cardíaco) apresenta 6,67% de probabilidade de ser causada pela droga *Tylenol*, segundo o K2, e 5,56%, de acordo com o DMBC.

Ao analisar a Tabela 10, é possível notar que, para as 20 instâncias mostradas, ambas Redes Bayesianas induzidas pelo K2 e pelo DMBC conseguiram inferir corretamente o valor da variável de saída. A rede Bayesiana induzida pelo K2 conseguiu acertar com um valor de probabilidade melhor que o DMBC na maioria dos casos. Esta taxa de acerto obtida era esperada, visto que estes valores já existiam na base de conhecimento do NELL. Isto demonstra também que as Redes Bayesianas induzidas são capazes de representar adequadamente este conhecimento aprendido pelo NELL.

N	Variável de entrada	Variável de saída	K2 (%)	DMBC (%)
01	Depression	sidedrug: Accutane	17,2	15,2
02	Bronchitis	sidedrug: Singulair	59,5	40,7
03	Heart_attack	sidedrug: Tylenol	6,67	5,56
04	Bladder_cancer	sidedrug: Actos	54,7	34,5
05	Anxiety	sidedrug: Remeron	86,8	79,7
06	Diabetes	drug: Avandia	14,5	8,39
07	Depression	drug: Ssris	3,99	3,66
08	Infection	drug: Levaquin	80,5	62,8
09	Allergy	drug: Claritin	34,7	29,6
10	Bipolar_disorder	drug: Seroquel	52,1	40,1
11	Breast_cancer	fooddec: mushrooms	4,19	4,70
12	Hypertension	fooddec: vegetables	5,29	6,15
13	Obesity	fooddec: vegetables	6,76	7,49
14	Cancer	fooddec: broccoli	5,66	1,13
15	Diabetes	fooddec: coffee	31,1	28,9
16	Depression	foodcan: sugar	41,2	45,2
17	Inflammation	foodcan: sugar	7,35	7,82
18	Cancer	foodcan: meats	2,60	2,68
19	Heart_disease	foodcan: junk_food	7,38	8,36
20	Blood_pressure	foodcan: chocolate	0,98	1,08

Tabela 10 – Valores inferidos (e presentes no NELL) a partir das Redes Bayesianas induzidas por K2 e DMBC.

A Tabela 11 exibe resultados de inferência interessantes para as variáveis de saída *bevsoft*, *sidedrug*, *foodcan* e *symptoms*. A partir dos valores da variável de entrada *disease*, foram inferidos valores que não existiam na base do NELL para estas variáveis de saída. Com base nisto, acredita-se que as Redes Bayesianas induzidas por K2 e DMBC são capazes de utilizar o conhecimento já aprendido para uma categoria para aprender outra. E assim, é possível identificar novos fatos que podem ser inseridos na base de NELL.

N	Variável de entrada	Variável de saída	K2	DMBC
01	Alzheimer	bevsoft: Pomegranate juice	36,9	32,0
02	Cardiovascular_problems	bevsoft: Cranberry	88,8	67,4
03	Heart_attack	bevsoft: Red wine	90,6	74,9
04	Atherosclerosis	bevsoft: Orange juice	82,6	59,4
05	Cancer	bevsoft: Pomegranate juice	49,3	46,1
06	Depression	sidedrug: Cipro	16,1	14,4
07	Fatigue	sidedrug: Neurontim	20,1	14,0
08	Asthma	sidedrug: Singulair	42,4	36,5
09	Colitis	sidedrug: Aldara	48,1	32,3
10	Digestive_disorder	sidedrug: Celexa	37,3	19,0
11	Insomnia	foodcan: drink	5,44	5,69
12	Anxiety_disease	foodcan: coffee	21,6	22,2
13	Chronic_pain	foodcan: junk_food	80,7	84,1
14	Schizophrenia	foodcan: glucose	70,9	75,7
15	Celulitis	foodcan: coffee	4,90	5,47
16	Asthma	symptoms: cough	5,39	6,55
17	Anxiety	symptoms: insomnia	8,56	8,84
18	Crohn_disease	symptoms: abdominal_pain	7,36	1,03
19	Fibromyalgia	symptoms: fatigue	3,36	3,83
20	Pneumonia	symptoms: fever	4,77	5,20

Tabela 11 – Valores inferidos (e não presentes no NELL) a partir das Redes Bayesianas induzidas por K2 e DMBC.

Por exemplo, na linha 01, a doença *Alzheimer* apresenta 36,9% de probabilidade de ser suavizada pela bebida *Pomegranate juice* (*suco de romã*) segundo o K2, e 32,0%, de acordo com o DMBC. Não consta na base de dados do NELL o *Pomegranate juice* (*suco de romã*) como sendo uma bebida que ameniza a doença *Alzheimer*. Porém, os algoritmos

DMBC e K2 aprenderam a partir dos valores existentes na base e conseguiram encontrar novos valores, no caso o *Pomegranate juice (suco de romã)*, assim como outros valores relacionados na Tabela 11.

5.3.1 Inferência da variável classe

Os algoritmos K2 e DMBC também foram comparados a outros dois classificadores do estado-da-arte, Naive Bayes e Árvore de Decisão² (NORVIG; RUSSELL, 2013), para avaliar a inferência da variável classe. Neste caso, todos os algoritmos foram treinados e testados usando a estratégia de treinamento *hold-out*, separando o conjunto de dados em 70% para treinamento e 30% para teste. Os resultados de classificação obtidos pelos algoritmos são exibidos na Tabela 12. Observando-se os resultados de inferência para a variável *disease* como classe, percebe-se que as Redes Bayesianas induzidas por K2 e DMBC também obtêm uma boa taxa de acerto. Os resultados obtidos por todos os algoritmos são muito próximos e superiores a 90%, o que indica que o conjunto de dados construído têm um conhecimento preciso.

Métrica	Árvore de Decisão (%)	Naive Bayes (%)	K2 (%)	DMBC (%)
acurácia	94,8	92,4	94,4	94,3
<i>recall</i>	94,8	92,5	94,3	93,6

Tabela 12 – Comparação dos resultados de inferência para a variável classe *disease*

² Árvore de Decisão são estruturas onde cada nó interno representa um teste em um atributo, cada ramo representa uma saída de cada teste e os nós folha representam as classes. Como vantagem, elas resultam em simples regras de classificação e permitem fácil interpretabilidade do modelo criado.

6 Conclusão e Trabalhos Futuros

Este trabalho permitiu avaliar o conhecimento adquirido por um sistema de aprendizado sem fim chamado NELL para induzir Redes Bayesianas, as quais podem ser aplicadas para ajudar o sistema NELL a estender e popular sua base de conhecimento e auxiliar no desenvolvimento de um sistema especialista para diagnóstico de doenças.

A metodologia apresentada para a construção de um conjunto de dados a partir das relações semânticas existentes na base de conhecimento do NELL mostrou-se promissora, fornecendo dados confiáveis sobre o domínio especificado. Desta forma, permitiu a aplicação dos algoritmos de aprendizado de Redes Bayesianas K2 e DMBC, os quais obtiveram bons resultados.

Algumas considerações sobre os algoritmos K2 e DMBC são interessantes de serem feitas. O K2 trabalha essencialmente por indução irrestrita (com foco em encontrar e formar novas arestas) e, através dos experimentos, foi possível concluir que o K2 atingiu esse propósito. E ao utilizá-lo também ficou evidente que a ordenação dos atributos interfere diretamente na construção de uma melhor ou pior rede Bayesiana. Já o algoritmo DMBC, que trabalha como um classificador Bayesiano, ao ser aplicado à base construída conseguiu induzir a rede Bayesiana com um desempenho computacional melhor que o K2, encontrando também 14 relações (das 16 já existentes) assim como o K2 (melhor rede) encontrou.

As estruturas de Redes Bayesianas induzidas pelos algoritmos K2 e DMBC representaram 87,5% das relações semânticas existentes no NELL, o que sugere que as novas relações identificadas podem ser usadas com segurança para estender a ontologia inicial do NELL. Isto reforça a hipótese de que as Redes Bayesianas são capazes de identificar relacionamentos entre os atributos que representam relações semânticas existentes e sugerir novas relações para o NELL. Além disso, estas Redes Bayesianas induzidas por ambos algoritmos fornecem bons resultados de inferência, sendo capazes também de deduzir novos fatos a partir do conhecimento existente no NELL, melhorando sua capacidade de aprendizado.

No entanto, o conhecimento armazenado pelo NELL ainda é incompleto e muito diversificado. Muitas categorias e relações têm poucos valores aprendidos e assim geram atributos com muitos valores ausentes. Outro problema é a geração de atributos com muitos valores diferentes e pouco frequentes. Estas dificuldades exigem a exploração de técnicas de pré-processamento para melhorar a qualidade do conjunto de dados construído.

Os algoritmos de aprendizado e inferência de Redes Bayesianas também sofrem com o aumento do conjunto de dados, exigindo mais memória e processamento. Assim,

à medida que o conjunto de dados cresce, os algoritmos exigem maior capacidade de processamento e mais memória disponível para melhor definição dos seus parâmetros.

6.1 Trabalhos futuros

Mesmo tendo vantagens sobre a utilização de teorias formais de primeira ordem (como a lógica de primeira ordem, por exemplo) a utilização de Redes Bayesianas (que não possuem um formalismo de primeira ordem) pode trazer algumas restrições (como por exemplo, a necessidade de adaptação e filtragem dos dados a serem utilizados para o aprendizado). Desta forma, pretende-se investigar e propor formas de utilização do formalismo de primeira ordem no algoritmo DMBC (que tende a exigir um esforço computacional bem menor que o exigido por algoritmos tradicionais), ao invés de se trabalhar com algoritmos tradicionais.

Deseja-se construir também um sistema especialista a partir da metodologia utilizada neste trabalho. O objetivo é a criação de uma arquitetura de rede Bayesiana a partir de dados do NELL para diagnosticar tipos diferentes de câncer, utilizando algoritmos de inferência Bayesiana. O propósito é evidenciar a eficiência do uso de Redes Bayesianas no diagnóstico desses tipos de doenças.

Também é de interesse ampliar o conjunto de dados de doenças, construindo um novo conjunto, desta vez, sem um domínio específico, mas incluindo todos domínios aprendidos pelo NELL. O objetivo é identificar novas relações entre domínios de conhecimento diferentes através da aplicação dos algoritmos K2 e DMBC, e fatos novos e existentes, através de algoritmos de inferência, buscando definir uma porcentagem de novas relações e fatos encontrados. Desta forma, será possível avaliar um conjunto de dados com uma quantidade maior de variáveis, o que proporcionará uma análise comparativa mais robusta dos algoritmos aplicados, verificando a possibilidade de relacionamentos entre variáveis de diferentes domínios.

6.2 Contribuições do projeto

Este projeto de pesquisa contribuiu para o aprimoramento do aprendizado do sistema NELL, realizando um estudo sobre uma forma de fornecer novas relações semânticas para sua ontologia e novas instâncias para sua base de conhecimento através das Redes Bayesianas. A metodologia desenvolvida mostrou-se promissora e, desta forma, após ser aperfeiçoada, pode se tornar mais um módulo de aprendizado a ser incorporado aos subsistemas do NELL.

A partir dos resultados obtidos, um artigo científico foi produzido e submetido ao periódico IEEE Latin America, o qual aguarda o resultado das avaliações dos revisores.

Referências

- APPEL, A. P.; JUNIOR, E. R. H. Prophet – a link-predictor to learn new rules on nell. In: *2011 IEEE 11th International Conference on Data Mining Workshops*. [S.l.: s.n.], 2011. p. 917–924. Citado 2 vezes nas páginas 13 e 31.
- BARCHI, P. H. *Expansão de ontologia através de leitura de máquina continua*. Tese (dissertation) — Universidade Federal de São Carlos: UFSCar, 2014. Citado 3 vezes nas páginas 13, 21 e 31.
- CARLSON, A. et al. Toward an architecture for never-ending language learning. In: *Proc. of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. [S.l.: s.n.], 2010. Citado 4 vezes nas páginas 12, 13, 18 e 33.
- CASTILLO, E.; GUTIERREZ, J. M.; HADI, A. S. *Expert systems and probabilistic network models*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 23.
- CLARK, P. et al. Reading to learn: An investigation into language understanding. In: *AAAI Spring Symposium: Machine Reading*. [S.l.: s.n.], 2007. p. 29–35. Citado na página 18.
- COOPER, G.; HERSKOVITS, E. A bayesian method for the induction of probabilistic networks from data. *Machine learning*, Springer, v. 9, n. 4, p. 309–347, 1992. Citado 4 vezes nas páginas 8, 14, 24 e 26.
- DUARTE, M. C. *Leitura da web em português em ambiente de aprendizado sem-fim*. Tese (dissertation) — Universidade Federal de São Carlos: UFSCar, 2016. Citado 5 vezes nas páginas 8, 18, 19, 20 e 32.
- FACELI, K. et al. *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC, 2017. Citado 4 vezes nas páginas 12, 16, 17 e 18.
- FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian network classifiers. *Machine Learning*, Kluwer Academic Publishers, Boston, v. 29, p. 131–163, 1997. Citado na página 30.
- GARDNER, M. et al. Improving learning and inference in a large knowledge-base using latent syntactic cues. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. [S.l.: s.n.], 2013. p. 833–838. Citado na página 20.
- GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. *Data Mining*. Rio de Janeiro: Elsevier, 2015. Citado 4 vezes nas páginas 12, 17, 18 e 29.
- HSU, W. H. Genetic wrappers for feature selection in decision tree induction and variable ordering in bayesian network structure learning. *Inf. Sci.*, v. 163, n. 1-3, p. 103–122, 2004. Disponível em: <<https://doi.org/10.1016/j.ins.2003.03.019>>. Citado na página 24.
- KUROSE, J. F. et al. *Redes de Computadores e a Internet: uma abordagem top-down*. São Paulo: Person Addison Wesley, 2006. (3). Citado na página 12.

- MANSANO, A. F. et al. *Aprendizado de máquina multivisão aplicado à análise de correferência em um sistema de aprendizado sem fim*. Tese (dissertation) — Universidade Federal de São Carlos: UFSCar, 2018. Citado 2 vezes nas páginas 13 e 20.
- MARQUES, R. L.; DUTRA, I. Redes bayesianas: o que são, para que servem, algoritmos e exemplos de aplicações. *Coppe Sistemas—Universidade Federal do Rio de Janeiro, RJ, Brasil*, 2002. Citado 2 vezes nas páginas 21 e 23.
- MIANI, R. G. L.; JUNIOR, E. H. Exploring association rules in a large growing knowledge base. *International Journal of Computer Information Systems and Industrial Management Applications*, v. 7, p. 106–114, 2015. Citado 2 vezes nas páginas 18 e 32.
- MITCHELL, T. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN 978-0-07-042807-2. Citado 4 vezes nas páginas 8, 12, 21 e 22.
- MITCHELL, T. et al. Never-ending learning. *Communications of the ACM*, ACM New York, NY, USA, v. 61, n. 5, p. 103–115, 2018. Citado 5 vezes nas páginas 8, 13, 19, 20 e 31.
- MOHAMED, T.; HRUSCHKA, E.; MITCHELL, T. Discovering relations between noun categories. In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. [S.l.: s.n.], 2011. p. 1447–1455. Citado na página 13.
- NETO, P. L. de O. C.; CYMBALISTA, M. *Probabilidades*. [S.l.]: Editora Blucher, 2006. Citado na página 24.
- NORVIG, P.; RUSSELL, S. *Inteligência Artificial*. 3ª. ed. Rio de Janeiro: Elsevier, 2013. ISBN 978-85-352-3701-6. Citado na página 50.
- REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. Barueri-SP: Editora Manole Ltda, 2003. Citado 3 vezes nas páginas 8, 16 e 17.
- SANTOS, E. B. dos. *A ordenação das variáveis no processo de otimização de classificadores bayesianos: Uma abordagem evolutiva*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de São Carlos, São Carlos, SP, 2007. Citado na página 24.
- SANTOS, E. B. dos. *Aprendizado Indutivo de Redes Bayesianas: além da precisão na tarefa de classificação*. Tese (Doutor em Engenharia Civil) — Universidade Federal do Rio de Janeiro - Coope/UFRJ, 2011. Citado 10 vezes nas páginas 8, 22, 23, 24, 25, 26, 27, 28, 29 e 30.
- SANTOS, E. B. dos et al. Bayesian network classifiers: Beyond classification accuracy. *Intell. Data Anal.*, v. 15, n. 3, p. 279–298, 2011. Disponível em: <[https://doi.org/10-3233/IDA-2010-0468](https://doi.org/10.3233/IDA-2010-0468)>. Citado na página 14.
- SILVA, F. N. d. *Modelagem de Emoções Utilizando Redes Bayesianas*. Dissertação (Mestrado em Modelagem Computacional) — Universidade Federal do Rio Grande, Rio Grande: RS, 2014. Citado na página 21.
- VERMA, T.; PEARL, J. Equivalence and synthesis of causal models. In: BONISSONE, P. P. et al. (Ed.). *UAI*. Elsevier, 1990. p. 255–270. ISBN 0-444-89264-8. Disponível em: <<http://dblp.uni-trier.de/db/conf/uai/uai1990.html#VermaP90>>. Citado na página 24.

WIJAYA, D. T.; MITCHELL, T. 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*. [S.l.: s.n.], 2016. p. 818–827. Citado na página [32](#).