

Aprendizado Automático de Sum-Product Networks

Relatório final - MAC0215 (Atividade Curricular em Pesquisa)
Aluno: Renato Lui Geh (Bacharelado em Ciência da Computação)
Orientador: Denis Deratani Mauá

1 INTRODUÇÃO

O intuito deste projeto de Iniciação Científica, como visto no relatório inicial do projeto [4], era aprender a estrutura e propriedades de uma Sum-Product Network, um Modelo Gráfico Probabilístico; entender a motivação por trás do estudo de tais redes probabilísticas, como Redes Bayesianas; adquirir toda a base matemática por trás destas redes e finalmente conseguir entender como aplicar aprendizado e inferência em Sum-Product Networks.

Com o tempo disposto neste semestre, conseguiu-se estudar grande parte da área teórica por trás de Sum-Product Networks, lendo-se a base matemática por trás de Redes Bayesianas e Sum-Product Networks – como será mostrado nas seções seguintes – e entendendo conceitos de Teoria de Grafos e probabilidade lendo-se tanto Darwiche[2] quanto estudando-se a matéria dada em MAC0425 Inteligência Artificial; estudou-se definições, teoremas e propriedades de Sum-Product Networks em dois diferentes artigos[11][10]; também foi estudado métodos de aprendizagem, como Regressão Linear, Naive Bayes, Nearest Neighbour e Decision Trees tanto por estudo próprio quanto pelas aulas dadas em MAC0425; e finalmente foi estudado um método de aprendizado em Sum-Product Networks em [11].

Como a maioria dos estudos feitos para este projeto são de caráter teórico, as evidências de estudo mais concretas são os relatórios publicados semanalmente. Na seção seguinte detalhar-se-á os estudos citados nesta seção.

2 HISTÓRICO

Para cada subseção nesta seção existe um relatório associado explicitando o trabalho em maiores detalhes. Cada subseção equivale a uma semana. Cada semana possui um relatório. O intuito de se fazer um relatório para cada semana é manter um resumo de toda a matéria vista na semana, além de servir como “notas de aula” sobre o que foi lido. Como o que é transcrito para os relatórios é um resumo da matéria, este também serve como um jeito de servir como referência posteriormente. Adicionalmente, tais relatórios também servem como uma forma gradual e didática de se aprender Sum-Product Networks, simulando os meus “passos”.

Cada subseção está organizada em diferentes blocos enumerados a seguir:

- Breve resumo da semana/relatório.
- Enumeração dos assuntos vistos.
- Resumo dos assuntos.

A breve descrição da semana/relatório consiste apenas na explicação do que foi feito na semana e eventualmente resumido no relatório. A discussão detalhada dos assuntos entrará em maiores detalhes sobre o estudo, explicando o que foi feito de concreto na semana. O resumo do relatório é mais técnico e entra em mais detalhes na matéria.

2.1 APRENDIZADO AUTOMÁTICO DE SUM-PRODUCT NETWORKS (SPN)

Nesta semana foi feita a definição do projeto (relatório inicial).[4]

Durante essa semana pesquisou-se o que Sum-Product Networks de fato são, foi estudada a base mínima de probabilidade necessária e foram lidas as aplicações de Sum-Product Networks. Os seguintes tópicos foram estudados:

- Distribuições de probabilidade multivariadas.
- Escopo de uma distribuição.
- Espaço de uma distribuição multivariada.
- Inferência.
- Definição de uma Sum-Product Network.
- Vantagens de Sum-Product Networks.
- Aplicações de Sum-Product Networks.

O relatório associado a essa semana que contém os tópicos citados acima é *Aprendizado Automático de Sum-Product Networks (SPN)*[4].

Esta semana foi importante para ter a base matemática para entender Sum-Product Networks. SPNs são redes probabilísticas que podem ser representadas como um digrafo acíclico enraizado com folhas representando probabilidades. Como cada folha é uma distribuição monovariável, então uma SPN que tenha mais de uma folha e tenha ao menos um nó produto é uma distribuição de probabilidade multivariada. Isso ocorre por causa da definição de consistência de uma SPN, que pode ser lida em [9]. Além disso este relatório representa os primeiros contatos com SPNs, estudando-se definição (superficialmente), vantagens de SPNs sobre outras redes e aplicações mais práticas.

2.2 MODELING AND REASONING WITH BAYESIAN NETWORKS: COMPILING BAYESIAN NETWORKS

Esta semana foi a primeira das muitas em que foi lido o livro *Modeling and Reasoning with Bayesian Networks*[2]. Nesta semana particularmente foi lido o capítulo 12 “Compiling Bayesian Networks”.

Podemos enumerar os tópicos estudados nesta semana da seguinte forma:

- Redes Bayesianas.
- CPT, MPE e MAP.
- Consistência de variáveis dada uma evidência.
- Network polynomial.
- Circuit propagation e derivadas parciais.

O relatório associado a essa semana que contém os tópicos citados acima é *Modeling and Reasoning with Bayesian Networks: Compiling Bayesian Networks*[5].

O capítulo “Compiling Bayesian Networks” pode ser visto como um precursor para a criação de Sum-Product Networks. Neste capítulo Darwiche propõe um método de se representar uma Rede Bayesiana como um digrafo acíclico enraizado onde as folhas são parâmetros ou indicadores e os nós internos são somas ou produtos de tal forma que o resultado deste circuito aritmético na raiz é a network polynomial da distribuição.

Poon e Domingos expandem esta ideia em seu artigo *Sum-Product Networks: A New Deep Architecture*[11] com mais formalização e com a diferença das folhas, ao invés de serem ou parâmetros ou indicadores, serem distribuições monovariáveis com os pesos nas arestas originárias de nós somas.

Também foi estudado o que significa uma variável da rede estar consistente com um conjunto de indicadores. Esse conjunto de indicadores pode ser chamado de evidência, e representa o conhecimento *a priori* de mundo. O conceito de consistência de uma variável dada uma evidência é muito importante para SPNs.

Neste capítulo temos o primeiro contato com network polynomials, um conceito fundamental para o entendimento de Sum-Product Networks. Além disso, foram estudados algumas nomenclaturas usadas em probabilidade, como CPT, MPE e MAP (seus significados podem ser vistos no relatório). Além disso foi estudado Redes Bayesianas superficialmente para podermos entender como compila-las em circuitos aritméticos.

Outro ponto fundamental foi o estudo de propagação do circuito. Propagação do circuito significa usar as derivadas parciais das variáveis de tal forma que não precisemos recalcular os parâmetros do circuito aritmético se mudarmos a evidência. No final do relatório também analisamos a complexidade dos algoritmos de *bottom-up* (calcular inferência) e *top-down* (calcular as derivadas parciais).

2.3 MODELING AND REASONING WITH BAYESIAN NETWORKS: INFERENCE BY VARIABLE ELIMINATION 6.1-6.5

Nesta semana foi lida a primeira parte do capítulo 6 do livro *Modeling and Reasoning with Bayesian Networks*[2]. A primeira parte consiste nas subseções 1-5.

Os tópicos 6.1-6.5 dizem respeito aos seguintes assuntos:

- Factors.
- Soma e multiplicação de factors.
- Eliminação de variáveis.
- Ordem de eliminação.

O relatório associado a essa semana que contém os tópicos citados acima é *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.1-6.5*[6].

Eliminação de variáveis é um método por estrutura de simplificar uma rede para torna-la mais tratável. Ao contrário da estrutura local (que será visto mais a frente), métodos por estrutura não usam os parâmetros dados pela query na hora da inferência. Isso as torna mais gerais, mas não possuem melhor performance quando comparadas a métodos que usam estrutura local.

Factors podem ser vistos como funções que mapeiam cada instância das variáveis em um número não-negativo. Também chamados de potenciais, factors servem para representar um jeito mais compacto de se fazer operações em CPTs.

Soma (*summing out*) de uma variável X é uma operação para se tentar eliminar a referência de uma variável em uma CPT. Multiplicação pode ser visto como a operação de união de variáveis.

Eliminação de variáveis consiste no uso de multiplicação e soma de variáveis para se descartar uma variável da Rede Bayesiana. Foram vistos algumas propriedades, algoritmos e a definição de se eliminar variáveis.

A ordem da eliminação de cada variável influencia na complexidade do algoritmo. Quer-se sempre eliminar variáveis em uma ordem menos custosa. Foi visto um algoritmo bem simples mas intuitivo de como achar a menor ordem de eliminação, mas este algoritmo era custoso, já que precisava calcular cada ordem individualmente.

2.4 MODELING AND REASONING WITH BAYESIAN NETWORKS: INFERENCE BY VARIABLE ELIMINATION 6.6-6.9

Nesta semana foi lida a segunda parte do capítulo 6 do livro *Modeling and Reasoning with Bayesian Networks*[2]. A segunda parte consiste nas subseções 6-9.

Os tópicos 6.6-6.9 dizem respeito aos seguintes assuntos:

- Escolhendo uma ordem de eliminação.
- Computando marginais posteriores.
- Estrutura e complexidade da rede.
- Estrutura e complexidade da query.

O relatório associado a essa semana que contém os tópicos citados acima é *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.6-6.9*[7].

Além do que foi lido no livro, também foram pesquisados os seguintes assuntos:

- Subgrafos, subgrafos induzidos, cliques, spanning subgrafos.
- Tree networks, polytrees e multiply connected.
- P, NP, NP-dificuldade e NP-completude.

Conforme os tópicos 6.6-6.9 foram lidos, assuntos que fossem dependências do que o autor citava foram pesquisados e adicionados no relatório. Portanto, os itens da segunda lista dessa seção estão presentes no relatório, com explicações e definições.

Primeiro foram vistos dois algoritmos não tão triviais para se achar uma melhor ordem de eliminação. Os dois algoritmos são similares, suas únicas diferenças sendo a heurística usada. A primeira tenta construir uma ordem que tenha menor fator possível. A segunda tenta adicionar o menor número possível de arestas no grafo de interação.

Para entender o grafo de interação, foi preciso primeiro pesquisar sobre subgrafos, subgrafos induzidos, cliques e spanning subgrafos. Um grafo de interação tem seus nós como variáveis que aparecem nos factors. Existe uma aresta entre dois nós se as duas variáveis aparecem no mesmo factor. Pode-se ver este grafo como um clique, tornando o problema de se achar uma ordem um problema de busca e eliminação de cliques. No entanto, achar uma boa ordem é NP-difícil.

Para se entender o que significa NP-difícil, foi pesquisado P, NP, NP-dificuldade e NP-completude. Este estudo foi adicionado no apêndice do relatório por ser mais geral que no caso do estudo sobre subgrafos.

O próximo tópico do livro era como se computar marginais posteriores. Foi então pesquisado o que significavam marginais posteriores e marginais conjuntas. Para isso foi usado o próprio livro do Darwiche[2] que definia-as. Em seguida foi visto como usar eliminação de variáveis para computar marginais conjuntas. Também foram vistas várias notações.

O seguinte assunto era sobre estrutura e complexidade da rede. A complexidade de duas Redes Bayesianas não se restringe apenas ao número de variáveis, mas também a sua estrutura. Um mesmo número de variáveis pode ter ordens ótimas de eliminação com diferentes complexidades. Isso se deve ao fator de treewidth da rede.

Para se entender treewidth foi preciso estudar tree networks, polytrees e multiply connected. O livro menciona de forma superficial o que elas são, sendo suas definições postas no relatório.

Em seguida foi visto um jeito de se tentar melhorar a estrutura de uma rede para simplificar a complexidade. Foi lido sobre como se efetuar um corte em uma árvore, o que inclui aparar tanto nós quanto arestas da Rede Bayesiana.

2.5 MODELING AND REASONING WITH BAYESIAN NETWORKS: INFERENCE WITH LOCAL STRUCTURE 13.1-13.3

Nesta semana foram lidos os tópicos 1-3 do capítulo 13 do livro *Modeling and Reasoning with Bayesian Networks*[2].

Os seguintes assuntos foram estudados:

- Comparação entre estrutura local e baseado em estrutura.
- Impacto de estrutura local na complexidade da inferência.
- Independência contexto-específica.
- Determinismo.
- Evidência em estrutura local.

O relatório associado a essa semana que contém os tópicos citados acima é *Modeling and Reasoning with Bayesian Networks: Inference with Local Structure 13.1-13.3*[8].

Estrutura local é o uso dos parâmetros da rede para melhorarmos a complexidade da inferência. Em baseado em estrutura, inferência de uma rede é limitado pela treewidth da rede. No entanto, com estrutura local, podemos computar em tempo tratável uma rede com treewidth intratável.

Foram lidos alguns exemplos de uso de estrutura local para deixar a rede mais simples. Uma delas é com o uso de restrições lógicas. Se um nó pai tem uma relação de ou-lógico com seus filhos, então podemos simplificar a rede fatorando a network polynomial de tal forma que tenhamos um circuit aritmético tratável, mesmo que o número de filhos seja exponencial.

Em seguida foi lido sobre independência contexto-específica. A relação entre nós de tal forma que podemos calcular a probabilidade de uma variável independentemente do valor dos outros nós se chama independência contexto-específica.

O próximo assunto tratava sobre determinismo de uma rede, que é uma relação de estrutura local onde os parâmetros da rede tem valor zero.

Finalmente, foi visto sobre como usar evidência como um jeito de se simplificar a rede. Usar a evidência de uma rede é um jeito de se fazer estrutura local.

2.6 UMA INTRODUÇÃO A SUM-PRODUCT NETWORKS

Nesta semana dediquei-me a escrita de um relatório onde, baseado em dois papers (Domingos, Poon [11] e Domingos, Gens [10]), pretende introduzir Sum-Product Networks de forma que, lendo-se todos os relatórios anteriores, consiga entender o que são Sum-Product Networks. A ênfase desta semana foi na definição e nas propriedades mais importantes de uma Sum-Product Network. Posteriormente planeja-se completar o relatório com os outros assuntos mencionados nesses e em outros papers.

Os tópicos incluídos neste relatório foram:

- Distribuição normalizada de produtos de factors.
- Função de partição.
- Definições de Sum-Product Networks.
- Validade, completude e consistência de uma Sum-Product Network.

O relatório associado a essa semana que contém os tópicos citados acima é *Uma Introdução a Sum-Product Networks*[9].

Foram lidos os papers *Sum-Product Networks: A New Deep Architecture*[11] de P. Domingos e H. Poon e *Learning the Structure of Sum-Product Networks* [10] de R. Gens e P. Domingos. Em particular, foi lida a introdução, definição e em seguida propriedades de Sum-Product Networks dos dois papers. Em Domingos e Poon a definição de SPNs é mais geral, enquanto que Gens e Domingos assumem uma SPN que seja sempre válida. No relatório foi escrita primeiro a definição em Domingos e Poon e, após termos uma noção do que são SPNs, a de Gens e Domingos.

Antes de entrarmos em detalhes sobre SPNs, precisou-se estudar o que são distribuições normalizadas de produtos de factors e funções partição. Foi lida a introdução contida em Poon e Domingos como referência. Após estudarmos a base, foram escritas as duas definições de Sum-Product Networks.

Em seguida foi lido sobre as três principais propriedades em SPNs: validade, completude e consistência. Uma SPN válida é importante pois computa a probabilidade de uma evidência em tempo linear e de forma correta. Uma SPN inválida sempre tem valor aproximado.

Finalmente, leu-se sobre o primeiro teorema em Sum-Product Networks, que dita que uma SPN é válida se ela é completa e consistente. Em seguida foi feita a prova deste teorema usando o paper de Poon e Domingos como ponto de partida.

2.7 APRENDIZADO AUTOMÁTICO DE SUM-PRODUCT NETWORKS (PÔSTER)

Nesta semana foi feito o pôster para MAC0215.

- Motivações.
- Definições.
- Propriedades.
- Aprendizado.
- Algoritmo de Aprendizado.
- Experimentos.

O pôster associado aos tópicos acima é *Aprendizado Automático de Sum-Product Networks* [3].

Para fazer o pôster releu-se os dois artigos [11] e [10]. Também foi feita uma breve pesquisa na motivação dos criadores de se criar Sum-Product Networks. Em seguida foi relida as definições dadas pelos dois artigos e transcrita de maneira mais sucinta possível para o pôster. Tentou-se manter as propriedades claras mas ao mesmo tempo pequenas e fáceis de ler e em seguida foi dada uma explicação da razão pela qual deseja-se validade em uma SPN.

Até este ponto do projeto não havia sido lido nada sobre aprendizado em Sum-Product Networks. Antes de entrar neste ponto era preciso saber toda a base que foi visto nos relatórios anteriores. Para fazer o pôster, no entanto, era preciso ter a ideia de aprendizado de SPNs em mente, e portanto foi lida a parte sobre aprendizado em Poon e Domingos [11] e também em Gens e Domingos [10].

Foi feita uma divisão entre dois tipos de aprendizado em SPNs. A primeira refere-se ao aprendizado dos pesos de uma SPN já existente. Já na segunda aprende-se tanto a estrutura quanto os parâmetros da SPN. O primeiro algoritmo de aprendizado em SPNs pertence a primeira classe e foi dada em [11]. Gens e Domingos criaram o primeiro algoritmo da segunda classe em [10].

Decidiu-se estudar mais profundamente o algoritmo dado por Poon e Domingos por ser mais simples e também por ter experimentos mais concretos. Foi lido sobre Gradient Descent e Expectation-Maximization, duas técnicas de otimização usadas para atualizar e inferir os pesos da SPN durante o aprendizado. Em seguida foi lido sobre como o algoritmo funcionava.

Para ter resultados mais concretos para o pôster, foram simulados os experimentos citados em [11]. As imagens sobre compleção de imagem e a tabela de taxa de acerto em reconhecimento de imagens no poster fazem parte do conjunto de experimentos feitos no artigo.

Finalmente, foi discutido com o orientador sobre quais seriam os futuros trabalhos que poderiam ser feitos no projeto, adicionando isso ao pôster.

2.8 PALESTRAS E RELATÓRIO FINAL

Nesta semana foram assistidas palestras na matéria de MAC0425 Inteligência Artificial e foi escrito o relatório final.

- *Aprendizagem de redes bayesianas* - Walter Perez
- *Redes neurais e Deep Learning* - Jessica de Souza
- *Sum-Product Networks* - Julissa Villanueva
- *Algoritmos genéticos e programação genética* - Nelson Lago
- *Programação dinâmica e Aprendizado por reforço aplicado ao jogo Tetris* - Suelen Carvalho
- Relatório final

Para esta semana planejou-se assistir as palestras dos alunos de mestrado em MAC0425 Inteligência Artificial, em especial as palestras citadas acima. Além disso essa semana foi reservada para o relatório final da matéria de MAC0215.

3 RESULTADOS

Nesta seção irei primeiro descrever os resultados pessoais desta matéria e do projeto e em seguida irei abordar os resultados acadêmicos do projeto.

Durante o semestre aprendi muito sobre Inteligência Artificial. É uma área que pretendo seguir no futuro, especialmente com relação a pesquisa, e gostei bastante do assunto que estou estudando. Por isso, acho que meus estudos em MAC0215 impactaram fortemente na minha produtividade da Iniciação Científica. Com *deadlines* impostos, me senti pressionado a estudar e a apresentar resultados. Não conseguiria apresentar os resultados de forma tão produtiva se não tivesse me inscrito nesta matéria.

Como a área que estou estudando – principalmente neste estágio dos meus estudos – é muito teórica, apresentar resultados concretos é difícil. Meus resultados neste projeto foram mais em adquirir conhecimento e construir uma base para que, em seguida, consiga atingir resultados mais concretos. Além disso, por essa área em especial ser relativamente nova (o primeiro paper em SPNs é de 2011, enquanto que a primeira vaga ideia de um circuito aritmético similar a SPNs data de 2003[1]), o conteúdo base por trás de toda esta teoria é relativamente grande, já que se trata de tecnologia de ponta. Portanto, os únicos resultados que posso apresentar são os relatórios que escrevi, os quais planejo continuar escrevendo mesmo após a conclusão da matéria, já que são um jeito muito bom de tanto aprender a matéria quanto arquivar para futuras consultas.

4 CONCLUSÃO

Cada semana foi dedicado sábado e domingo para se escrever os relatórios, enquanto que de segunda a sexta eram lidos os artigos ou capítulos do livro. No entanto, em casos em que houvesse EPs ou provas, foi reservada a semana para estudo ou para os EPs. Por causa disso existem algumas semanas que não possuem relatório. Em casos em que a matéria lida era muito grande e não era possível escrever o relatório a tempo, juntou-se as duas semanas em um relatório.

Apesar de tudo isso, acredito que estudei no mínimo 100 horas. Os relatórios ajudam tanto a memorizar quando a revisar o conteúdo visto durante a semana, e portanto acredito serem uma forma de estudo.

Adicionalmente, as aulas de MAC0425 (Inteligência Artificial) contribuíram bastante no estudo da Iniciação Científica, já que o conteúdo de aula e o que foi visto no projeto tem partes iguais, principalmente os assuntos sobre Aprendizado e Incerteza na aula de MAC0425 são diretamente relacionados ao projeto.

REFERÊNCIAS

- [1] Adnan Darwiche. “A Differential Approach to Inference in Bayesian Networks”. Em: (2003).
- [2] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. 1st Edition. Cambridge University Press, 2009.
- [3] Renato Lui Geh. *Aprendizado Automático de Sum-Product Networks*. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/poster/poster.pdf>.
- [4] Renato Lui Geh. *Aprendizado Automático de Sum-Product Networks (SPN)*. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/project/relatorio.pdf>.
- [5] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Compiling Bayesian Networks*. 1. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week1/relatorio.pdf>.
- [6] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.1-6.5*. 2. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week2/relatorio.pdf>.
- [7] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.6-6.9*. 3. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week5/relatorio.pdf>.
- [8] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference with Local Structure 13.1-13.3*. 4. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week8/relatorio.pdf>.
- [9] Renato Lui Geh. *Uma Introdução a Sum-Product Networks*. 5. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week9/relatorio.pdf>.
- [10] Robert Gens e Pedro Domingos. “Learning the Structure of Sum-Product Networks”. Em: *International Conference on Machine Learning* 30 (2013).
- [11] Hoifung Poon e Pedro Domingos. “Sum-Product Networks: A New Deep Architecture”. Em: *Uncertainty in Artificial Intelligence* 27 (2011).