UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

APRENDIZADO EM FLUXOS CONTÍNUOS DE DADOS POR MEIO DE TÉCNICAS SEMISSUPERVISIONADAS

PRISCILLA DE ABREU LOPES

ORIENTADORA: PROFA. DRA. HELOISA DE ARRUDA CAMARGO

São Carlos – SP

Março/2015

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

APRENDIZADO EM FLUXOS CONTÍNUOS DE DADOS POR MEIO DE TÉCNICAS SEMISSUPERVISIONADAS

PRISCILLA DE ABREU LOPES

Qualificação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação, área de concentração: Inteligência Artificial

Orientadora: Profa. Dra. Heloisa de Arruda Camargo

São Carlos – SP

Março/2015

RESUMO

Palavras-chave: aprendizado semissupervisionado, fluxos contínuos de dados, fuzzy

ABSTRACT

....

Keywords: semi-supervised learning, data streams, clustering, fuzzy

LISTA DE FIGURAS

Lista de Algoritmos

LISTA DE TABELAS

ACRÔNIMOS E SIGLAS

FCD – *Fluxos ContÃnuos de Dados*

FCM – *Fuzzy C-Means*

SUMÁRIO

Capítulo 1

INTRODUÇÃO

Este capítulo introduz o contexto e a motivação que levaram à elaboração de uma proposta ...

Capítulo 2

CONCEITOS GERAIS

verde - estranho, deve mudar

azul - referência?

vermelho - reescrever

laranja - importante

...

Neste capítulo são apresentados conceitos gerais a respeito de aprendizado semissupervisionado e de aprendizado em fluxos contínuos de dados.

2.1 Aprendizado Semissupervisionado

Aprendizado de máquina refere-se à investigação de métodos computacionais capazes de adquirir conhecimento de forma automática. Um dos mecanismos utilizados para derivar conhecimento novo é por meio de inferência indutiva sobre um conjunto de dados ou exemplos. O aprendizado indutivo pode ser dividido em três abordagens: supervisionada, não supervisionada e semissupervisionada.

2.1.1 Aprendizado Supervisionado e Não Supervisionado

Abordagens supervisionadas são aquelas que realizam a extração de conhecimento pelo desenvolvimento de um modelo geral baseado em um conjunto de dados totalmente rotulado. Um exemplo de um conjunto de dados é dito rotulado se a classe à qual pertence é conhecida.

Métodos de classificação e regressão tipicamente pertencem a esta categoria e são amplamente utilizadas por produzirem bons resultados.

A maioria dos métodos de classificação utilizam-se de um conjunto de exemplos de treinamento para a construção de um classificador, comumente constituído de um conjunto de regras ou uma estrutura da qual possam ser extraídas regras de classificação. Um conjunto de exemplos de teste independente do conjunto de treinamento é aplicado ao classificador no intuito de verificar a qualidade do resultado obtido na etapa de construção. Se a avaliação for satisfatória, o classificador poderá ser aplicado a conjuntos de novos exemplos. Alguns métodos podem requerer um ajuste do classificador após um período de tempo ou o aumento do volume de dados. Esse ajuste é, por vezes, realizado pela aplicação dos passos de criação e verificação de um classificador.

Aplicações de árvores de decisão (QUINLAN, 1986), redes neurais (BISHOP, 1995) e métodos estatísticos (DUDA; HART, 1973) fazem parte do conjunto de tentativas para a resolução do problema de classificação (MITCHELL, 1997). Existem métodos, como o *K-Nearest Neightbors* (COVER; HART, 1967), que não geram classificadores, mas utilizam a informação de rótulos para classificar novos exemplos, atribuindo classes por meio de métricas de similaridade.

Variações de métodos de classificação baseados na teoria de conjuntos *fuzzy* (ZADEH, 1965) podem realizar a indução de regras que permitem a representação de conhecimento impreciso a partir de um conjunto de dados (PEDRYCZ; GOMIDE, 1998). Sistemas neuro-*fuzzy* (KLOSE et al., 2001) se utilizam de algoritmos de aprendizado derivados da teoria de redes neurais para gerar regaras *fuzzy*. Outras bordagens são baseadas em árvores de decisão, que podem ser induzidas e, posteriormente, ter regras extraídas da estrutura resultante (QUINLAN, 1993). Propostas para extensões chamadas árvores de decisão *fuzzy* também podem ser encontradas na literatura (JANIKOW, 1998; CINTRA; MONARD; CAMARGO, 2012).

Estratégias evolutivas, como Algoritmos Genéticos, são utilizados na otimização e criação de sistemas *fuzzy*. Sua habilidade é otimizar a estrutura e parâmetros de modelos, enquanto grande parte das estratégias de otimização é capaz apenas de adaptar parâmetros de um modelo (KLOSE; KRUSE, 2005).

Apesar dos bons resultados produzidos por técnicas supervisionadas, é possível que as classes não estejam disponíveis para determinados domínios, impedindo sua aplicação. Neste contexto normalmente são aplicadas técnicas não supervisionadas de aprendizado.

Agrupamento de dados é uma típica técnica não supervisionada, ou seja, um processo capaz de realizar aprendizagem a partir de um conjunto de dados não rotulado. A aplicação de

agrupamento tem como objetivo definir uma possível partição dos dados em grupos, de forma que exemplos semelhantes pertençam a um mesmo grupo e exemplos distintos pertençam a grupos distintos. Essa divisão dos dados é baseada em métricas que determinam a relação de dissimilaridade ou similaridade entre diferentes exemplos.

As diferentes técnicas de agrupamento podem ser divididas nas seguintes categorias (HAN; KAMBER; PEI, 2012):

Hierárquico: cria uma decomposição hierárquica de um conjunto de exemplos de acordo com algum critério (DAY; EDELSBRUNNER, 1984; KAUFMAN; ROUSSEEUW, 1990; ZHANG; RAMAKRISHNAN; LIVNY, 1996);

Particional: constrói uma partição inicial de um conjunto de exemplos e, por meio de um processo iterativo, busca melhorar a partição, mudando exemplos de grupo baseado, geralmente, em uma medida de distância (MACQUEEN, 1967; BEZDEK, 1981; KAUFMAN; ROUSSEEUW, 1990);

Baseado em Densidade: baseado em funções densidade, é capaz de criar uma partição ou uma decomposição hierárquica de um conjunto de exemplos (ESTER et al., 1996; HINNEBURG; KEIM, 1998; ANKERST et al., 1999);

Baseado em Grades: todas as operações de agrupamento são realizadas dentro de uma estrutura de grades (*grid*), que é uma divisão do espaço dos exemplos em um número finito de células (WANG; YANG; MUNTZ, 1997; SHEIKHOLESLAMI; CHATTERJEE; ZHANG, 1998);

É relevante mencionar que dentro dos conjuntos descritos é possível encontrar técnicas que utilizam métodos de computação flexível, como redes neurais (KOHONEN, 1990), algoritmos genéticos (referências) e teoria de conjuntos *fuzzy*. O *Fuzzy C-Means* (FCM) (BEZDEK, 1981), por exemplo, implementa uma extensão *fuzzy* do algoritmo *k-means*. No FCM a partição é composta por grupos que podem ser não disjuntos e cada exemplo do conjunto possui um grau de pertinência para cada k grupo.

2.1.2 Técnicas de Aprendizado Semissupervisionado

O crescimento acelerado de conjuntos de dados em muitos domínios torna a rotulação manual e total dos dados onerosa. O aprendizado semissupervisionado tem como base técnicas supervisionadas ou não supervisionadas, adaptadas a fim de realizar a aprendizagem utilizando

conjuntos parcialmente rotulados e/ou alguma informação prévia já disponível, sendo mais explorado nos últimos 10 anos.

visão geral e referências.

As técnicas de aprendizado citadas e referenciadas nesta seção consideram características particulares para os dados disponíveis. Para essas propostas assume-se que o conjunto de dados é finito, os exemplos seguem uma distribuição estática e estão disponíveis para acesso sempre que necessário durante o processo de aprendizagem.

A evolução da tecnologia, a internet e o aumento significativo de seu número de usuários permitiu o desenvolvimento de domínios para os quais as características assumidas pelas abordagens mais clássicas de aprendizado não são verdadeiras.

2.2 Aprendizado em Fluxos Contínuos de Dados

Existe hoje uma variedade de sistemas que produzem grande quantidade de dados em curto espaço de tempo. Estes conjuntos de dados têm tamanho indefinido, potencialmente infinito, e podem gerar exemplos com distribuição estatística mutável de acordo com o tempo.

O surgimento e crescimento deste tipo de sistemas impulsionaram a pesquisa por técnicas que pudessem realizar a aprendizagem considerando as características específicas por estes domínios, referidos como Fluxos Contínuos de Dados (FCD) (em inglês *Data Streams* ou *Streaming Data*). A Figura 2.1 traz um gráfico que demonstra uma visão geral do crescimento no número de publicações sobre aprendizado/mineração em FCD.

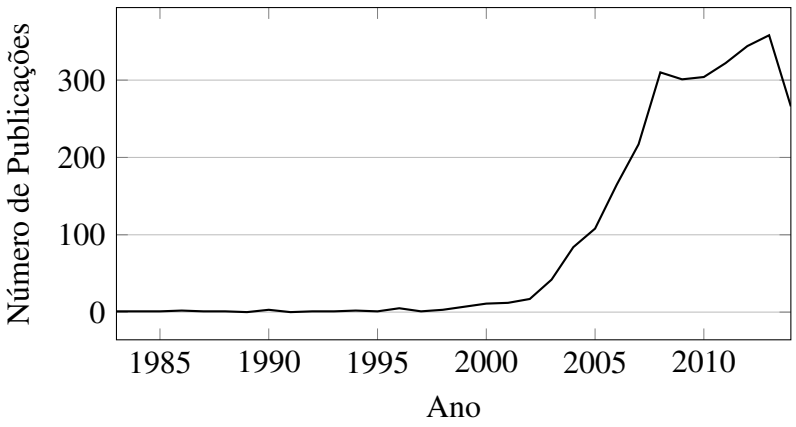

Figura 2.1: Progressão anual do número de publicações em inglês considerando o resultado de busca realizada na base Scopus pela combinação dos termos *learning/mining* e *data streams/streaming data*

(GAMA; GABER, 2007) (AGGARWAL, 2007) (AGGARWAL et al., 2007) (LESKOVEC; RAJARA-

MAN; ULLMAN, 2014)

2.2.1 Técnicas

2.3 Considerações Finais

Capítulo 3

APRENDIZADO SEMISSUPERVISIONADO EM FLUXOS CONTÍNUOS DE DADOS

verde - comentário geral sobre a referência

azul - referência?

vermelho - reescrever

laranja - importante

3.1 Técnicas Semissupervisionadas...

3.1.1 #sqn (?)

3.1.1.1 *Anomaly Intrusion Detection for Evolving Data Stream Based on Semi-supervised Learning* Yu2009 $\longrightarrow$ SPRINGER

Técnica: agrupamento semissupervisionado para *estender o conjunto de dados rotulado* e *aprendizado supervisionado em FCD* (?), com mecanismo de "esquecimento" de dados mais antigos.

Aplicação: *Sistemas de Detecção de Intrusos* em redes de computadores.

Resumo: Proposta de um algoritmo baseado em aprendizado semissupervisionado para detecção de anomalias em FCD, dentro do contexto de segurança em redes de computadores. O algoritmo se utiliza de técnicas de atenuação para resolver o problema de mudança no padrão de tráfego, diminuindo a importância dos dados mais antigos, e um conjunto rotulado estendido, gerado por aprendizado semissupervisionado, para treinar o modelo de detecção. O algoritmo

manifesta resultados experimentais de melhor acurácia comparado a algoritmos baseados em histórico completo e totalmente rotulado.

Comentário: A semissupervisão, aparentemente, é utilizada como parte de rotulação automática e estática para extensão de conjunto de treinamento para construção supervisionada de modelo em FCD. O objetivo desta pesquisa é encontrar métodos que incorporem semissupervisão no processo de aprendizagem em FCD de forma mais intrínseca.

3.1.2 *Semi-supervised learning for cyberbullying detection in social networks* Nahar2014 $\longrightarrow$ SPRINGER

Técnica: ensemble classifier supervisionado + abordagem semissupervisionada para aumentar conjunto de treinamento rotulado.

Aplicação: Detecção de cyberbulling em redes sociais

Resumo: Proposta de uma abordagem semissupervisionada para aumentar a quantidade de instâncias de treinamento e aplicação de um algoritmo de SVM fuzzy. Um conjunto de treinamento rotulado inicial é utilizado para rotulação automática de novas instâncias do stream, por meio de um ensemble classifier. O algorithm de SVM fuzzy é utilizado para ponderar o espaço de atributos. A avaliação mostra a superioridade da proposta em diferentes cenários.

Comentário: ensemble classifiers não são foco da pesquisa e a proposta não utiliza semissupervisão no processo de aprendizagem

3.1.3 ... Baseadas em Agrupamento

3.1.3.1 *C-DenStream: Using Domain Knowledge on a Data Stream* (RUIZ; MENASALVAS; SPILIOPOULOU, 2009)

Técnica: agrupamento semissupervisionado (informação no formato de restrições), baseado em *densidade*, para aprendizado em FCD.

Aplicação: Conjuntos reais e sintéticos.

Relevância: *To our knowledge, this is the first approach to* ***include domain knowledge in clustering for data streams***.

Resumo: 1ª (?) extensão do paradigma de aprendizado semissupervisionado (agrupamento) estático para FCD. Apresentação do C-DenStream, algoritmo de agrupamento baseado em densidade para FCD, que usa informação de domínio em formato de restrições. Proposta

de novo método para utilização de conhecimento prévio em FCD. Estudo de performance em conjuntos reais e sintéticos demonstra efetividade e eficiência do método.

Comentário: Este método deve ser melhor investigado, pois utiliza uma extensão semissupervisionada (formato de restrições) para realizar agrupamento em FCD.

3.1.3.2 *Clustering evolving data stream with affinity propagation algorithm* (ATWA; LI, 2014)

Técnica: extensão semissupervisionada do agrupamento *Affinity Propagation* (FREY; DUECK, 2007) para uso em FCD.

Aplicação: ???

Resumo: Proposta de um algoritmo de agrupamento semissupervisionado que extende *Affinity Propagation (AP)* para lidar com FCD. Um conjunto de instâncias rotuladas é incorporado para detecção de mudança in the generative process underlying the data stream (desvio de conceito ?), que requer a atualização do modelo o mais rápido possível. Experimentos comparativos com outros métodos de agrupamento em FCD demonstram efetividade e eficiência do método proposto.

Comentário: Este método deve ser melhor investigado, pois utiliza uma extensão semissupervisionada (formato de rótulos/sementes) para realizar agrupamento em FCD.

3.1.4 ... Baseadas em Classificador

3.1.4.1 *Active Learning Framework Combining Semi-Supervised Approach for Data Stream Mining* (KHOLGHI; KEYVANPOUR, 2011)

Técnica: active learning (seleção de instâncias "importantes" para rotulação) + influência de instâncias não rotuladas para aprendizado semissupervisionado em FCD. $\longrightarrow$ não está muito claro como ocorre a integração pelos trechos disponíveis do artigo, então talvez seja interessante ler outros trechos do artigo.

Aplicação: conjuntos reais.

Relevância: *To the best of our knowledge, no* ***semi-supervised Active Learning*** *combination exists for* ***data streams***.

Resumo: Construir um modelo para predição de rótulos de instâncias futuras com grande acurácia. Proposta de um framework que combina semissupervisão por meio de active learning e consideração de influência de instâncias não rotuladas a fim de melhorar a performance

de aprendizagem. Visa solucionar os "problemas" de active learning (desconsideração de influência de instâncias não rotuladas) e desafios de aprendizado em FCD. Experimentos provam efetividade do framework proposto.

Comentário: Trata-se de um método de classificação adaptado para semissupervisão (influência de dados não rotulados). Aparentemente, não é uma adaptação "fake" como em 3.1.1.1, mas é uma solução baseada em classificadores, não agrupamento.

3.1.5 ... Baseadas em Ensemble de classificadores

3.1.5.1 *Semi-supervised ensemble learning of data streams in the presence of concept drift* Ahmadi2012 ⟶ SPRINGER

Técnica: aprendizado ensemble para rotulação de instâncias não rotuladas e posterior atualização dos modelos.

Aplicação: ???

Resumo: Apresentação de um novo algoritmo de aprendizado ensemble (que traduz para combinação? comitê?) semissupervisionado em FCD. "Voto da maioria" para rotular instâncias não rotuladas. Estudos demonstram que o algoritmo proposto é comparável a outros algoritmos de aprendizado semissupervisionado online (em FCD).

Comentário: Esta pesquisa não prioriza métodos ensemble classifier.

3.1.5.2 *A new semi-supervised learning based ensemble classifier for recurring data stream* Zhang2014 ⟶ SPRINGER

Técnica: Classificador ensemble, com algoritmos base semissupervisionados.

Aplicação: ???

Resumo: Proposta de um classificador ensemble para aprendizado semissupervisionado para solução do problema de desvio de conceito em FCD. Algoritmos base utilizam instâncias rotuladas e não rotuladas como conjunto de treinamento para obter melhor aprendizado, informação histórica é mantida como parte de peso no fator de decisão quando construído o classificador ensemble. Nova abordagem melhor que o "modelo ensemble geral" e pode ser utilizado em FCD.

Comentário: Esta pesquisa não prioriza métodos ensemble classifier.

3.1.5.3 *Detecting cyberbullying in social networks using multi-agent system* Nahar2014a ⟶ IOS PRESS

Técnica: ensemble classifier que considera instâncias não rotuladas.

Aplicação: Texto (Detecção de cyberbulling em redes sociais.

Resumo: Proposta de um framework baseado em sessão para detecção automática de cyberbulling dentro do grande volume de FC de texto não rotulado. Com a incorporação de um classificador ensemble de uma classe ao framework. O processamento do FCD é feito em ambiente distribuído multi-agente para processar multiplas fontes de redes sociais. Apenas algumas instâncias positivas de cyberbulling estão disponíveis para treinamento inicial. Contribuição maior é detecção de cyberbulling quando não há rótulos disponíveis. Experimentos indicam que a proposta obtém melhor resultados que outros métodos.

Comentário: Esta pesquisa não prioriza métodos ensemble classifier.

3.1.6 Aplicações diferenciadas (não dado/texto)

3.1.6.1 *Online detection of concept drift in visual tracking* Liu2014 ⟶ SPRINGER

Técnica: detecção de desvio de conceito com método de boosting semissupervisionado

Aplicação: Visual Tracking (tradução??)

Resumo: Apresentação de um framework que combina detecção de desvio de conceito com método de boosting semissupervisionado para construir um visual tracker.

Comentário: Esta pesquisa atualmente prioriza métodos aplicados a mineração de dados ou texto, devido peculiaridades de trabalho em outros tipos de domínios. Ainda, trata-se de uma proposta que utiliza mecanismo de boosting semissupervisionado, não priorizada neste momento.

3.2 Técnicas Outras

3.2.1 Multiple data streams

3.2.1.1 *Semi-supervised learning algorithm for online electricity data streams* Patil2015 ⟶ SPRINGER

Técnica: Agrupamento adaptativo para dados não rotulados e classificador adaptativo para dados rotulados.

Aplicação: Preços e demanda no fornecimento de eletricidade. (Multiple data Stream)

Resumo: Apresentação de um modelo de aprendizagem adaptativo para o domínio pela detecção e adaptação a mudanças em tendências e valores. Desafios principais: predições, sumarização do FCD, mudança de tendência em grupos do FCD e adaptatividade do modelo. Método de similaridade baseado em correlação é usado para produzir concept clusters para as instâncias não rotuladas e análise de tendência, detecção de tipos de mudanças entre clusters antigos e atuais, e predição. E um algoritmo de classificação adaptativo para avaliação da habilidade de predição do conjunto de testes. O método proposto é aplicável para conjuntos com 80-85% de instâncias não rotuladas.

Comentário: Trata-se de uma proposta para agrupamento de múltiplos data streams, não considerado foco desta pesquisa.

3.3 Na Fila

3.3.1 2002

(BABCOCK et al., 2002)

3.3.2 2005

(GABER; ZASLAVSKY; KRISHNASWAMY, 2005)

3.3.3 2006

** ***Clustering-training for data stream mining*** **(WU; YANG; ZHOU, 2006)**

(TJHI; CHEN, 2006)

3.3.4 2007

**** Online co-localization in indoor wireless networks by dimension reduction*** **(PAN; YANG; PAN, 2007)**

Applying lazy learning algorithms to tackle concept drift in spam filtering **(FDEZ-RIVEROLA et al., 2007)**

(HO; WECHSLER, 2007)

(HORE; HALL; GOLDGOF, 2007b)

(HORE; HALL; GOLDGOF, 2007a)

3.3.5 2008

**** A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data*** **(MASUD et al., 2008a)**

Learning from Testing Data: A New View of Incremental Semi-Supervised Learning **(CAO; HE, 2008)**

Dynamic integration of classifiers for handling concept drift **(TSYMBAL et al., 2008)**

Data Streaming with Affinity Propagation **(ZHANG; FURTLEHNER; SEBAG, 2008)**

(XINQUAN, 2008)

(HORE et al., 2008)

(MASUD et al., 2008b)

(AGGARWAL; YU, 2008)

3.3.6 2009

Incremental Semi-supervised Clustering In A Data Stream With A Flock Of Agents **(BRUNEAU; PICAROUGNE; GELGON, 2009)**

An incremental extremely random forest classifier for online learning and tracking **(WANG et al., 2009)**

Mining data streams with labeled and unlabeled training examples **(ZHANG; ZHU; GUO, 2009)**

Ambiguous decision trees for mining concept-drifting data streams **(LIU; LI; ZHONG, 2009)**

Mining decision rules on data streams in the presence of concept drifts **(TSAI; LEE; YANG, 2009)**

(HUANG; SANG; TANG, 2009)

(WU; YE; FU, 2009)

(FORESTIERO; PIZZUTI; SPEZZANO, 2009)

(SHI et al., 2009)

(CHAOVALIT; GANGOPADHYAY, 2009)

(AGGARWAL, 2009)

3.3.7 2010

***** Semi-supervised Classification Method for Dynamic Applications*** **(MOUCHAWEH, 2010)**

Semi Supervised Multi Kernel (SeSMiK) graph embedding: Identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy **(TIWARI; KURHANEWICZ, 2010)**

Evolving granular neural network for semi-supervised data stream classification **(LEITE; COSTA; GOMIDE, 2010)**

Using correlation based subspace clustering for multi-label text data classification **(AHMED; KHAN; RAJESWARI, 2010)**

Application of Compound Gaussian Mixture Model clustering in the data stream **(GAO; LIU; GAO, 2010)**

3.3.8 2011

3.3.8.1 *Classifying evolving data streams with partially labeled data* (BORCHANI; nAGA; BIELZA, 2011)

***** Clustering Feature Decision Trees for Semi-supervised Classification from High-speed Data Streams*** **(XU; QIN; CHANG, 2011)**

**** Concurrent Semi-supervised Learning of Data Streams*** **(NGUYEN et al., 2011)**

**** Label-based Semi-supervised Fuzzy Co-clustering for Document Categorization*** **(YAN; CHEN, 2011)**

**** Learning to Group Web Text Incorporating Prior Information*** **(CHENG et al., 2011)**

**** MINETRAC: Mining Flows for Unsupervised Analysis & Semi-supervised Classification*** **(CASAS; MAZEL; OWEZARSKI, 2011)**

On-line learning from streaming data with delayed attributes: A comparison of classifiers and strategies **(MILLáN-GIRALDO; SáNCHEZ; TRAVER, 2011)**

Semi-supervised learning in nonstationary environments **(DITZLER; POLIKAR, 2011)**

A semi-supervised boosting algorithm for mining time-changing data streams **(HUANG; SHA; MA, 2011)**

Semi-supervised approach to handle sudden concept drift in Enron data **(KMIECIAK; STEFANOWSKI, 2011)**

Semi-supervised classification of characterized patterns for demand forecasting using smart electricity meters **(De Silva et al., 2011)**

A Study of Density-Grid based Clustering Algorithms on Data Streams **(AMINI et al., 2011)**

Learning recurring concepts from data streams with a context-aware ensemble **(GOMES;**

(HAMASUNA; ENDO, 2011)

(CHENG et al., 2011)

(MAGDY; YOUSRI; EL-MAKKY, 2011)

3.3.9 2012

***** Mining Recurring Concept Drifts with Limited Labeled Streaming Data*** **(LI; WU; HU, 2012)**

Extensão de *Mining Recurring Concept Drifts With Limited Labeled Streaming Data* (LI; WU; HU, 2010b)

Semi-supervised Data Stream Ensemble Classifiers Algorithm Based On Cluster Assumption **(WANG, 2012)**

**** A Semi-supervised Incremental Clustering Algorithm for Streaming Data*** **(HALKIDI; SPILIOPOULOU; PAVLOU, 2012)**

Semi-supervised classification for reducing false positives **(HUANG; WANG; LI, 2012)**

Particle competition and cooperation in networks for semi-supervised learning with concept drift **(BREVE; ZHAO, 2012)**

A framework for application-driven classification of data streams **(ZHANG et al., 2012)**

***** Learning From Concept Drifting Data Streams with Unlabeled Data*** **(WU; LI; HU, 2012)**

Extensão de *Learning from concept drifting data streams with unlabeled data* (LI; WU; HU, 2010a)

***** Facing the Reality of Data Stream Classification: Coping with Scarcity of Labeled Data*** **(MASUD et al., 2012)**

Robust Re-identification Using Randomness and Statistical Learning Quo-vadis **(NAPPI; WECHSLER, 2012)**

Partially labeled data stream classification with the semi-supervised K-associated graph **(BERTINI; LOPES; ZHAO, 2012)**

Semi-supervised learning techniques in artificial olfaction: A novel approach to classification problems and drift counteraction **(De Vito et al., 2012)**

**** Data Understanding Using Semi-Supervised Clustering*** **(BHATNAGAR et al., 2012)**

Thermal modeling of power transformers using evolving fuzzy systems **(SOUZA et al., 2012)**

Learning very fast decision tree from uncertain data streams with positive and unlabeled samples **(LIANG et al., 2012)**

(ZHANG et al., 2012)

(HAVENS et al., 2012)

3.3.10 2013

3.3.10.1 ** CE-Stream: Evaluation-based Technique for Stream Clustering with Constraints* (SIRAMPUJ; KANGKACHIT; WAIYAMAI, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.2 *** Concurrent Semi-supervised Learning with Active Learning of Data Streams* (NGUYEN; NG; WOON, 2013)

Técnica:

Aplicação:

Resumo:

Comentário: *Concurrent Semi-supervised Learning of Data Stremas* (NGUYEN et al., 2011)

3.3.10.3 *On Achieving Semi-supervised Pattern Recognition By Utilizing Tree-based SOMs* (ASTUDILLO; OOMMEN, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.4 *A comparison of two purity-based algorithms when applied to semi-supervised streaming data classification* (BERTINI; ZHAO, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.5 *Semi-supervised learning with concept drift using particle dynamics applied to network intrusion detection data* (BREVE; ZHAO, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.6 *Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition* (KASABOV et al., 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.7 *On-line laplacian one-class support vector machines* (FRANDINA et al., 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.8 *A semi-supervised ensemble approach for mining data streams* (LIU et al., 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.9 *A stream-based semi-supervised active learning approach for document classification* (BOUGUELIA; BELAID; BELAID, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.10 *Evolving fuzzy pattern trees for binary classification on data streams* (SHAKER; SENGE; HüLLERMEIER, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.11 *Fuzzy Passive?Aggressive classification: A robust and efficient algorithm for online classification problems* (WANG; JI; JIN, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.12 *Adaptive fault detection and diagnosis using an evolving fuzzy classifier* (LEMOS; CAMINHAS; GOMIDE, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.13 *Online extraction of main linear trends for nonlinear time-varying processes* (KALHOR; ARAABI; LUCAS, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.14 *On-line dynamic adaptation of fuzzy preferences* (MARIN et al., 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.15 *Online stream clustering using density and affinity propagation algorithm* (ZHANG et al., 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.10.16 (SILVA et al., 2013)

3.3.10.17 (YOGITA; TOSHNIWAL, 2013)

3.3.10.18 (CHEN; CHEN; SHENG, 2013)

3.3.11 2014

3.3.11.1 *Compose: A semisupervised learning framework for initially labeled nonstationary streaming data* (DYER; CAPO; POLIKAR, 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.2 *A pattern query strategy based on semi-supervised machine learning in distributed WSNs* (LI, 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.3 *GT2FC: An online growing interval type-2 self-learning fuzzy classifier* (BOUCHACHIA; VANARET, 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.4 *On Density-Based Data Streams Clustering Algorithms: A Survey* (AMINI; WAH; SABOOHI, 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.5 *D-FICCA: A density-based fuzzy imperialist competitive clustering algorithm for intrusion detection in wireless sensor networks* (SHAMSHIRBAND et al., 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.6 *Online fuzzy medoid based clustering algorithms* (LABROCHE, 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.7 *Clustering data streams using grid-based synopsis* (BHATNAGAR; KAUR; CHAKRAVARTHY, 2013)

Técnica:

Aplicação:

Resumo:

Comentário:

3.3.11.8 *Data Stream Clustering With Affinity Propagation* (ZHANG et al., 2014)

Técnica:

Aplicação:

Resumo:

Comentário:

3.4 Considerações Finais

Capítulo 4

PROPOSTA DE TRABALHO

4.1 Atividades Principais

4.2 Cronograma de Atividades

4.3 Contribuições Esperadas

4.4 Considerações Finais

REFERÊNCIAS

AGGARWAL, C. A Framework for Clustering Massive-Domain Data Streams. In: *2009 IEEE 25th International Conference on Data Engineering*. [S.l.]: IEEE, 2009. p. 102–113.

AGGARWAL, C. C. An Introduction to Data Streams. In: AGGARWAL, C. C. (Ed.). *Data Streams - Models and Algorithms*. [S.l.]: Springer, 2007. p. 1–8.

AGGARWAL, C. C. et al. On Clustering Massive Data Streams: A Summarization Paradigm. In: AGGARWAL, C. C. (Ed.). *Data Streams - Models and Algorithms*. [S.l.]: Springer, 2007. p. 9–38.

AGGARWAL, C. C.; YU, P. S. A Framework for Clustering Uncertain Data Streams. In: *2008 IEEE 24th International Conference on Data Engineering*. [S.l.]: IEEE, 2008. v. 00, p. 150–159.

AHMED, M. S.; KHAN, L.; RAJESWARI, M. Using Correlation Based Subspace Clustering for Multi-label Text Data Classification. In: *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*. [S.l.]: IEEE, 2010. p. 296–303.

AMINI, A.; WAH, T. Y.; SABOOHI, H. On Density-Based Data Streams Clustering Algorithms: A Survey. *Journal of Computer Science and Technology*, v. 29, n. 1, p. 116–141, 2014.

AMINI, A. et al. A study of density-grid based clustering algorithms on data streams. In: *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. [S.l.]: IEEE, 2011. p. 1652–1656.

ANKERST, M. et al. Optics: Ordering points to identify the clustering structure. In: *ACM Sigmod Record*. [S.l.: s.n.], 1999. p. 49–60.

ASTUDILLO, C. A.; OOMMEN, B. J. On achieving semi-supervised pattern recognition by utilizing tree-based SOMs. *Pattern Recognition*, v. 46, n. 1, p. 293–304, jan. 2013.

ATWA, W.; LI, K. Clustering Evolving Data Stream with Affinity. In: *Database and Expert Systems Applications*. [S.l.]: Springer International Publishing, 2014. p. 446–453.

BABCOCK, B. et al. Models and issues in data stream systems. In: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '02*. New York, New York, USA: ACM Press, 2002. p. 1–16.

BERTINI, J. a. R.; LOPES, A. D. A.; ZHAO, L. Partially labeled data stream classification with the semi-supervised K-associated graph. *Journal of the Brazilian Computer Society*, v. 18, n. 4, p. 299–310, abr. 2012.

BERTINI, J. R.; ZHAO, L. A Comparison of Two Purity-Based Algorithms When Applied to Semi-supervised Streaming Data Classification. In: *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence*. [S.l.]: IEEE, 2013. p. 21–27.

BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

BHATNAGAR, V. et al. Data Understanding using Semi-Supervised Clustering. In: *2012 Conference on Intelligent Data Understanding*. [S.l.]: IEEE, 2012. p. 118–123.

BHATNAGAR, V.; KAUR, S.; CHAKRAVARTHY, S. Clustering data streams using grid-based synopsis. *Knowledge and Information Systems*, v. 41, n. 1, p. 127–152, jun. 2013.

BISHOP, C. M. *Neural Networks for Pattern Recognition*. [S.l.]: Oxford University Press, 1995.

BORCHANI, H.; nAGA, P. L.; BIELZA, C. Classifying evolving data streams with partially labeled data. *Intelligent Data Analysis*, v. 15, n. 5, p. 655–670, 2011.

BOUCHACHIA, A.; VANARET, C. GT2FC: An Online Growing Interval Type-2 Self-Learning Fuzzy Classifier. *IEEE Transactions on Fuzzy Systems*, v. 22, n. 4, p. 999–1018, ago. 2014.

BOUGUELIA, M.-R.; BELAID, Y.; BELAID, A. A Stream-Based Semi-supervised Active Learning Approach for Document Classification. In: *2013 12th International Conference on Document Analysis and Recognition*. [S.l.]: IEEE, 2013. p. 611–615.

BREVE, F.; ZHAO, L. Particle competition and cooperation in networks for semi-supervised learning with concept drift. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. [S.l.]: IEEE, 2012. p. 1–6.

BREVE, F.; ZHAO, L. Semi-supervised Learning with Concept Drift Using Particle Dynamics Applied to Network Intrusion Detection Data. In: *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence*. [S.l.]: IEEE, 2013. p. 335–340.

BRUNEAU, P.; PICAROUGNE, F.; GELGON, M. Incremental semi-supervised clustering in a data stream with a flock of agents. In: *2009 IEEE Congress on Evolutionary Computation*. [S.l.]: IEEE, 2009. p. 3067–3074.

CAO, Y.; HE, H. Learning from testing data: A new view of incremental semi-supervised learning. In: *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. [S.l.]: IEEE, 2008. p. 2872–2878.

CASAS, P.; MAZEL, J.; OWEZARSKI, P. MINETRAC: Mining flows for unsupervised analysis & semi-supervised classification. In: *Proceedings of the 23rd International Teletraffic Congress*. [S.l.: s.n.], 2011. p. 87–94.

CHAOVALIT, P.; GANGOPADHYAY, A. A method for clustering transient data streams. In: *Proceedings of the 2009 ACM symposium on Applied Computing - SAC '09*. New York, New York, USA: ACM Press, 2009. p. 1518.

CHEN, J.; CHEN, P.; SHENG, X. A Sketch-based Clustering Algorithm for Uncertain Data Streams. *Journal of Networks*, v. 8, n. 7, p. 1536–1542, jul. 2013.

CHENG, Y. et al. Learning to Group Web Text Incorporating Prior Information. In: *2011 IEEE 11th International Conference on Data Mining Workshops*. [S.l.]: IEEE, 2011. p. 212–219.

CINTRA, M. E.; MONARD, M. C.; CAMARGO, H. A. FuzzyDT - A Fuzzy Decision Tree Algorithm Based on C4. 5. In: *CBSF - Brazilian Congress on Fuzzy Systems*. [S.l.: s.n.], 2012. p. 199–211.

COVER, T. M.; HART, P. E. Nearest Neighbor Pattern Classification. *IEEE Transactions in Information Theory*, IT-13, n. 1, p. 21–27, 1967.

DAY, W. H. E.; EDELSBRUNNER, H. Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification*, v. 1, n. 1, p. 7–24, 1984.

De Silva, D. et al. Semi-supervised classification of characterized patterns for demand forecasting using smart electricity meters. In: *2011 International Conference on Electrical Machines and Systems*. [S.l.]: IEEE, 2011. p. 1–6.

De Vito, S. et al. Semi-Supervised Learning Techniques in Artificial Olfaction: A Novel Approach to Classification Problems and Drift Counteraction. *IEEE Sensors Journal*, v. 12, n. 11, p. 3215–3224, nov. 2012.

DITZLER, G.; POLIKAR, R. Semi-supervised learning in nonstationary environments. In: *The 2011 International Joint Conference on Neural Networks*. [S.l.]: IEEE, 2011. p. 2741–2748.

DUDA, R. O.; HART, P. E. *Pattern classification and scene analysis*. [S.l.]: John Wiley and Sons, 1973.

DYER, K. B.; CAPO, R.; POLIKAR, R. COMPOSE: A semisupervised learning framework for initially labeled nonstationary streaming data. *IEEE transactions on neural networks and learning systems*, v. 25, n. 1, p. 12–26, jan. 2014.

ESTER, M. et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: *Second International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 1996. p. 226–231.

FDEZ-RIVEROLA, F. et al. Applying lazy learning algorithms to tackle concept drift in spam filtering. *Expert Systems with Applications*, v. 33, n. 1, p. 36–48, jul. 2007.

FORESTIERO, A.; PIZZUTI, C.; SPEZZANO, G. FlockStream: A Bio-Inspired Algorithm for Clustering Evolving Data Streams. In: *2009 21st IEEE International Conference on Tools with Artificial Intelligence*. [S.l.]: IEEE, 2009. p. 1–8.

FRANDINA, S. et al. On-Line Laplacian One-Class Support Vector Machines. In: *Artificial Neural Networks and Machine Learning (ICANN2013)*. [S.l.]: Springer Berlin Heidelberg, 2013. p. 186–193.

FREY, B. J.; DUECK, D. Clustering by Passing Messages Between Data Points. *Science*, v. 315, n. 5814, p. 947–949, fev. 2007.

GABER, M. M.; ZASLAVSKY, A.; KRISHNASWAMY, S. Mining data streams: a review. *ACM SIGMOD Record*, v. 34, n. 2, p. 18, jun. 2005.

GAMA, J.; GABER, M. M. (Ed.). *Learning from Data Streams: Processing Techniques in Sensor Networks*. [S.l.]: Springer, 2007.

GAO, M.-m.; LIU, J.-z.; GAO, X.-x. Application of Compound Gaussian Mixture Model clustering in the data stream. In: *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*. [S.l.]: IEEE, 2010. p. V7–172–V7–177.

GOMES, J. a. B.; MENASALVAS, E.; SOUSA, P. a. C. Learning recurring concepts from data streams with a context-aware ensemble. In: *Proceedings of the 2011 ACM Symposium on Applied Computing - SAC '11*. New York, New York, USA: ACM Press, 2011. p. 994.

HAHSLER, M.; DUNHAM, M. H. Temporal Structure Learning for Clustering Massive Data Streams in Real-Time. In: *Proceedings of the 2011 SIAM International Conference on Data Mining*. [S.l.]: Society for Industrial and Applied Mathematics, 2011. p. 664–675.

HALKIDI, M.; SPILIOPOULOU, M.; PAVLOU, A. A semi-supervised incremental clustering algorithm for streaming data. *Advances in Knowledge Discovery and Data Mining*, v. 7301, p. 578–590, 2012.

HAMASUNA, Y.; ENDO, Y. On semi-supervised fuzzy c-means clustering with clusterwise tolerance by opposite criteria. In: *2011 IEEE International Conference on Granular Computing*. [S.l.]: IEEE, 2011. p. 225–230.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. [S.l.]: Morgan Kaufmann Publishers, 2012. 744 p. (Data Management Systems Series).

HAVENS, T. C. et al. Fuzzy c-Means Algorithms for Very Large Data. *IEEE Transactions on Fuzzy Systems*, v. 20, n. 6, p. 1130–1146, dez. 2012.

HINNEBURG, A.; KEIM, D. A. An Efficient Approach to Clustering in Large Multimedia Databases with Noise. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*. [S.l.: s.n.], 1998. v. 5865, p. 58–65.

HO, S.-s.; WECHSLER, H. Detecting Changes in Unlabeled Data Streams using Martingale. In: *Proceedings of the 2007 International Joint Conference on Artificial Intelligence (IJCAI07)*. [S.l.: s.n.], 2007. p. 1912–1917.

HORE, P. et al. Online fuzzy c means. In: *NAFIPS 2008 - 2008 Annual Meeting of the North American Fuzzy Information Processing Society*. [S.l.]: IEEE, 2008. p. 1–5.

HORE, P.; HALL, L. O.; GOLDGOF, D. B. A fuzzy c means variant for clustering evolving data streams. In: *2007 IEEE International Conference on Systems, Man and Cybernetics*. [S.l.]: IEEE, 2007. p. 360–365.

HORE, P.; HALL, L. O.; GOLDGOF, D. B. Creating Streaming Iterative Soft Clustering Algorithms. In: *NAFIPS 2007 - 2007 Annual Meeting of the North American Fuzzy Information Processing Society*. [S.l.]: IEEE, 2007. p. 484–488.

HUANG, R.; SANG, N.; TANG, Q. Segmentation via Incremental Transductive Learning. In: *2009 Fifth International Conference on Image and Graphics*. [S.l.]: IEEE, 2009. v. 1, n. c, p. 213–216.

HUANG, S.; SHA, A.; MA, S. A Semi-Supervised Boosting Algorithm for Mining Time-Changing Data Streams. *Journal of Information & Computational Science*, v. 13, p. 2807–2814, 2011.

HUANG, S.; WANG, K.; LI, T. Semi-supervised Classification for Reducing False Positives. *Journal of Computational Information Systems*, v. 13, n. 8, p. 5327–5334, 2012.

JANIKOW, C. Z. Fuzzy decision trees: issues and methods. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, v. 28, n. 1, p. 1–14, 1998.

KALHOR, A.; ARAABI, B. N.; LUCAS, C. Online extraction of main linear trends for nonlinear time-varying processes. *Information Sciences*, Elsevier Inc., v. 220, p. 22–33, jan. 2013.

KASABOV, N. et al. Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition. *Neural networks : the official journal of the International Neural Network Society*, Elsevier Ltd, v. 41, n. 1995, p. 188–201, maio 2013.

KAUFMAN, L.; ROUSSEEUW, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. [S.l.]: John Wiley and Sons, 1990. 368 p.

KHOLGHI, M.; KEYVANPOUR, M. Active Learning Framework Combining Semi-Supervised Approach for Data Stream Mining. In: *Intelligent Computing and Information Science*. [S.l.]: Springer Berlin Heidelberg, 2011, (Communications in Computer and Information Science, v. 135). p. 238–243.

KLOSE, A.; KRUSE, R. Semi-supervised learning in knowledge discovery. *Fuzzy Sets and Systems*, v. 149, p. 209–233, 2005.

KLOSE, A. et al. Data mining with neuro-fuzzy models. In: KANDEL, A.; LAST, M.; BUNKE, H. (Ed.). *Data Mining and Computational Intelligence*. Heidelberg, Germany: Physica-Verlag GmbH, 2001. p. 1–35.

KMIECIAK, M. R.; STEFANOWSKI, J. Semi-supervised approach to handle sudden concept drift. *Control and Cybernetics*, v. 40, n. 3, p. 667–695, 2011.

KOHONEN, T. The self-organizing map. *Proceedings of the IEEE*, v. 78, n. 9, p. 1464–1480, 1990.

LABROCHE, N. Online fuzzy medoid based clustering algorithms. *Neurocomputing*, Elsevier, v. 126, p. 141–150, fev. 2014.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural network for semi-supervised data stream classification. In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*. [S.l.]: IEEE, 2010. p. 1–8.

LEMOS, A.; CAMINHAS, W.; GOMIDE, F. Adaptive fault detection and diagnosis using an evolving fuzzy classifier. *Information Sciences*, Elsevier Inc., v. 220, p. 64–85, jan. 2013.

LESKOVEC, J.; RAJARAMAN, A.; ULLMAN, J. D. *Mining of Massive Datasets*. [S.l.]: Cambridge University Press, 2014. 476 p.

LI, F. A Pattern Query Strategy Based on Semi-supervised Machine Learning in Distributed WSNs. *Journal of Information and Computational Science*, v. 11, n. 18, p. 6447–6459, dez. 2014.

LI, P.; WU, X.; HU, X. Learning from Concept Drifting Data Streams with Unlabeled Data. In: *Proceedings of the National Conference on Artificial Intelligence*. [S.l.: s.n.], 2010. p. 1945–1946.

LI, P.; WU, X.; HU, X. Mining Recurring Concept Drifts with Limited Labeled Streaming Data. In: *JMLR: Workshop and Conference Proceedings 13*. [S.l.: s.n.], 2010. v. 3, n. 2, p. 241–252.

LI, P.; WU, X.; HU, X. Mining Recurring Concept Drifts with Limited Labeled Streaming Data. *ACM Transactions on Intelligent Systems and Technology*, v. 3, n. 2, p. 1–32, fev. 2012.

LIANG, C. et al. Learning very fast decision tree from uncertain data streams with positive and unlabeled samples. *Information Sciences*, Elsevier Inc., v. 213, p. 50–67, dez. 2012. ISSN 00200255.

LIU, J.; LI, X.; ZHONG, W. Ambiguous decision trees for mining concept-drifting data streams. *Pattern Recognition Letters*, Elsevier B.V., v. 30, n. 15, p. 1347–1355, nov. 2009.

LIU, J. et al. A Semi-supervised Ensemble Approach for Mining Data Streams. *Journal of Computers*, v. 8, n. 11, p. 2873–2879, nov. 2013.

MACQUEEN, J. B. Some Methods for Classification and Analysis of MultiVariate Observations. In: *In Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.]: University of California Press, 1967. v. 1, p. 281–297.

MAGDY, A.; YOUSRI, N. a.; EL-MAKKY, N. M. Discovering Clusters with Arbitrary Shapes and Densities in Data Streams. In: *2011 10th International Conference on Machine Learning and Applications and Workshops*. [S.l.]: IEEE, 2011. p. 279–282.

MARIN, L. et al. On-line dynamic adaptation of fuzzy preferences. *Information Sciences*, Elsevier Inc., v. 220, p. 5–21, jan. 2013.

MASUD, M. M. et al. A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data. In: *2008 Eighth IEEE International Conference on Data Mining*. [S.l.]: IEEE, 2008. p. 929–934.

MASUD, M. M. et al. *A Practical Approach To Classify Evolving Data Streams: Training With Limited Amount Of Labeled Data*. [S.l.], 2008. 11 p.

MASUD, M. M. et al. Facing the reality of data stream classification: coping with scarcity of labeled data. In: *Knowledge and Information Systems*. [S.l.: s.n.], 2012. v. 33, n. 1, p. 213–244.

MILLáN-GIRALDO, M.; SáNCHEZ, J. S.; TRAVER, V. J. On-line learning from streaming data with delayed attributes: a comparison of classifiers and strategies. *Neural Computing and Applications*, v. 20, n. 7, p. 935–944, jun. 2011.

MITCHELL, T. *Machine Learning*. [S.l.]: McGraw-Hill Education (ISE Editions), 1997.

MOUCHAWEH, M. S. Semi-supervised classification method for dynamic applications. *Fuzzy Sets and Systems*, Elsevier, v. 161, n. 4, p. 544–563, fev. 2010.

NAPPI, M.; WECHSLER, H. Robust re-identification using randomness and statistical learning: Quo vadis. *Pattern Recognition Letters*, Elsevier B.V., v. 33, n. 14, p. 1820–1827, out. 2012.

NGUYEN, H.; NG, W.; WOON, Y. Concurrent Semi-supervised Learning with Active Learning of Data Streams. In: *Transactions on Large-Scale Data- and Knowledge-Centered Systems VIII*. [S.l.]: Springer Berlin Heidelberg, 2013. v. 7790, p. 113–136.

NGUYEN, H.-l. et al. Concurrent Semi-supervised Learning of. In: *Data Warehousing and Knowledge Discovery*. [S.l.]: Springer Berlin Heidelberg, 2011. p. 445–459.

PAN, J.; YANG, Q.; PAN, S. Online co-localization in indoor wireless networks by dimension reduction. In: *Proceedings of the National Conference on Artificial Intelligence*. [S.l.: s.n.], 2007. p. 1102–1107.

PEDRYCZ, W.; GOMIDE, F. *An Introduction to Fuzzy Sets: Analysis and Design*. [S.l.]: MIT Press, 1998. (A Bradford book).

QUINLAN, J. R. Induction of decision trees. *Machine Learning*, v. 1, n. 1, p. 81–106, 1986.

QUINLAN, J. R. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

RUIZ, C.; MENASALVAS, E.; SPILIOPOULOU, M. C-DenStream: Using domain knowledge on a data stream. In: *Discovery Science*. [S.l.]: Springer Berlin Heidelberg, 2009. p. 287–301.

SHAKER, A.; SENGE, R.; HüLLERMEIER, E. Evolving fuzzy pattern trees for binary classification on data streams. *Information Sciences*, Elsevier Inc., v. 220, p. 34–45, jan. 2013.

SHAMSHIRBAND, S. et al. D-FICCA: A density-based fuzzy imperialist competitive clustering algorithm for intrusion detection in wireless sensor networks. *Measurement*, Elsevier Ltd, v. 55, p. 212–226, set. 2014.

SHEIKHOLESLAMI, G.; CHATTERJEE, S.; ZHANG, A. Wavecluster: A multi-resolution clustering approach for very large spatial databases. In: *Proceedings of the International Conference on Very Large Data Bases*. [S.l.: s.n.], 1998. p. 428–439.

SHI, X. et al. An incremental affinity propagation algorithm and its applications for text clustering. In: *2009 International Joint Conference on Neural Networks*. [S.l.]: IEEE, 2009. p. 2914–2919.

SILVA, J. A. et al. Data stream clustering: A survey. *ACM Computing Surveys*, v. 46, n. 1, p. 1–31, out. 2013.

SIRAMPUJ, T.; KANGKACHIT, T.; WAIYAMAI, K. CE-Stream : Evaluation-based technique for stream clustering with constraints. In: *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. [S.l.]: IEEE, 2013. p. 217–222.

SOUZA, L. et al. Thermal modeling of power transformers using evolving fuzzy systems. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 25, n. 5, p. 980–988, ago. 2012.

TIWARI, P.; KURHANEWICZ, J. Semi supervised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2010)*. [S.l.]: Springer Berlin Heidelberg, 2010. p. 666–673.

TJHI, W.-C.; CHEN, L. Flexible Fuzzy Co-clustering with Feature-cluster Weighting. In: *2006 9th International Conference on Control, Automation, Robotics and Vision*. [S.l.]: IEEE, 2006. p. 1–6.

TSAI, C.-J.; LEE, C.-I.; YANG, W.-P. Mining decision rules on data streams in the presence of concept drifts. *Expert Systems with Applications*, Elsevier Ltd, v. 36, n. 2, p. 1164–1178, mar. 2009.

TSYMBAL, A. et al. Dynamic integration of classifiers for handling concept drift. *Information Fusion*, v. 9, n. 1, p. 56–68, jan. 2008.

WANG, A. et al. An incremental extremely random forest classifier for online learning and tracking. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. [S.l.]: IEEE, 2009. p. 1449–1452. ISBN 978-1-4244-5653-6.

WANG, L.; JI, H.-B.; JIN, Y. Fuzzy Passive-Aggressive classification: A robust and efficient algorithm for online classification problems. *Information Sciences*, Elsevier Inc., v. 220, p. 46–63, jan. 2013.

WANG, W.; YANG, J.; MUNTZ, R. STING: A statistical information grid approach to spatial data mining. In: *Proceedings of International Conference on Very Large Data*. [S.l.: s.n.], 1997. p. 1–18. ISBN 1558604707. ISSN 10477349.

WANG, X. Semi-supervised Data Stream Ensemble Classifiers Algorithm Based on Cluster Assumption. In: *Software Engineering and Knowledge Engineering: Theory and Practice*. [S.l.: s.n.], 2012. v. 1, p. 713–721.

WU, Q.-Y.; YE, Y.; FU, J. Learnable topical crawler through online semi-supervised clustering. In: *2009 International Conference on Machine Learning and Cybernetics*. [S.l.]: IEEE, 2009. p. 231–236.

WU, S.; YANG, C.; ZHOU, J. Clustering-training for Data Stream Mining. In: *Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)*. [S.l.]: IEEE, 2006. p. 653–656. ISBN 0-7695-2702-7.

WU, X.; LI, P.; HU, X. Learning from Concept Drifting Data Streams with Unlabeled Data. *Neurocomputing*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 92, p. 145–155, 2012.

XINQUAN, C. Weighted Clustering and Evolutionary Analysis of Hybrid Attributes Data Streams. *Journal of Computers*, v. 3, n. 12, p. 60–67, 2008.

XU, W.-h.; QIN, Z.; CHANG, Y. Clustering feature decision trees for semi-supervised classification from high-speed data streams. *Journal of Zhejiang University SCIENCE C*, SP Zhejiang University Press, v. 12, n. 8, p. 615–628, 2011.

YAN, Y.; CHEN, L. Label-based semi-supervised fuzzy co-clustering for document categoraization. In: *2011 8th International Conference on Information, Communications & Signal Processing*. [S.l.]: IEEE, 2011. p. 1–5.

YOGITA, Y.; TOSHNIWAL, D. Clustering techniques for streaming data - a survey. In: *2013 3rd IEEE International Advance Computing Conference (IACC)*. [S.l.]: IEEE, 2013. p. 951–956.

ZADEH, L. A. Fuzzy sets. *Information and Control*, v. 8, n. 3, p. 338–353, 1965.

ZHANG, D. et al. A Clustering Algorithm Based on Density-Grid for Stream Data. In: *2012 13th International Conference on Parallel and Distributed Computing, Applications and Technologies*. [S.l.]: IEEE, 2012. p. 398–403.

ZHANG, J.-p. et al. Online stream clustering using density and affinity propagation algorithm. In: *2013 IEEE 4th International Conference on Software Engineering and Service Science*. [S.l.]: IEEE, 2013. p. 828–832.

ZHANG, P. et al. A framework for application-driven classification of data streams. *Neurocomputing*, Elsevier, v. 92, p. 170–182, set. 2012.

ZHANG, P.; ZHU, X.; GUO, L. Mining Data Streams with Labeled and Unlabeled Training Examples. In: *2009 Ninth IEEE International Conference on Data Mining*. [S.l.]: IEEE, 2009. p. 627–636.

ZHANG, T.; RAMAKRISHNAN, R.; LIVNY, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In: *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*. New York, NY, USA: ACM, 1996. (SIGMOD '96), p. 103–114.

ZHANG, X. et al. Data Stream Clustering With Affinity Propagation. *IEEE Transactions on Knowledge and Data Engineering*, v. 26, n. 7, p. 1644–1656, jul. 2014.

ZHANG, X.; FURTLEHNER, C.; SEBAG, M. Data streaming with affinity propagation. In: *Machine Learning and Knowledge ...* . [S.l.]: Springer Berlin Heidelberg, 2008. p. 628–643.