Seleção Local de Características em Agrupamento Hierárquico de Documentos

Marcelo N. Ribeiro [1] **, Manoel J. R. Neto**[2] **, Ricardo B. C. Prudêncio**[1]

[1]Centro de Informática, Universidade Federal de Pernambuco, Recife – PE – Brasil

[2]Instituto de Computação, Universidade Federal de Alagoas, Maceió – AL – Brasil

mnr@cin.ufpe.br,mjrn@tci.ufal.br,rbcp@cin.ufpe.br

Abstract. *Feature selection has improved the performance of text clustering. Global feature selection tries to identify a single subset of features which are relevant to all clusters. However, the clustering process might be improved by considering different subsets of features for locally describing each cluster. In this work, we introduce the method ZOOM-IN to perform local feature selection for partitional hierarchical clustering of text collections. The proposed method explores the diversity of clusters generated by the hierarchical algorithm, selecting a variable number of features according to the size of the clusters. Experiments were conducted on Reuters collection, by evaluating the bisecting K-means algorithm with both global and local approaches to feature selection. The results of the experiments showed an improvement in clustering performance with the use of the proposed local method.*

Resumo. *O uso de seleção de características é capaz de melhorar a precisão e tempo de execução dos algoritmos de agrupamento de documentos. A seleção global de características tenta identificar um único subconjunto de características que é relevante para todos os grupos. No entanto, o processo de agrupamento pode ser melhorado considerando diferentes subconjuntos de características que descrevam localmente cada grupo. Neste trabalho, é introduzido o método ZOOM-IN para realizar seleção local de características para agrupamento hierárquico divisivo de documentos. O método proposto explora a diversidade de grupos retornados por um algoritmo hierárquico, selecionando um número variável de características de acordo com o tamanho dos grupos. Experimentos foram conduzidos na base Reuters, avaliando o algoritmo bisecting K-means com ambas as abordagens global e local para seleção de características. Os resultados dos experimentos mostraram uma melhora no desempenho do agrupamento com o uso do método local proposto.*

1. Introdução

Algoritmos de agrupamento vêm sendo aplicados para dar suporte ao acesso de informação em grandes coleções de documentos [Steinbach et al. 2000]. Tais técnicas organizam documentos similares em grupos associados a diferentes níveis de especificidade e diferentes contextos. A estrutura de grupos, devidamente etiquetados, oferece uma visão de quais tipos de questões podem ser respondidas pelo resultado da consulta em um modelo de recuperação de informação.

De forma a realizar o processo de agrupamento de documentos, os documentos são representados, na maioria dos casos, como um conjunto de *termos de indexação*, que formam um vetor de características associadas a pesos numéricos. Considerar todos as características existentes em uma coleção traz algumas dificuldades ao algoritmo de agrupamento. De fato, quando o espaço de características é muito grande, a distância entre pontos similares não é muito diferente que a distância entre pontos distantes, fenômeno este chamado de "praga da dimensionalidade" [Tang et al. 2005].

Considerando o contexto em discussão, agrupamento de documentos usualmente contém uma fase de redução da dimensionalidade dos vetores que representam os documentos. Estas características no espaço reduzido podem corresponder a um subconjunto das características originais (o que é realizado por métodos de seleção de características [Dy and Brodley 2004]), ou elas podem ser criadas pela combinação das características originais (o que é realizado por métodos de extração de características [Tang et al. 2005]). Em agrupamento de documentos, a extração de características possui uma desvantagem comparada a seleção de características, uma vez que cada nova característica não está mais associada com um termo ou palavra existente, o que torna os grupos formados menos auto-descritíveis [Tang et al. 2005].

A seleção de características pode ser classificada como global ou local [Li et al. 2008]. A abordagem global visa selecionar um único subconjunto de características que são relevantes para derivar todos os grupos [Li et al. 2008]. Apesar do uso frequente dos métodos globais na literatura, dependendo do problema, é possível que exista vários subconjuntos de características que revelem bons grupos. De maneira a superar esta limitação, a seleção local de características, por sua vez, tenta identificar diferentes subconjuntos de características associados a cada grupo formado. Embora trabalhos recentes tenham obtido bons resultados avaliando a seleção local de características em dados não-textuais (ver [Li et al. 2008]), não há pesquisa do uso da seleção de características para agrupamento de documentos.

Neste trabalho, é proposto o ZOOM-IN, um método de seleção local de características para agrupamento hierárquico divisivo de documentos. Neste método, todos os documentos são inicialmente alocados para um grupo-raiz que é recursivamente dividido em subgrupos menores. Em cada passo de divisão, um critério de relevância das características é aplicado para escolher as características que são mais relevantes somente considerando o grupo a ser dividido. O número de características selecionadas é definido de acordo com o tamanho do grupo. O resultado do método proposto é uma hierarquia de grupos em que cada grupo é representado por um subconjunto de características diferente. Experimentos foram realizados com o uso da base Reuters [Lewis 1999], comparando o algoritmo *bisecting K-means* (um algoritmo hierárquico divisivo) [Steinbach et al. 2000], com ambas as abordagens global e local de seleção de características. Os resultados revelaram uma melhora na precisão quando a abordagem local foi comparada com a abordagem global.

A Seção 2 traz uma breve introdução sobre agrupamento de documentos. A Seção 3 apresenta abordagens de seleção de características aplicadas para agrupamento de documentos, seguida da Seção 4, que apresenta o método proposto. A Seção 5 descreve os experimentos realizados e resultados alcançados. Finalmente, a Seção 6 apresenta algumas considerações finais e possibilidades de trabalho futuro.

2. Agrupamento de documentos

Agrupamento de documentos é a atividade de separar grupos de documentos similares, de maneira a melhor discriminar documentos pertencentes a classes diferentes. Um documento é descrito por um conjunto de palavras-chave, os chamados *termos de indexação*, definido a partir do vocabulário presente na coleção de textos. O modelo mais usado com sucesso para agrupamento de documentos é o Modelo de Espaço Vetorial [Salton et al. 1975]. Um peso é associado a cada termo de indexação, o que define um vetor de características que representa o documento. A forma como os pesos são calculados é definida pelo modelo de recuperação de informação usado.

Um algoritmo de agrupamento hierárquico organiza os dados em uma hierarquia de grupos. Para agrupamento de documentos, a solução hierárquica tem maiores vantagens em relação a abordagem *flat*, pois divide a coleção de documentos em vários níveis de especificidade, com diferentes granularidades, proporcionando uma melhor visão da coleção [Sahoo et al. 2006].

Algoritmos de agrupamento hierárquicos podem ainda ser classificados em aglomerativos ou divisivos. O processo aglomerativo é uma abordagem ascendente (*bottom-up*) e começa afetando cada documento a um grupo distinto e prossegue combinando os documentos em grupos mais similares, até que todos os documentos sejam alocados a um único grupo, ou outro critério de parada seja alcançado. O processo divisivo, por sua vez, é uma abordagem descendente (*top-down*) e começa considerando todos os documentos em um único grupo, escolhendo um grupo, particionando-o em outros grupos e prossegue escolhendo e dividindo até que cada grupo terminal da árvore possua somente um documento, ou outro critério de parada seja alcançado. Como observado em [Zhao and Karypis 2002], para os algoritmos aglomerativos falta a visão global de possíveis grupos coesos, e eventuais decisões de combinação erradas, no começo da execução do algoritmo, tendem a se multiplicar à medida que o agrupamento é executado. Algoritmos divisivos, por sua vez, têm uma melhor visão global de possíveis grupos coesos, e por isso, serão o foco deste trabalho. Um algoritmo divisivo bastante difundido é o *bisecting K-means* [Steinbach et al. 2000], em que o simples algoritmo *K-means* é usado para dividir cada grupo em dois sub-grupos em cada passo de divisão. O *bisecting K-means* mostrou ser um concorrente competitivo quando comparado aos algoritmos aglomerativos [Steinbach et al. 2000].

3. Seleção de características em agrupamento de documentos

A seleção de características em agrupamento de documentos é a tarefa de desconsiderar características irrelevantes e redundantes nos vetores que representam os documentos, com o objetivo de achar o menor subconjunto de características que revelem grupos "naturais" dos documentos. O uso do menor subconjunto possível irá tornar o tempo de computação da tarefa de agrupamento menor, ao mesmo tempo que evita a praga da dimensionalidade.

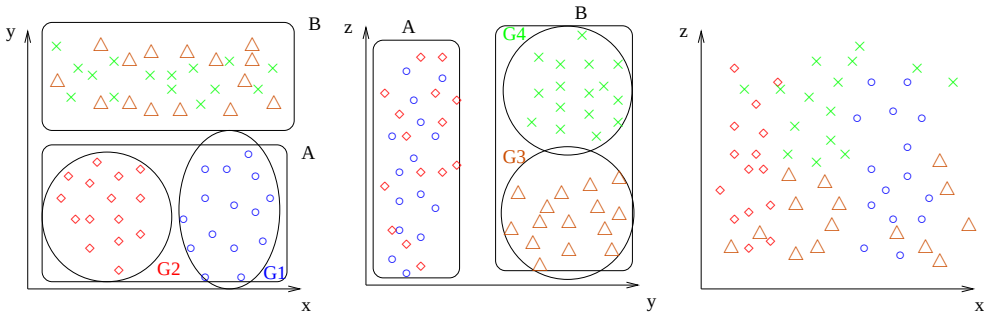
Métodos para realizar seleção podem ser classificados em **filtros**, quando usam alguma propriedade estatística dos dados para ordenar as características pela qualidade de cada termo [Dhillon et al. 2003, Tang et al. 2005] (veja Seção 5.1.1), ou ***wrappers***, quando fazem uma busca por características usando como critério uma função de avaliação interna do agrupamento. Como os métodos *wrappers* exigem várias execuções

do algoritmo de agrupamento, a sua aplicação para agrupamento *online* de documentos torna-se inviável. Portanto, boa parte dos métodos utilizados para seleção são métodos filtros, onde a seleção é feita com o uso de um limiar ou um número fixo de quantas características são desejadas.

3.1. Seleção global e local de características

A seleção de características pode ocorrer de forma global ou local. A seleção global de características é o processo que seleciona uma vez as características e considera sempre as mesmas características no processo de descoberta dos grupos, sendo a forma mais pesquisada até então [Dash et al. 2002, Dy and Brodley 2004, Tang et al. 2005]. Na seleção local de características, um subconjunto de características é escolhido para cada grupo, considerando a suposição que cada grupo possui características mais importantes, que ajudam a discriminar cada grupo dos demais grupos.

A Figura 1 ilustra um conjunto de objetos descritos pelas características x, y e z. Os grupos G1 e G2 são somente revelados quando os atributos x e y são considerados, isto é, o atributo z é irrelevante para discriminar entre G1 e G2 (ver Figura 1(a)). A Figura 1(b), por sua vez, mostra que as características y e z são relevantes para identificar os grupos G3 e G4, isto é, a característica x é irrelevante neste contexto. Finalmente, os atributos x e z correspondem a um subconjunto irrelevante de características (Figura 1(c)). Nesta situação, qualquer subconjunto com duas 2 características eventualmente retornado por um método global não é capaz de identificar os quatro grupos existentes. É necessário examinar uma característica no contexto de diferentes subconjuntos antes de afirmar que uma característica é realmente irrelevante [Dy and Brodley 2004].



(a) em x e y (b) em y e z (c) em x e z

Figura 1. Dados dos grupos G1, G2, G3 e G4 para diferentes características.

Comparado à abordagem global, há poucos trabalhos relevantes na literatura de agrupamento que investigam a seleção local de características. Em [Li et al. 2008], por exemplo, os autores propõem um método de seleção local para agrupamento, buscando vários subconjuntos de características relacionados à formação de grupos diferentes, escolhendo os grupos mais coesos baseado em um critério de avaliação do agrupamento. Em [Li et al. 2008], experimentos foram realizados para avaliar o método local proposto para o algoritmo *K-means*. Com o uso do método local, os autores obtiveram uma melhoria na precisão do agrupamento para dados não-textuais do repositório de aprendizagem de máquina da Universidade da Califórnia – Irvine (UCI). Foi vislumbrado aqui que, até o momento, a abordagem local para seleção de características ainda não foi aplicada em qualquer trabalho sobre agrupamento de documentos. Como o método de [Li et al. 2008]

usa *wrapppers*, alternativa muito custosa computacionalmente, sua aplicação direta em agrupamento *online* de documentos é impraticável. Na próxima seção será descrito o método proposto neste trabalho na tentativa de alcançar os benefícios esperados com o uso de uma abordagem local, sem recorrer ao uso de *wrappers*.

4. Método proposto

Neste trabalho, é proposto um algoritmo de agrupamento hierárquico divisivo com seleção local de atributos, onde é esperado que a privilegiada visão global em algoritmos divisivos possa tomar proveito de uma visão local proporcionada pela seleção de características local, e que a variedade de subconjuntos de características selecionados a cada divisão dos grupos possa revelar grupos ocultos por características ruidosas. O algoritmo proposto para seleção de características local em agrupamento hierárquico divisivo utilizando o *bisecting K-means* segue os passos:

1. Escolha um grupo para dividir, considerando um grupo inicial contendo todos os dados.
2. Faça a seleção de características do grupo escolhido, usando como critério a escolha de n características ou de um limiar τ (**passo de seleção local**).
3. Construa 2 sub-grupos usando o algoritmo *K-means* (passo de divisão)
4. Repita o passo 3 por $ITER$ vezes e fique com a partição com melhor valor do critério interno.
5. Repita os passos 1, 2, 3 e 4 até que o número de grupos requerido seja alcançado.

O problema da Figura 1 pode ser resolvido começando por um subconjunto de características que melhor revele grupos nos dados (pode ser o caso da Figura 1(a)) e particionando todos os dados em grupo A (constituem os dados de G1 e G2) e grupo B (os dados de G3 e G4). Então é realizado uma nova seleção de características para os dados de cada grupo, o grupo A permanece com as características x e y e grupo B com y e z. Os grupos A e B agora podem ser facilmente discriminados em G1 e G2 (filhos do grupo A), G3 e G4 (filhos do grupo B), revelando assim todos os grupos presentes nestes dados.

Um aspecto importante a ser considerado neste algoritmo é o número N de características a serem selecionados para cada grupo. À medida que o algoritmo é executado, os grupos criados ficam cada vez menores e, com isto, o número de características distintos presentes nos documentos também diminui. Desta forma, a escolha de um número constante de características para cada grupo a ser particionado tende a perder seu potencial seletivo (capacidade de selecionar características relevantes), pois o número de características selecionadas irá se aproximar do número de características distintas. Uma alternativa seria escolher um número pequeno de características, mas assim se perde informação quando os grupos ainda são grandes.

Uma solução para os problemas acima é a escolha de um número de características variável de acordo com o tamanho dos grupos e o número de características distintas. De maneira simples, neste trabalho é proposto a escolha do número de características n_i para o grupo i igual a:

$$n_i = \left\lfloor \frac{N_T}{N_C} \cdot m_i \right\rfloor \quad (1)$$

onde N_T é o número de características distintas em toda coleção de documentos, N_C é o tamanho da coleção de documentos e m_i é o tamanho do grupo i. N_T/N_C é a proporção de características reveladas distintas em cada documento da coleção. Como o procedimento de diminuir o número de características selecionadas localmente a cada divisão de grupo lembra o ajuste de um binóculo, este método será referido neste trabalho como método ZOOM-IN.

Este procedimento de selecionar características localmente em agrupamento de documentos pode ser visto como análogo ao proposto nos trabalhos [Esuli et al. 2008, Koller and Sahami 1997] para classificação de documentos, onde é chamado de "glocal", já que é realizada uma seleção *global* de características a cada divisão dos grupos e é *local* pois ele continua a fazer a seleção de características nos grupos gerados a cada divisão, de maneira que cada conjunto de nós-irmãos no dendograma formado é representado por um subconjunto de características diferente.

O algoritmo *bisecting K-means* reduz a quantidade de dados envolvida nos cálculos, a medida que seleciona os grupos a serem particionados. Com a introdução da seleção local de características, o número de características também é reduzido durante a execução do agrupamento, o que contrabalanceia o custo computacional adicional em realizar a seleção de características para cada partição de um grupo.

Como será visto, foram realizados experimentos com seleção local para um número constante de características selecionadas e com ZOOM-IN para decidir o número de características selecionadas a cada iteração.

5. Experimentos e resultados

A Seção 5.1 descreve os experimentos realizados para avaliar a viabilidade do método proposto. A Seção 5.2, por sua vez, apresenta os resultados obtidos.

5.1. Descrição dos experimentos

Nos experimentos, foi utilizado um subconjunto da base de documentos Reuters-21578 [Lewis 1999], considerando somente os documentos com um único tópico, o que constitui a classe, representando um número total de 1228 documentos associados a 42 classes, tendo sido selecionado no máximo 30 documentos aleatoriamente para cada classe. Os documentos coletados, que são todos escritos em Inglês, foram processados de maneira a remover *stopwords* (preposições e palavras comuns). Também foi aplicado uma função de *stemming* com o algoritmo de Porter [Oleander Solutions] e remoção de características que ocorram menos que 5 vezes em toda a base. Os documentos são representados por vetores de características com valores calculados usando o difundido método TF-IDF.

Nos experimentos, foi adotada uma abordagem de validação externa para avaliar a qualidade dos agrupamentos. Os resultados de agrupamento foram comparados usando a micro-média de precisão, também utilizada, por exemplo, no trabalho de [Slonim et al. 2002]. Os grupos gerados são comparados com as classes conhecidas *a priori*. A micro-média de precisão assume que cada grupo formado pelo algoritmo de agrupamento possui uma classe c representante que é majoritária. Considerando T o conjunto de grupos e C o conjunto de classes, a micro-média de precisão é dada por [Slonim et al. 2002]:

$$P(T) = \frac{\sum_{c \in C} \alpha(c, T)}{\sum_{c \in C} \alpha(c, T) + \beta(c, T)}$$

onde $\alpha(c, T)$ é o número de documentos corretamente afetados para c e $\beta(c, T)$ é o número de documentos de documentos incorretamente afetados para c.

Para avaliação do agrupamento hierárquico, os grupos considerados são aqueles presentes nas folhas do dendograma. Os números de grupos especificados para execução dos algoritmos foi igual ao número de classes da base. O número ITER do algoritmo *bisecting K-means* foi atribuído com valor 5, a medida de similaridade utilizada foi a cosseno e todos os valores de precisão apresentados neste trabalho são a média de 30 execuções dos algoritmos para os mesmos parâmetros.

5.1.1. Critérios de ordenação de características

Nesta seção, são citados os critérios que irão ser aplicados nos experimentos realizados e que apresentaram bons resultados em trabalhos anteriores.

- **Freqüência em documentos (DF).** O valor DF do termo t é definido como o número de documentos em que o termo t ocorre ao menos uma vez na coleção de documentos [Tang et al. 2005].
- **Variância de freqüência do termo (TfV).** Sendo tf_j a freqüência do termo t no documento d_j, a qualidade do termo t é calculada por:

$$TfV_t = \sum_{j}^{n} tf_j^2 - \frac{1}{n}\left[\sum_{j}^{n} tf_j\right]^2 \quad (2)$$

  onde n é o número de documentos na base. Nos experimentos realizados em [Dhillon et al. 2003], o método TfV consegue manter a precisão do processo de agrupamento com até 15% do número total de características.
- **Média do TF_IDF (TI).** TI é um método proposto em [Tang et al. 2005], calculado pela média do valor de TF_IDF para todos os documentos $(j = 1, \ldots, n)$ na base. O método TI mostrou ter um desempenho superior ao DF e similar ao TfV.

5.2. Resultados e discussão

Na Figura 2, são apresentados os resultados obtidos para a base Reuters utilizando *bisecting K-means* com seleção de características global, que é a maneira tradicional, e local, utilizando o método proposto, com o número de características constante, para os critérios de ordenação DF, TfV e TI. Como pode ser visto, para todos os métodos globais, quando poucas características são selecionadas a precisão do método global cai. O critério de ordenação que obtém os melhores valores de precisão é o TfV, com desempenho similar ao TI e melhor que o DF, como já foi observado no trabalho de [Tang et al. 2005]. Pode-se concluir que para poucas características selecionadas, há pouca informação nos documentos, o que deteriora a precisão do agrupamento.

Por outro lado, a abordagem local consegue manter a precisão até com uma quantidade de características muito pequena, com exceção do método DF. Além disso, foi

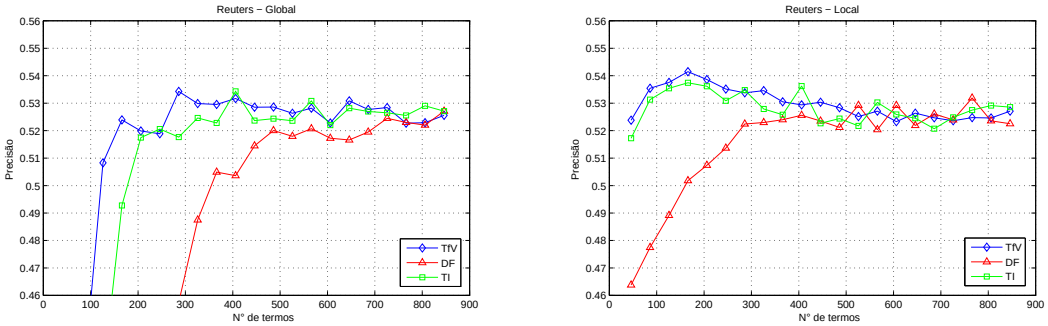


Figura 2. Micro-média de precisão em relação ao número de características utilizadas para a base Reuters, com seleção de características global e local

observado um fenômeno interessante para um número relativamente pequeno de características, onde a precisão aumenta. Isto se deve ao fato que para grandes valores do número de características selecionadas, o potencial seletivo diminui localmente, pois de acordo com que os grupos vão ficando menores, existem menos características distintas. Com um valor pequeno de características selecionadas, o potencial seletivo consegue se manter durante as divisões dos grupos e, de maneira interessante, a precisão melhora. Isso indica que a seleção local ajuda para que as divisões dos grupos não venham a separar grupos coesos.

No entanto, um valor pequeno de características selecionadas pode comprometer a quantidade de informação necessária no início da atividade de agrupamento, quando os grupos ainda são grandes e a quantidade de características distintas também. É necessário selecionar as características que refletem um real benefício ao agrupamento, função bem desempenhada por métodos *wrappers* [Dy and Brodley 2004]. Como o custo computacional do uso de *wrappers* não é prático para agrupamento *on-line* de textos, o uso de uma quantidade variável de características selecionadas localmente, que aqui é chamado método ZOOM-IN, objetiva manter o mesmo potencial seletivo durante o agrupamento. Nos experimentos realizados com o método ZOOM-IN, apresentados na Tabela 1, é comprovado um bom desempenho desta abordagem, com valores de precisão equivalentes ao caso de poucas características selecionadas localmente com o uso dos métodos TfV, DF e TI.

Tabela 1. Micro-média de precisão com o método ZOOM-IN

Base	Sem seleção	TfV	DF	TI
Reuters	0.527117	0.540988	0.527362	0.541395

6. Conclusões

Neste trabalho, foi proposto o uso de seleção local de características para agrupamento hierárquico divisivo de documentos. Cada conjunto de nós-irmãos derivados pelo método proposto é representado por um subconjunto diferente de características. Nos experimentos realizados, a abordagem local foi comparada com a abordagem global de seleção de características para o algoritmo *bisecting K-means*. Foi observado que a abordagem

local obtém uma boa precisão até para poucas características selecionadas. Também foram realizados experimentos usando o método ZOOM-IN para automaticamente definir o número de características selecionadas em cada iteração do algoritmo divisivo. Os resultados obtidos com o método ZOOM-IN foram satisfatórios, pois mostraram os benefícios em selecionar características localmente.

Como trabalho futuro, é intencionado avaliar outros critérios para ordenar características, que usam informação da similaridade entre os documentos, tal como a ordenação baseada em entropia [Dash et al. 2002], além da avaliação do método proposto em outras coleções de documentos.

Como os métodos *wrappers* são uma alternativa muito custosa de selecionar características, sua aplicação em agrupamento de documentos não possui trabalhos relevantes que analisam o custo computacional em relação ao ganho de precisão e em que condições métodos *wrappers* podem ser aplicáveis. Investigar o uso de wrappers para realizar seleção local de características em agrupamento de documentos pode constituir um avanço na área.

Referências

Dash, M., Choi, K., Scheuermann, P., and Liu, H. (2002). "Feature selection for clustering - a filter solution". In *Proceedings of the 2002 IEEE International Conference on Data Mining*, pages 115–122, Washington, DC, USA. IEEE Computer Society.

Dhillon, I., Kogan, J., and Nicholas, C. (2003). "Feature selection and document clustering". In Berry, M. W., editor, *Survey of Text Mining*, pages 73–100. Springer.

Dy, J. G. and Brodley, C. E. (2004). "Feature selection for unsupervised learning". *Journal of Machine Learning Research*, 5:845–889.

Esuli, A., Fagni, T., and Sebastiani, F. (2008). "Boosting multi-label hierarchical text categorization". *Information Retrieval*, 11(4):287–313.

Koller, D. and Sahami, M. (1997). "Hierarchically classifying documents using very few words". In *ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning*, pages 170–178, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Lewis, D. D. (1999). "Reuters-21578 text categorization test collection distribution 1.0". `http://www.daviddlewis.com`.

Li, Y., Dong, M., and Hua, J. (2008). "Localized feature selection for clustering". *Pattern Recognition Letters*, 29(1):10–18.

Oleander Solutions. "Oleander Stemming Library". `http://www.oleandersolutions.com/stemming/stemming.html`.

Sahoo, N., Callan, J., Krishnan, R., Duncan, G., and Padman, R. (2006). "Incremental hierarchical clustering of text documents". In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 357–366, New York, NY, USA. ACM.

Salton, G., Wong, A., and Yang, C. S. (1975). "A vector space model for automatic indexing". *Communications of the ACM*, 18(11):613–620.

Slonim, N., Friedman, N., and Tishby, N. (2002). "Unsupervised document classification using sequential information maximization". In *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 129–136, New York, NY, USA. ACM.

Steinbach, M., Karypis, G., and Kumar, V. (2000). "A comparison of document clustering techniques". Technical report, Department of Computer Science and Engineering, University of Minnesota.

Tang, B., Shepherd, M., Milios, E., and Heywood, M. I. (2005). "Comparing and combining dimension reduction techniques for efficient text clustering". In *International Workshop on Feature Selection for Data Mining*.

Zhao, Y. and Karypis, G. (2002). "Evaluation of hierarchical clustering algorithms for document datasets". In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 515–524, New York, NY, USA. ACM.