

Formação de *clusters* para o gerenciamento da cadeia de suprimentos em operações humanitárias

Formation of clusters for supply chain management in humanitarian operations

Fabiana Santos Lima

Mestre em Engenharia de Produção, Doutoranda do Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina – UFSC.
Florianópolis, SC [Brasil]
fsantoslima1@gmail.com

Daniel de Oliveira

Mestre em Computação Aplicada, Doutorando do Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina – UFSC.
Florianópolis, SC [Brasil]

Mirian Buss Gonçalves

Mestre em Matemática e Computação Científica, Doutora do Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina – UFSC.
Florianópolis, SC [Brasil]

Resumo

Neste trabalho, tem-se como objetivo apresentar uma metodologia para formar *clusters* por combinação de frequência de desastres no estado de Santa Catarina – Brasil, utilizando mineração de dados. Pode-se assim auxiliar na coordenação estratégica, na definição de prioridades e no compartilhamento de experiências entre os municípios. Diante do exposto, a metodologia proposta nesta pesquisa é verificar a frequência por tipo de desastre em todos os municípios e identificar regiões semelhantes para poder padronizar e sugerir um método de prevenção e, com isso, melhorar e auxiliar os processos de tomada de decisão diante dos eventos de logística humanitária. Para cada um dos *clusters*, as experiências e conhecimentos podem ser compartilhados, o que poderá servir de base para o planejamento integrado, visando a preparar os processos logísticos na iminência dos desastres. Um experimento computacional, aplicando a metodologia proposta, foi realizado, utilizando o algoritmo *k-means*. Os resultados obtidos são apresentados e analisados ao final.

Palavras-chave: Cadeia de suprimentos. *Clusters*. Desastres naturais. Logística humanitária.

Abstract

This paper aims to present a methodology using data mining for clustering by combination frequency of disasters in the state of Santa Catarina – Brazil. This can help strategic coordination in setting priorities and sharing experience among municipalities. Hence, the methodology proposed in this research is to check the frequency by type of disaster in all municipalities and identify similar regions to standardize and suggest a method of prevention, thereby improving and assisting in the decision-making processes in the face of events involving humanitarian logistics. For each cluster, the experience and knowledge can be shared, which could provide a basis for integrated planning in order to prepare the logistics processes when disaster is imminent. A computational experiment applying the proposed methodology was carried out using the k-means algorithm. The results are presented and analyzed at the end.

Key words: Clusters. Humanitarian logistics. Natural disasters. Supply chain.

1 Introdução

Organizações de ajuda humanitária coordenam bilhões de dólares anualmente no alívio às vítimas de catástrofes naturais, conflitos civis e guerras. Sua tarefa principal é a mobilização oportuna de financiamento e de bens de doadores internacionais e administração de alívio para os beneficiários vulneráveis em locais de desastres em todo o globo. Como tal, a logística é fundamental para suas atividades e estratégica para as suas missões.

Segundo pesquisa apresentada por Sapir (2011), no ano de 2010, ocorreram, no mundo, 373 desastres naturais, tais como secas, terremotos, altas temperaturas, enchentes, movimentos de terra seca e de terra molhada, tempestades, vulcões e incêndios, provocando 296.800 mortes e um total de 207 milhões de pessoas afetadas, com um prejuízo estimado em 109 bilhões de dólares. Comparando o número desses desastres, no referido ano, com a média da década de 2000 a 2009, constata-se que não houve crescimento das catástrofes mundiais. Entretanto, analisando um estudo mais amplo de mortes, verifica-se um aumento representativo de óbitos nos últimos dez anos.

Segundo Chade (2009), o Brasil, em 2008, foi o 13º país mais afetado por desastres naturais, que atingiram 2 milhões de pessoas, principalmente, pela ação das chuvas; e 1,5 milhão de pessoas sofreram com as catástrofes em Santa Catarina. Conforme estatísticas da Defesa Civil desse estado, somente nos meses de janeiro até março de 2011 foram registrados 152 eventos, como alagamentos, erosões, pragas, ressacas, vendavais e enxurradas, sendo este último o evento predominante, com 143 ocorrências.

Em todos estes casos o atendimento aos desastres foi difícil e complicado. Muitos fatores, como o grande número de atores envolvidos na ajuda humanitária e a falta de recursos suficientes, contribuíram para as dificuldades de coordenação de so-

corro (FENTON, 2003 apud BALCIK et al., 2010). Existem poucas histórias de sucesso de coordenação, de modo que “[...] a coordenação continua a ser a fraqueza fundamental da ação humanitária [...]” (Rey, 2001 apud BALCIK et al., 2010, p. 22).

O termo coordenação tem interpretações variadas dentro do ambiente de alívio ao desastre. Por exemplo, pode referir-se à coordenação de recursos e partilha de informação, centralização das decisões, a realização de projetos comuns, a divisão regional de tarefas, ou um sistema baseado em *clusters*, em que cada *cluster* representa uma área de setor diferente (alimentos, água, saneamento e tecnologia da informação). Trabalha-se também na coordenação de *clusters* por região, que permite analisar a frequência e o alcance geográfico dos eventos, resultando num perfil de classificação dos desastres (ALMEIDA; ZARATE, 2007; SOUZA, 2004).

Diante do contexto, neste estudo, objetiva-se contribuir para a logística humanitária por meio de uma metodologia voltada a formar *clusters* por combinação de frequência de desastre com o intuito de auxiliar na coordenação estratégica, na definição de prioridades e no compartilhamento de experiências em determinadas regiões.

Verificar a frequência por tipo de catástrofe, possibilita identificar regiões semelhantes para então poder padronizar um método na prevenção e resposta que pode, assim, auxiliar os órgãos operacionais competentes na tomada de decisão referente ao tipo de capacitação necessária e investimentos em infraestrutura. Esta metodologia será desenvolvida utilizando a técnica de Mineração de Dados (MD).

Cabe ressaltar que a pesquisa realizada, que resultou neste artigo, tem por objetivo a continuidade do apresentado em Lima, Oliveira e Gonçalves (2011), em que se verificou a frequência por tipo de desastre utilizando a divisão política das microrregiões no estado de Santa Catarina. Já

neste trabalho o foco é para formação de *clusters* por municípios, com o intuito de eliminar possíveis ruídos que podem ter surgido em virtude da divisão política das microrregiões, que nem sempre refletem a realidade geográfica a ser considerada para cada município impactado.

O artigo inicia-se abordando os conceitos associados à logística humanitária, ao gerenciamento da cadeia de suprimentos em operações humanitárias, e são referenciados alguns modelos voltados à gestão em desastres. Em seguida, contextualiza-se e referencia-se mineração de dados e destaca-se seu principal objetivo e características. Por fim, apresenta-se a metodologia proposta com uma aplicação para o estado de Santa Catarina. Nos itens seguintes serão relatadas as considerações finais e bibliografia.

2 Logística humanitária

O conceito de logística humanitária foi desenvolvido a partir dos objetivos logísticos de vencer o tempo e a distância na movimentação de materiais e serviços de forma eficiente e eficaz. Nesse sentido, surge a logística humanitária, um conceito que vem sendo desenvolvido e aplicado, principalmente, em países da Europa e nos Estados Unidos, mas que ainda é muito recente no Brasil (NOGUEIRA; GONÇALVES, 2010).

Logística humanitária se refere aos processos e sistemas envolvidos na mobilização de pessoas, recursos, habilidades e conhecimento para ajudar as pessoas vulneráveis afetadas por desastres naturais e emergências complexas. Engloba uma série de atividades, incluindo compras, transporte, detecção e acompanhamento, desembarço aduaneiro, transporte interno, armazenamento e entrega de última milha (THOMAS, 2007).

Embora a logística humanitária tenha que tratar de circunstâncias especiais e de enormes de-

safios, os princípios básicos da logística comercial permanecem válidos e podem ser adaptados e aplicados. A característica específica neste contexto é o envolvimento da vida humana, a qual depende do desempenho eficiente e eficaz das operações de resposta, o que significa que na logística humanitária o auxílio deve chegar ao seu destino de maneira correta e em tempo oportuno, sempre com foco no alívio do sofrimento e na preservação da vida (NOGUEIRA et al., 2009).

Nesse sentido, é importante destacar algumas das características que apontam grandes desafios à logística humanitária (MEIRIM, 2007):

- **Materiais:** O que é necessário? Para onde deve ser enviado? Acúmulo de doações nas primeiras semanas, gerando, assim, desperdícios e avarias, devido a itens inadequados.
- **Ausência de processos coordenados:** informações, pessoas e materiais.
- **Infraestrutura:** na maioria dos casos, está destruída, o que dificulta o acesso, a chegada de recursos e a saída de pessoas.
- **Recursos humanos:** excesso de pessoas (voluntários) sem treinamento adequado, heróis que agem somente com a emoção, celebridades que só querem aparecer nesse momento, pessoas que vão para o local e não conhecem a magnitude do problema.

Em resumo, pode-se dizer que a logística humanitária propõe o uso efetivo dos conceitos logísticos adaptados às especificidades da cadeia de assistência humanitária. Esses conceitos podem ser o grande diferencial no sentido de minimizar ações de improvisação, muito comuns nessas ocorrências, maximizando a eficiência e o tempo de resposta à situação de emergência.

Para melhor entender a cadeia de assistência humanitária, é fundamental trabalhar os conceitos de gerenciamento da cadeia de abastecimento,

já reconhecidos e consolidados cientificamente, sendo um instrumento fundamental na operacionalização do processo de assistência humanitária.

2.1 Gerenciamento da cadeia de suprimentos em operações humanitárias

As agências humanitárias precisam olhar para além da logística básica e usar a abordagem de gerenciamento da cadeia de suprimentos para coordenar os diferentes cenários envolvidos em uma operação de socorro. As técnicas de gerenciamento dessa cadeia tornaram-se mais comuns na área humanitária com o cruzamento de atividades e aproximação entre as agências.

A coordenação de fluxos de materiais e de informação entre os intervenientes da cadeia de abastecimento tem sido amplamente abordada no domínio da gestão da cadeia de abastecimento comercial. Foi demonstrado que a falta de coordenação entre os membros da cadeia aumenta os custos de estoque, alongam os prazos de entrega e comprometem o serviço ao cliente (SIMATUPANG et al. 2002 apud BALCIK et al., 2010). Conforme Tomasini e Wassenhove (2009), 80% dos custos nas operações de socorro são devidos à logística; portanto, a coordenação da cadeia nesse tipo de operação é fundamental para melhorar o seu desempenho.

O desastre do tsunami, por exemplo, reforça o que foi levantando anteriormente. Grande número de aviões com suprimentos sendo desviados, atrasos na distribuição, agências de assistência se esforçando para localizar depósitos e acomodar pessoas. Segundo relatos de pessoas envolvidas em operações de emergência, verifica-se que os muitos dos envolvidos não compreendem a complexidade logística até o momento de ocorrência de uma catástrofe (NOGUEIRA et al., 2009).

Para Tomasini e Wassenhove (2009), o gerenciamento da cadeia de suprimentos em emergência é também um resultado dos seguintes fatores:

- A revolução da informação: a informação é tão rápida que em tempo real a notícia já está no ar.
- O pedido perfeito: a cadeia de fornecimento humanitária deve ser mais reativa à mudança súbita. Ineficiências e erros nessa cadeia são menos tolerados. Existe uma pressão para executar pedidos perfeitos (*perfect orders*).
- Relacionamentos interorganizacionais: o aumento da pressão de pedidos perfeitos força as agências humanitárias independentes a quebrarem seus “silos” e a construir novas relações estratégicas entre os setores envolvidos.

Essas novas formas de relações interorganizacionais são fundamentais para gerir eficazmente uma demanda crescente da sociedade civil de ajuda em situações de emergência. A coordenação interagências é fundamental para prevenir um sistema entupido com boas intenções, mas, por vezes, mal geridas e descoordenadas.

Pode-se definir, segundo Tomasini e Wassenhove (2009), 5Bs que auxiliam na classificação do Gerenciamento da Cadeia de Suprimentos: *Boxes*: fluxo de produtos físicos; *Bytes*: transmissão de pedidos, coordena o fluxo de produtos; *Bucks*: fluxo financeiro (doações ou pagamentos); *Bodies*: pessoas envolvidas na cadeia e *Brains* (conhecimento e habilidades): necessidade de rapidez no reconfiguramento da cadeia. A Figura 1 ilustra esta relação.

Os 5Bs dessa cadeia necessitam de um modelo coerente de negócios e um *design* apropriado para o sistema, executado do início ao fim da cadeia.

Assim como na atual visão da Supply Chain Management, que representa o esforço de integração dos diversos participantes do canal de distribuição, por meio de administração compartilhada de processos-chave de negócios, que interligam as diversas unidades organizacionais e membros do canal, a cadeia humanitária preserva esta abordagem sistêmica, o que implica em alta interação entre os participantes.

As considerações de “última milha” também são extremamente importantes. Em outras palavras: como fazer chegar a ajuda ao povoado mais distante, ou a um lugar sem infraestrutura e sem transporte. Em alguns casos, o volume da ajuda oferecida é muito grande, mas não há como fazê-la chegar aos lugares atingidos. A distribuição final, talvez seja um dos desafios logísticos mais cruciais, já que grande parte da infraestrutura encontra-se destruída ou danificada. Uma maneira de superar esse problema, de certa forma, consiste em recorrer à ajuda dos meios de comunicação e informação, de modo que os itens prioritários cheguem aos lugares onde há maior necessidade deles (NOGUEIRA et al., 2007).

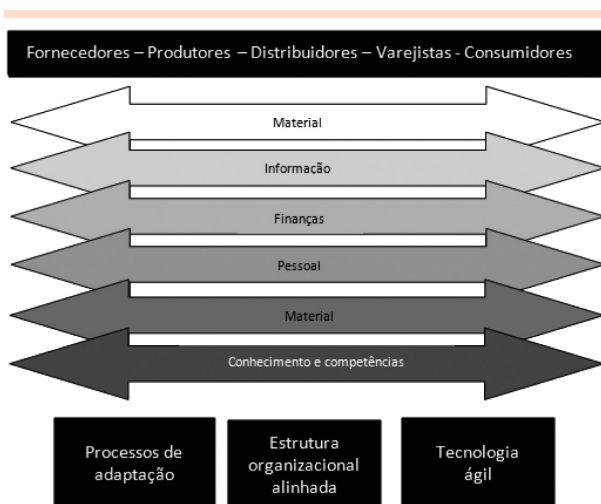


Figura 1: Fluxos da cadeia de abastecimento

Fonte: Adaptado de Wassenhove e Tomasini, 2009.

Conforme Tomasini e Wassenhove (2009), a chave para a competitividade na cadeia humanitária é estender a preocupação com a eficiência de custos e velocidade, de modo a incluir:

- **Agilidade:** capacidade de fornecer uma resposta rápida às mudanças de demanda ou perturbações externas ao fornecimento.
- **Adaptabilidade:** habilidade de se ajustar para atender às mudanças estruturais nos

mercados e modificar a rede de abastecimento se necessário.

- **Alinhamento:** cria condições para um melhor desempenho e requer intercâmbio de informações com todos os parceiros relevantes. Aqui se definem as responsabilidades de todos os stakeholders para criar o senso de unidade e identificação.

Podem-se citar algumas características na cadeia humanitária de suprimentos (TOMASINI; WASSENHOVE, 2009):

- **Objetivos ambíguos:** dificuldade de avaliar o nível de comprometimento dos diversos atores e seu relacionamento na cadeia.
- **Recursos limitados:**
 - **Humano:** alta rotatividade, pesadas exigências físicas e emocionais, poucas pessoas qualificadas.
 - **Capital:** nem sempre disponível no tempo e local necessário.
 - **Infraestrutura:** frequentemente afetada por catástrofes, inexistente ou insuficiente para a magnitude das necessidades.
- **Alta incerteza:** constante mudança de fornecedores, incerteza nas quantidades doadas por cada agente.
- **Urgência:** intervenções após desastres são caracterizadas pela urgência.
- **Ambiente politizado:** as ações humanitárias são altamente políticas em toda a cadeia de abastecimento, a partir de doações para distribuição. É difícil manter e proteger o espaço humanitário, no qual os humanitaristas podem fazer seu trabalho de apoio independentemente das pressões externas.

Sabe-se que nenhuma crise envolve apenas um fator ou necessidade, surgindo daí a necessidade de que essas agências trabalhem juntas

para atingirem resultados sustentáveis e efetivos (TOMASINI; WASSENHOVE, 2009).

Considerando a importância da cooperação na logística humanitária, apresentam-se, no próximo item, alguns modelos utilizados na gestão em desastres.

2.2 Gestão em desastres

A gestão de desastres pode ser definida como o conjunto de atividades destinadas a manter o controle sobre as situações de desastre e emergência e fornecer uma estrutura para ajudar pessoas em risco, para evitar ou recuperar os efeitos do evento danoso. Essa gestão trata de situações que ocorrem antes, durante e depois do desastre. Seus objetivos são: evitar ou reduzir as perdas humanas, físicas e econômicas sofridas pelas pessoas, pela sociedade e pelo país em geral; reduzir o sofrimento dos indivíduos; acelerar a recuperação da área afetada (SCHULTZ, 2008).

2.2.1 Fases e Ciclos de alívio de desastres

Pesquisas referentes à gestão de desastres estão sendo desenvolvidas com o intuito de diminuir as dificuldades encontradas na coordenação dos diversos tipos de catástrofes.

Consideram-se três fases, ilustradas na Figura 2, que tradicionalmente envolvem a gestão em situações emergenciais:

1. Pré-desastre (prevenção, mitigação e preparação);
2. Resposta (advertência, impacto e resposta de emergência);
3. Pós-desastre (transição, reabilitação e reconstrução).

O desempenho de qualquer resposta ao desastre depende do nível de preparação, ou seja, da fase de pré-desastre. Segundo um modelo de simulação criado pelo Massachusetts Institute of Technology (MIT), as pequenas despesas relativas à preparação para as catástrofes diminuí significativamente o tempo e o custo da logística de resposta (GRALLA, 2007 apud SCHULTZ, 2008).

Uma vez que uma melhor prevenção e mitigação dos desastres pode e deve fazer parte da etapa de reconstrução e desenvolvimento, o modelo das três fases pode ser convertido em um ciclo contínuo de gestão de desastres, como mostrado na Figura 3

Segundo Nogueira e Gonçalves (2010), aspectos ligados à infraestrutura, localização de centrais de assistência, distribuição de recursos, coordena-

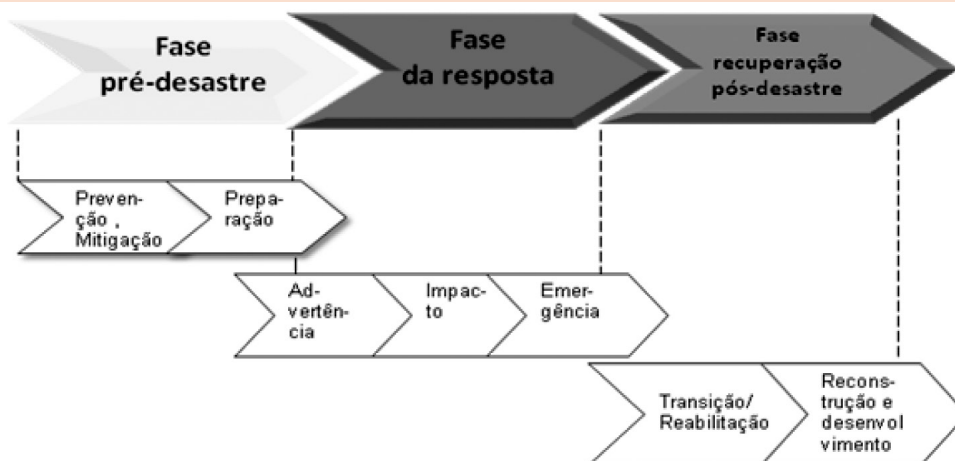


Figura 2: Modelo das três fases

Fonte: Adaptado de Tufinkgi, 2006.



Figura 3: Ciclo contínuo de gestão de desastres

Fonte: Adaptado de Schultz, 2008.

ção de processos (pessoas, suprimentos, informações, materiais) merecem destaque nos processos logísticos sistematizados na logística humanitária.

Ainda para as autoras, no momento do desastre ou de sua iminência, precisa-se considerar: onde localizar uma (ou várias) central (ais) de recebimento, controle e distribuição de materiais; como organizar e controlar o recebimento e a distribuição desses materiais; como fazer a distribuição na rede. Assim, é proposta a implantação de uma central de inteligência e suporte, de caráter permanente, que atue como indutora de políticas de preparação para situações de desastre. A Figura 4 mostra os principais problemas na iminência do desastre, conforme indicado pelas autoras.

A abordagem deste trabalho está relacionada com o momento pré-desastre, podendo ser uma das políticas de preparação para a situação de desastres da Central de Inteligência de Suporte de Caráter Permanente.

Trabalhos como este podem contribuir na iminência do desastre e também auxiliar na contribuição dos três problemas.

Considera-se que um mesmo desastre pode atingir diferentes municípios ao mesmo tempo, exigindo a coordenação entre eles para o desenvolvimento de ações de socorro ou auxílio. Além disso, o impacto de uma catástrofe sobre um município específico pode exigir o auxílio de outro, seja na forma de unidades de resgate, seja em empréstimo de equipamentos, assessoria de especia-

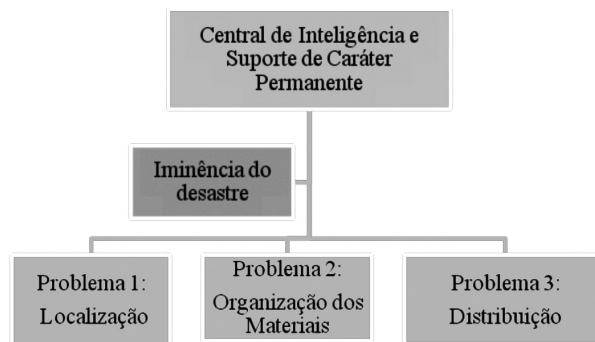


Figura 4: Principais problemas na iminência do desastre

Fonte: Adaptado de Nogueira e Gonçalves, 2010.

listas ou recolhimento de doativos. As operações nessas situações podem ter peculiaridades muito específicas. São características que podem tornar a coordenação dessas operações bastante difícil se não houver um sistema padronizado, comum para o planejamento e implementação das ações.

Referencia-se neste artigo o trabalho desenvolvido por Tufinkgi (2006) que descreve processos de modelagem logística em desastres, utilizando um Modelo de Processo Logístico de Socorro em Catástrofes Internacionais (MPLSCI), em que considera o gerenciamento logístico de desastres sob dois focos: de um lado não específico (genérico), e de outro específico, considerando suas interdependências e referenciando a interação dos dois lados (genérico e específico). O MPLSCI proposto por Tufinkgi (2006) é definido em sete etapas, em que as quatro primeiras etapas estão relacionadas a casos genéricos e são pré-requisitos para as três etapas restantes que estão relacionadas a casos específicos de desastres. A Figura 5 apresenta este modelo.

Assim, este artigo volta-se para o primeiro item do MPLSCI demonstrado na Figura 5, ou seja, com foco nas funções dadas pela região de clusterização e análise de risco, que são as seguintes:

- a) Estabelecer o perfil da extensão de cada desastre tomando como base a análise de dados

obtidos a partir de catástrofes no passado, isto é, por meio de uma análise histórica.

- b) Identificar regiões ameaçadas e principais ameaças, resultando em diferentes cenários com relação ao impacto dos diferentes tipos de desastres, os recursos e as necessidades previstas.
- c) A combinação das informações obtidas a partir do desempenho das duas primeiras funções permite criar perfis de necessidades de recursos, dependendo da região, bem como do tipo de catástrofe. Diferentes regiões com níveis similares de critérios serão reunidas para formar *clusters*.

Quando se trabalha com um sistema de *clusters* pretende-se auxiliar na coordenação estratégica, na definição de prioridades e no compartilhamento de experiências. Uma abordagem comum é a análise de risco, que leva a vantagem de que cada região ou município contribui com seu conhecimento e experiência para que, finalmente, um entendimento comum para o planejamento (conjunto) possa ser aplicado em várias regiões.

Entretanto, há outra abordagem que trata as metodologias de agrupamento com uma perspectiva computacional. Neste sentido, as metodologias de *clusters* são descritivas e destinam-se a identificar, dada uma determinada base de dados, um conjunto finito de *clusters* naturais chamados conjuntos com base na similaridade das propriedades de cada uma das amostras disponíveis (KOGAN et al., 2006). Em um levantamento bibliográfico mais específico, foram encontradas poucas pesquisas em que se usou a metodologia de *clusters* por meio de dados históricos em logística humanitária. Destacam-se os trabalhos apresentados por Liu et al. (2013), Li et al. (2013), Chu et al. (2012), Chang et al. (2010), Wan (2012), Acosta et al. (2011) e Dalal et al. (2007). A análise desses documentos permitiu a identificação de diferentes ferramentas utilizadas para o desenvolvimento de *clusters* com base em dados de desastres naturais. Entre elas, os mais citados são: *k-means fuzzy clustering*, *C-means* e *hierarchical methods*. Nesse contexto, destaca-se o estudo de Jahre e Navangul (2011), no qual, os meios de oferta e de-

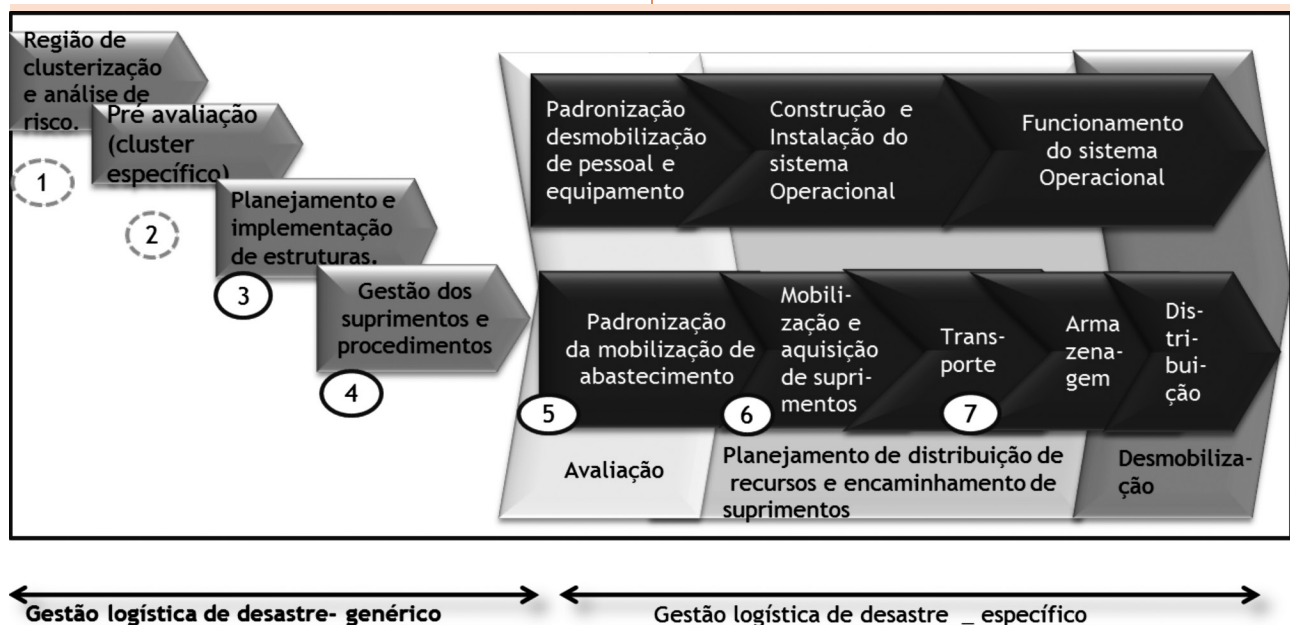


Figura 5: Modelo de processo logístico de socorro em catástrofes internacionais

Fonte: Adaptado de Tufinkgi (2006).

manda de serviços de previsão foram elaboradas no contexto da logística humanitária. Assim, os dados históricos relacionados à informação sobre a assistência humanitária foram considerados, o que permitiu compreender a situação para prever a demanda futura e identificar os padrões-chave em termos de necessidades e resposta à demanda.

Este artigo visa a contribuir para a logística humanitária por meio de uma metodologia voltada a formar *clusters* por combinação de frequência de desastre na região de Santa Catarina, utilizando para tal, a técnica de mineração de dados. Com isso, pode-se verificar a frequência por tipo de desastre em determinada região e identificar regiões semelhantes, para, então, poder padronizar um método na prevenção.

3 Mineração de dados

Segundo Fayyad et al., (1996) Mineração de Dados (MD) é um processo de extração do conhecimento de bases de dados que tem o objetivo de encontrar o conhecimento a partir de um conjunto de dados para ser utilizado no processo decisório. A MD é uma atividade multidisciplinar e engloba o desenvolvimento de modelos aplicáveis a fenômenos em engenharia nas diversas escalas: espacial, temporal e multivariada.

A MD é considerada como sinônimo de Knowledge Discovery in Databases (KDD) ou Descoberta de Conhecimento em Banco de Dados. No entanto, o KDD é um processo mais amplo, em que a MD está inserida, e consiste nas seguintes etapas ilustradas na Figura 6 (SOUZA, 2004):

- Limpeza dos dados: etapa em que são eliminados ruídos e dados inconsistentes.
- Integração dos dados: nessa etapa diferentes fontes de dados podem ser combinadas produzindo um único repositório de dados.
- Seleção: são selecionados os atributos que interessam ao usuário.
- Transformação dos dados: etapa em que os dados são transformados num formato apropriado para a aplicação de algoritmos de mineração (p.e. agrupamento de dados – *clustering*).
- Mineração: etapa essencial do processo, consiste na aplicação de técnicas inteligentes a fim de se extrair os padrões de interesse.
- Avaliação ou pós-processamento: nesta etapa são identificados os padrões interessantes de acordo com algum critério do usuário.
- Visualização dos resultados: nesta etapa são utilizadas técnicas de representação de conhecimento a fim de apresentar ao usuário o conhecimento minerado.

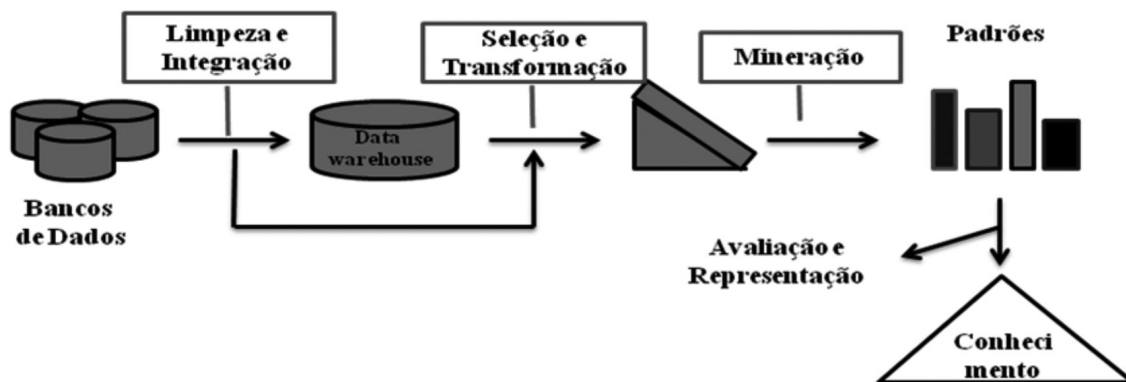


Figura 6: Etapas do processo KDD

Este artigo está focado nas técnicas frequentemente utilizadas na etapa de mineração do processo KDD. Considera-se que os dados foram selecionados e transformados, integrados num armazém de dados (*data warehouse*) e deles foram eliminados ruídos que possam afetar o processo de descoberta de conhecimento.

A MD reúne várias técnicas, entre as quais, pode-se destacar (FAYYAD et al., 1996): classificação e predição, análise de *outliers* (exceções) e agrupamento de dados (*clustering*). Neste trabalho, a técnica aplicada será a de *cluster*.

Os agrupamentos de dados ou *clustering* são atividades em mineração de dados, que consistem em agrupar os dados dentro de classes ou *clusters* tais que os objetos em uma classe tenham alta similaridade em comparação com outro objeto dessa classe (HAN; KLAMBER, 2000), mas tenham baixa similaridade a objetos de outras classes. A árvore de agrupamento (ou árvore hierárquica) usa as dissimilaridades ou distâncias entre os objetos para formar as classes. O cálculo das distâncias euclidianas é o método mais direto de calcular as distâncias entre os objetos num espaço multidimensional.

A seguir, na próxima sessão, descreve-se o método de agrupamento *k-means*, utilizado neste trabalho.

3.1 Método *k-means* para análise de clusters

O método *k-means*, para análise de *clusters*, recebe como *input* um banco de dados de objetos e um número *k* que representa o número de *clusters* que se deseja formar entre os objetos do banco de dados (SOUZA, 2004).

O banco de dados é apresentado em forma de uma matriz de dissimilaridade entre os objetos. Nessa matriz, o elemento da coluna *j* e linha *i* da matriz é o número $d(i, j)$ que representa a distância entre os objetos *i* e *j*. Várias funções distância po-

dem ser utilizadas, destacando-se: distância euclidiana, de Manhattan e a de Minkowski (SOUZA, 2004). A técnica para construir a matriz de dissimilaridade depende do tipo de dados presentes no banco de dados.

O algoritmo *k-means* realiza uma sequência de procedimentos iterativos que formam *k clusters* esféricos compactos com o objetivo de minimizar a Equação 1.

$$J = \sum_{j=1}^n \sum_{k=1}^K \mu_{kj} D^2(z_k, x_j) \quad (1)$$

em que $D^2(z_k, x_j)$, no atual trabalho, é a distância euclidiana de x_j para z_k ; e z_1, z_2, \dots, z_k são os centroides dos *clusters*.

De acordo com Bandyopadhyay (2011), o algoritmo, em primeiro lugar, inicializa aleatoriamente os *k* centroides. Esses são utilizados para dividir os dados, atribuindo cada ponto para o seu centro mais próximo. Sucessivamente, para cada *cluster*, o valor médio de todos os pontos de dados que lhe foram atribuídas é calculado, e isto é considerado como o novo centro. Isso completa uma iteração do *k-means clustering*. Em seguida, a iteração usa os centros recém-calculados para redistribuição dos pontos de dados aos *k clusters* e o processo continua até que a medida de *J* cai abaixo de um determinado limite ou um número máximo de iterações já foram executados. O algoritmo é descrito como segue:

Passo 1: Obter *k* centroides, selecionados aleatoriamente.

Passo 2: Atribuir cada objeto no banco de dados para o centro do *cluster*.

Passo 3: Calcular os novos centros dos *k clusters* pela média dos pontos de dados atribuídos a cada grupo.

Etapa 4: Atualizar os centros dos *clusters* com a média.

Passo 5: Calcular J dado pela Equação 1.

Passo 6: Se um critério de parada é satisfeito, pare, caso contrário, volte ao Passo 2.

A grande vantagem do método é sua eficiência em tratar extensos conjuntos de dados e suas desvantagens são: o fato de o usuário fornecer o parâmetro k de *clusters*, não descobrir *clusters* de formatos não convexos e de ser sensível a ruídos. Assim, objetos com valores altos podem causar uma grande alteração no centro da gravidade dos *clusters* e distorcer a distribuição dos dados.

A seguir, são apresentadas essas etapas aplicadas na descoberta de conhecimento na base de dados.

4 metodologia proposta e experimentos computacionais

Nesta sessão do artigo, será apresentada a metodologia proposta para a formação dos *clusters*, bem como o experimento computacional para validação e verificação do modelo apresentado, utilizando para tal dados reais sobre desastres naturais do estado de Santa Catarina.

4.1 Metodologia

Foi utilizado o gerenciador de banco de dados PostgreSQL 8.2, para organização da base de dados, juntamente com a linguagem de programação Java 1.6. Os mapas apresentados são os georeferenciados pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2011), e, para a formação dos *clusters*, utilizou-se a ferramenta computacional RapidMiner 5 (RAPIDMINER 5, 2011).

A Figura 7 esquematiza as etapas da metodologia proposta neste trabalho.



Figura 7: Esquema dos passos da metodologia proposta

4.2 Experimento computacional

Para testes da metodologia proposta, foram utilizados dados reais de desastres ocorridos entre os anos de 2000 e 2010 no estado de Santa Catarina. Esses dados foram levantados por meio da base de dados do Centro Universitário de Estudos e Pesquisas sobre Desastres da Universidade Federal de Santa Catarina (CEPED/UFSC) (BRASIL, 2011), a qual utilizou documentos oficiais, como o formulário de Avaliação de Danos (AVADANS); a Notificação Preliminar de Desastre (NOPRED); o Decreto, Decreto Municipal; as portarias; os relatórios de danos e outros considerados não oficiais, como as notícias de jornal. Estes documentos estão sendo referenciados neste artigo como eventos registrados.

Inicialmente, as ocorrências de desastres naturais foram separadas de acordo com os municípios do estado. Em seguida, como os dados do CEPED incluíam vários tipos de catástrofes, foram filtrados da base de dados somente tipos de desastres naturais conforme classifica a Secretaria de Defesa Civil Nacional (BRASIL, 2011), resultando, assim em 3.247 ocorrências selecionadas. Pôde-se, com isso, constatar a distribuição anual dos eventos registrados por desastres naturais (Figura 8), e por tipo de desastre (Figura 9).

Não foi necessário pré-processamento ou limpeza dos registros (incompletos, repetidos etc.) do CEPED antes do início da mineração de dados. No entanto, para o objetivo deste trabalho, foi necessária uma conversão dos dados a fim de definir a base de dados e os atributos que deveriam ser explorados pelo algoritmo *k-means*. Com isso, relacionou-se numa matriz cada município com sua

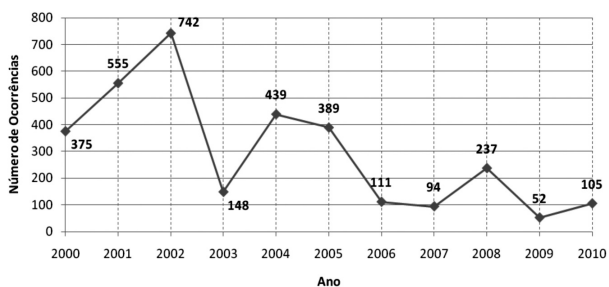


Figura 8: Distribuição anual de eventos registrados em Santa Catarina

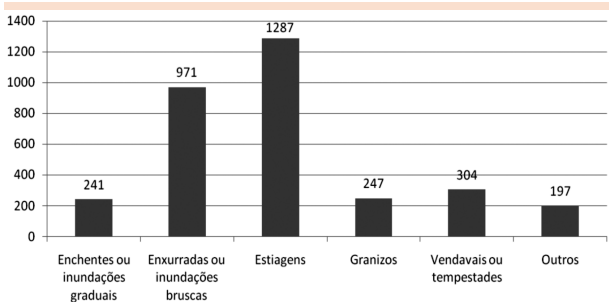


Figura 9: Soma absoluta dos eventos registrados por tipo de desastre

devida frequência para cada tipo de desastre natural. Ou seja, definiu-se como atributos a frequência que cada município de Santa Catarina possuiu em dez anos para vendaval, granizo, enchente, enxurrada e estiagem.

5 Resultados obtidos

Para definir o número de *clusters* foram analisados a média da distribuição centroide e o cálculo do índice Davies-Bouldin (HALKIDI et al., 2001), que é baseado na medida de dispersão e uma medida de similaridade entre os *clusters*. Com base nisso, decidiu-se que o número de três *clusters* responde ao teste computacional esperado neste artigo. Logo, executando o algoritmo *k-means* com $k = 3$, chegou-se ao resultado sumarizado na Tabela 1 e demonstrado graficamente na Figura 10.

Constatou-se que os *clusters* formados apresentam uma grande aproximação referente ao perfil nos municípios impactados. Verifica-se que no

Tabela 1: Resultado do algoritmo *k-mean* com $k = 3$

Cluster	Predominância	
	Menor representatividade	Maior representatividade
0	Estiagens; granizos; vendavais e enchentes.	Enxurradas.
1	Granizos; vendavais, enchentes e estiagens.	Enxurradas.
2	Enchente; granizo; vendaval e enxurrada.	Estiagens.

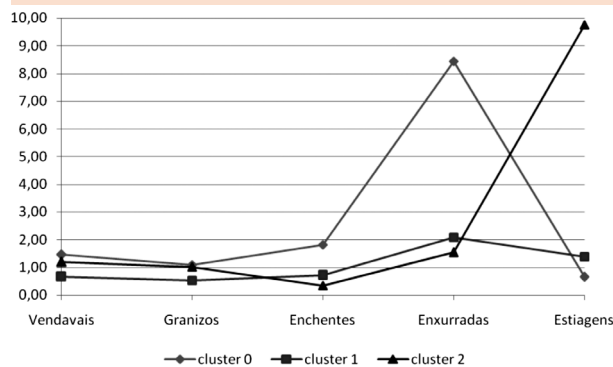


Figura 10: Representação gráfica dos clusters, baseada na matriz de dissimilaridade dada pelo método *k-means*

cluster 0 e no 1 enxurradas é o tipo de desastre mais representativo e observa-se sua maior proximidade do litoral. Contudo, cabe ressaltar que o *cluster* 1 apresenta uma diversidade de desastres com pouca diferença de frequência, como pode-se constatar na Figura 10, considerando-se, assim, a estiagem como um desastre representativo. No 2, a estiagem foi mais representativa e, analisando a Figura 11, observa-se a sua localização no oeste de Santa Catarina, onde a estiagem é realmente mais típica da região.

6 Considerações finais

Neste artigo, foram apresentados alguns conceitos envolvidos na logística humanitária com o objetivo de determinar o foco da pesquisa. Foram abordadas definições e características de um sistema *clusters* tendo como intuito verificar a aplicação deste na logística humanitária. Dessa

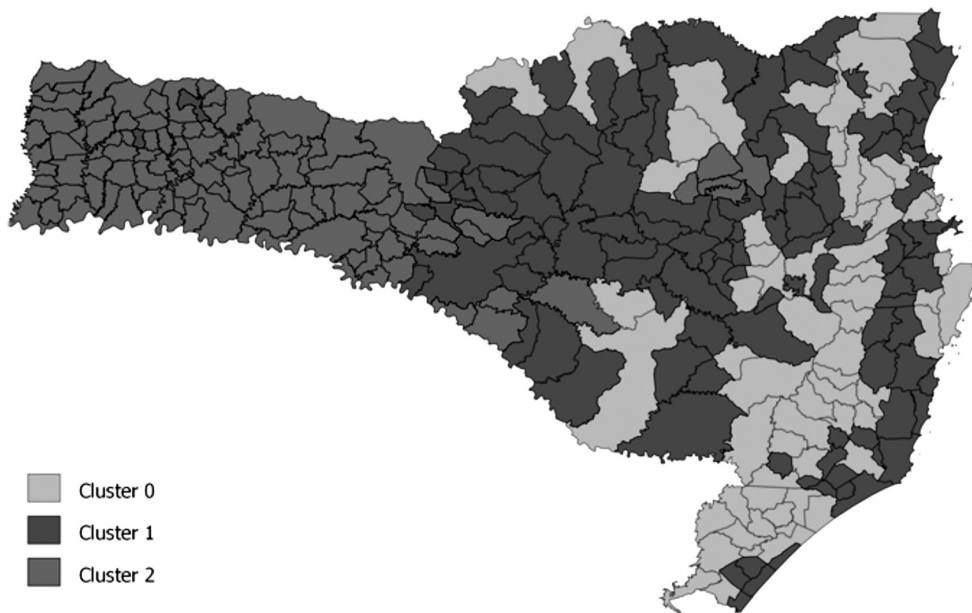


Figura 11: Clusters resultantes do experimento

forma, a contribuição deste trabalho consiste em propor uma metodologia de clusterização, que atua no tratamento de dados históricos de desastres, a fim de formar padrões de *clusters*. Esta metodologia deve ser considerada em conjunto com uma sequência de procedimentos que incluem atividades de mapeamento mediante a preparação para um desastre.

Agradecimentos

Os autores agradecem ao corpo de especialistas do CEPED/UFSC e às agências brasileiras de pesquisa, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

Referências

ACOSTA, M. et al. CAREY: *Climatological control of emergency regions*. OTM Workshops. Berlin Heidelberg: Springer-Verlag, 2011. LNCS 7046, 494-503.

ALMEIDA, S. J.; ZARATE, L. H. Identificação de grupos de estados brasileiros similares em relação a índices de mortalidade. In: WORKSHOP EM ALGORITMOS E APLICAÇÕES DE MINERAÇÃO DE DADOS, 3., 2007, João Pessoa. *Anais...* João Pessoa: WAAMD, 2007.

BALCIK, B. et al. Coordination in humanitarian relief chains: practices, challenges and opportunities. *International Journal Production Economics*. Science Direct, v.126, n. 1, p. 22-34, 2010.

BANDYOPADHYAY, S. Genetic algorithms for clustering and fuzzy clustering. *WIREs Data Mining and Knowledge Discovery*, p. 524-531, 2011.

BRASIL. Ministério da Integração Nacional. Secretaria Nacional de Proteção e Defesa Civil. Defesa Civil, 2011. Disponível em: <www.defesacivil.gov.br>. Acesso em: 20 jun. 2011.

CHADE J. Especialista em desastres naturais da ONU critica o Brasil. Estadão Planeta. O Estado S. Paulo. 22 de janeiro de 2009. Disponível em: <<http://www.estadao.com.br/noticias/vidae,especialista-em-desastres-naturais-da-onu-critica-o-brasil,311350,0.htm>>. Acesso em: 5 jun. 2010.

CHANG, L. C. et al. Clustering-based hybrid inundation model for forecasting flood inundation depths. *Journal of Hydrology*, 385, 257-268, 2010.

CHU, H. J. et al. Integration of fuzzy cluster analysis and kernel density estimation for tracking typhoon trajectories in the Taiwan region. *Expert Systems with Applications*, v. 39 n. 10, p. 9451-9457, 2012.

- DALAL, J.; MOHAPATRA, P. K. J.; MITRA, G. C. Locating cyclone shelters: a case. *Disaster Prevention and Management*, v. 16, n. 2, p. 235-244, 2007.
- FAYYAD, U. M. et al. *Advances in knowledge discovery and data mining*. Menlo Park, CA : AAAI Press, 1996.
- HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On clustering validation techniques. *Journal of Intelligent Information Systems*, v. 17, n. 2/3, p. 107-145, 2001.
- HAN, J.; KAMBER, M. *Data mining: concepts and techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.
- IBGE – Instituto Brasileiro de Geografia e estatística. 2011. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 27 maio 2011.
- JAHRE, M.; JENSEN, M. Coordination in humanitarian logistics through clusters. *International Journal of Physical Distribution & Logistics Management*, v. 40, n. 8/9, p. 657-674, 2010.
- JAHRE, M.; NAVANGUL, A. K. Predicting the unpredictable – demand forecasting in international humanitarian response. In: ANNUAL NOFOMA CONFERENCE, 23rd., 2011. *Proceedings...* Norway: Harstad, p. 265-281, 2011.
- KOGAN, J.; NICHOLAS, C.; TEBoulLE, M. *Grouping multidimensional data: recent advances in clustering*. Berlin: Springer, 2006.
- LI, B. et al. Theoretical and Applied Climatology. *Springer Vienna*, v. 111, n. 3, p. 401-415, 2013.
- LIMA, F. S.; OLIVEIRA, D.; GONÇALVES, M. B. A formação de clusters na logística humanitária utilizando mineração de dados. In: ANPET – CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 25., 2011, Belo Horizonte. *Anais...* Belo Horizonte: Anpet, 2011.
- LIU, C.; KUO, Y.; CHEN, C.-W. Emergency responses to natural disasters using Formosat-2 high-spatiotemporal-resolution imagery: forest fires. *Natural Hazards*, v. 66, n. 2, p. 1037-1057, 2013.
- MEIRIM, H. Logística humanitária e logística empresarial, mmrbrasil, 2007. Disponível em: <www.mmrbrasil.com.br>. Acesso em: 15 abr. 2010.
- NOGUEIRA, C. W.; GONÇALVES, M. B.; NOVAES A. G. Logística humanitária e logística empresarial: relações, conceitos e desafios. Artigo. In: CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 21., 2007, Rio de Janeiro. *Anais...* Rio de Janeiro: Anpet, 2007.
- NOGUEIRA, C.W.; GONÇALVES, M.B. ; OLIVEIRA, D. O Enfoque da logística humanitária no desenvolvimento de uma rede dinâmica para situações emergenciais: o caso do vale do Itajaí em Santa Catarina. In: CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 23., 2009, Vitória. *Anais...* Vitória: Anpet, 2009.
- NOGUEIRA, C. W.; GONÇALVES, M. B. O Desenvolvimento e aplicação de um modelo para a localização de uma central de inteligência e suporte para recebimento, controle e distribuição de recursos em situações emergenciais com foco na logística humanitária. In: CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, Salvador, Bahia, 24., 2010. *Anais...* Rio de Janeiro: Anpet, 2010.
- RAPIDMINER 5. Disponível em: <<http://www.rapid-i->>. Acesso em: 31 mar. 2011.
- SAPIR , G. D. *Disasters in numbers 2010*. Belgium: CRED, Catholic University of Louvain, Brussels, Jan. 24, Geneva, 2011.
- SCHULTZ, S. F. *Disaster relief logistics: benefits of and impediments to horizontal cooperation between humanitarian organizations*. Tese. Technischen Universität Berlin, Berlin, 2008.
- SOUZA, F. T. *Predição de escorregamentos das encostas do município do rio de janeiro através de técnicas de mineração de dados*. tese (doutorado em Engenharia)– Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.
- THOMAS, A. Enabling disaster response. Institute Fritz, 2007. Disponível em: <<http://www.fritzinstitute.org/index.htm>>. Acesso em: 20 jan. 2012.
- TOMASINI, R.; Van Wassenhove, L. *Humanitarian logistics*. London: Palgrave MacMillan, 2009 (Insead Business Press).
- TUFINKGI, P. *Logistik im kontext internationaler katastrophenhilfe: entwicklung eines logistischen referenzmodells für katastrophenfälle*. Bern. Stuttgart Wien: Ed. Haupt Verlag, 2006.
- WAN, S. Entropy-based particle swarm optimization with clustering analysis on landslide susceptibility mapping. *Environment Earth Science*, 68, n. 5, p. 1349-1366, 2012.

Recebido em 2 dez. 2013 / aprovado em 18 mar. 2014

Para referenciar este texto

LIMA, F. S.; OLIVEIRA, D.; GONÇALVES, M. B. Formação de clusters para o gerenciamento da cadeia de suprimentos em operações humanitárias. *Exacta – EP*, São Paulo, v. 12, n. 1, p. 55-68, 2014.