

# Crenças de Grupo como Instrumento de Formação da Reputação: uma Abordagem de Agentes e Artefatos

Tiago L. Schmitz<sup>1</sup>, Jomi F. Hübner<sup>1</sup>, Carine G. Webber<sup>2</sup>

<sup>1</sup>Departamento de Engenharia de Automação e Sistemas - UFSC  
PO Box 476, Florianópolis, SC, CEP 88040-900, Brasil

<sup>2</sup>Centro de Computação e Tecnologia da Informação - UCS  
R. Francisco Getúlio Vargas, 1130, Caxias do Sul, RS, CEP 95070-560, Brasil

{tlschmitz,jomi}@das.ufsc.br, cgwebber@ucs.br

**Abstract.** *A qualidade das interações entre agentes é um fator determinante na eficiência de um sistema multiagente (SMA). Uma importante ferramenta utilizada para melhorar a qualidade das interações é a reputação. Dentre os diversos modelos para a reputação, o modelo ForTrust destaca-se por sua excelente fundamentação teórica. O modelo ForTrust apresenta uma fundamentação bem definida que possibilitou a construção de uma arquitetura de implementação usando Agentes e Artefatos. Nela, artefatos são estruturas passivas e orientadas a funções compartilhadas por um grupo de agentes como mecanismo de cooperação e coordenação na formação da reputação. A proposta de tais artefatos é ilustrada no cenário de teste, onde a reputação é utilizada como auxílio aos patrulheiros da Wikipédia. A arquitetura apresentada neste trabalho se mostra de fácil utilização uma vez que obedece a uma série de requisitos específicos observados nesta pesquisa. Seu funcionamento é corroborado pela sua implementação no auxílio aos patrulheiros da Wikipédia.*

## 1. Introdução

Em sistemas multi-agentes o comportamento coletivo de agentes autônomos é utilizado para resolver problemas. A metáfora de sociedade observada no comportamento coletivo dos agentes traz consigo conceitos de outras áreas, tais como: sociologia, psicologia e economia. Um dos conceitos mais importantes herdados é a interação entre os participantes, função elementar para o funcionamento de um sistema multi-agente (SMA). A qualidade dessas interações é decisiva na concepção da solução, pois um comportamento desleal de um participante ou uma falha na comunicação tem potencial para afastar o sistema de uma solução. A questão que permeia este assunto é como realizar uma avaliação das informações e a escolha de parceiros em uma sociedade de altruísmo limitado, ou seja, onde os agentes podem agir de má fé. Dois conceitos oriundos das ciências sociais podem auxiliar a responder esta questão: a confiança e a reputação.

O termo reputação possui várias conceituações nas mais diversas áreas do conhecimento. Todas elas observam aspectos diferentes do mesmo objeto de estudo, existindo assim diversos mecanismos de reputação. Boa parte destes mecanismos adotam uma representação quantitativa, onde a reputação é reduzida a uma representação numérica, abstraindo importantes propriedades da sua concepção [6]. Estas propriedades poderiam fornecer mais subsídios aos agentes para determinar um estado de confiança em um agente alvo. Em contrapartida existem modelos que provêm uma análise qualitativa da reputação baseada nos ingredientes. Dentre esses, destaca-se o modelo ForTrust que possui uma formalização fortemente inspirada no modelo de confiança social [3].

O modelo ForTrust constrói os conceitos de confiança e reputação sobre quatro ingredientes: o objetivo, a capacidade, a competência e a intenção. A confiança e

a reputação encontram-se em diferentes escopos, na crença individual e na crença de grupo (CG), respectivamente [6]. O ForTrust não define a concepção e a localização das crenças de grupo e esses elementos são importantes na definição de uma arquitetura viável de implementação. O presente trabalho tem por objetivo apresentar uma arquitetura de implementação para as crenças de grupo que viabilize o desenvolvimento de uma implementação do modelo ForTrust, obedecendo os requisitos de visibilidade, dinamicidade, flexibilidade, reusabilidade, heterogeneidade e tratamento da credibilidade, desejáveis em um sistemas multiagente abertos conforme apontado por outros modelos de reputação.

O cenário de teste utilizado é o problema de classificação das atualizações da Wikipédia. Ambientes colaborativos abertos, tais como as Wikis, tem como propósito a livre criação e edição de artigos por seus usuários e podem ser concebidos como um SMA aberto. Políticas de uso são atribuídas para tais ambientes para regulamentar as interações dos agentes e evitar excessos, tais como: propagandas, ataques pessoais, palavras de baixo calão etc. Em ambientes onde há um número massivo de alterações como a Wikipédia, o volume de inserções e alterações executadas diariamente torna complexa a tarefa de validar as novas informações. Esta característica torna o ambiente propício ao surgimento de atos de vandalismo, imprecisão de informações, propagandas, entre outros. Para tanto existe dentro da Wikipédia uma entidade denominada patrulheiro, para analisar as alterações e para evitar estas desconformidades. Este trabalho propõe também a aplicação da reputação na Wikipédia como forma de auxílio aos patrulheiros.

## 2. ForTrust

O ForTrust é baseado no modelo de confiança de Castelfranchi e Falcone [2] e não se resume a uma mera probabilidade atualizada a cada interação dos agentes. A confiança é construída através de uma série de crenças que quando reunidas levam a um estado interno que permite ou não confiar em um determinado agente [5].

No modelo ForTrust, as crenças que fundamentam uma relação de confiança são: objetivo, capacidade, competência e intenção. Sendo assim, para que um agente  $A$  confie em um agente  $B$  para fazer uma tarefa  $J$  é necessário que:  $A$  tenha um objetivo  $X$ ;  $A$  acredite que  $B$  é capaz de fazer  $J$ ;  $A$  acredite que  $B$  tem competência para atingir  $X$  fazendo  $J$  e  $A$  acredite que  $B$  tem a intenção de fazer  $J$ . O agente  $A$ , tendo essas crenças, pode confiar em  $B$  [5]. Tomando como exemplo um time de futebol. O técnico tem o objetivo de levar a equipe à vitória, para tanto deve escalar 11 jogadores que acredita serem capazes de levar o time a atingir este objetivo. Ele escolherá para cada posição um jogador que ele acredite competente para a posição; seja capaz de desempenhar o papel satisfatoriamente e tenha a intenção de desempenhar a função.

Para a reputação o caso é similar, contudo faz se uso do conceito de crenças de grupo. Uma CG é um conjunto de crenças partilhadas por um grupo, logo, não é mais um agente  $A$  que tem uma crença sobre um agente  $B$ , mas um grupo  $I$  que tem uma crença sobre um agente  $B$  [5]. Sendo assim, para que o grupo  $I$  crie uma reputação para um agente  $B$  para fazer uma tarefa  $J$  é necessário que:  $I$  tenha um objetivo  $X$ ;  $I$  acredite que  $B$  é capaz de fazer  $J$ ;  $I$  acredite que  $B$  tem competência para atingir  $X$  fazendo  $J$ ;  $I$  acredite que  $B$  tem a intenção de fazer  $J$ . O grupo  $I$  tendo essas crenças pode estabelecer uma reputação para  $B$  [5]. Considere o seguinte exemplo: na escolha de um representante de uma agremiação os sócios nem sempre conhecem diretamente os candidatos a representante. O grupo de sócios acaba por escolher o representante através de sua reputação, sendo assim, o grupo acredita que: o candidato pode exercer a função

de representante; tem competência para representar satisfatoriamente o grupo e tem a intenção de representar o grupo. O fato do grupo ter essas crenças aumenta o grau de complexidade da concepção da arquitetura do sistema, pois não temos hoje uma proposta de como as CGs são instituídas e concretizadas em sistemas computacionais.

### 3. Trabalhos relacionados

Esse trabalho observou diferentes modelos de reputação e modelos de crença de grupo. Nessa seção serão brevemente apresentados alguns trabalhos relacionados as contribuições desse artigo.

Primeiramente, a representação das crenças de grupo, na qual as crenças são entidades pertencentes ao ambiente. O trabalho [14], proposto pelo filósofo Tuomela propõe uma representação das CG muito similar, argumentando que o modo como as crenças de grupo são constituídas é dependente do comportamento do grupo. As candidatas a crenças de grupo são anunciadas a qualquer instante por qualquer elemento do grupo, mas são cristalizadas como crença do grupo quando esse passa a agir em acordo com o que rege a candidata a crença de grupo. Contudo a discussão de Tuomela não avança em sociedades virtuais e sistema multiagentes, servindo de inspiração para o nosso trabalho.

A segunda contribuição, extensão do ForTrust, que apresenta a reputação como a composição de ingredientes dotados de certeza e força. O modelo proposto em [16] é amplamente utilizado em sistema de comércio eletrônico e possui uma estrutura centralizada. A reputação é composta por apenas um valor assim como em [9, 12] enquanto no modelo proposto nesse trabalho é apresentada uma composição de ingredientes dotados de certeza e força que permite uma avaliação mais completa da reputação pelo agente. Por fim em [4], é apresentado uma solução similar utilizando o modelo ForTrust com adição de um grau de certeza aos ingredientes. Em comparação com esse modelo nossa proposta apresenta uma maior riqueza de informação uma vez que além da certeza é adicionado um valor para a força dessa certeza.

### 4. Composição das Crenças de Grupo

A composição de uma crença de grupo, contribuição deste artigo, segue um modelo centralizado para a concepção das crenças de grupo, permitindo o isolamento das regras de formação das crenças do agente e do grupo, ou seja, as regras de formação das crenças individuais se encontram dentro da mente dos agentes e as regras que formam as crenças de grupo se encontram em uma entidade externa à mente dos agentes. A comunicação que alimenta a formação da crença de grupo é o anúncio. As informações são enviadas a um componente que as agrega formando uma crença de grupo.

A figura 1 ilustra o modo como os agentes fazem seus anúncios a uma estrutura central que através de regras de **síntese** dos anúncios estabelece as crenças de grupo.

Os anúncios são compostos por um predicado, um grau de certeza e uma força. O predicado é uma sentença sugerida pelo agente anunciante. O grau de certeza é um valor entre 0 e 1 que quantifica a veracidade que agente associa para o predicado anunciado. A força é um valor natural que corresponde à experiência que o agente tem sobre o predicado. Por exemplo, no anúncio *Capacidade* (“Agente X”, 50) com um grau de certeza 0,7 e força 5, o

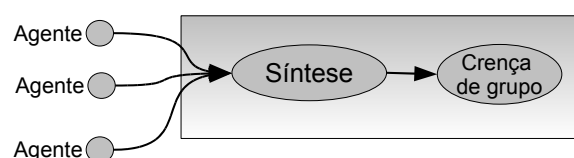


Figura 1. Modelo de Crença de grupo

agente anunciante tem uma boa certeza que o agente X tem capacidade de atender 50 itens por ter 5 experiências onde constatou que o agente X foi capaz de atender tal demanda. Definindo formalmente os anúncios são triplas:

$$A \stackrel{def}{=} \{ \langle p, c, s \rangle \mid p \in P, c \in [0..1], s \in \mathbb{N} \}$$

onde  $A$  é o conjunto de todos os anúncios,  $P$  é o conjunto de todos os predicados e  $p$ ,  $c$  e  $s$  são respectivamente o predicado, o grau de certeza e a força de um anúncio.

Na formação das crenças de grupo é estabelecido um método de síntese que possui as regras que avaliam os anúncios e formam as crenças de grupo. A crença de grupo e a função de síntese podem ser definidas como:

$$G \stackrel{def}{=} \{ \langle p, c, s \rangle \mid p \in P, c \in [0..1], s \in \mathbb{N} \}$$

$$sin : 2^A \rightarrow 2^G$$

onde  $2^G$  é um subconjunto de crenças de grupo composto por tuplas  $\langle p, c, s \rangle \in G$  representando respectivamente o predicado, o grau de certeza e a força de uma crença de grupo  $G$ .  $sin$  é a função de síntese que determina as crenças de grupo para um conjunto de anúncios.

Nesse trabalho foi usado o conceito de crença de grupo proposto pelo filósofo Tuomela [13] e algumas características dos modelos de reputação, pois a reputação pode ser vista como uma crença de grupo. Não existe nenhum método de síntese de crença de grupo que seja absoluto, pois existem diversos tipos de sociedades, cada qual com suas propriedades. Por exemplo, sociedades totalmente igualitárias onde a força da decisão parte da maioria, ou sociedades autoritárias com hierarquia de poder onde a decisão não necessariamente reflete na vontade da maioria são exemplos que demonstram a necessidade de ter diferentes tipos de síntese para as crenças de grupo, cabendo a sociedade ou ao projetista do SMA estabelecer a função mais apropriada para sua composição. Para tanto, é desejável um grau de modularidade que permita ao projetista estabelecer novas funções de síntese com facilidade. A seguir são propostos três tipos de síntese de crença de grupo que ilustram a sua diversidade.

**Síntese ponderada:** os anúncios são sintetizados de maneira a buscar um termo intermediário entre os anúncios. A função de síntese tem como regra uma ponderação relacionando o grau de certeza com força apresentada a seguir. Onde  $sinpon(A')$  é a função de síntese ponderada,  $A'$  é o conjunto de anúncios de um predicado,  $c'$  é o grau de certeza calculado e  $s'$  é a força calculada. Note que  $c$  é o grau de certeza de um anúncio em  $A'$  e de forma análoga define-se  $s$  como a força de um anúncio.

$$sinpon(A') = \{ \langle p, c', s' \rangle \in G \mid \text{onde: } c' = \frac{\sum_{\langle p, c, s \rangle \in A'} c \times s}{\sum_{\langle p, c, s \rangle \in A'} s} \quad s' = \frac{\sum_{\langle p, c, s \rangle \in A'} s}{|A'|}$$

Este tipo de síntese beneficia anúncios com maior força, aproximando a crença de grupo dos anúncios considerados mais experientes. Por exemplo, um usuário X de um SMA aberto anuncia que o usuário Y é incompetente no ingrediente capacidade através de um grau de certeza 0,4 obtida de uma única experiência, ou seja, força igual a 1. O usuário Z anuncia que a Y é competente no ingrediente capacidade e por isso da um grau de certeza de 0,8 baseada em 3 experiências, logo a força do anúncio é 3. Os cálculos

a seguir apresentam como é calculado o grau de certeza ( $c$ ) e a força ( $s$ ) neste exemplo. Observa-se que a função de síntese considera o anúncio do usuário Z é 3 vezes mais forte que o anúncio de X, criando uma crença de grupo sobre Y com grau de certeza de 0,7 e força 2;

**Entr.:**  $\text{sinpon}(\{< \text{capacidade}(Y), 0.4, 1 >; < \text{capacidade}(Y), 0.8, 3 >\})$

**Cálculo:**  $c = \frac{(0.4 \times 1) + (0.8 \times 3)}{1+3}$   $s = \frac{1+3}{1+1}$

**Saída:**  $< \text{capacidade}(Y), 0.7, 2 >$

**Síntese pelo mínimo:** a síntese seleciona o anúncio de menor certeza como crença de grupo. Este tipo de síntese é interessante em sociedades pessimistas, onde a menor certeza de um anúncio sobre um predicado deve ser a crença do grupo. A função de síntese tem como regra a seleção do anúncio de menor grau de certeza de um predicado, tal como  $\text{sinmin}(A') = \arg \min_{< p, c, s > \in A'} c$ . Onde  $\text{sinmin}(A')$  é a função de síntese mínimo,  $\arg \min_{< p, c, s > \in A'} c$  é a seleção do anúncio de menor certeza sobre um predicado.

Por exemplo, o caso dos usuários apresentado anteriormente, o anúncio do usuário X é tomado como crença de grupo por ser o anúncio com menor grau de certeza do predicado. Sendo assim a crença de grupo é sintetizada com uma certeza de 0,4, representando que Y é uma empresa de qualidade questionável no ingrediente capacidade;

**Síntese pelo máximo:** a síntese seleciona o anúncio de maior certeza como crença de grupo. Este tipo de síntese é interessante em sociedades otimistas, onde o pior caso é desinteressante, interessando apenas o anúncio sobre um predicado de maior certeza. A função de síntese tem como regra a seleção do anúncio de maior grau de certeza de um predicado, tal como:  $\text{sinmax}(A') = \arg \max_{< p, c, s > \in A'} c$ . Onde  $\text{sinmax}(A')$  é a função de síntese máximo,  $\arg \max_{< p, c, s > \in A'} c$  é a seleção do anúncio de maior certeza sobre um predicado.

Por exemplo, no caso dos usuários o anúncio de Z é tomado como crença de grupo por ser o anúncio mais positivo do predicado. Sendo assim a crença de grupo é sintetizada com uma certeza de 0,8 representando que a empresa presta o serviço de forma satisfatória no ingrediente capacidade.

Existem ainda muitos outros métodos de síntese que podem estabelecer as crenças de grupo. Esses métodos estão intimamente relacionadas com as normas que regem a sociedade. Como Tuomela [13] apresenta os anúncios que são feitos pelos agentes e a síntese destes anúncios em crenças de grupo podem descrever o comportamento social do grupo, tornando a forma de síntese um critério fundamental das crenças de grupo.

## 5. Formalização da Reputação com grau de certeza e força

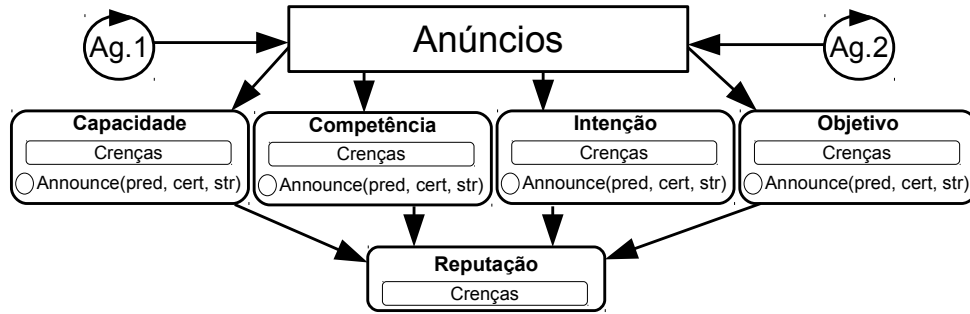
A segunda contribuição desse trabalho visa adicionar os elementos grau de certeza e força à formalização do ForTrust. A reputação observada originalmente no ForTrust pode ser descrita como um conjunto de crenças de grupo que formam a reputação de um indivíduo. Dessa maneira o ForTrust permite elaborar apenas dois estados, possui ou não reputação. Com a adição do grau de certeza e da força ao ForTrust será possível quantificar o quanto se acredita em uma reputação e qual a intensidade desta crença. A sentença a seguir representa como uma reputação é formalizada com a adição do grau de certeza e da força. Onde  $R$  é o conjunto de todas as reputações e  $p$ ,  $c$  e  $s$  são, respectivamente, o predicado, o grau de certeza e a força da reputação.

$$R \stackrel{def}{=} \{< p, c, s > \mid p \in P, c \in [0..1], s \in \mathbb{N}\}$$

Por exemplo, a tupla  $< \text{reputacao}('bob', 'cantar'), 0.7, 7 >$  informa que a reputação de bob para cantar é 0.7 com força 7.

A reputação originalmente no ForTrust é um estado mental no qual todos os ingredientes que a compõem estão presentes. A função  $sinr : 2^G \times 2^G \times 2^G \times 2^G \rightarrow 2^R$ , tal que  $sinr(G_{comp}, G_{cap}, G_{int}, G_{obj})$  mapeia as crenças do grupo sobre competências ( $G_{comp}$ ), capacidades ( $G_{cap}$ ), intenção ( $G_{int}$ ) e objetivos ( $G_{obj}$ ) para um conjunto de reputações sintetizadas desses 4 ingredientes.

Conforme ilustrado na figura 2 os agentes emitem anúncios que são sintetizados em crenças de grupo, que alimentam a reputação e assim estabelecem as crenças de grupo sobre a reputação.



**Figura 2. Exemplificação da arquitetura**

Nessa proposta foram adicionados componentes contínuos, grau de certeza e força, logo foram necessários ajustes na função de síntese da reputação para tratar desses elementos. A nova função de síntese estabelece valores de certeza e força baseado nos valores obtidos nas crenças de grupo. Os métodos adotados nesse trabalho para calcular a certeza e força da reputação seguem o mesmo princípio das descritas anteriormente para a formação das crenças de grupo (ponderada, mínima e máxima), contudo fazendo uso de CG ao invés dos anúncios dos agentes.

Por exemplo, um grupo  $X$  tem o seguinte conjunto de crenças a serem sintetizadas em reputação.

**Entr.:**  $sinr(\{< capacidade(Y, cantar), 0.4, 1 >\}, \{< competência(Y, cantar), 0.8, 3 >\}, \{< intenção(Y, cantar), 0.3, 2 >\}, \{< objetivo(Y, cantar), 0.7, 6 >\})$

Nesse exemplo para sintetizar a reputação foi escolhido o método de síntese pelo mínimo, onde a crença de menor valor de certeza é tomada como reputação, criando uma reputação sobre o agente  $Y$  cantar com certeza de 0.3 e força 2.

**Cálculo:**  $c = \min(0.4, 0.8, 0.3, 0.7)$

**Saída:**  $< reputação(Y, cantar), 0.3, 2 >$

## 6. Implementação

Para implementar as crenças de grupo de forma global, como proposto nesta pesquisa foi utilizado o conceito de Agentes e Artefatos [11]. Modelando as crenças de grupo na forma de artefatos obtêm-se estruturas que são modulares e amplamente manipuláveis pelos agentes uma vez que estas herdaram as características dos artefatos.

O ambiente de trabalho proposto pelo meta-modelo de agentes e artefatos (AA) [11] é definido como parte do SMA, o qual é planejado e construído pelos agentes como suporte às suas atividades. Aos programadores dos SMA fica a tarefa de conceber os tipos de artefatos que serão dinamicamente instanciados e cooperativamente utilizados

pelos agentes. A noção de artefato é o ponto central do modelo de AA. Sua representação é de uma estrutura fora da mente do agente que é compartilhada por todos os agentes presentes no ambiente de trabalho, podendo ser criada, utilizada e, por vezes, removida pelos agentes durante a execução de tarefas.

Desta maneira os artefatos costumam encapsular algum tipo de função e também se estruturam de forma similar aos artefatos utilizados pelo ser humano. A estrutura genérica dos artefatos é composta por interfaces de uso (gatilhos que disparam as funções do artefato), propriedades observáveis (informações que o artefato deixa visível aos agentes), eventos observáveis (informações efêmeras), interface de comunicação (comunicação entre artefatos) e manual (informações de utilização dos recursos do artefato)

A implementação utilizou a plataforma CArtAgO [10] que fornece suporte aos conceitos de Agentes e Artefatos. O diagrama de classe da figura 3 apresenta a arquitetura de implementação para o modelo estendido do ForTrust proposto nesta pesquisa.

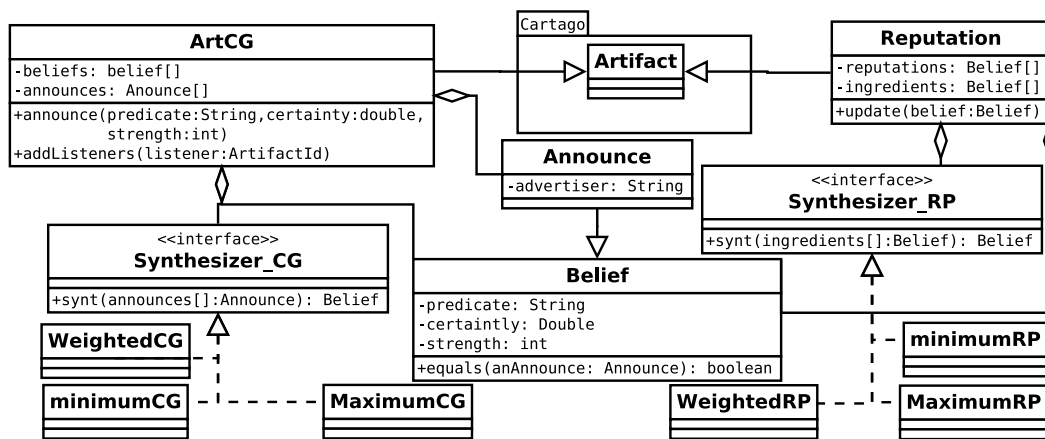


Figura 3. Diagrama de classe da reputação

O artefato CG possui as crenças do grupo como propriedades observáveis, sendo que cada crença corresponde a uma propriedade observável. Os anúncios são tratados como uma operação na interface de uso, no qual é passado como parâmetro o predicado anunciado, um grau de certeza sobre a crença e a força da certeza. Ele ainda possui uma interface para o acoplamento de outros artefatos representados na interface de comunicação, que permite a transmissão das crenças de grupo para outros artefatos.

A Classe **ArtCG** representa o artefato de crença de grupo. Esta classe herda da **Artifact** as primitivas dos artefatos, além de agregar as classes **Announce**, **Belief** e a interface **synthesizer**. O método **announce** corresponde à operação de anúncio do artefato de crença de grupo e recebe como parâmetro um predicado, um grau de certeza e uma força. O artefato ao receber um anúncio adiciona a lista de anúncios ou substitui o anúncio antigo do agente pelo recebido. Atualizada a lista de anúncios é sintetizada a crença de grupo do predicado anunciado. Em seguida, a crença de grupo é enviada para o artefato de reputação. O método **addListeners** é o método que adiciona os artefatos que receberão a notificação quando uma nova crença for sintetizada, neste caso o artefato ligado é o de reputação.

A Classe **Belief** representa uma crença de grupo composta pela tripla: predicado, grau de certeza e força. A Classe **Announce** representa os anúncios pelos agentes. Esta classe herda a classe **Belief** e adiciona o atributo **advertiser** que representa o agente que

fez o anúncio.

O modo de síntese da reputação difere das outras crenças de grupo uma vez que essa é composta pelos ingredientes. Sendo assim a reputação foi implementada em uma classe separada nomeada **Reputation**. Essa classe possui apenas um método, o `update`, no qual é recebido as crenças enviadas pelo artefato de CG. Quando o método `update` é disparado a crença recebida é atualizada na lista de ingredientes e é recalculado a reputação da crença correspondente.

## 7. Experimento

A Wikipédia, enciclopédia livre, permite a seus usuários a livre criação e edição de artigos contanto que a sua manipulação esteja em acordo com as políticas de uso da ferramenta. O grande domínio abrangido pela enciclopédia, o volume de inserções e alterações executadas diariamente torna complexa a tarefa de validar as novas informações. Esta característica torna o ambiente propício ao surgimento de atos de vandalismo maliciosos ou não, imprecisão de informações, propagandas, entre outros. Para analisar as alterações e para evitar estas desconformidades, a Wikipédia implementa diversos dispositivos. Dentre eles é previsto o uso de uma entidade denominada patrulheiro. Os patrulheiros são usuários normais que assumem uma função de analisar, voluntariamente, a lista de alterações em busca destas desconformidades [7].

Como auxílio aos patrulheiros é proposto o uso de um SMA para a criação de um sistema de reputação dentro da comunidade da Wikipédia. O uso desse tipo de sistema permite estabelecer um grau de confiança entre o patrulheiro e o usuário contribuidor, mesmo que um determinado patrulheiro nunca tenha analisado a contribuição de um referido usuário. A reputação perante a comunidade permite ao patrulheiro inferir um grau de confiança no usuário, facilitando assim a análise das suas alterações. O patrulheiro pode assim se focar nos usuários com reputação duvidosa e analisar suas contribuições [7].

O objetivo deste experimento é ilustrar a funcionalidade e aplicação da arquitetura proposta. Para tanto, foram efetuados dois ensaios. O primeiro tem o objetivo de ilustrar a importância do componente força na formalização proposta. O segundo apresenta a aplicação da heurística de revisão proposta que faz uso do conceito de quartis [8].

A inviabilidade de inserir essa estrutura na Wikipédia para fim de testes fez com que nesse trabalho os patrulheiros da Wikipédia fossem substituídos por agentes que simulam suas atividades. Os agentes, dotados de diferentes heurísticas, percorrem um log de contribuições, estabelecem análises estatísticas sobre essas e emitem os anúncios ( $\langle p, c, s \rangle$ ) que alimentam o sistema de reputação proposto. As heurísticas utilizadas são fortemente baseadas nos trabalhos [1] e [15] e são caracterizadas como:

**Agente Avaliador Textual.** Observa o crescimento do texto com o passar das contribuições. Ele crê que informações novas são adicionadas quando há um aumento no volume de texto. É observado que o crescimento do texto em uma página na Wikipédia não se dá de forma linear. Inicialmente, o crescimento é muito grande e tende a diminuir com o passar das contribuições [1].

**Agente Avaliador de Reversão.** Um patrulheiro tem a possibilidade de ao encontrar uma contribuição de um vândalo fazer uma reversão para uma contribuição anterior (ponto de retorno) que ele considera de qualidade. Desta maneira, este agente avalia as contribuições baseado no histórico de reversões, punindo os usuários que tem suas contribuições revertidas e recompensando usuários que tem suas contribuições como ponto de retorno de uma reversão.



**Agente Avaliador Temporal.** Este agente avalia o tempo de persistência da contribuição. Contribuições com baixa qualidade ou ações de vandalismo são geralmente identificadas em um período de 48 horas. Já contribuições de boa qualidade encontram-se em páginas maduras, que já atingiram uma estabilidade e raramente terão uma nova contribuição. Neste caso o período de persistência de uma página é muitas vezes superior a um mês (720 horas).

O dados utilizados nos experimentos descritos a seguir foram extraídos de uma versão da Wikipédia feita para testes conhecida como simplewiki. O dados desta Wiki são periodicamente copiados para arquivos XML disponibilizados para download<sup>1</sup>. Estes arquivos possuem informações referentes as contribuições dos usuários dispostos nos seguintes campos: usuário, timestamp, comentário e conteúdo da contribuição. Para fazer os experimentos ordenou-se os registros de contribuições em ordem decrescente de data de publicação, reservando uma fração de 10% das contribuições mais recentes para testes (dados de teste), simulando contribuições a serem encontradas pelos patrulheiros. A outra fração correspondente a 90% da lista de contribuições é utilizada para treinamento do sistema de reputação (dados de treinamento).

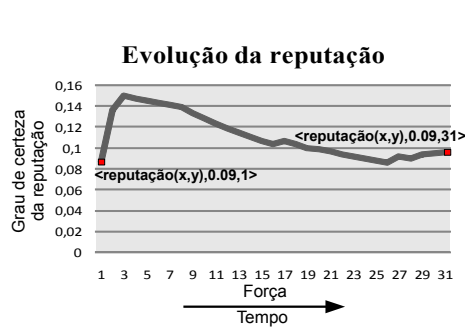
O primeiro experimento observa a relação entre a força e o grau de certeza da reputação. Como cenário para ilustrar esta relação foi selecionado um usuário aleatoriamente. O gráfico da figura 4(a) representa a relação da força com o grau de certeza. Inicialmente, o usuário tem uma reputação de 0,09 e uma força 1, ou seja, teve apenas uma contribuição revisada. Depois de um tempo ele tem 31 contribuições revisadas e os mesmos no grau de certeza 0,09. O fator diferencial destes dois momentos é a força. No segundo momento a crença reputação tem uma veracidade maior por ter sua análise composta por 30 revisões a mais que no primeiro momento. Desta maneira pode ser observado que a força é um fator relevante na análise de uma crença de grupo do tipo reputação.

O segundo experimento apresenta a aplicação da heurística proposta para o patrulheiro revisar as contribuições da Wikipédia, usando um conceito da estatística descritiva conhecido como quartil [8]. Este conceito é a representação de três valores que dividem uma amostra em quatro partes iguais. Desta maneira ao utilizar estes valores como ponto de corte é possível classificar os usuários como ruins, medianos e bons, sendo os usuários ruins aqueles que têm um grau de certeza inferior ao primeiro quartil, os usuários medianos são aqueles que têm um grau de certeza entre o primeiro e o terceiro quartil e os bons usuários são aqueles que têm grau de certeza superior ao terceiro quartil. O gráfico da figura 4(b) representa respectivamente a divisão dos usuários (Losangos) pelos quartis (linhas Horizontais) como proposto na heurística para um tópico da wikipédia.

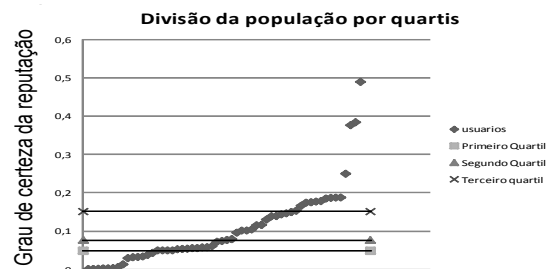
Com a classificação dos usuários efetuada foi atribuído os conceitos aos usuários contribuintes presentes nos dados de teste. Nesta página exemplo foram classificados 28,84% dos usuários como ruins e portanto necessitam revisão. 28,84% dos usuários foram considerados medianos e tiveram a revisão recomendada. 3,84% dos usuários foram classificados como bons e portanto recebem uma menor necessidade de revisão. Por fim, 38,46% são usuários desconhecidos do sistema de reputação. Através desse experimento observa-se que o uso da reputação, juntamente com a heurística proposta, pode ser um instrumento de classificação dos usuários da Wikipédia para o auxílio dos patrulheiros durante a revisão das contribuições.

---

<sup>1</sup><http://download.wikipedia.org/simplewiki>



(a) Gráfico da evolução da reputação de um usuário



(b) Usuários particionados pelos quartis

Figura 4. Gráficos dos experimentos

## 8. Análise e Discussão

Através de uma análise do sistema proposto observou-se uma série de características que são desejáveis em um sistema de reputação para um SMA aberto.

A primeira delas diz respeito à **visibilidade**, ou seja, as crenças de grupo têm de estar disponível a todos os agentes. Por exemplo, todos os agentes do grupo têm de conhecer o estado atual da reputação. No caso dos SMA abertos, no qual agentes entram e saem a qualquer momento do sistema, essa característica ganha uma complexidade maior uma vez que agentes que recém ingressaram no sistema precisam conhecer as crenças do grupo. Para o sistema proposto possuir essa característica as crenças de grupo seguem um modelo global. Esse modelo é uma simplificação do comportamento social humano uma vez que as mensagens de testemunho deixam de circular de agente em agente para ser recebido por uma estrutura externa. A simplificação proposta por essa visão ainda se apresenta eficiente para sociedades virtuais, visto a quantidade de sistemas de reputação que seguem essa abordagem.

A segunda característica observada é a **dinamicidade** que diz respeito à periodicidade da atualização das crenças do grupo. Com a livre entrada e saída dos agentes do sistema o conteúdo de uma crença de grupo está em constante mudança. Nesse sentido para manter essa dinamicidade elevada foi utilizado um modelo de alimentação por anúncios o que permite uma elevada taxa de atualização da crença e com um baixo número de mensagens, uma vez que os agentes que concordam com o as crenças de grupo não precisam se manifestar.

A terceira característica observada é a **flexibilidade** de como as crenças de grupo são sintetizadas. Diferentes tipos de aplicações contemplam diferentes maneiras de sintetizar as suas crenças de grupo. Nessa direção à proposta apresentada oferta uma arquitetura que permite aos agentes determinarem as regras que serão utilizadas para sintetizar as crenças de grupo.

A quarta característica observada diz respeito ao suporte a **heterogeneidade** dos agentes. Em SMA abertos com o livre ingresso dos agentes do sistema é possível que esses sejam compostos por diferentes tecnologias. Nesta direção a proposta usa do conceito de Agentes e Artefatos [11] através da biblioteca CarTaGo que dá suporte a essa questão.

A quinta característica refere-se à **reusabilidade**. Sistemas de reputação têm sua aplicação em diferentes contextos, para tanto eles devem ser genéricos o suficiente para que sua estrutura possa ser utilizada em diferentes contextos. Neste sentido essa proposta

oferece uma arquitetura baseada em artefatos que permite sua utilização em diversos contextos fazendo apenas algumas parametrizações.

A sexta característica diz respeito ao **tratamento da credibilidade**. Sistemas de reputação tem que lidar com diferentes informações, cada qual, com sua credibilidade. Esta credibilidade pode ser decomposta em dois itens: o grau de certeza e a força. O grau de certeza fornece a possibilidade de o grupo quantificar o quanto acredita em uma determinada reputação. A força é um valor que permite quantificar a relevância da crença. Nesta direção foi observado o ForTrust, que em sua proposta trata apenas dois estados mentais onde os agentes possuem ou não reputação. Com a finalidade de tratar a credibilidade essa proposta estendeu o modelo ForTrust adicionando dois elementos, o grau de certeza e a força.

## 9. Conclusão

A composição, proposta nessa pesquisa, para as crenças de grupo permitiu a construção de uma arquitetura de implementação para o modelo de reputação ForTrust. Apesar da complexidade inerente a composição das crenças de grupo e por conseguinte sua arquitetura, essa proposta tornou possível a composição de uma arquitetura que atendesse as principais características de um SMA aberto. Para o ForTrust, a contribuição apresentada por este trabalho foi a adição de dois novos conceitos no modelo, o grau de certeza e a força. Esta contribuição permite quantificar a reputação e qual a credibilidade desta. Para o ForTrust ainda foi concebido neste trabalho uma arquitetura de implementação baseada em artefatos de crenças de grupo. Outra contribuição desse trabalho foi a heurística de classificação utilizando quartis. Esta heurística, juntamente com o ForTrust demonstrou que é possível fornecer um auxílio aos patrulheiros na revisão das contribuições com o uso de um sistema de reputação.

No decorrer desse trabalho surgiram algumas questões que não foram tratadas devido a delimitação do escopo proposto. Estas questões representam os principais temas de continuação para este trabalho, são elas: (i) o uso dos artefatos de crença de grupo em diferentes escopos que não a reputação; (ii) o modo como os agentes racionalizam as crenças de grupo, tal estudo pode elaborar estratégias para os agentes processarem as informações vindas das crenças de grupo; (iii) a busca por diferentes métodos de síntese de crenças, explorando abordagens oriundas das ciências sociais e das técnicas de inteligência computacional para elevar a qualidade das crenças de grupo obtidas e aprimorar a adaptabilidade dos artefatos de crença de grupo em diferentes escopos; (iv) aplicação efetiva do sistema de reputação na Wikipédia; o que permitiria a utilização da ferramenta por usuários reais, desta maneira fornecendo um resultado mais preciso da eficiência da ferramenta.

## Referências

- [1] B. T. Adler, L. de Alfaro, I. Pye, and V. Raman. Measuring author contributions to the wikipedia. *The Internacional Symposium on Wikis and Open Collaboration*, 1, 2008.
- [2] C. Castelfranchi and R. Falcone. Principles of trust for MAS: Cognitive anatomy, social importance and quantification. *ICMAS*, 1:72–79, 1998.
- [3] C. Castelfranchi, R. Falcone, B. Firozabadi, and Y. Tan. Special issue on trust, deception and fraud in agent societies. *Applied Artificial Intelligence Journal*, 1:763–768, 2000.

- [4] J. L. D. Coi and L. Vercouter. Trust put to the test: A testcase for a cognitive trust model. In O. Boissier, J. Bradshaw, L. Cao, K. Fischer, and M.-S. Hacid, editors, *IAT*, pages 248–255. IEEE Computer Society, 2011.
- [5] A. Herzig, E. Lorini, J. F. Hübner, and L. Vercouter. A logic of trust and reputation. *Logic Journal of the IGPL*, 18(1):214–244, Feb. 2010. Special Issue “Normative Multiagent Systems”.
- [6] A. Herzig, E. Lorini, J. F. Hübner, J. Ben-Naim, C. Castelfranchi, R. Demolombe, D. Longin, and L. Vercouter. Prolegomena for a logic of trust and reputation. In *4th For-Trust Meeting*, 2008.
- [7] Y. Krupa, L. Vercouter, J. F. Hübner, and A. Herzig. Trust based evaluation of wikipedia’s contributors. In *Proceedings of the 10th International Workshop on Engineering Societies in the Agents World X*, ESAW ’09, pages 148–161, Berlin, Heidelberg, 2009. Springer-Verlag.
- [8] C. LOESCH and C. E. STEIN. *Estatistica Descritiva E Teoria Das Probabilidades*, volume 1. EDIFURB, 1 edition, 2008.
- [9] G. Muller and L. Vercouter. Decentralized monitoring of agent communications with a reputation model. In R. Falcone, S. Barber, J. Sabater-Mir, and M. Singh, editors, *Trusting Agents for Trusting Electronic Societies*, volume 3577 of *Lecture Notes in Computer Science*, pages 99–99. Springer Berlin / Heidelberg, 2005.
- [10] A. Ricci, M. Viroli, and A. Omicini. Cartago: An infrastructure for engineering computational environments. In *3rd Inter. Workshop environments for Multi-Agent Systems*, pages 102–119, 2006.
- [11] A. Ricci, M. Viroli, and A. Omicini. The A&A programming model and technology for developing agent environments in MAS. *ProMAS*, 1:89–106, 2008.
- [12] J. Sabater and C. Sierra. Regret: A reputation model for gregarious societies. *Proc. of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, 1:61–69, 2001.
- [13] R. Tuomela. ‘can collectives have beliefs’? In *Language, Knowledge, and Intentionality: Perspectives on the Philosophy of Jaakko Hintikka*, *Acta Philosophica Fennica* 48, 1990.
- [14] R. Tuomela. On the structural aspects of collective action and free-riding. *Theory and Decision* 32, 1:165–202, 1992.
- [15] A. G. West, sampath kannan, and I. Lee. Detecting wikipedia vandalism via spatio-temporal analysis of revision metadata. *Proceedings of the Third European Workshop on System Security*, 2010.
- [16] G. Zacharia. Collaborative reputation mechanisms for online communities. Master’s thesis, Massachusetts Institute of Technology, 1999.