

MAL-FITT: MyAnimeList Forum Interpreter Through Text

Gianluca Lodron Zuin¹, Luiz Felipe Gonçalves Magalhães¹, Túlio Corrêa Loures¹

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

Abstract. *The great amount of themes being discussed simultaneously in discussion forums can present an extra difficulty for outsiders. With this in mind, we develop the method MAL-FITT which returns the main comment of a discussion thread, and explores different ranking metrics developed. We avoided the use of generic "relevance" indicators, such as "likes" and "upvotes". The thread and its posts were represented using TF-IDF and sentiment polarity based models, and then applied in our method. Fifteen threads were approached and the method's output was compared to the one given by three human evaluators. The results were interesting, like the idea itself, being mostly unexplored so far.*

Resumo. *A grande quantidade de temas sendo discutidos simultaneamente em fóruns de discussão pode apresentar uma dificuldade extra a pessoas externas. Com isso em vista, foi elaborado o método MAL-FITT, que retorna o comentário central de um tópico de discussão, e explora o uso de diferentes métricas de ranking desenvolvidas. Evitamos o uso de indicadores genéricos de "relevância" como curtidas e upvotes. Cada tópico e suas postagens foram representados com auxílio dos modelos de TF-IDF e polaridade de sentimento, sendo então aplicado em nosso método. Quinze tópicos são abordados e a saída do método é comparada com a de três avaliadores humanos. Os resultados são interessantes, assim como a ideia, tendo sido pouco explorada até então.*

1. Introdução

Em uma discussão, tem-se frequentemente grande quantidade de temas e tópicos abordados. Para alguém externo, isso pode significar uma barreira para o entendimento da ideia geral que está sendo tratada; logo, seria interessante poder apresentar para ele, rapidamente e automaticamente, a ideia central da discussão, aquilo que a resume.

Esse é um ponto especialmente frequente em diversos fóruns de discussões presentes na Web: tem-se grande quantidade de postagens em um mesmo tópico, muitas vezes com conteúdos extremamente divergentes. O objetivo desse trabalho é elaborar um método que retorne ao usuário o comentário central, que melhor resume a ideia de determinado tópico de um fórum de discussão. Uma motivação é evitar o uso de indicadores de "importância", como curtidas (*facebook*) e *upvotes* (*reddit*), pois, muitas vezes, retratam postagens mais populares, mas não necessariamente mais representativas do tópico.

Esse trabalho foi avaliado por meio do uso de alguns fóruns de discussão do portal *MyAnimeList*¹, uma página dedicada à listagem e discussão dos *animes* e *mangas* os quais seus usuários pretendem assistir, estão assistindo ou já os assistiram.

¹<http://myanimelist.net/>

A ideia central deste estudo é representar o tópico da discussão pelo método *TF-IDF* [Salton and Buckley 1988], e realizar uma análise da polaridade associada a cada entidade referenciada nas postagens usando o *VADER* [Hutto and Gilbert 2014]. Partindo dessa representação, utiliza-se a *Similaridade Cosseno* [Baeza-Yates et al. 1999, Sidorov et al. 2014] de formas variadas para encontrar aquela que seria a postagem central da discussão. Os métodos propostos são avaliados por meio de uma comparação com um conjunto de dados rotulados manualmente.

Para o restante do artigo, uma breve discussão de trabalhos semelhantes é apresentada na Seção 2. A Seção 3 detalha a base de dados utilizada para os experimentos enquanto a 4 discute as técnicas usadas para a representação dos dados e uso dos mesmos. A Seção 5 descreve como foram feitos os experimentos e os resultados encontrados. Finalmente, a Seção 6 traz a conclusão e projeta trabalhos futuros.

2. Trabalhos Relacionados

O portal *MyAnimeList* já foi foco de outros trabalhos científicos, em particular de sistemas de recomendação [Miller and Southern 2014, Vie et al. 2015]. Miller estuda diferentes combinações de métodos e técnicas para obter resultados precisos e rápidos, baseando-se, principalmente, em filtragem colaborativa. Vie relata a arquitetura básica de seu sistema, o *Mangaki*, que é baseado em elicitación de preferências sobre um grupo de exemplos apresentados, associando uma polaridade às diferentes possíveis respostas e usando-as para calcular a similaridade entre usuários.

Ao sumarizar um tópico, primeiramente obtém-se uma representação intermediária do texto, que capta os temas discutidos na entrada ou cobre um conjunto diversificado de possíveis indicadores de importância. As sentenças presentes são, então, ordenadas. Um resumo é produzido selecionando as frases de maneira gulosa ou a partir da otimização do melhor conjunto global de sentenças. Nenkova faz uma revisão extensa dos principais métodos utilizados nesse domínio, abordando desde algoritmos baseados em frequência (como o *TF-IDF* utilizado nesse trabalho) até os baseados em aprendizado de máquina [Nenkova and McKeown 2012]. Este autor também discute que, dependendo do contexto da entrada avaliada, informações extras latentes podem ser utilizadas para guiar o processo de sumarização. Aqui, exploramos o uso da polaridade em uma análise de sentimentos associada às postagens.

Trabalhos contendo diferentes métodos foram realizados na área de sumarização de discussões online. Zhou levanta uma discussão sobre as particularidades de se trabalhar com esse tipo de dados [Zhou and Hovy 2006]. A utilização de *Support Vector Machines* (SVM) para obter as seções mais relevantes é uma das estratégias abordadas [Huang et al. 2007, Zhou et al. 2005]. Auria levanta algumas das desvantagens da utilização de SVM [Auria and Moro 2008]. No presente trabalho buscou-se uma representação mais simples sem a necessidade de modificar a dimensionalidade dos dados.

Demais métodos de sumarização criam resumos extraíndo frases relevantes de múltiplos documentos criando um novo sintético [Mani and Maybury 1999, Salton et al. 1997, Gong and Liu 2001]. Diferente dos demais, esse estudo buscou identificar as postagens que representem o tópico e o sumarie, assim como realizado nos métodos baseados em SVM.

3. Base de Dados

O portal MyAnimeList.net (MAL) possui uma listagem de séries de animação orientais (animes), onde cada uma possui a uma página própria. A partir dela, é possível acessar seu respectivo fórum de discussão. Esses fóruns são compostos por diversos tópicos. Os principais tópicos de cada fórum são aqueles referentes às discussões dos episódios individuais de seus respectivos animes (seguindo o formato $\langle anime \rangle$: *Episode* $\langle numero \rangle$ *Discussion*). Além desses, existem também outros tipos de tópicos, dentre eles, discussões da série como um todo, personagens favoritos, detalhes técnicos, entre outros. O esquema estrutural do MAL pode ser visualizado na Figura 1.

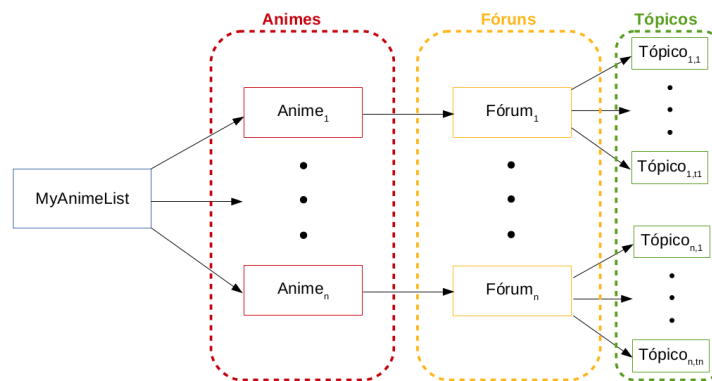


Figura 1. Estrutura do portal MyAnimeList.net

Atualmente, o portal possui mais 3.000.000 de usuários, contribuindo ativamente em mais de 655.000 tópicos de discussão de 48.000 séries diferentes (estatísticas de Dezembro de 2015). Algumas das funcionalidades disponíveis para os usuários são a citação, na qual os usuários podem referenciar postagens de outros usuários, *spoiler*, onde usuários podem esconder parte de suas postagens por conterem conteúdo que pode revelar fatos da narrativa, que podem arruinar a experiência de outros usuários, e denúncia, onde postagens possivelmente abusivas são enviadas aos moderadores para avaliação.

3.1. Tratamento dos Dados

Os dados utilizados para os testes foram obtidos por meio da API² disponibilizada pelo MAL. A análise do VADER usa as postagens originais, sem nenhum pré-processamento, apenas segmentando-as em sentenças e palavras. No *TF-IDF*, as postagens sofrem um tratamento para gerar a coleção de dados. Inicialmente são retirados todos os caracteres especiais (não-alfanuméricos) e as palavras são normalizadas para letras minúsculas. Além disso, utiliza-se uma lista de *stopwords* e um algoritmo de *stemming*.

3.1.1. Stopwords

Foram retiradas do texto palavras consideradas pouco relevantes, que não agregam valor semântico à futura coleção. Para isso, foi utilizada a lista SMART de *stopwords* [Lewis et al. 2004] acrescida de outras palavras específicas ao contexto do trabalho:

²Informações sobre a API, incluindo link para cadastro na *whitelist* e fórum de discussão, podem ser encontradas em <http://myanimelist.net/clubs.php?cid=13727>

episode anime end show character watch serie manga people
scene forum discuss myanimelist mal

Embora a utilização de *stopwords* possa simplificar a análise do texto, filtrá-las pode causar problemas. Um deles é o de impedir a distinção de nomes que fazem uso delas para formar alguma entidade representativa, como em '*Firefox OS*'. Outro problema está no fato de que alguns termos como '*best*', '*better*' ou '*unlikely*' estão presentes na lista SMART, e a análise de polaridade das sentenças pode ser afetada pela ausência desse tipo de palavra. Além disso, grande parte das postagens do MAL são extremamente curtas, contendo apenas algumas poucas palavras. Propomos, então, testes mantendo e filtrando *stopwords*, os quais serão retratados na seção 5.3.

3.1.2. Stemming

Todas as palavras, antes de serem inseridas na coleção, passaram por um processo de *stemmização* usando o algoritmo *Porter Stemming* [Porter 1997]. Esse algoritmo transforma formas variantes de palavras em uma única representação concisa, denominada *stem*. A ideia do processo segue o Princípio da Anotação [Stefik 1995], no qual o objetivo é mapear, para uma forma unívoca, objetos que possuem uma mesma etimologia.

3.2. Montagem da Coleção

Com os textos tratados, criou-se uma coleção de dados. Nesse estudo, uma coleção representa o tópico de discussão tratado, e suas postagens são ditos documentos de dada coleção. A Seção 4 descreve os métodos usados para representar os dados da coleção e mostra a representação usada no contexto desse trabalho. Inicialmente os dados da coleção são colocados no formato de *bag-of-words*, com a contagem de cada palavra presente em cada postagem do tópico.

4. Métodos Utilizados

O método *MAL-FITT* proposto utiliza a coleção formada pelo tópico e suas postagens representada pelo *TF-IDF* ou pela polaridade. A ideia consiste em, a partir dessas representações, utilizar uma das técnicas de *ranking* (detalhadas nas Seções 4.1 e 4.2) para montar uma ordenação de documentos. O documento do topo do ranking seria aquele que melhor representa a coleção em questão.

4.1. TF-IDF

O *TF-IDF* (*Term Frequency-Inverse Document Frequency*) é uma medida estatística que tem o intuito de indicar a importância de uma palavra de um documento em relação a uma coleção de documentos. O valor de uma palavra, nessa representação, aumenta de acordo com o número de ocorrências do termo no documento e a raridade do termo na coleção.

Dada a representação formada no *TF-IDF*, formulou-se três métodos de *rankings* de documentos baseados na métrica de similaridade de cosseno entre eles:

- Maior similaridade com a concatenação dos documentos (*TFIDF_COMM*)
- Maior similaridade média com todos os outros documentos (*TFIDF_AVGC*)
- Maior similaridade mínima com todos os outros documentos (*TFIDF_MAXC*)

Para o primeiro método, foi inserido, ao fim da coleção, um documento que é a concatenação de todos os documentos anteriores, com o objetivo de que ele represente a coleção como um todo. A hipótese é que postagens próximas a ele são, possivelmente, as mais prováveis de representarem um resumo da discussão. O método de *ranking TFIDF_COMM* compara esse novo documento com os demais, buscando o mais similar a ele.

O segundo e o terceiro métodos calculam a similaridade par-a-par entre os documentos, mas eles diferem na forma como é decidido o documento mais relevante. Enquanto o *TFIDF_AVGC* (Equação 1) escolhe aquele com maior similaridade média para todos os outros documentos, o *TFIDF_MAXC* (Equação 2) seleciona o documento com a máxima similaridade mínima para qualquer outro documento.

$$f_{avgc}(T = \text{topico}) = \max_{i \in T} \left(\frac{1}{|T|} * \sum_{j \in T, j \neq i} \text{similaridade}(i, j) \right) \quad (1)$$

$$f_{maxc}(T = \text{topico}) = \arg\max_{i \in T} (\min_{j \in T, j \neq i} (\text{similaridade}(i, j))) \quad (2)$$

4.2. Polaridade

No método de *Polaridade (POL_SENT)*, observa-se cada entidade foco como um ponto no espaço tridimensional. Suas coordenadas são derivadas de sua polaridade positiva, negativa e neutra, oriundas da análise do *VADER* para a sentença que ela se encontra. Calcula-se a distância de cosseno das coordenadas de um determinado termo em uma deliberada sentença à média das polaridades desse termo nas demais (Equação 3). Foi criado um *ranking* das postagens mais relevantes a partir dos termos que as compõem.

Classificou-se como foco de uma frase os substantivos (*nouns*) presentes na mesma. A classificação do foco de uma sentença é feita utilizando o classificador nativo da NLTK³ para *part-of-speech*. Uma vez classificados os termos da frase, buscam-se entidades nomeadas (como personagens ou locais) e estas são adicionadas como foco das sentenças. Logo, uma sentença possui um conjunto de focos. Para cada um deles, atribuímos a polaridade da sentença.

Por exemplo, para a postagem da Figura 2, tem-se duas frases: *"To be honest, I actually think everything was ok about the show."* e *"I've read the manga too."*. A primeira possui dois focos, *'honest'* e *'show'* e a análise do *VADER* retorna que ela é 0.621 neutra e 0.379 positiva. Calcula-se a similaridade entre os termos na frase e a média em todo o tópico. Denomina-se como *fitness* do termo o quão apto ele é em resumir o tópico. Partindo de um princípio similar ao do *TF-IDF*, assumimos que termos que aparecem frequentemente são candidatos interessantes no resumo da discussão. O *fitness* do termo é o produto simples da sua similaridade pela sua frequência no tópico (Equação 4).

$$dist_{x=termo} = \cos(\theta) = \frac{\sum pol_x pol_{\bar{x}}}{\sqrt{\sum pol_x^2} \sqrt{\sum pol_{\bar{x}}^2}} \quad (3)$$

$$fitness(x = termo) = sim_x * frequencia_x = (1 - dist_x) * frequencia_x \quad (4)$$

onde *pol* denota as saídas da análise do *Vader*, *x* a palavra na sentença atual e \bar{x} a média da polaridade da palavra em todas as sentenças do tópico.

³Natural Language Toolkit - <http://www.nltk.org/>

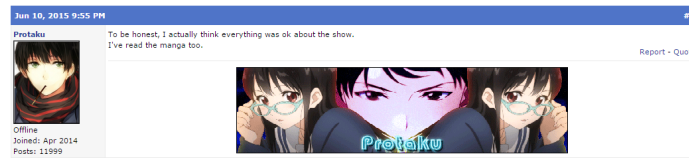


Figura 2. Postagem do usuário Protaku no fórum de discussão do anime Akame Ga Kill discutindo sobre o que ele mudaria na série.

A aptidão de uma postagem como resumo é encontrada pelo somatório do *fitness* de cada um de seus focos dividido pelo número de sentenças presentes. Sentenças em uma mesma postagem normalmente são complementares e seu conjunto tende a abranger os temas principais, o que justifica o numerador da fórmula. Uma postagem com poucas sentenças é preferível à uma com muitas, se ambas resumem os mesmos assuntos, visto que postagens muito grandes tendem a abordar diversos assuntos, fazendo com que os temas discutidos acabem difundidos. De maneira geral, o *fitness* de uma postagem pode ser encontrado pela média de suas sentenças:

$$fitness(p = \text{postagem}) = \frac{\sum_{s \in S} fitness(s)}{|S|} = \frac{\sum_{s \in S} \sum_{f \in F} fitness(f)}{|S|} \quad (5)$$

onde S denota o conjunto de sentenças na postagem p e F o conjunto de focos em cada sentença s , que quando agrupados se tornam o conjunto de focos na postagem p .

Para a postagem exemplo, 'show' aparece várias vezes ao longo do tópico, porém, os demais termos empregados são pouco utilizados. Apesar da distância de cosseno retornar que as polaridades são próximas, o *fitness* da postagem é relativamente baixo. Para fins comparativos, a postagem mais relevante possui três sentenças e *fitness* 17.896, contendo 2 dos 5 termos mais usados: ('Tatsumi' e 'death'), enquanto a postagem exemplo possui *fitness* 7.863.

5. Experimentos

Foram avaliados quinze (15) tópicos, cada um deles retirado do fórum de um anime diferente. Os tópicos foram selecionados tentando-se manter um conjunto de temas distintos na coleção, a fim de evitar que a análise ficasse enviesada a algum estilo específico de discussão. Foram observadas discussões sobre a trama geral do anime, votações de personagens favoritos, opiniões sobre algum evento da série e conversas sobre produção e mercado. Para cada tópico escolhido, foram selecionadas as dez (10) postagens mais relevantes. Essa seleção foi realizada por meio de uma votação entre 3 avaliadores. A figura 3 ilustra os tópicos selecionados⁴.

5.1. Métodos de Avaliação

Para avaliar o quanto a postagem retornada pelo *MAL-FITT* é uma boa representação do tópico como um todo, utilizou-se uma série de métricas de qualidade de ranqueamento. Foram calculados os valores de Precisão, Revocação, F1 e *Mean Average Precision* (MAP) para a lista ranqueada das cinco (5) postagens selecionadas como mais relevantes pelo *MAL-FITT*, tomando-se os dados descritos na tabela de validação como *ground truth*.

⁴Tabela completa pode ser encontrada no link: <http://goo.gl/IbP4YA>

- *POL_SENT* (maior similaridade com a média das polaridades)
- *TFIDF_COMM* (maior similaridade com a concatenação dos documentos).
- *TFIDF_AVGC* (maior similaridade média com todos os outros documentos).
- *TFIDF_MAXC* (maior similaridade mínima dentre todos os documentos).

TOPICO	ANIME	1	2	3	4	5	6
?topicid=1393108	Akame	17	19	13	16	15	29
?topicid=717919	Harlock	29	14	7	6	33	4
?topicid=49381	GTO	31	32	7	11	2	18
?topicid=1121883	D-Frag	17	23	24	27	35	31
?topicid=1449160	Glasslip	6	11	22	23	37	9
?topicid=1398255	Nisekoi	7	14	17	4	5	3
?topicid=254671	MInami-ke	4	6	14	12	16	20
?topicid=4407096	SAD	6	4	5	24	44	6

Figura 3. Parte da Tabela de Validação, mostrando os números das postagens votadas como mais relevantes em cada tópico analisado. Animes destacados são aqueles nos quais o método *TF-IDF* AVGC teve piores resultados.

5.2. Mean Average Precision (MAP)

A métrica utilizada na avaliação da qualidade da postagem para representar os tópicos é denominada *Mean Average Precision* (MAP) [Voorhees and Harman 2005]. O MAP determina o quão bem colocados foram os documentos relevantes na lista de documentos retornados.

5.3. Análise dos Resultados

Os gráficos da figura 6 ilustram a performance da cada um dos métodos na ausência e presença de *stopwords*. Em todas as métricas, a performance do método de polaridade se mostrou semelhante quando comparada com a do *TF-IDF*. Neste sentido, infere-se que o sentimento envolvido nas postagens deu pouca informação ao método. A escolha das postagens mais relevantes acabou se baseando principalmente na frequência de cada entidade foco. Como o algoritmo do *TF-IDF* parte de um pressuposto similar, os resultados próximos são justificáveis. Uma verificação das polaridades das sentenças justifica essa afirmação. O desvio na polaridade de cada postagem é pequeno e, de maneira geral, tendencioso ao neutro, como ilustrado na figura 4.

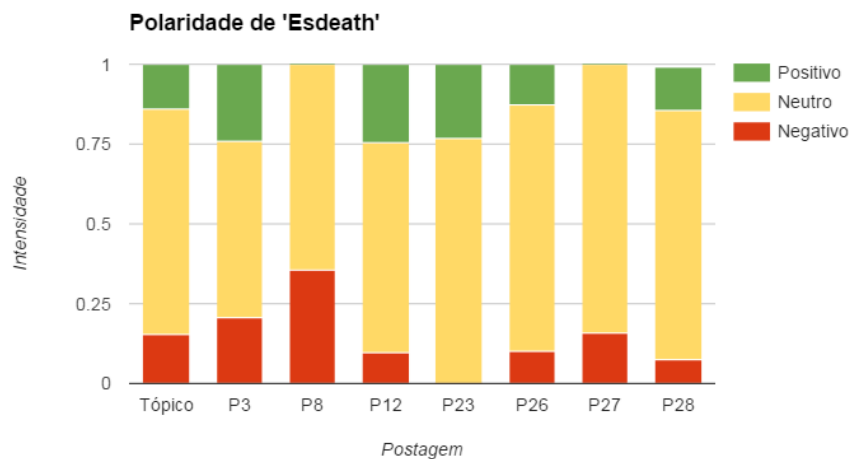


Figura 4. Polaridade do termo 'Esdeath' (um dos personagens principais) no tópico sobre o anime Akame Ga Kill.

Dentre as métricas de similaridade empregadas que utilizam o *TF-IDF*, o *TFIDF_AVGC* teve o melhor resultado. No *TFIDF_COMM*, nota-se uma redução do valor retornado pelo *TF-IDF* para as palavras mais frequentes. Isto ocorre porque, na etapa *IDF*, apenas é verificado se um termo existe ou não nesse novo documento. Ao se calcular a média das similaridades, valorizamos palavras frequentes, uma vez que documentos contendo os mesmos termos possuem um valor próximo, remediando uma das falhas do *TF-IDF* nesse cenário e, portanto, tendo um resultado superior ao *TFIDF_MAXC* e *TFIDF_COMM*.

No gráfico da figura 5, é possível visualizar o impacto das *stopwords* na performance dos métodos em cada tópico analisado. Houve casos em que a remoção de *stopwords* piorou a resposta da ferramenta mas, na maior parte dos tópicos, ocorreu uma melhora, com exceção no método de *Polaridade*. A remoção de *stopwords* faz com que a pequena quantidade de focos em cada sentença fique potencialmente menor ainda. A perda de informação acaba debilitando o método no caso geral. A performance ruim do *TF-IDF*, em alguns tópicos, pode ser justificada pelo uso de palavras similares. O tópico 2, por exemplo, discute a qualidade gráfica da série. Apesar do tema ser bem definido, uma diversidade grande de palavras é utilizada como '*graphic*' '*CGI*' '*CG*' '*animation*'. Todas essas palavras possuem um sentido e contexto similar, contribuindo como um todo para o tema do tópico. Todavia, o método *TF-IDF* não possui a capacidade de abstrair o sentido das palavras e acaba tendo um resultado ruim em casos como esse.

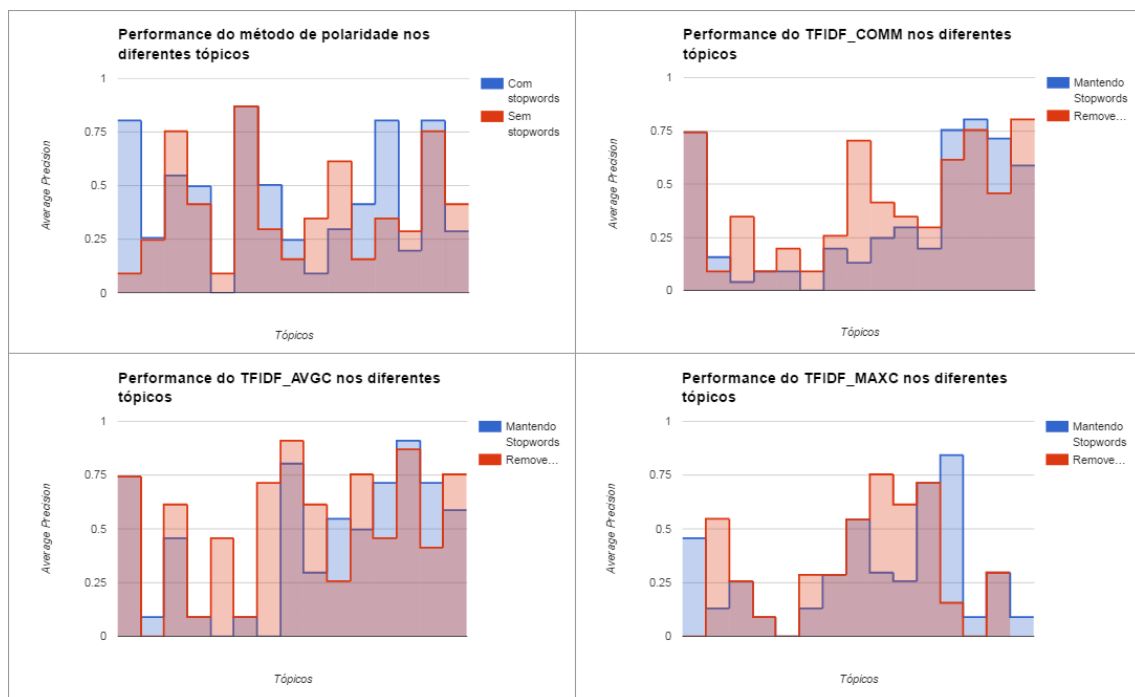


Figura 5. Performance dos diferentes métodos sobre todos os tópicos avaliados usando a métrica de *average precision*.

Uma alternativa para tentar contornar o problema de sentenças com focos diferentes, mas que seus termos compartilham de uma semântica similar, é utilizar a WordNet⁵.

⁵<https://wordnet.princeton.edu/>

Iterando pelos diversos possíveis significados de cada termo, tentou-se colapsar entidades semelhantes usando o algoritmo de similaridade de *Wu-Palmer* [Wu and Palmer 1994]. Os resultados obtidos foram inferiores aos do método padrão, com uma queda de precisão na maioria dos tópicos. Isso ocorre, principalmente, devido ao colapso indevido de algumas entidades. No tópico de *Akame Ga Kill*, por exemplo, 'character' e 'death' foram colapsadas com um índice de similaridade de 0.88 devido aos significados:

- *Character: an imaginary person represented in a work of fiction (play or film or story).*
- *Death: the personification of death.*

que de fato são similares no contexto do teatro e da dramaturgia, mas não para o tópico em questão. Dessa forma, realizar uma análise de contexto seria necessário para utilizar a WordNet com eficácia.

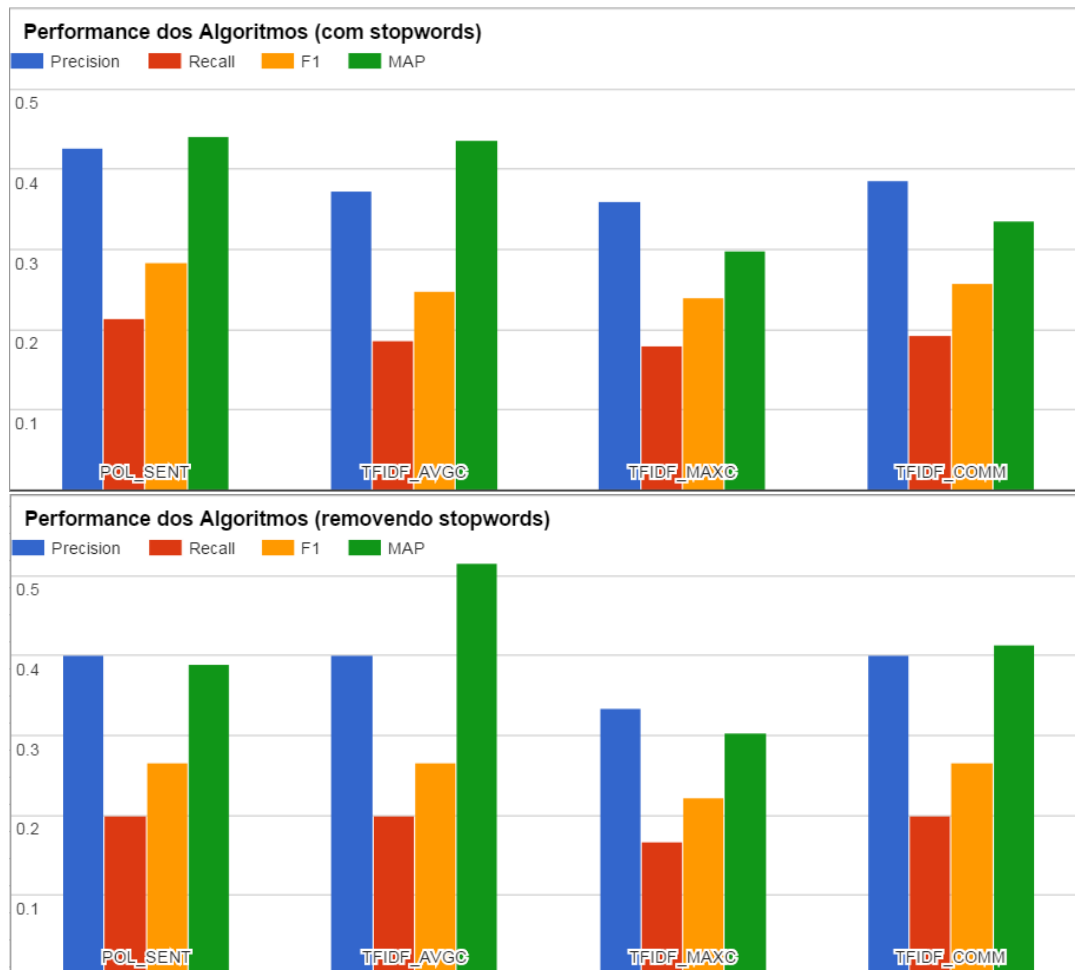


Figura 6. Avaliação dos resultados dos métodos utilizados após a execução dos 15 tópicos filtrando e mantendo as stopwords.

6. Conclusão e Trabalhos Futuros

Propôs-se nesse trabalho um arcabouço, denominado *MAL-FITT*, que permite a identificação da postagem central de um tópico em um fórum de discussão. Para isso,

o tópico e suas postagens foram representados através do *TF-IDF* e por meio de uma análise de sentimentos, métricas de similaridade, utilizando essas representações, permitiram julgar aquela postagem mais relevante.

O problema de identificar os tópicos mais relevantes é qualitativo, pois, diferentes avaliadores podem ter diferentes opiniões. De fato, como ilustrado na figura 7, na média, os humanos concordam entre si 60% das vezes. O método de análise de polaridade concorda com os humanos, na média, em 43% das postagens. Considerando os problemas apresentados, sua performance atingiu valores interessantes: 73% do considerado ótimo, que seria a análise humana.

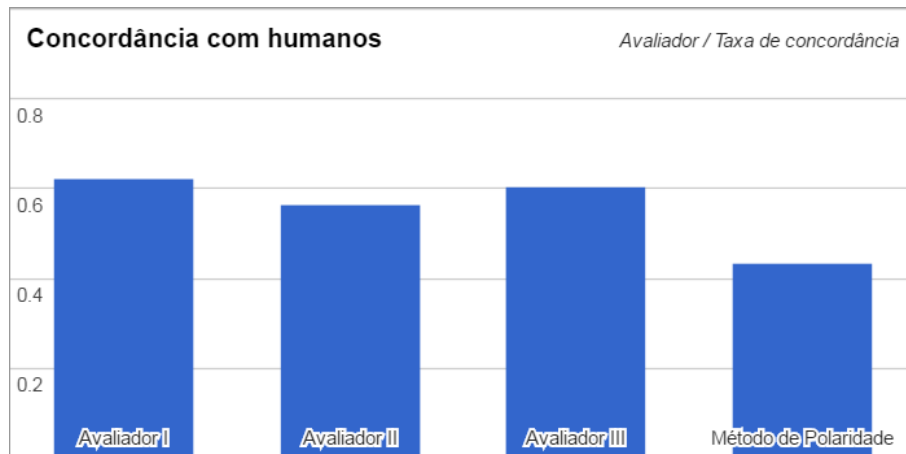


Figura 7. Concordância entre humanos e do método de análise de polaridade. Na média os humanos concordam em 59.8% do tempo (taxa de concordância média de 0.598). O método proposto atinge 72.9% desse valor, com uma taxa de concordância de 0.436.

Os métodos do *MAL-FITT* apresentaram resultados aceitáveis com algumas de suas variações de similaridade. É possível notar que os melhores valores foram, em geral, com o uso do método de polaridade. Dentre as representações do *TF-IDF*, o melhor foi obtido com o uso da maior similaridade média (*TFIDF_AVGC*).

Como esperado em todos os métodos, nos casos onde termos similares eram utilizados, não obteve-se bons resultados (exemplo: tópico que se debatia o visual de uma animação, com foco em termos como 'cgi', 'graphics', 'visuals', etc.). Isso ocorre porque as representações não trabalham com a semântica das palavras, então 'CG' ou 'CGI' são considerados completamente distintos. Caso fossem consideradas como sinônimos, o método possivelmente apresentaria um resultado superior.

A alternativa usando WordNet não se mostrou efetiva em tentar solucionar esse problema. Isso ocorre, em parte, devido aos métodos de similaridades nativos da WordNet. Ela é calculada em função da distância de dois termos na árvore léxica e não, necessariamente, pelo seu *gloss*, pelo seu sentido semântico. Isso faz com que as métricas de similaridade, na verdade, meçam o parentesco entre palavras. 'Love' e 'Hate' tem uma similaridade maior que 'Love' e 'Romance' por exemplo. Ao tentar colapsar os termos próximos, relações, como antônimos, acabam sendo ressaltados, o que prejudica o desempenho do método. Além disso, a falta de contexto no método implementado causa o colapso indevido de alguns termos, como 'character' e 'death'.

Em contrapartida, nos tópicos mais simples, como o caso de discussões sobre qual é o melhor personagem de uma série, os métodos se mostraram extremamente efetivos (Tópico 13 da avaliação: 0.87 AP no *TFIDF_AVGC* removendo *stopwords*). Isso acontece, pois, a maioria das postagens são simples e normalmente contém apenas o nome do personagem favorito em questão, sem variações como foi no caso descrito anteriormente.

Como trabalhos futuros, pretende-se realizar um estudo da implicação da presença e ausência de citações, pois elas podem ter um peso significativo para seleção de uma dada postagem. Pretende-se também realizar um maior e melhor processo avaliativo, com mais pessoas e mais tópicos sendo avaliados. Outra ideia é a de utilizar métricas de avaliação mais adequadas ao problema, como a proposta em [Cremonesi et al. 2010].

Finalmente, propõe-se a utilização de outras formas de representação dos documentos usando outras possíveis características relevantes. A utilização de métodos de similaridade entre palavras, baseados no seu sentido semântico, certamente auxiliaria no desempenho do *MAL-FITT*. Comparar a similaridade das frases do *gloss* de cada termo é uma hipótese.

Agradecimentos

Gostaríamos de agradecer ao MyAnimeList.net e a todos os envolvidos na manutenção do site, especialmente o administrador conhecido como *Xinil*⁶. O site não permite que seus dados sejam baixados normalmente (proteção utilizando *Incapsula*), então foi necessário que ele nos fornecesse uma permissão para realizar a coleta das informações. Agradecemos também ao usuário *Protaku*⁷ pela participação.

Referências

- Auria, L. and Moro, R. (2008). *Support Vector Machines (SVM) as a Technique for Solvency Analysis*. Discussion Papers. Deutsches Institut für Wirtschaftsforschung.
- Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- Cremonesi, P., Koren, Y., and Turrin, R. (2010). Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, pages 39–46, New York, NY, USA. ACM.
- Gong, Y. and Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 19–25. ACM.
- Huang, J., Zhou, M., and Yang, D. (2007). Extracting chatbot knowledge from online discussion forums. In *IJCAI*, volume 7, pages 423–428.
- Hutto, C. J. and Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014*.

⁶<http://myanimelist.net/profile/Xinil>

⁷<http://myanimelist.net/profile/Protaku>

- Lewis, D. D., Yang, Y., Rose, T. G., and Li, F. (2004). Rcv1: A new benchmark collection for text categorization research. *J. Mach. Learn. Res.*, 5:361–397.
- Mani, I. and Maybury, M. T. (1999). *Advances in automatic text summarization*, volume 293. MIT Press.
- Miller, J. and Southern, G. (2014). Recommender system for animated video. *Issues in Information Systems*, 15(2):321–327.
- Nenkova, A. and McKeown, K. (2012). A survey of text summarization techniques. In *Mining text data*, pages 43–76. Springer.
- Porter, M. F. (1997). Readings in information retrieval. chapter An Algorithm for Suffix Stripping, pages 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523.
- Salton, G., Singhal, A., Mitra, M., and Buckley, C. (1997). Automatic text structuring and summarization. *Information Processing & Management*, 33(2):193–207.
- Sidorov, G., Gelbukh, A., Gómez-Adorno, H., and Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, 18(3):491–504.
- Stefik, M. (1995). *Introduction to knowledge systems*. Morgan Kaufmann.
- Vie, J.-J., Lailly, C., and Pichereau, S. (2015). Mangaki: an anime/manga recommender system with fast preference elicitation. *ACM Conference on Recommender Systems*.
- Voorhees, E. M. and Harman, D. K. (2005). *TREC: Experiment and Evaluation in Information Retrieval (Digital Libraries and Electronic Publishing)*. The MIT Press.
- Wu, Z. and Palmer, M. (1994). Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138. Association for Computational Linguistics.
- Zhou, L. and Hovy, E. H. (2006). On the summarization of dynamically introduced information: Online discussions and blogs. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, page 237.
- Zhou, L., Shaw, E., Lin, C.-Y., and Hovy, E. (2005). Classsummary: introducing discussion summarization to online classrooms. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 4–5. Association for Computational Linguistics.