

7 Conclusões

O meta-alinhamento de ontologias é um campo promissor para o problema de alinhamento de ontologias, uma vez que não existem alinhadores capazes de prover alinhamentos ótimos para todos os cenários. Este trabalho tratou o problema de calibragem de alinhadores partindo de um conjunto pequeno de alinhamentos conhecidos fornecidos por engenheiros de ontologias.

Foi utilizada uma abordagem heurística uni-objetivo baseada em populações, uma vez que metaheurísticas baseadas em populações exploram o espaço de soluções através de uma diversificação das soluções iniciais. Para problemas em que podem existir muitos ótimos locais, a exploração do espaço de soluções através de um processo de diversificação pode gerar melhores resultados do que através de um processo de intensificação da solução (como é o caso das metaheurísticas baseadas em soluções únicas).

A seção 7.1 descreve as contribuições alcançadas por essa pesquisa. Em seguida, a seção 7.2 discute as limitações da proposta e possíveis soluções. Por fim, novas direções de pesquisa para esse trabalho são apresentadas na seção 7.3.

7.1 Contribuições

Em (Shvaiko & Euzenat, 2012), os autores analisam o campo de alinhamento de ontologias e identificam oito desafios que necessitam ser resolvidos para que esse campo de pesquisa consiga alcançar melhores resultados. Este trabalho contribui diretamente para dois desses desafios: a seleção e calibragem de alinhadores e a criação de alinhadores com alto desempenho.

Este trabalho contribui para esses desafios ao apresentar uma abordagem para calibragem de alinhadores que não necessita conhecer o alinhamento final esperado e que, embora necessite de uma fase de treinamento do algoritmo, consegue resolver o problema com o fornecimento de poucos exemplos de correspondências entre os dois modelos. Tal característica dessa abordagem faz com que a abordagem tenha um bom desempenho, permitindo com que possa

ser utilizada para alinhamentos em larga escala e em sistemas que necessitam realizar a operação de alinhamento em tempo de execução.

Os resultados alcançados pelo GNoSIS+ em relação aos demais meta-alinhadores mostram que os resultados estão próximos do estado-da-arte das ferramentas que utilizam computação evolucionária, o qual atualmente pertence ao meta-alinhador GOAL. Porém, o custo computacional exigido pelo GNoSIS+ é menor em relação aos demais meta-alinhadores, uma vez que o GNoSIS+ consegue calibrar os alinhadores de forma satisfatória mesmo com poucas correspondências de exemplo. Além disso, o GNoSIS+ implementa uma abordagem mais aplicável, uma vez que não é necessário conhecer o alinhamento esperado para calibragem dos alinhadores, como é pré-requisito no GOAL. Assim, considera-se que o GNoSIS+ possa ser utilizado em cenários reais, onde é possível, com pouco esforço, determinar algumas poucas ocorrências esperadas entre os modelos a serem alinhados. Ainda, ao contrário do eTuner, o GNoSIS+ é aplicável para alinhamentos em larga escala, visto que o custo computacional desta abordagem não aumenta conforme o tamanho do modelo, como é o caso da geração de esquemas sintéticos do eTuner.

Tecnicamente, esta pesquisa contribuiu ao fornecer uma ferramenta para alinhar ontologias em formato OWL ou RDF, além de uma API para que a ferramenta possa ser integrada a outros sistemas. A ferramenta permite diversas adaptações e inclusões de novos alinhadores e estratégias.

Embora o foco deste trabalho tenha sido o alinhamento de ontologias, esta abordagem pode ser adaptada para alinhar outros tipos de modelos com alinhadores aplicáveis a esses modelos.

7.2

Limitações

Dois problemas enfrentados neste trabalho são discutidos nesta seção: a dificuldade de se analisar o desempenho da proposta em relação a outros trabalhos na literatura e a importância de um conjunto de correspondências fornecidas pelo engenheiro de ontologias. Também são discutidas as formas de se lidar com tais limitações.

A primeira limitação diz respeito à análise de desempenho desta proposta e das demais existentes na literatura. Parte desse problema é consequência da falta de um *benchmark* específico para analisar desempenho de alinhadores¹.

¹Na sua edição 2012, a OAEI criou um desafio específico para pesquisadores que possuem abordagens para alinhamento de larga escala. Embora ainda seja uma trilha específica, espera-se que pesquisadores submetam seus alinhadores para mais de uma trilha e, assim, nos próximos anos a comparação de desempenho entre as abordagens venha a ser facilitada.

Infelizmente, os autores das demais propostas não disponibilizam suas implementações para que possam ser comparadas em relação ao tempo gasto para calibrar e alinhar modelos. Dessa forma, foi realizada uma comparação conforme os critérios e a abordagem proposta por Martinez-Gil & Aldana-Montes (2012). Tal análise, contudo, está sujeita a críticas.

A abordagem de calibragem apresentada neste trabalho é dependente de um conjunto de correspondências conhecidas fornecidas por engenheiros de ontologias. Uma vez que a abordagem tenta calibrar com um número muito pequeno de correspondências, é necessário que estas correspondências representem as características dos modelos que serão analisados. Por exemplo, caso o engenheiro de ontologias forneça somente duas correspondências $c_1 = (e_1, e'_1, =, s)$ e $c_2 = (e_2, e'_2, =, s)$, onde, para a correspondências c_1 , o termo que descreve a entidade e_1 seja igual ao termo que descreve e'_1 como, por exemplo, duas classes com mesmo nome. Caso o mesmo ocorra para a correspondência c_2 , uma solução possível para calibrar um conjunto de alinhadores é a solução que possui um peso igual a zero a todos os alinhadores, com exceção do peso aplicado a um alinhador baseado em técnicas sintáticas, como em distâncias de edição.

Assim, para evitar que o algoritmo genético gere uma solução que privilegie um único alinhador, é necessário fornecer um conjunto de ocorrências de exemplo que represente a heterogeneidade existente nos modelos. Esse problema pode ocorrer em menor frequência ao aumentar o número de ocorrências fornecidas ao algoritmo genético. Algumas soluções para tratar essa limitação são propostas na seção seguinte.

7.3

Trabalhos futuros

Esta seção objetiva fornecer algumas direções para tratar as limitações descritas na seção 7.2, além de apontar outros caminhos de pesquisa que podem gerar bons resultados no campos de pesquisa de alinhamento de ontologias.

A geração de soluções com valores de pesos mais distribuídos pode gerar resultados melhores que soluções que privilegiam um único alinhador. Esse comportamento pode ser introduzido na abordagem com uma alteração na função de aptidão 3-5 (ver seção 3.2). Uma outra alternativa seria a introdução de um processo de armazenamento das soluções com valores próximos da solução vencedora e a posterior escolha da solução que gera um melhor alinhamento. Essa solução, contudo, pode diminuir o desempenho da abordagem.

Para mais detalhes, ver <http://oaei.ontologymatching.org/>.

Ainda em relação a novas heurísticas, foi verificado que o processo de intensificação, realizado por uma heurística de busca local, faz com que novas soluções vencedoras sejam geradas, melhorando a convergência da abordagem. Assim, pretende-se criar uma abordagem para calibragem que seja baseada somente no processo de intensificação de soluções iniciais, utilizando a metaheurística VNS (*Variable Neighborhood Search*) (Hansen & Mladenović, 2003). A metaheurística VNS é um algoritmo estocástico que explora sucessivamente um conjunto pré-definido de vizinhanças para fornecer uma melhor solução. Esta abordagem explora sistematicamente ou de forma randômica um conjunto de vizinhanças para recuperar diferentes ótimos locais. Assim, essa metaheurística explora o fato de que a utilização de várias vizinhanças na busca local pode gerar diferentes ótimos locais e que o ótimo global é um ótimo local para uma dada vizinhança. Como esta não é uma metaheurística baseada em populações, há um menor gasto de memória para sua execução. Assim, uma abordagem VNS pode ser adequada para meta-alinhadores a serem utilizados em alinhamentos em larga escala.

Outra abordagem a ser verificada é a definição de uma fase de pré-processamento das ontologias. O pré-processamento pode ser utilizado para verificar os tipos de heterogeneidade que as ontologias podem apresentar ou, por outro lado, o quão homogênea são as estruturas. Essa informação pode ajudar na escolha de soluções que privilegiem alinhadores que resolvem certas heterogeneidades, em detrimento de outros. O sistema RIMOM (Li et al, 2009), por exemplo, é um alinhador que realiza o alinhamento das ontologias com dois alinhadores pré-definidos, onde um alinhador analisa heterogeneidades sintáticas e o outro alinhador analisa heterogeneidades estruturais. O sistema realiza um pré-processamento das ontologias para determinar dois fatores denominados fator de similaridade de rótulos e fator de similaridade de estrutura. O primeiro fator é proporcional ao número de entidades que possuem mesmo nome nas duas ontologias, enquanto o segundo fator é proporcional ao número de entidades que formam um mesmo grafo com relações de hierarquia. Com base nestes fatores, o RIMOM define um peso para cada alinhador. Para meta-alinhadores, a estratégia do RIMOM poderia ser generalizada para criação de demais fatores de similaridade. Tais fatores podem conduzir o algoritmo genético na fase de criação da população inicial.

A necessidade de treinamento do algoritmo com correspondências fornecidas por engenheiros de ontologias pode deixar de ser um requisito da proposta caso uma abordagem de auto-alimentação seja implementada. Este pré-processamento realizado através da análise dos alinhamentos gerados por cada

alinhador antes da calibragem dos alinhadores e a seleção de, por exemplo, correspondências comuns, pode gerar um conjunto de correspondências para alimentar o algoritmo genético. Estratégias baseadas em aprendizado de máquina, por exemplo, possuem uma abordagem parecida (Duchateau et al, 2008, 2009; Mao et al, 2008; Spohr et al, 2011). Contudo, este tipo de pré-processamento pode diminuir o desempenho do meta-alinhador por necessitar de um tempo de treinamento maior ou de interação com o usuário (como o caso das abordagens que treinam o meta-alinhador através de *feedback* do usuário (Ehrig et al, 2005)).

Alguns meta-alinhadores são capazes de otimizar a precisão ou cobertura dos alinhamentos, como o sistema GOAL. Contudo, pode-se facilmente obter uma alta precisão em detrimento da cobertura (e vice-versa). Alguns sistemas tentam otimizar a medida F, a qual é uma média entre precisão e cobertura. Dessa forma, espera-se evitar o problema de privilegiar uma medida ou outra. Porém, esse tipo de abordagem não evita o desvio indesejado entre a precisão e a cobertura. Por exemplo, a melhor medida F pode ser composta de uma alta precisão e uma baixa cobertura ou vice-versa. Segundo Martinez-Gil & Aldana-Montes (2012), esse problema pode ser tratado com a aplicação de algoritmos multi-objetivos. Almeja-se tratar essa questão alterando as heurísticas do GNoSIS+ para levar em consideração uma nova função objetivo (Nebro et al, 2008).

Por fim, uma vez que a OAEI criou uma trilha específica para comparar o desempenho de alinhadores, é esperado submeter essa proposta para avaliar quantitativamente o seu desempenho em relação aos alinhadores mais modernos e realizar um experimentos com um sistemas reais que necessitem de alinhamentos em larga escala.