

Montando equipes usando grandes modelos de linguagem e similaridade de texto

Miguel Scatolin Teixeira¹

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

Resumo. *Esta monografia investiga a aplicação de grandes modelos de linguagem, especificamente o Transformador Pré-treinado Gerativo (GPT), combinado com técnicas de similaridade de texto, para a formação de equipes no contexto organizacional. Com foco em um estudo de caso envolvendo a Dexco, a abordagem buscou alinhar habilidades e experiências de funcionários com requisitos específicos de projetos. Os resultados revelaram que o modelo foi eficaz na identificação de candidatos adequados para diversas funções, demonstrando precisão notável na correspondência de empregados às necessidades dos projetos. Embora evidenciando o potencial da IA na gestão de recursos humanos, os resultados também destacaram áreas para refinamento adicional, oferecendo insights valiosos para a integração da IA na formação eficaz de equipes em ambientes corporativos.*

1. Introdução

Em um mundo empresarial cada vez mais complexo e dinâmico, a flexibilidade e agilidade são fundamentais para o sucesso. A formação de equipes multidisciplinares para gerenciar projetos diversos tornou-se uma necessidade crítica para atender a essas demandas. Tradicionalmente, a criação dessas equipes depende amplamente do networking e do conhecimento direto dos gestores sobre as competências e experiências dos funcionários. [Masuda et al. 2016] Este processo, muitas vezes, é demorado e sujeito a limitações de percepção e conhecimento pessoal. Com o avanço da inteligência artificial e a crescente capacidade de análise de grandes volumes de dados, novas abordagens para otimizar a formação de equipes estão emergindo.

Neste contexto, o presente trabalho propõe a utilização de grandes modelos de linguagem, explorando seu potencial "criativo", para a geração de tarefas e habilidades necessárias em projetos específicos e, posteriormente, a identificação de funcionários adequados por meio de técnicas de similaridade de texto. Esta abordagem inovadora visa auxiliar no processo de formação de equipes, superando as limitações dos métodos tradicionais e promovendo insights que podem levar a composições de equipe mais eficazes e menos intuitivas.

Organizações estão se voltando para a área de dados para obter vantagens competitivas, enquanto isso a maioria dos dados disponíveis nas empresas é composto por texto não estruturado. [Kobayashi et al. 2018] O objetivo deste estudo é demonstrar como a inteligência artificial, particularmente os modelos de linguagem como o GPT, pode ser aplicada juntamente com a análise de dados não estruturados para auxiliar no processo de planejamento e alocação de recursos humanos. Espera-se que este trabalho ofereça uma metodologia eficaz e eficiente para a formação de equipes, contribuindo para a melhoria dos processos de gestão em ambientes corporativos como a Dexco, onde a criação rápida e eficaz de equipes é essencial para a execução de projetos inovadores e de sucesso.

2. Referencial Teórico

A área de aprendizado de máquina aplicada à recursos humanos está incrivelmente ativa e abrange diversos temas, como recrutamento, treinamento e desenvolvimento, gerência de performance e turnover. [Garg et al. 2021] Grande parte dos artigos focam em características dos indivíduos e dados demográficos das equipes, mas também é possível encontrar estudos buscando entender a dinâmica das equipes e as relações entre indivíduos. [Leonardi and Contractor 2018]

O tema de formação de equipes em contextos organizacionais é amplo e abrange desde a recomendação de equipes colaborativas até a análise de compatibilidade entre indivíduos e equipes. Estudos recentes enfatizam a importância de recomendar equipes que funcionem bem em conjunto, baseando-se em avaliações de comportamentos passados, expertise e carga de trabalho. [Arias et al. 2016] Paralelamente, há uma crescente atenção para a adequação de candidatos não apenas às funções, mas também às dinâmicas de equipe, destacando a relevância de características relacionais na formação de equipes eficazes. [Malinowski et al. 2008] Por fim, o desenvolvimento de métodos para identificar a composição ideal de equipes, incluindo papéis e habilidades, ressalta a necessidade de compreender a dinâmica e a estrutura das equipes para melhorar a alocação de recursos humanos. [Datta et al. 2014]

A maior parte dos trabalhos no tema atualmente lançam mão de dados mais estruturados, tanto sobre as habilidades específicas dos funcionários quanto sobre tarefas e projetos passados. Além disso esses trabalhos também partem de times e equipes muito bem definidas e delimitadas. O presente trabalho busca inovar tanto na parte de criação da definição dos cargos necessários para um time quanto na utilização de dados não estruturados mais facilmente encontrados nas organizações.

3. Desenvolvimento

3.1. Análise de Dados dos Funcionários

A primeira etapa essencial do projeto envolveu uma análise dos dados disponíveis dos funcionários. Para isso, foram utilizados dois datasets distintos, cada um fornecendo perspectivas diferentes sobre as competências e experiências dos funcionários.

3.1.1. Experiências Passadas dos Funcionários

O dataset de experiência é composto por 6183 registros, correspondentes a 1259 funcionários únicos. Cada registro inclui informações vitais como matrícula, experiência, empresa, área, meses de experiência, cargo e data de contrato. A coluna 'experiencia', que descreve as experiências profissionais dos funcionários, é de particular interesse para o nosso estudo. A análise desta coluna revelou uma ampla variação no comprimento dos textos, com um tamanho mínimo de 1 palavra, um máximo de 696 palavras e uma média de 64,19 palavras por entrada. Além disso, ao considerar todos registros de um mesmo funcionário encontramos um novo máximo de 1956 palavras e uma média de 315,02 palavras. Essa variação sugere uma diversidade significativa nas experiências e no nível de detalhamento fornecidos pelos funcionários.

3.1.2. Habilidades dos Funcionários

O segundo dataset, contendo informações sobre as experiências dos funcionários, conta com 3620 registros para 1203 funcionários. As colunas incluem 'matricula' e 'descricao', esta última oferecendo comentários sobre as habilidades dos funcionários feitas por líderes diretos. A análise desta base de dados mostrou que o texto na coluna 'descricao' varia consideravelmente em comprimento, com um mínimo de 1 palavra, um máximo de 802 palavras e uma média de 14,09 palavras. Além disso, ao considerar todos registros de um mesmo funcionário encontramos o mesmo mínimo e máximo, mas uma média de 42,33 palavras. Essa discrepância nos comprimentos dos textos indica a variedade na forma como as habilidades são descritas pelos funcionários, variando de breves menções a descrições mais extensas.

3.1.3. Considerações sobre os dados disponíveis

A análise preliminar desses datasets foi fundamental para compreender a natureza e a estrutura dos dados disponíveis. A diversidade no comprimento e na detalhação das descrições de experiências e habilidades apresenta tanto um desafio quanto uma oportunidade para a aplicação de técnicas de processamento de linguagem natural. Os insights obtidos desta análise serão utilizados para informar as etapas subsequentes do projeto, particularmente na seleção e configuração do modelo de linguagem a ser empregado para a geração de tarefas e habilidades, bem como na filtragem de funcionários para a formação de equipes.

3.2. Escolha do Modelo de Linguagem

A seleção do modelo de linguagem adequado é um passo crucial no projeto "Montando equipes usando grandes modelos de linguagem e similaridade de texto". O objetivo central é aproveitar a "criatividade" e o vasto conhecimento incorporado nos Grandes Modelos de Linguagem. Para isso, focamos na escolha de modelos que não apenas ofereçam capacidade avançada de geração de texto, mas também sejam viáveis em termos de recursos de armazenamento e processamento.

3.2.1. Considerando o Modelo Llama-2-70B

Uma das opções consideradas foi o Llama-2-70B, uma versão pré-treinada com um peso considerável de 129GB. Apesar de sua capacidade avançada e volume substancial de conhecimento, a utilização deste modelo apresenta desafios significativos relacionados a requisitos de armazenamento e processamento. O tamanho do modelo exige uma infraestrutura robusta e capacidade de computação elevada, o que pode ser impraticável ou economicamente inviável para muitos contextos, incluindo o nosso projeto.

3.2.2. Optando pelo GPT e Sua API

Diante dessas considerações, a escolha recaiu sobre o uso do modelo GPT (Generative Pre-trained Transformer) por meio de sua API. O GPT é conhecido pela sua eficácia em

tarefas de geração de texto e por sua capacidade de "criatividade", sendo uma escolha ideal para a geração de descrições de funções e habilidades de equipe. Além disso, a API do GPT oferece vantagens significativas:

A API é de fácil utilização, permitindo uma integração rápida e eficiente no fluxo de trabalho do projeto. Experimentos preliminares com o GPT mostraram resultados encorajadores na geração de tarefas e habilidades para equipes, indicando que o modelo é capaz de atender às necessidades do projeto com eficiência.

3.2.3. Escolha final

Considerando todos esses fatores, a escolha do GPT e sua API representa um equilíbrio ideal entre capacidade de geração de texto criativa e viabilidade prática. Esta decisão alinha-se com os objetivos do projeto de explorar o potencial dos Grandes Modelos de Linguagem de maneira eficiente e eficaz, maximizando os resultados enquanto minimiza os requisitos de recursos.

3.3. Geração de Cargos e Habilidades

A etapa de geração de cargos e habilidades constituiu um componente chave do projeto, focando na utilização de um modelo de linguagem avançado para criar uma lista de funções e cargos adequados para o projeto em questão. Este processo envolveu o desenvolvimento de prompts, uma prática conhecida como "prompt engineering", para direcionar o modelo a gerar saídas específicas e relevantes.

O processo de prompt engineering foi crucial para garantir que o modelo GPT entendesse e respondesse com a saída desejada. Um desafio significativo foi a necessidade de especificar claramente o formato de saída desejado. Para este projeto, o formato escolhido foi o JSON, devido à sua estrutura clara e facilidade de integração com outras etapas do processo, como a filtragem dos funcionários. Foi necessário experimentar diferentes formas de prompts até que todas as respostas geradas pelo modelo estivessem consistentemente no formato JSON especificado. Esses prompts foram cuidadosamente elaborados para instruir o modelo a gerar saídas que incluíssem nome do cargo, descrição, responsabilidades e habilidades necessárias. Essas propriedades foram pensadas para gerar dados que fosse semelhantes aos tipos de dados disponíveis sobre os funcionários, aumentando a efetividade do processo de filtragem.

Outro aspecto importante desta etapa foi a preocupação em manter o número de cargos gerados ao mínimo necessário para reduzir a redundância. Isso não apenas ajudou a manter a eficiência do processo de formação de equipe, mas também garantiu que cada cargo gerado fosse distinto e relevante para o projeto. Ao limitar o número de cargos, pudemos focar em funções essenciais e evitar a superposição de habilidades e responsabilidades, o que poderia levar a uma alocação ineficaz de recursos.

Esta etapa ocorreu em um processo iterativo e informativo de tentativa e erro. O uso bem-sucedido de prompt engineering e a especificação precisa do formato de saída permitiram a criação de uma lista de funções bem definida e estruturada. Esta lista foi fundamental para a etapa subsequente de filtragem dos funcionários, permitindo uma formação de equipe mais alinhada e eficaz.

3.4. Filtragem dos Funcionários Usando as Tarefas Geradas

Após a etapa de geração de tarefas e habilidades com o modelo GPT, a filtragem dos funcionários foi realizada com base na similaridade entre as características dos funcionários e as descrições das funções. Esta etapa é crucial para garantir que os indivíduos selecionados sejam os mais adequados para as funções e equipes propostas.

O método TF-IDF (Term Frequency-Inverse Document Frequency) foi utilizado para calcular a similaridade entre os textos. Esse método é eficaz para atribuir pesos às palavras, considerando tanto a frequência de uma palavra em um documento específico quanto a sua frequência em todo o corpus. Palavras que aparecem frequentemente em muitos documentos recebem um peso menor, enquanto termos mais raros, mas potencialmente mais significativos, recebem pesos maiores. Ao combinar a frequência das palavras com essa ponderação, o TF-IDF é uma ferramenta poderosa para destacar as características mais distintas de cada texto.

Para cada função gerada pelo modelo, selecionamos os candidatos mais compatíveis com base em um limiar de similaridade calculado pelo TF-IDF. Em média, foram escolhidos os 10 candidatos com as maiores pontuações de similaridade para cada função. Essa seleção foi realizada utilizando a feature de descrição das habilidades e cargo e experiência da base de dados de experiências dos funcionários.

3.5. Discussão de resultados

Para avaliar o modelo foi gerada uma lista de dez projetos/times hipotéticos, buscando abranger diversos casos de uso relevantes para a empresa. As principais informações registradas dessa rodada de testes podem ser conferidas no Apêndice 2. Para cada um destes projetos foram analisados principalmente os melhores candidatos para cada cargo, realizando uma análise qualitativa considerando a adequação dos usuários aos cargos.

3.5.1. Análise dos resultados

Foram registrados os tempos de resposta da API do GPT e o tempo total para a geração da lista de candidatos para o projeto. Os tempos encontrados foram de 29,11 segundos de resposta da API e 64,98 de tempo total (geração dos cargos e filtragem da lista de candidatos). O tempo nos testes foi com o maior tempo total sendo de 80 segundos. Apesar disso, o tempo escalará linearmente com o tamanho dos datasets, sendo assim a solução pode não ser ideal para bases de dados muito grandes.

A qualidade das recomendações de candidatos foi avaliada com base em uma análise subjetiva da relevância das habilidades e experiências dos candidatos selecionados para as funções geradas. Em geral, as recomendações foram de boa qualidade, com muitos candidatos exibindo um forte alinhamento com as exigências das funções propostas.

Cada um dos melhores candidatos foi analisado e classificado como "Incrível", "Adequado", "Inadequado" ou "Parcialmente inadequado". No total foram classificados como "Incrível" 14 candidatos, que apresentavam experiências e habilidades completamente alinhadas com as necessidades do cargo relacionado. Na categoria "Adequado" foram encontrados 16 candidatos, que tinham experiências e habilidades relevantes para o cargo em questão. 12 candidatos foram enquadrados na categoria "Inadequado",

tendo experiências não relacionadas com as necessárias para exercer o cargo. Por último, 4 dos candidatos caíram na categoria de "Parcialmente adequado", possuindo algumas das experiências pedidas para o cargo, mas faltando algum conhecimento específico importante.

Os projetos "Iniciativa de Sustentabilidade Ambiental" e "Projeto de Desenvolvimento de Parcerias Estratégicas" foram os que tiveram as piores recomendações de funcionários, com a maioria de seus candidatos sendo inadequados aos cargos. Para a iniciativa de sustentabilidade parecem faltar funcionários com foco no tema. Com relação ao desenvolvimento de parcerias é possível que uma caracterização melhor dos cargos teria gerado recomendações de candidatos melhores.

Uma das limitações identificadas é a geração e a recomendação para alguns cargos "genéricos", sem especificações de qualificações específicas, como Gerente de Projetos. Por exemplo, para o cargo de Líder de Projeto para o "Projeto de Responsabilidade Social Corporativa" foi recomendado um funcionário com extensa experiência em gerência, mas nenhuma experiência no tema específico do projeto.

A filtragem apresenta uma dificuldade em lidar com termos menos específicos e que se apresentam em vários contextos. Alguns exemplos são as palavras desenvolvimento e analista, que são usados tanto com um significado mais amplo e geral quanto em contextos técnicos mais específicos.

3.5.2. Possíveis explicações e melhorias

O método TF-IDF, embora eficaz para a análise de similaridade textual, tem limitações particularmente no que diz respeito a capturar nuances e contextos específicos. O TF-IDF é excelente para identificar palavras-chave relevantes, mas não é capaz de entender o significado ou a relevância contextual dessas palavras. A utilização de modelos de linguagem, capazes de capturar essas nuances, se apresenta como uma opção, mas essas também apresentam limitações, relacionadas a seu treinamento e execução.

Com relação à cargos pouco específicos ou relevantes é possível que o problema esteja na formulação dos projetos e seus objetivos. Uma maior contextualização e mais melhorias no prompt de geração possivelmente direcionariam melhor a geração dos cargos. Além disso, a utilização de versões futuras do GPT e a aplicação de fine-tuning do modelo usando os dados disponíveis dos funcionários e exemplos de definições de equipes existentes também se apresentam como estratégias para melhorar os resultados.

Algumas recomendações inadequadas parecem ter acontecido devido a experiências muito antigas dos funcionários, não sendo mais tão relevantes no momento atual da carreira. Uma alternativa possível seria atribuir pesos diferentes para as diversas experiências diminuindo sua influência de acordo com a distância temporal. Em contrapartida, essa solução perder parcialmente em uma das vantagens do modelo atual, que a sua capacidade de recomendações não triviais devido a experiências relevantes passadas dos funcionários.

4. Conclusões

Em resumo, os resultados obtidos oferecem uma visão promissora do potencial dos grandes modelos de linguagem para auxiliar gerentes de recursos humanos na formação de equipes. Apesar de algumas limitações e desafios, o modelo demonstrou ser viável para sugerir cargos e candidatos para projetos. Os aprendizados adquiridos através deste trabalho podem servir de base para aprimorar o modelo e a metodologia, melhorando a relevância das recomendações de cargos e candidatos em trabalhos futuros.

Referências

- Arias, M., Munoz-Gama, J., and Sepúlveda, M. (2016). A multi-criteria approach for team recommendation. In *Lecture Notes in Business Information Processing*, vol 281, pages 384–396. Springer.
- Datta, A., Yong, J. T. T., and Braghin, S. (2014). The zen of multidisciplinary team recommendation. In *Journal of the Association for Information Science and Technology*, pages 2518–2533. Wiley-Blackwell.
- Garg, S., Sinha, S., and Kar, A. K. (2021). A review of machine learning applications in human resource management. In *International Journal of Productivity and Performance Management*, pages 1590–1610. Emerald Publishing.
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., and Hartog, D. N. D. (2018). Text mining in organizational research. In *Organizational Research Methods*, pages 733–765. Sage Journals.
- Leonardi, P. and Contractor, N. (2018). Better people analytics measure who they know, not just who they are. In *HARVARD BUSINESS REVIEW*, pages 70–81. HARVARD BUSINESS SCHOOL PUBLISHING CORPORATION.
- Malinowski, J., Weitzel, T., and Keim, T. (2008). Decision support for team staffing: An automated relational recommendation approach. In *Decision Support Systems*, vol 45, pages 429–447. Elsevier.
- Masuda, A., Morimoto, C., Matsuodani, T., and Tsuda, K. (2016). Construction of the collaboration skills knowledge in software development. In *Procedia Computer Science*, pages 1129–1136. Elsevier.