

iSent: Classificador do Itaú de Sentimento do Banco Central

- ▶ Este relatório apresenta o iSent, o índice do Itaú de sentimento do Banco Central para Brasil e Chile, que representa um classificador de sentimento baseado no GPT-4, desenvolvido pela nossa equipe de ciência de dados usando frases publicadas em documentos oficiais dos bancos centrais, rotuladas por nossos economistas.
- ▶ O iSent-BCB tem uma boa aderência às mudanças atuais e futuras da taxa de juros no Brasil (correlação em torno de 0,8). O índice identifica um *shift hawkish* na comunicação recente do BCB que seria consistente com aumento da taxa de juros, o que não é a nossa expectativa atual.
- ▶ Para o Chile, o iSent-BCCh também adere aos movimentos de política atuais e futuros (correlação de 0,7). O índice agora mostra uma mudança para um tom neutro, sugerindo que o BCCh provavelmente fará uma pausa no curto prazo, consistente com a nossa expectativa.
- ▶ Também realizamos um *backtest* simples usando o iSent-BCB como um sinal de negociação para o mercado de DI brasileiro e mostramos que o índice tem um *information ratio* positivo na amostra como um todo.

Contexto

As comunicações dos bancos centrais (BC) são cruciais para orientar expectativas (especialmente em economias que seguem o sistema de metas de inflação), gerenciar a economia e implementar a política monetária. Uma vasta literatura tem explorado o uso de Processamento de Linguagem Natural (NLP) para analisar as comunicações do Banco Central. De fato, bons modelos de NLP têm o potencial de resolver vários problemas enfrentados pelos *central bank watchers* – eles podem mitigar vieses humanos, automatizar a classificação de documentos escritos em vários idiomas e fusos horários, além de criar uma avaliação rápida e comparável historicamente de mensagens complexas.

As primeiras versões de captura de sentimento sistemática de documentos de bancos centrais baseavam-se em algoritmos conhecidos como 'bag of words' (BoW) (por exemplo, [Correa et al., 2017](#)). Algoritmos BoW contam quantas vezes uma palavra (ou expressão) aparece em um documento. Para tarefas de análise de sentimento, o algoritmo tipicamente usa um dicionário que associa cada palavra/expressão a uma classe de sentimento, conta o número de instâncias de cada classe e estima o sentimento no documento. Eles são simples e computacionalmente baratos, mas têm algumas desvantagens: 1) ignoram o contexto, ou seja, não consideram a ordem ou dependência entre palavras; e 2) não conseguem lidar com novas palavras/expressões.

Estudos recentes usam NLP baseado em Deep Learning, especificamente Modelos de Linguagem de Grande Escala (Large Language Models ou LLMs), que consideram nuances contextuais, trazendo uma grande melhoria de desempenho. De fato, em 2022, o Itaú lançou seu algoritmo para leitura das atas do Banco Central do Brasil usando o modelo RoBERTA (uma versão otimizada do modelo BERT do Google), com boa aderência ao cenário da taxa de juros (ver relatório [aqui](#)). Versões otimizadas desses modelos usam LLMs associados a um dicionário específico para a tarefa (por exemplo, [Xia, 2021](#)) ou sentenças rotuladas manualmente ([Pfeifer and Marohl, 2023](#)) para melhorar a precisão do modelo.

Para nós, o idioma e o contexto específico dos mercados emergentes (EM) eram barreiras. Enfrentamos algumas dificuldades para incorporar os avanços da literatura de NLP na leitura de sentimento a partir de documentos oficiais de países da América Latina e outros EMs. Primeiro, os modelos mais robustos pré-treinados em textos financeiros (como FinBERT) e dicionários focados em comunicações de bancos centrais são em inglês – e

queríamos construir uma estrutura que fosse flexível em termos de idioma. E, ainda mais importante, o *background* desses modelos pré-treinados e dicionários em inglês é a realidade das economias desenvolvidas, onde (pelo menos até a era pandêmica) a inflação é mais baixa e menos volátil do que nas economias emergentes, e a política fiscal, taxas de câmbio e o cenário externo tendem a ter um papel menos importante nas discussões de política monetária.

Com o surgimento do GPT e sua capacidade de dialogar em diferentes idiomas, tornou-se mais fácil ler o sentimento dos bancos centrais da América Latina, sem a necessidade de re-treinar modelos no idioma local.

Com o objetivo de melhorar a precisão do modelo, construímos nossa própria base constituída por sentenças rotuladas pelos nossos economistas especializados em acompanhar a política monetária do Brasil e do Chile, classificando uma amostra de sentenças de comunicados e atas de reuniões de política monetária dos respectivos bancos centrais. Os economistas não foram informados sobre a data a que os documentos se referiam. Nosso conjunto de dados rotulados é composto por cerca de mil frases de documentos oficiais do Banco Central do Brasil e outras ~1.000 frases de documentos oficiais do Banco Central do Chile. Cada sentença foi classificada como *dovish*, *neutral*, *hawkish* ou *fora de contexto* e, posteriormente, foram utilizadas para aprimorar o modelo de sentimento.

Metodologia

Nossa unidade de classificação é a frase. Uma classe de sentimento (*dovish*, *neutral*, *hawkish* ou *fora de contexto*) é atribuída a cada sentença e o índice é construído com base na presença relativa de cada classe dentro do documento.

Nossa estratégia consistiu em construir um *prompt* GPT que pudesse fazer uma boa classificação, buscando minimizar a contaminação de dados. Assim, construímos uma *persona* descrita como um observador de banco central e cenários básicos que devem ser vistos como *hawkish* (inflação alta, atividade forte, etc.) ou *dovish* (inflação baixa, atividade fraca, etc.).

Para a classificação das sentenças usando o GPT-4, empregamos uma abordagem inspirada nos conceitos de *Retrieval-Augmented Generation* (RAG) ([Lewis, Patrick, et al. 2020](#)). Construímos uma base de dados externa usando o sistema de indexação FAISS ([Johnson, Douze and Jégou, 2017](#)), contendo nossos dados rotulados. Nesse conjunto de dados, as frases são representadas em vetores, e vetores semelhantes são definidos como aqueles que estão próximos no espaço euclidiano. Para classificar novas frases, utilizamos as capacidades de busca por similaridade do FAISS para extrair as frases pré-rotuladas mais relevantes no nosso conjunto de dados. Os exemplos recuperados são sistematicamente integrados ao *prompt* de entrada para o GPT-4, melhorando assim o embasamento contextual do modelo antes da classificação. Esse método melhora a compreensão contextual, permitindo que o modelo infira a classe de forma mais precisa, com base em um conjunto direcionado de frases de exemplo relacionadas à frase em questão.^{1,2}

¹ O Fine-Tuning (FT) também é um método comum onde os pesos de um modelo pré-treinado são atualizados por meio de treinamento em um conjunto de dados supervisionado específico para a tarefa, geralmente envolvendo um grande conjunto (de milhares a centenas de milhares) de exemplos rotulados. A principal vantagem é o forte desempenho em muitos benchmarks, mas as desvantagens incluem a necessidade de grandes conjuntos de dados e o potencial de má generalização fora da distribuição. Veja Brown et al. 2020 para uma discussão.

² Se a sentença a ser classificada aparecer no conjunto rotulado, ela não é usada como exemplo rotulado – escolhemos outras sentenças diferentes (mas semelhantes). Com isso, evitamos que o corpus do prompt possa *ground truth information*. Ainda assim, estamos cientes de que modelos LLM pré-treinados têm problemas de contaminação de dados, e nosso modelo não é diferente. Para uma discussão interessante, veja [Jiang et al. \(2024\)](#)

Erro! Fonte de referência não encontrada. Erro! Fonte de referência não encontrada. **apresenta a estrutura do prompt utilizado.**

Estrutura do Prompt

1	Persona and Task Description
2	Examples
3	Output Format

O índice é calculado seguindo a fórmula abaixo. O índice é um valor entre -1 e 1 e é maior quando o tom percebido é mais *hawkish*.

$$iSent_{hawk/dove} = \frac{\# \text{ frases hawk} - \# \text{ frases dove}}{\# \text{ frases hawk} + \# \text{ frases dove} + \# \text{ frases neutral}}$$

Desempenho do Modelo vs. Decisão de Política

Tanto o iSent-BCB quanto o iSent-BCCh têm boa aderência às mudanças de política monetária. De fato, testamos a capacidade de nossos classificadores de sentimento na tarefa de antecipar mudanças de política monetária, encontrando uma correlação de ~0,8 para o Brasil e ~0,7 para o Chile, conforme mostrado nas tabelas abaixo.

Tabela 1. Matriz de Correlação - Brasil

	Δ Taxa Selic, t	Δ Taxa Selic, t+1
iSent BCB	0.79	0.77
Δ Taxa Selic, t	1	0.92
Δ Taxa Selic, t+1	0.92	1

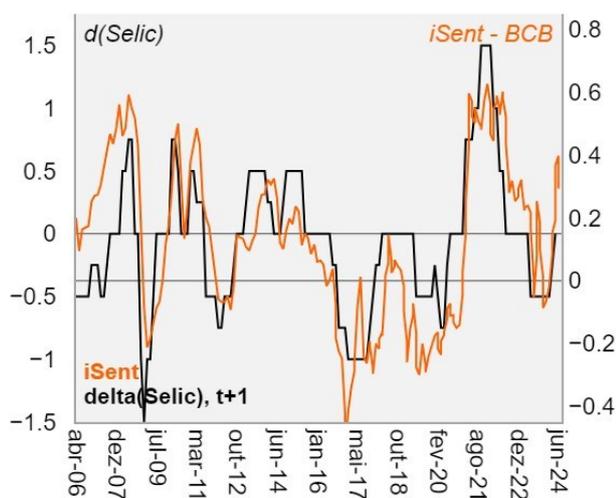
Tabela 2. Matriz de Correlação - Chile

	Δ TPM, t	Δ TPM, t+1
iSent BBCh	0.71	0.68
Δ TPM, t	1	0.83
Δ TPM, t+1	0.83	1

Para o Brasil, o índice sugere que a comunicação recente é consistente com um aumento da taxa de juros na próxima reunião, o que não é a nossa expectativa atual. Uma análise visual confirma uma boa aderência do índice e a mudança da taxa Selic uma reunião à frente. Na verdade, o índice capturou bem a maioria das mudanças nos últimos 18 anos, principalmente os ciclos de alta no final dos anos 2000 e início dos anos 2020. Mais recentemente, o sentimento capturado nos documentos da última reunião é consistente com uma alta de juros na próxima reunião do Copom, embora essa não seja nossa expectativa.

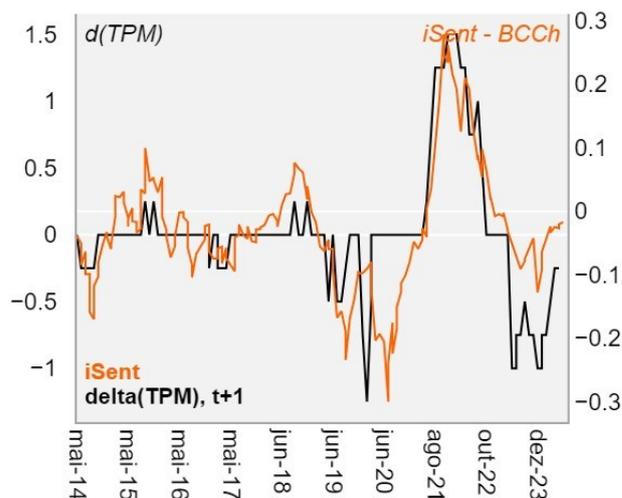
Para o Chile, o índice sugere que a comunicação recente sinaliza uma postura neutra. O índice iSent antecipa amplamente a direção das decisões de política do BCCh, apesar do orçamento dos ciclos de política ser mais contido. Naturalmente, o período da pandemia representa desafios para o índice, pois a taxa de política estava no limite inferior efetivo. O índice efetivamente corresponde ao ciclo de alta do BCCh de 2021-2022 e ao ciclo de afrouxamento pós-julho de 2023. O índice agora sinaliza um tom neutro, sugerindo que o BCCh provavelmente fará uma pausa no curto prazo, consistente com nossa expectativa.

iSent – Brasil



Fonte: BCB, Itaú

iSent – Chile



Fonte: BCCh, Itaú

iSent usado sistematicamente em uma estratégia de DI

Como vimos, variações em nosso indicador de sentimento têm uma forte correlação histórica com decisões futuras de política monetária. No entanto, a próxima pergunta que se deve fazer é se o indicador de sentimento extraído pelo modelo tem poder preditivo para o mercado de taxas. No caso do Brasil, as declarações e atas do Copom são sempre divulgadas ao público fora do horário do mercado. Assim, os participantes do mercado teoricamente têm a mesma quantidade de informação e tempo de processamento antes do início da próxima sessão de negociação. Por outro lado, as pessoas têm vieses cognitivos e comportamentais e os participantes do mercado não são diferentes. Uma vantagem chave dos modelos estatísticos e matemáticos é exatamente remover vieses humanos e extrair conclusões mais diretas a partir dos dados.

Para lançar luz sobre este assunto, realizamos um *backtest* simples usando nosso iSent-BCB como um sinal de negociação para o mercado de DI brasileiro. Primeiro, tomamos a média móvel do indicador (usando pelo menos 8 documentos) como uma forma de remover um possível viés de nível. Então, a partir de junho de 2006, calculamos a média móvel de 2 documentos do indicador (incluindo atas e comunicados quando disponíveis) para suavizar o sinal de negociação, já que a classificação documento por documento pode ser ruidosa. Em seguida, o sinal de negociação é o mais simples possível: quando este indicador está acima (abaixo) de zero, interpretamos como um sinal *hawkish* (*dovish*). Inicialmente, testamos duas estratégias, uma negociando o contrato de DI equivalente de 1 ano e outra negociando a inclinação entre os DI de 2 anos e 5 anos³. Sinais *hawkish* (*dovish*) são traduzidos em uma posição tomada (aplicada) em taxa no vértice de 1 ano e *flattening* (*steepening*) da curva. A cada período, a carteira é rebalanceada para atingir um VaR diário ex-ante de R\$1 milhão⁴. Supomos, naturalmente, que o sinal calculado após a divulgação do último documento só pode estar disponível na próxima sessão de negociação, e as negociações são sempre consideradas como executadas no fechamento do mercado⁵. Custos de corretagem padrão são considerados.

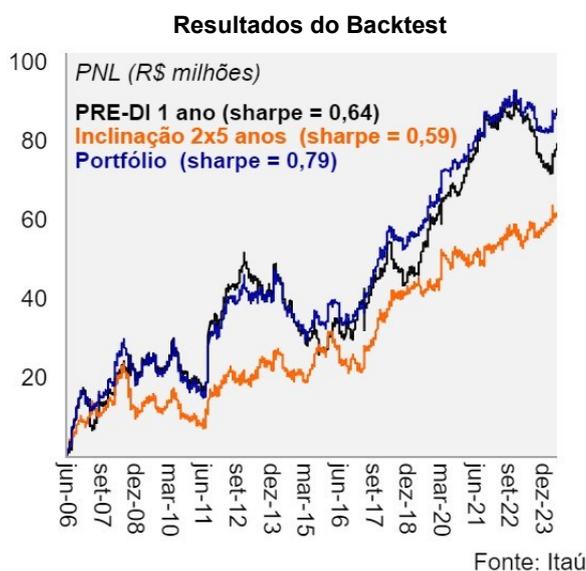
³ Construímos o equivalente a 1 ano e a inclinação rolando historicamente contratos de DI, interpolando linearmente as datas de vencimento para manter a duração aproximadamente constante. Devido a problemas de liquidez, negociamos apenas os vencimentos de janeiro e junho para o equivalente a 1 ano e apenas os vencimentos de janeiro para as durações mais longas.

⁴ Utilizamos a volatilidade realizada ewma dos contratos equivalentes.

⁵ As negociações ocorrem exatamente no fechamento do mercado na próxima sessão de negociação após a divulgação do documento. Assim, se for uma ata (disponível no início da manhã), a estratégia negocia no mesmo dia da divulgação. Se for um comunicado (divulgado após o fechamento do mercado), a estratégia é reequilibrada no próximo dia de negociação. Supomos preços exatamente no fechamento do mercado, portanto, não há *slippage*.

Ambas as estratégias têm *information ratio* positivo na amostra como um todo, embora com *drawdowns* significativos. A maior perda para a estratégia no 1 ano aconteceu durante os anos de 2012-2015, quando seguir o sinal percebido pelo nosso indicador sozinho não resultou em estratégias lucrativas. O último ciclo de afrouxamento monetário também causou uma perda notável para essa estratégia, mas o desempenho da estratégia de inclinação foi satisfatório. Notamos que, embora ambas as estratégias usem o mesmo sinal de negociação, a correlação histórica entre as duas é de apenas cerca de 0,2. Portanto, negociar uma carteira combinada de ambas as estratégias supera ambas.

Deve-se sempre lembrar que o desempenho passado não é garantia de desempenho futuro. Além disso, alguns documentos classificados foram usados por nossos analistas para reforçar o modelo GPT, que por si só também é um modelo treinado em inúmeros documentos ao redor do mundo, incluindo possivelmente comunicações passadas de bancos centrais. Mesmo assim, acreditamos que nosso exercício de *backtesting* pode fornecer *insights* intuitivos e fortalecer o argumento que favorece o aumento do uso de ferramentas quantitativas para aprimorar a análise econômica e de mercado.



Priscilla Burity
Dávila Patrícia Ferreira Cruz
Marcel Chamarelli Gutierrez
Andrés Perez Morales

Referências Bibliográficas

Correa, Ricardo, et al. "Sentiment in Central Banks' Financial Stability Reports." *Review of Finance*, vol. 25, no. 1, 2021, pp. 85-120. Oxford University Press, <https://academic.oup.com/rof/article/25/1/85/5824804>.

Itaú Unibanco. "Introduzindo um algoritmo para leitura de atas de bancos centrais: aplicação para o Brasil indica fim do ciclo." 23 Aug. 2022, https://www.itaubba.com.br/content/dam/ibba/analises-economicas/pdfs/macro-visao/23082022_MACRO_VISAO_NPL-BCB.pdf.

Xia, Ruizhe. "An Automated Approach On Generating Macro-Economic Sentiment Index Through Central Bank Documents." Imperial College, 2023, https://www.imperial.ac.uk/media/imperial-college/faculty-of-natural-sciences/department-of-mathematics/math-finance/Xia_Ruizhe_01201752.pdf.

Pfeifer, Marohl. "CentralBankRoBERTa: A Fine-tuned Large Language Model for Central Bank Communications." ScienceDirect, 2023, <https://www.sciencedirect.com/science/article/pii/S2405918823000302#dfg1>.

Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." 2020, <https://arxiv.org/abs/2005.11401>.

Johnson, Jeff, Matthijs Douze, and Hervé Jégou. "Billion-scale Similarity Search with GPUs." 2017, <https://arxiv.org/abs/1702.08734>.

Brown, Tom, et al. "Language Models are Few-Shot Learners." 2020, <https://arxiv.org/abs/2005.14165>.

Lee, Kenton, Jason Lee, Jing Zhao, and Zhouhan Zhao. "Investigating Data Contamination for Pre-training Language Models." 2024, <https://arxiv.org/pdf/2401.06059>.

Pesquisa macroeconômica – Itaú

Mario Mesquita – Economista-Chefe

Para acessar nossas publicações e projeções visite nosso site:

<https://www.itaubba.com.br/itaubba-pt/analises-economicas>



App Itaú Análises Econômicas
Nossos relatórios no seu celular.
Baixe agora na App Store ou no Google Play.

Informações Relevantes

1. Este relatório foi desenvolvido e publicado pelo Departamento de Pesquisa Macroeconômica do Itaú Unibanco S.A. ("Itaú Unibanco"). Este relatório não é um produto do Departamento de Análise de Ações do Itaú Unibanco ou da Itaú Corretora de Valores S.A. e não deve ser considerado um relatório de análise para os fins do artigo 1º da Instrução CVM n.º 20, de 2021.
2. Este relatório tem como objetivo único fornecer informações macroeconômicas e não constitui e nem deve ser interpretado como sendo uma oferta de compra e/ou venda ou como uma solicitação de uma oferta de compra e/ou venda de qualquer instrumento financeiro, ou de participação em uma determinada estratégia de negócios em qualquer jurisdição. As informações contidas neste relatório foram consideradas razoáveis na data em que o relatório foi divulgado e foram obtidas de fontes públicas consideradas confiáveis. Entretanto, o Itaú Unibanco não dá nenhuma segurança ou garantia, seja de forma expressa ou implícita, sobre a integridade, confiabilidade ou exatidão dessas informações. Este relatório também não tem a intenção de ser uma relação completa ou resumida dos mercados ou desdobramentos nele abordados. O Itaú Unibanco não possui qualquer obrigação de atualizar, modificar ou alterar este relatório e informar o respectivo leitor.
3. As opiniões expressas neste relatório refletem única e exclusivamente as visões e opiniões pessoais do analista responsável pelo conteúdo deste material na data de sua divulgação e foram produzidas de forma independente e autônoma, inclusive em relação ao Itaú Unibanco, à Itaú Corretora de Valores S.A. e demais empresas do grupo econômico do Itaú Unibanco.
4. Este relatório não pode ser reproduzido ou redistribuído para qualquer outra pessoa, no todo ou em parte, qualquer que seja o propósito, sem o prévio consentimento por escrito do Itaú Unibanco. Informações adicionais sobre os instrumentos financeiros discutidos neste relatório encontram-se disponíveis mediante solicitação. O Itaú Unibanco e/ou qualquer outra empresa de seu grupo econômico não se responsabiliza e tampouco se responsabilizará por quaisquer decisões, de investimento ou de outra, que forem tomadas com base nos dados aqui divulgados.

Observação Adicional: Este material não leva em consideração os objetivos, situação financeira ou necessidades específicas de qualquer cliente em particular. Os clientes precisam obter aconselhamento financeiro, legal, contábil, econômico, de crédito e de mercado individualmente, com base em seus objetivos e características pessoais antes de tomar qualquer decisão fundamentada na informação aqui contida. Ao acessar este material, você declara e confirma que compreende os riscos relativos aos mercados abordados neste relatório e às leis em sua jurisdição referentes a provisão e venda de produtos de serviço financeiro. Você reconhece que este material contém informações proprietárias e concorda em manter esta informação somente para seu uso exclusivo.

SAC Itaú: Consultas, sugestões, reclamações, críticas, elogios e denúncias, fale com o SAC Itaú: 0800 728 0728. Ou entre em contato através do nosso portal <https://www.itaú.com.br/atendimento-itaú/para-voce/>. Caso não fique satisfeito com a solução apresentada, de posse do protocolo, contate a Ouvidoria Corporativa Itaú: 0800 570 0011 (em dias úteis das 9h às 18h) ou Caixa Postal 67.600, São Paulo-SP, CEP 03162-971. Deficientes auditivos, todos os dias, 24h, 0800 722 1722.