

Introdução

A análise de sentimentos (AS), também chamada de Mineração de Opiniões (MO), é um processo de extração de informações de textos em linguagem natural. A técnica tem como foco principal as opiniões que expressam sentimentos positivos ou negativos.

A área de pesquisa em MO vem crescendo nos últimos anos, grande parte devido ao aumento no conteúdo independente publicado na Web, como opiniões, debates, entre outros

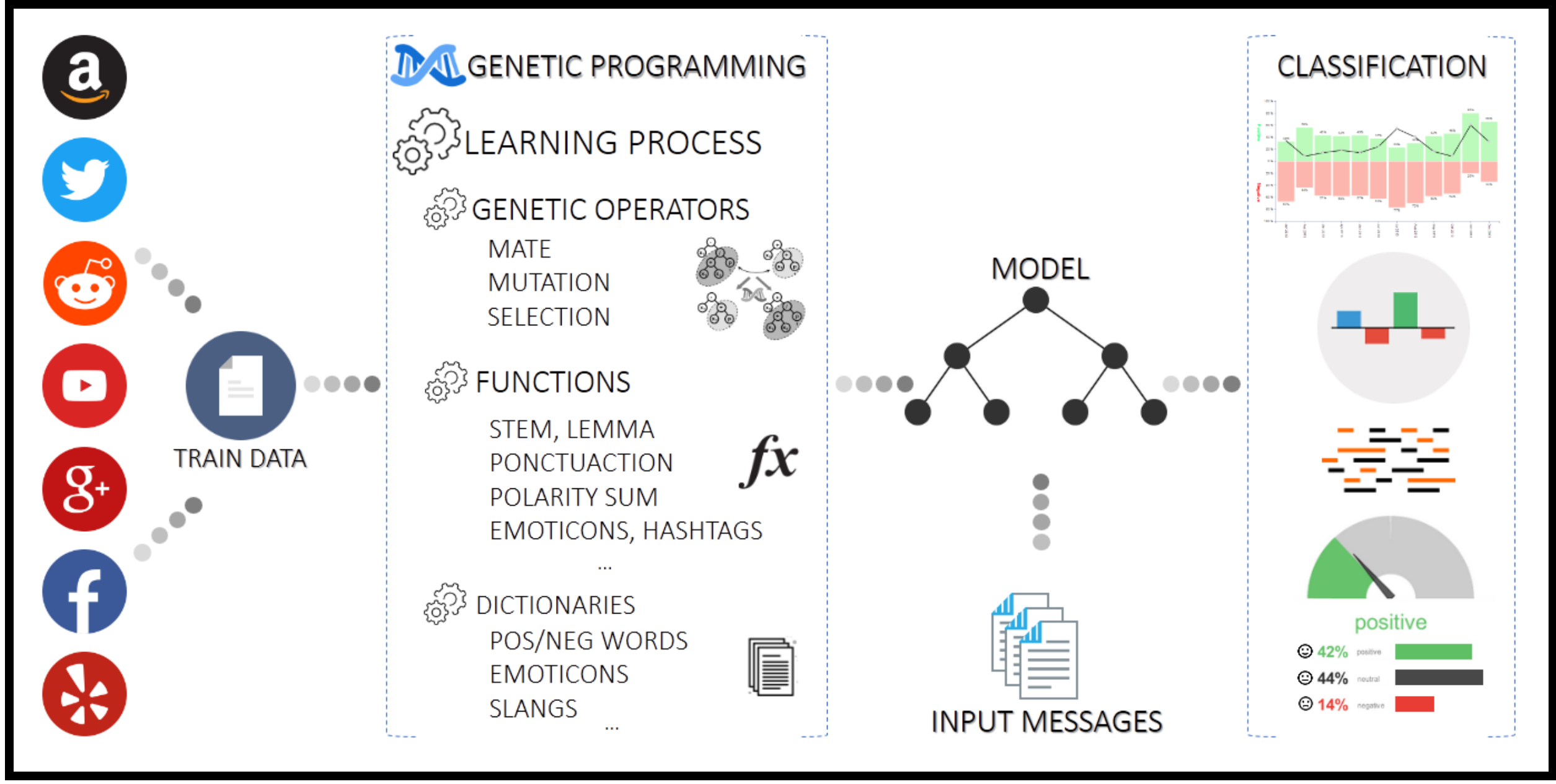


Figura 1: Visão geral do funcionamento do trabalho

A criação de classificadores de sentimentos de forma automatizada é, portanto, um desafio para a área de AS.

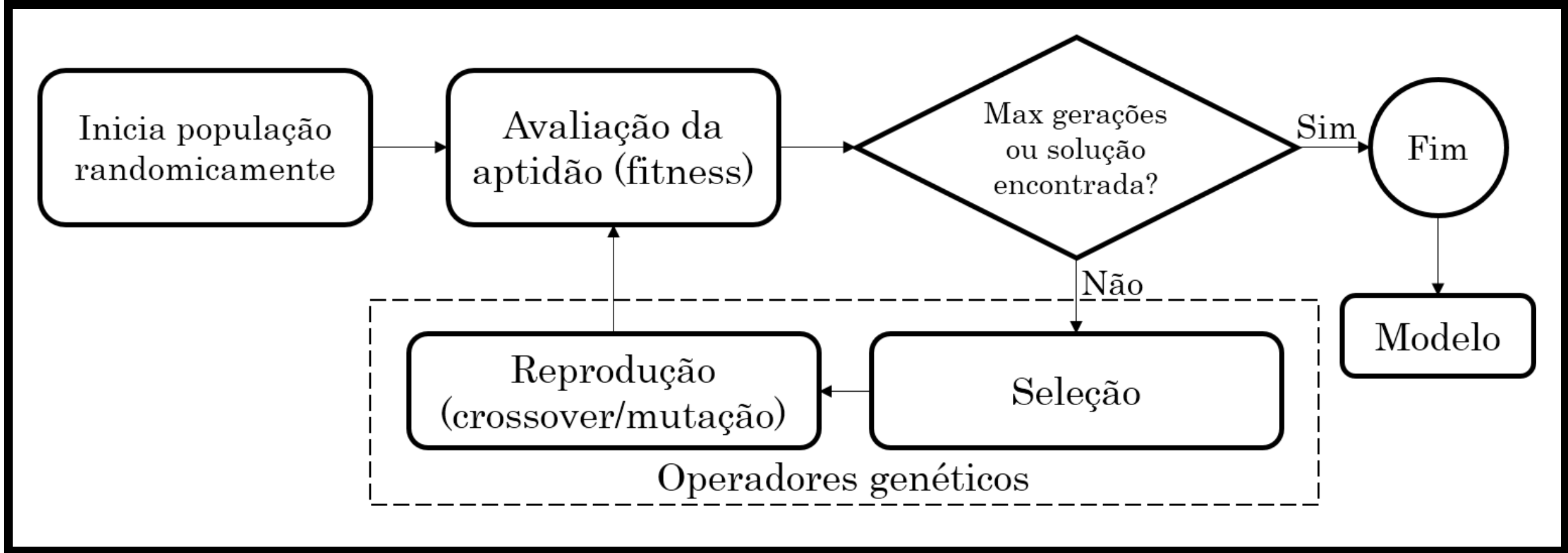


Figura 2: Visão geral do funcionamento da Programação Genética

Considerando a importância dos modelos léxicos de classificação de sentimento e o custo para gerá-los manualmente, o trabalho propõe a utilização da Programação Genética (PG). De forma geral, a PG possibilita a criação de programas de forma automatizada, iniciando com um conhecimento de alto nível sobre as regras gerais do problema.

Materiais e métodos

- Programação Genética para o treinamento e criação do modelo de classificação de sentimentos
- Biblioteca DEAP, do *Python*, para apoiar na implementação do algoritmo de aprendizado
- Dicionários léxicos: Palavras positivas/negativas, *Emoticons* positivos/negativos, *Hashtags* positivas/negativas, Palavras de negação

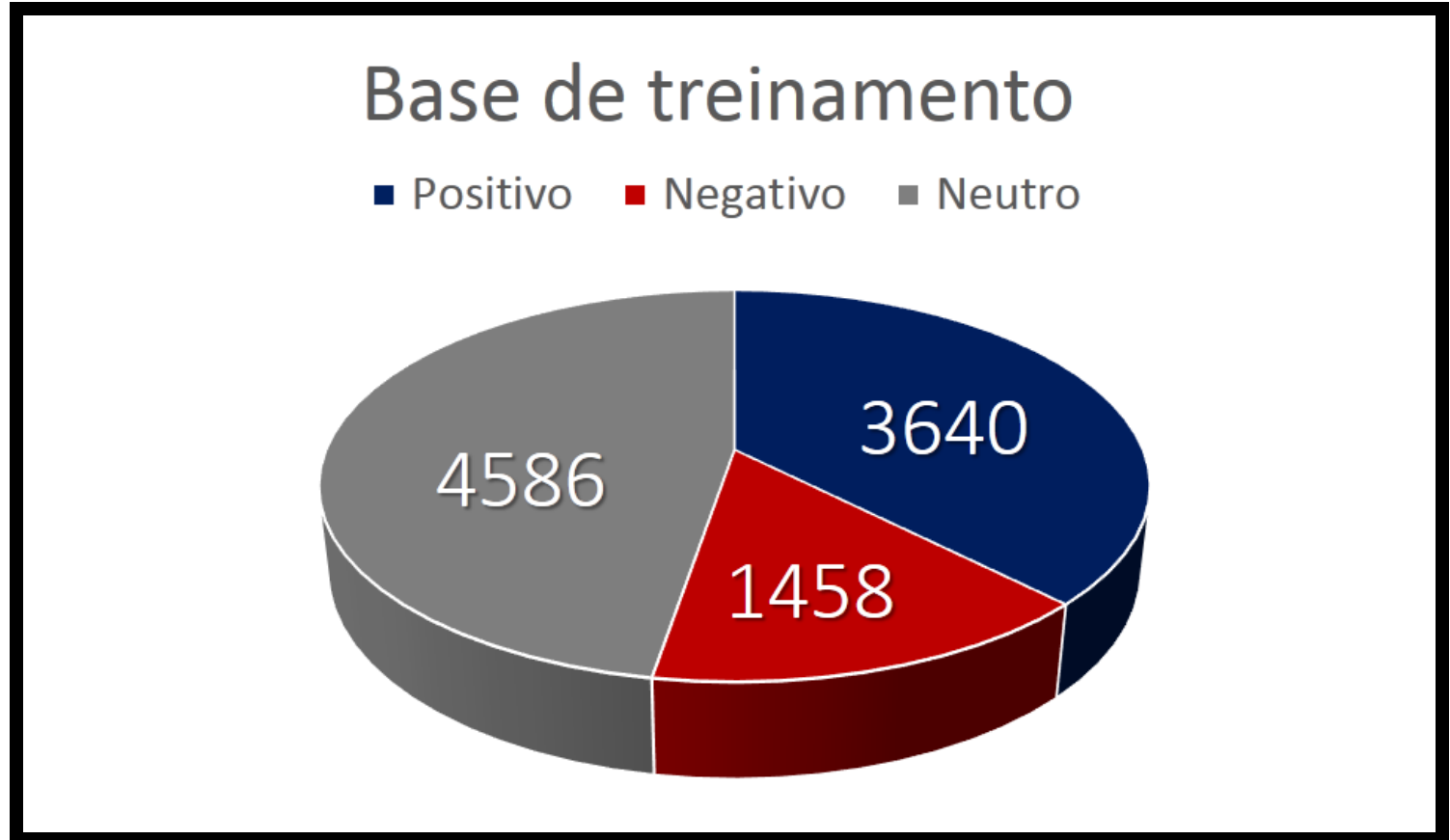


Figura 3: Mensagens para treinamento do modelo de Classificação de Opiniões

- Como forma de comparação dos modelos, foi considerado um *baseline* que faz a simples soma das polaridades das palavras

	Gerações	População	Crossover	Mutação
Modelo 1	600	50	95%	55%
Modelo 2	650	100	45%	25%

Tabela 1: Parâmetros usados na Programação Genética

Resultados

Para a validação da abordagem, foram construídos dois modelos pela PG utilizando uma base de treinamento do *benchmark* SemEval 2014 (figura

3). Posteriormente, os modelos foram avaliados utilizando a base de teste fornecida pelo mesmo evento (figura 4).

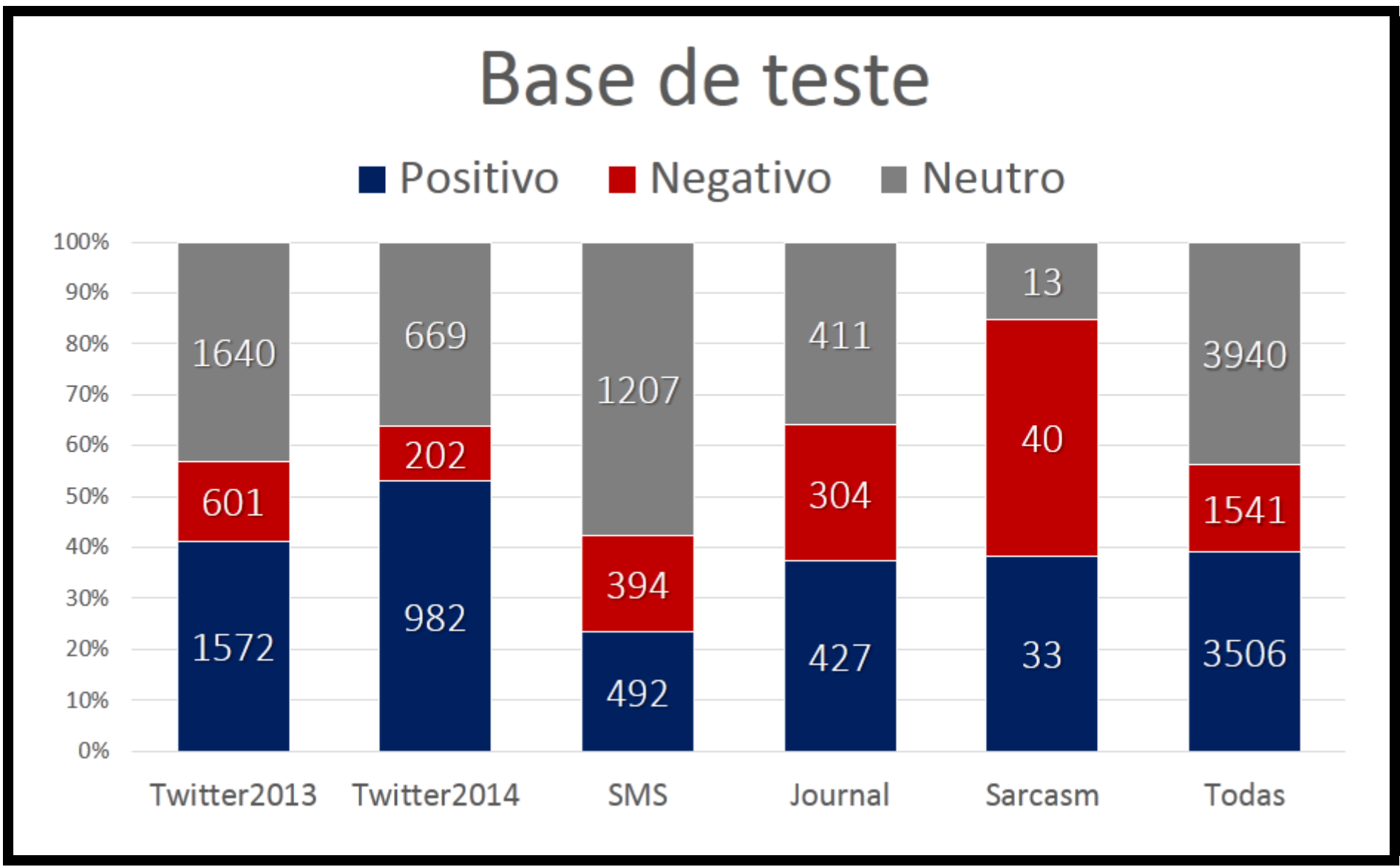


Figura 4: Mensagens de teste do modelo por polaridade

Em comparação com o modelo *baseline*, utilizado em vários trabalhos da literatura, o melhor classificador gerado pela PG obteve um ganho de 12% na média F1 de todas as bases. Destaca-se a melhoria de 45% na base Sarcasm do modelo 2. Os ganhos em relação ao *baseline* podem ser vistos na tabela (2)

Percebe-se que, entre os modelos, alguns são mais customizados à determinados tipos de mensagens do que outros, mas que no F1 médio pelo menos um deles teve desempenho superior ao *baseline* nas bases analisadas e, por isso, satisfatório.

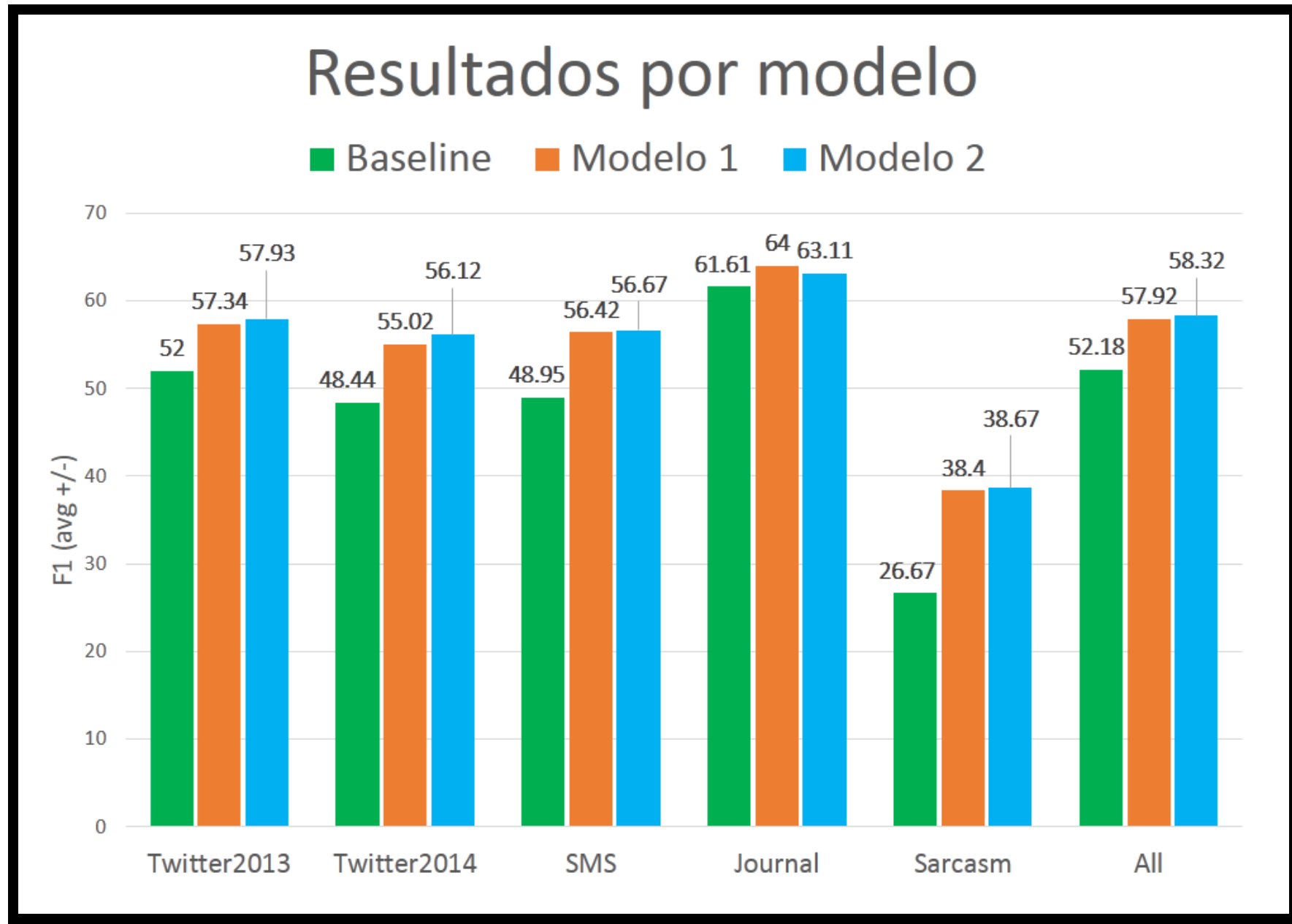


Figura 5: Resultados dos testes por modelo de classificação

	Modelo 1	Modelo 2
Twitter2013	10%	11%
Twitter2014	14%	16%
SMS	15%	16%
Journal	4%	2%
Sarcasm	44%	45%
All	11%	12%

Tabela 2: Ganhos em relação ao modelo baseline

Conclusões

Os resultados mostram que a abordagem utilizando a PG para a criação automatizada de modelos de classificação de sentimentos é promissora, superando o *baseline* e mostrando-se competitiva com outros trabalhos relacionados ao tema. Destaca-se o desempenho do trabalho na avaliação de mensagens da base LiveJournal do *benchmark*, com F1 médio de aproximadamente 64%, apenas 8 pontos abaixo do primeiro colocado na competição. Em todas as bases do *benchmark* a abordagem obteve resultados competitivos, posicionando-se dentro do ranking de trabalhos.

Destaca-se a média de F1 dos dois modelos para todas as mensagens de teste é de 59%, o que é satisfatório dada a dificuldade do *benchmark* utilizado.

Para melhorar a abordagem pretende-se, como trabalhos futuros, incluir novas funções para uso da PG, fazer modificações no *design* do algoritmo, ampliar o conjunto de dicionários utilizados e, por fim, usar bases de treinamento mais diversas.

Mais informações

- E-mail: airtonbjunior@gmail.com.
- Repositório: github.com/airtonbjunior/opinionMining