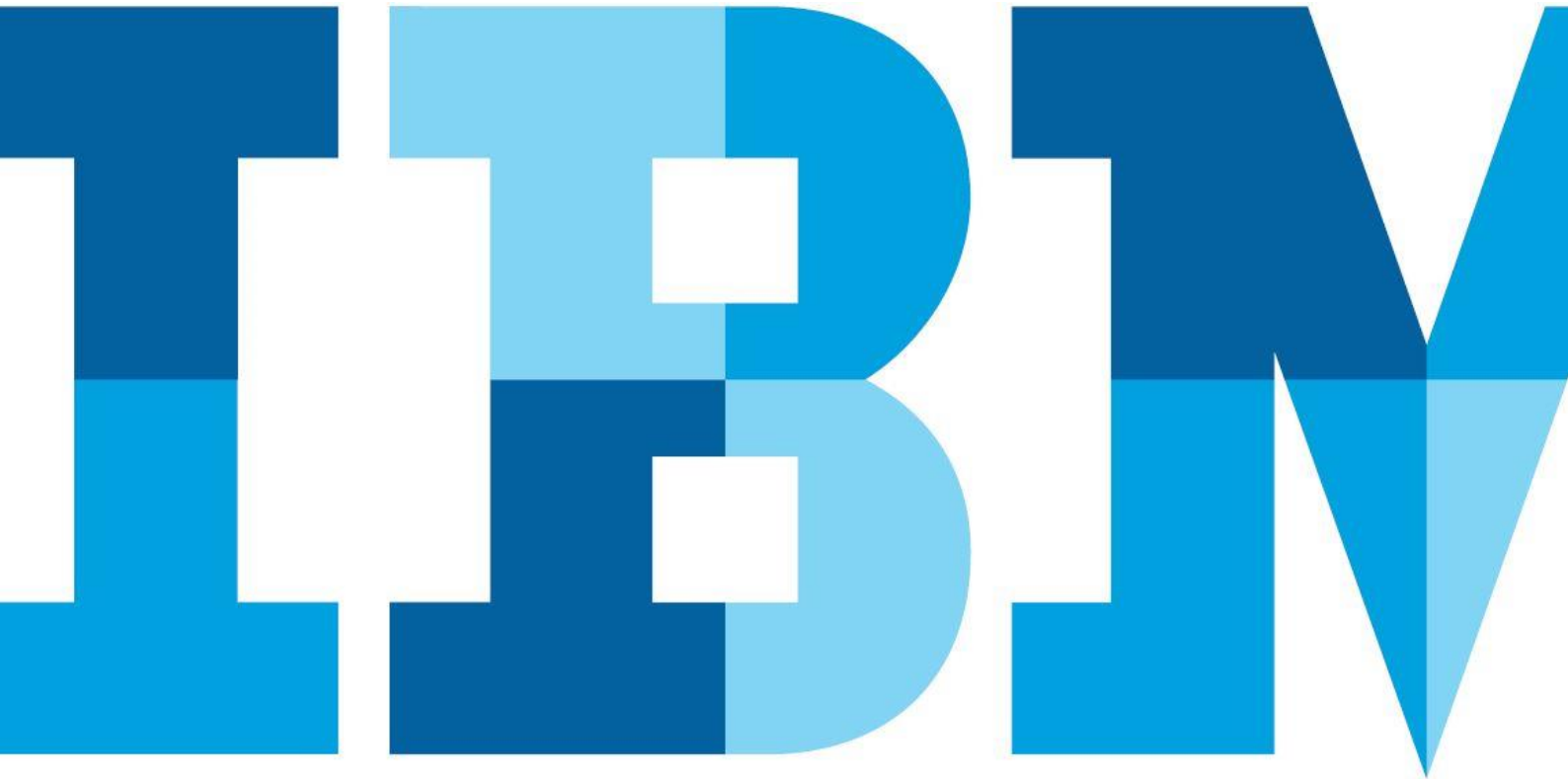


IBM Watson

ソート・リーダーシップ・ホワイト・ペーパー

AlchemyAPI による 感情分析: ハイブリッド手法

言語解析技術と統計分析技術の単一
システムへの統合と正確度の向上



IBM.

手法

ハイブリッド手法の利点

AlchemyAPI の自然言語処理の手法では、言語解析技術と統計分析技術が単一システムに統合されます。これら 2 つの技術にはコンテンツや個々のユース・ケースによって利点と欠点があるため、このハイブリッドな手法は業界最高レベルの利点を生み出します。言語解析は、単語がどのように組み合わせられて句になり、それらの句がどのように組み合わせられて文になっているかを理解する基本的な文法手法です。この手法は社説的なテキスト（ニュース記事やプレス・リリースなど）には効果的ですが、ユーザー生成コンテンツではスラング、つづりの誤り、イディオムが含まれることが多く、同様の効果は期待できません。一方で、統計分析は数学的観点から言語を理解し、「ノイズが多い」コンテンツ（ツイート、ブログ投稿、Facebook の状況アップデートなど）に対しても効果的に機能します。このように、これら 2 つの手法を組み合わせると、さまざまなコンテンツに関する正確度を向上させることができます。

大量の自動学習

AlchemyAPI では、大規模な学習によりデータの真の利点を発揮できるようになります。AlchemyAPI は Web 全体を継続的にクロールし、2000 億以上の単語を含む (Wikipedia の約 285 倍) トレーニング・セットを作成します。AlchemyAPI のシステムは学習アルゴリズムに基づいているため、システムに取り込まれるデータの分野の範囲が広くなればなるほど、新しい話し言葉、イディオム、微妙な表現をより多く学習できます。

言語は絶えず進化するため、さまざまなデータでシステムを継続的に再トレーニングし、最新の状態を保つ必要があります。月ごとの Web の再クロールに加え、AlchemyAPI は主要なソーシャル・メディア・プラットフォームのデータ・フィードを監視します。例えば、AlchemyAPI のデータ・セットには毎月 10 億以上のツイートが追加されます。また、AlchemyAPI では 30,000 人を超える API ユーザーのコミュニティーのデータも使用します。6 つの大陸と多数の業界から収集して生成されるこの大量のデータ・セットにより、AlchemyAPI は初めて扱うエンティティやトピックであってもより正確に分析できます。

AlchemyAPI は Web 全体を継続的にクロールし、2000 億以上の単語を含む (Wikipedia の約 285 倍) トレーニング・セットを作成します。

感情分析

テキスト分析では、「感情」は何か（人、製品、組織、場所など）に対して示された態度や意見を意味します。感情には、肯定的（「この車は素晴らしい」）、否定的（「この車は高すぎる」）、中立的（「この車は赤い」）があります。コンピューターがテキストの感情を判別するプロセスには、高度なアルゴリズムが必要となります。

正確度の向上

大量のテキストを正確に分析できるソフトウェアを作成するのは複雑な作業です。AlchemyAPI の感情分析が優れているのは、ハイブリッドな手法と、さまざまな分野から集められた大量のデータでトレーニングされた大規模な学習テクノロジーをシステムの基盤としているためです。これにより、感情をさらに深く理解し幅広い機能を実現できるだけでなく、最終的により正確なシステムに進化させていくことができます。

AlchemyAPI の優れた感情識別の正確度は、サード・パーティーの調査によって実証されています。例えば、Meehan 他の論文「Context-Aware Intelligent Recommendation System for Tourism¹」（2013 年）では、5,370 ツイートのコーパスに対して実施された手動のテストの結果、86.01% という高い正確度レベルが示されました。正確度は文書の種類や個々のユース・ケースによって異なるため、お客様は自分の文書を使用して感情を評価する必要があります。お客様が個々の実装の正確度を評価できるように、AlchemyAPI では無料でソフトウェア評価を行っています。

感情の深い理解

AlchemyAPI では、文書レベル、エンティティ・レベル、キーワード・レベル、方向性、関係性、ユーザー指定といった、感情分析に関する複数のターゲット・オプションが提供されます。このきめ細かな感情分析によって、より有用なビジネス・インテリジェンスが可能となります。

文書レベルの感情分析

文書レベルの感情は感情分析の最も一般的な形式の 1 つで、文書やテキスト・ブロックが肯定的、否定的、またはそれらの混合として分析されます。文書では 2 つ以上のことが書かれていたり感情が入り交じっていることが多いため、これはかなり大ざっぱな手法です。

エンティティ・レベルの感情分析

エンティティ抽出の追加により感情分析は一步前進し、検出されたエンティティごとに感情データを作成できるようになりました。例えば、ある文が人や製品に関して否定的な内容を示している、別のものに関して肯定的な内容を示していることもあります。Rizzo、Troncy (2011)²、Saif 他 (2012)³ などのサード・パーティーの調査により、AlchemyAPI のエンティティ抽出の優れた正確度が実証されています。

キーワード・レベルの感情分析

抽出されたキーワードごとに感情を評価できるため、例えば、ある製品や製品機能に関する感情の理解が可能です。

方向性レベルの感情分析

感情の方向性によって、誰が感情を発しているかが明らかになります。例えば、ある人が製品について否定的なことを話していた場合、製品が否定的に言及されたことだけでなく、誰が製品を否定的に言及したかも判別します。

関係性の感情分析

関係性の判別により、テキスト内の主語-述語-目的語の関係を識別し、主語または目的語ごとに感情を評価できます。エンティティとキーワードの抽出を組み合わせると、感情の関係性は、強力な検索アプリケーションや傾向分析アプリケーションの基盤とすることができます。

ユーザー指定のターゲット

開発するアプリケーションでは上記のターゲット・オプションのいずれも適切ではなかった場合、AlchemyAPI はユーザー指定のターゲットをサポートします。これにより、任意のターゲットに対する感情を評価できます。

感情の強度

AlchemyAPI のダイナミックな手法により、システムでテキストをより深く理解できるようになります。例えば、AlchemyAPI では以下が理解されます。

- **否定:** 感情と反対の意味の語。例えば、「この電話は素晴らしい」に対し「この電話は素晴らしくない」
- **増幅:** 感情の強度を強める語。「とても」など。
- **軽減:** 文の強度を弱める語。「やや」など。
- **強度:** 「否定」、「増幅」、「軽減」が、検出された感情の強度スコアを導き出します。「このコーヒーは少し甘い」と「このコーヒーはとてもまずい」の間には大きな違いがあり、強度スコアによってそれぞれの否定文の強度を示します。

AlchemyAPI のダイナミックな手法により、システムでテキストをより深く理解できるようになります。

展望

感情分析を適切に実行することは複雑なタスクですが、AlchemyAPI の自然言語処理の革新的な手法と大規模な学習テクノロジーの導入により、現在だけでなく将来においても、さまざまな業界やアプリケーションで高い正確度を維持できます。近日予定されている、ディープ・ラーニング機能を備えた AlchemyAPI の感情エンジンのアップデートでは、大量のデータ・セットが利用され、さらにスマートで正確なシステムへと進化します。また、もっと多くの業界や異なる国でのシステムの利用を可能にするために、追加の言語での開発も行われています。

脚注

¹ Meehan, K., Lunney, T., Curran, K., および McCaughey, A.; Context-Aware Intelligent Recommendation System for Tourism; 進行中のセッション - PerCom 2013、サンディエゴ (2013 年 3 月 19 日); School of Computing and Intelligent Systems, University of Ulster、デリー、北アイルランド、E メール: meehan-k@email.ulster.ac.uk; http://eprints.ulster.ac.uk/28277/1/Context-Aware_Intelligent_Recommendation_System_for_Tourism_-_PerCom_13.pdf

² Rizzo G., Troncy R. (2011)。NERD: Evaluating Named Entity Recognition Tools in the Web of Data. (ISWC' 11) Web Scale Knowledge Extraction (WEKEX' 11, ボン、ドイツ) でのワークショップ、10 月。pp. 1 ~ 16、Politecnico di Torino, Porto Institutional Repository; http://port.polito.it/2440793/1/wekex2011_submission_6.pdf

³ 4 Saif, H., He, Y., および Alani, H. (2012); Semantic Sentiment Analysis of Twitter; Knowledge Media Institute, The Open University, United Kingdom; {h.saif,y.he,h.alani}@open.ac.uk; <http://iswc2012.semanticweb.org/sites/default/files/76490497.pdf>

AlchemyAPI について

AlchemyAPI の目的は、ディープ・ラーニング・ベースの人工知能の進歩を一般に広めて、人間の言語や視覚を理解する新世代のスマート・アプリケーションを推進することです。リアルタイムのテキスト分析とコンピューター・ビジョンのための使いやすい高性能なクラウド・サービスにより、企業は、膨大な非構造化データをビジネスを推進するアクションに変換するのに必要なインテリジェンスを手にすることができます。AlchemyAPI のソリューションは、世界 36 カ国のさまざまな業界で働く 40,000 人を超える開発者によって、何十億ものテキストや画像の処理に毎月利用されています。詳細については、Web サイト www.alchemyapi.com をご覧ください。

IBM Watson について

IBM Watson の詳細については、www.ibmwatson.com をご覧ください。



© Copyright IBM Corporation 2015

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
June 2015

IBM、IBM ロゴ、ibm.com および Watson は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。本書の情報は最初の発行日の時点で得られるものであり、予告なしに変更される場合があります。すべての製品が、IBM が営業を行っているすべての国において利用可能なものではありません。

本書に掲載されている情報は特定物として現存するままの状態を提供され、第三者の権利の不侵害の保証、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任なしで提供されています。

IBM 製品は、IBM 所定の契約書の条項に基づき保証されます。

適切なセキュリティ実施について: IT システム・セキュリティには、企業内外からの不正アクセスの防止、検出、および対応によって、システムや情報を保護することが求められます。不正アクセスにより、情報の改ざん、破壊もしくは、悪用を招くおそれがあり、またはシステムの損傷や、他のシステムへの攻撃を含む悪用につながるおそれがあります。完全に安全と見なすことができる IT システムまたは IT 製品は存在せず、また単一の製品またはセキュリティ対策が、不正アクセスを防止する上で、完全に有効となることもありません。IBM のシステムおよび製品は、包括的なセキュリティの取り組みの一部となるように設計されており、これらには必ず追加の運用手順が伴います。また、最高の効果を得るために、他のシステム、製品、またはサービスを必要とする場合があります。IBM は、何者かの悪意のある行為または違法行為によって、システム、製品、またはサービスのいずれも影響を受けないことを保証していません。



Please Recycle