

# 1. الطرائق المعتمدة على GNNs

أ. 2024 Unsupervised Social Event Detection via Hybrid Graph Contrastive Learning and Reinforced Incremental Clustering

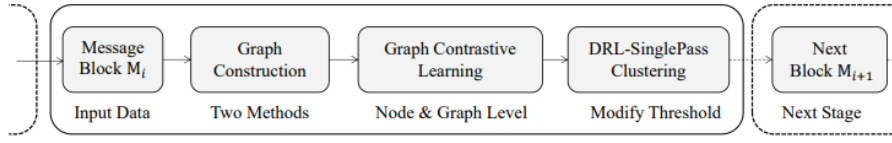
- الملخص:

التحدي الرئيسي لكشف الأحداث هو استخراج معلومات تمييزية من بيانات وسائل التواصل لإسناد البيانات إلى أحداث مختلفة. تبعاً للتباين والتحديث السريع والمتكرر للبيانات، فإن استخدام طرائق خاضعة للإشراف هنا هو شيء صعب التحقيق والوصول إليه. حتى الآن، الأعمال الحالية لاستكشاف تعلم معلومات تمييزية من البيانات *discriminative information* بالاستفادة من التعلم المقارن للبيان (GCL) *graph contrastive learning* والتضمين، للعقدة بطريقة غير خاضعة للإشراف. ومع ذلك، توجد مشكلتان جوهريتان في طرائق GCL: لا يمكن لـ GCL سوى استكشاف السمات الجزئية تقريباً، وبالتالي لا يتعلم المعلومات التمييزية للرسائل الاجتماعية بشكل كافٍ؛ بالنسبة لطرق GCL، يتم تجميع التضمينات المتعلمة في الفضاء الضمني من خلال الاستفادة من بعض المعرفة المحددة مسبقاً، مما يتعارض مع مبدأ التعلم غير الخاضع للإشراف.

تم في هذه الورقة اقتراح طريقة جديدة للكشف عن الأحداث بطريقة غير خاضعة للإشراف من خلال التعلم المقارن للبيان الهجين (hybrid graph contrastive learning) وعقدة تزايدية معززة (reinforced incremental clustering) (HCRC)، والذي يستخدم التعلم المقارن للتعلم الشامل للمعلومات التمييزية الدلالية والبنوية من الرسائل الاجتماعية والعقدة التزايدية المعززة لتنفيذ العقدة بشكل فعال وبطريقة غير خاضعة للإشراف. وأجريت تجارب شاملة لتقييم HCRC على مجموعات بيانات Twitter وMaven. بإعداد تزايدي تقليدي، تزايدي شبه خاضع للإشراف وغير خاضع للإشراف تماماً.

- المنهجية:

عندما تصل كتلة رسائل  $M_i$ ، يتم تنفيذ مسار العمل الموضح بالشكل أدناه، حيث نبدأ بإنشاء البيان من الرسائل (نبدأ من بيان رسائل غير متجانس يحوي الرسائل وكياناتها المسماة، ومن ثم يتم تحويله إلى بيان متجانس يحوي الرسائل فقط)، وبعد ذلك تبدأ عملية تعلم مقارن على البيان ( يتم هذا على مستويين -مستوى البيان ومستوى العقدة - في مستوى البيان يتم تشكيل Social Message Attribute Graph على شكل نجمة حيث لكل رسالة يكون في المركز الرسالة وترتبط مع الكيانات المسماة الموجودة بها ويتم تعزيز البيان بعد لك وإجراء تعلم مقارن من أجل تعلم التضمين، وفي مستوى العقدة يتم تشكيل Social Message Relation Graph يحوي الرسائل ووصلات بينها بناءً على مدى تشاركها بالكيانات المسماة وبعد ذلك يتم تعزيز البيان وإجراء تعلم مقارن لتعلم التضمين)، من ثم تجري عملية عقدة باستخدام خوارزمية SiglePass المعززة ويتم تعديل العتبة بناءً على تباعد العناقيد عن بعضها البعض.



- مجموعة البيانات: Events 2012, MAVEN

- الملاحظات:

هذه الورقة تركز على جانب كشف الأحداث فقط (القناة مفلترة مسبقاً)، لم تتطرق إلى بناء كامل النظام وإنما فقط على تحسين الأداء على مجموعة البيانات. والكود متوفر.

تمت مقارنة النموذج المطروح مع SOTA النماذج المطروحة وهي:

Word2vec, LDA, WMD, BERT, BiLSTM, PP-CGN, EventX, KPGNN, QSGNN

ب. 2023 Towards Cross-lingual Social Event Detection with Hybrid Knowledge Distillation

- الملخص:

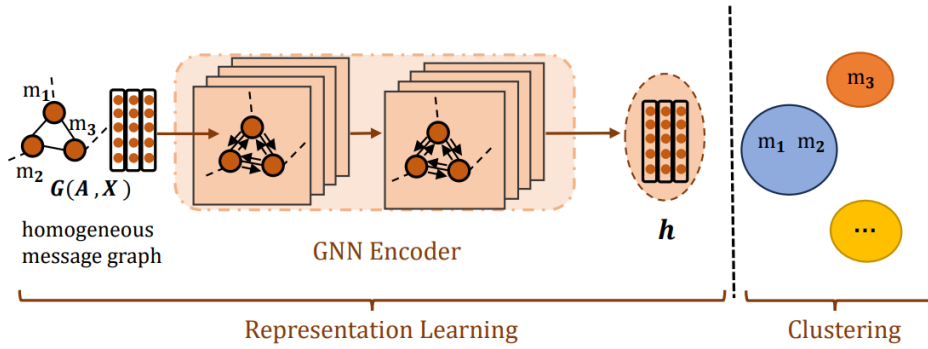
أبحاث GNN المنشورة حديثاً تظهر أداء واعد في مهمة كشف الأحداث الاجتماعية. ولكن معظم الأبحاث تتجه نحو بيانات من لغة واحدة. هذا جعل اللغات المحكية قليلاً، نسبياً غير مستكشفة.

تهدف هذه الورقة البحثية إلى تطوير إطار جديد يسمى CLKD (تقطير المعرفة متعددة اللغات) لكشف الأحداث الاجتماعية من وسائل التواصل الاجتماعي، خاصة في اللغات منخفضة الموارد مثل العربية والفرنسية. تعتمد الطريقة المقترحة على استخدام الشبكات العصبونية البيانية (GNN)، وتضمن الكلمات متعددة اللغات، لتجاوز مشكلة نقص البيانات المعلّمة في تلك اللغات. يقوم الإطار بنقل المعرفة المكتسبة من نموذج مدرب مسبقاً على لغة غنيّة (مثل الإنجليزية) إلى نموذج يعمل على لغة ضعيفة من خلال تقطير المعرفة على مستويين: سمات الرسائل وعلاقاتها، مع تطبيق وحدة ترجمة دلالية لتقليل الفجوة اللغوية. تم اختبار الإطار على ثلاث مجموعات بيانات من تويتر (إنجليزية، فرنسية، عربية)، وأظهرت النتائج أن CLKD يتفوق بشكل ملحوظ على الأساليب الحالية في كشف الأحداث، خصوصاً في اللغات ذات الموارد المحدودة.

- المنهجية:

يعتمد الإطار المقترح في هذا البحث، على منهجية تهدف إلى تحسين كشف الأحداث الاجتماعية من البيانات النصية بلغات منخفضة الموارد مثل العربية والفرنسية، من خلال الاستفادة من المعرفة المكتسبة مسبقاً من نماذج مدربة على لغات غنية. تبدأ المنهجية ببناء بيان متجانس يربط الرسائل من وسائل التواصل الاجتماعي بناءً على المستخدمين، الوسوم، والكيانات المشتركة، ثم يتم استخدام شبكة عصبونية بيانية (GNN) من نوع GAT لاستخراج تمثيل للرسائل. ولتجاوز فجوة اللغة، يتم إدخال وحدة

متعددة اللغات (Cross-Lingual Module) تعتمد على تحويل تمثيلات الكلمات من اللغة الهدف إلى الفضاء الدلالي للغة الإنجليزية باستخدام تقنيات التحويل الخطي (مثل MUSE) أو غير الخطي (مثل LNMAP). بعد ذلك، يتم تطبيق تقطير المعرفة على مستويين: الأول على مستوى السمات، حيث يتم نقل التمثيلات من نموذج المعلم إلى نموذج الطالب بعد تحويلها إلى نفس الفضاء؛ والثاني على مستوى العلاقات، حيث يتم نقل العلاقات بين الرسائل ذات الصلة، مع مراعاة نقل العلاقات "المفيدة" فقط لتفادي تأثير الأخطاء المحتملة في نموذج المعلم. يدعم الإطار وضعين تدريبيين: وضع المعلم والطالب للاستخدام Offline، والتعلم المتبادل بين نموذجين (أقران) في وضع Online ومستمر. كما يتيح الإطار كشفًا تزايدياً للأحداث عبر ثلاث مراحل: تدريب مبدئي، الكشف والصيانة، مع تحديث تدريجي للنموذج بمرور الوقت.



- مجموعة البيانات: Events 2012, Events 2018, Kawarith

- الملاحظات:

هذه الورقة تركز على جانب كشف الأحداث فقط (القناة مفلترة مسبقاً). والكود متوفر.

وتمت مقارنة النموذج المطروح مع: Word2vec, LDA, BERT, PP-CGN, EventX, KPGNN, QSGNN

ت. 2022 Reinforced, Incremental and Cross-lingual Event Detection From Social Messages

- الملخص:

تتناول الورقة مشكلة اكتشاف الأحداث الاجتماعية الهامة من دفق من رسائل الشبكات الاجتماعية (مثل الفضاءات السياسية أو الكوارث الطبيعية). وعلى الرغم من الاهتمام بنماذج الاكتشاف التزايدية لمعالجة طبيعة التدفق، تواجه الأساليب الحالية تحديات تتمثل في غموض سمات الحدث، وتشتت محتوى الرسائل، وتعدد اللغات، مما يؤدي إلى ضعف الدقة والقدرة على التعميم. لمواجهة هذه التحديات، تقدم الورقة بنية جديدة تسمى

.FinEvent (Reinforced, Incremental and Cross-lingual social Event detection)

تم تصميم FinEvent لتكون متوافقة مع طبيعة تدفق الرسائل والواقع العملي لاكتشاف الأحداث، مع التركيز على المعالجة الآنية، واكتساب الدلالات وتحديثها، والتعامل مع مشكلة الاكتشاف متعدد اللغات. حيث تعتمد آلية دورة حياة تزايدية وتتكون من خمس وحدات: المعالجة المسبقة، تضمين الرسائل، التدريب، الاكتشاف، والنقل. تشمل التقنيات الأساسية في FinEvent نمذجة الرسائل في بيان متعدد العلاقات مثل، للحفاظ على معلومات اتصال غنية، وشبكة عصبونية للبيان متعدد العلاقات متعدد العملاء معزز (MarGNN) لتعلم تضمين الرسائل باستخدام التعلم المعزز لاختيار عتبات التجميع المثلى، وآلية تعلم متباين، ونموذج عنقدة مكاني مستند إلى الكثافة وموجه بالتعلم المعزز (DRL-DBSCAN) لاختيار المعلمات المثلى للعنقدة، وطريقة تضمين رسائل عابرة للغات (Crlme) لنقل المعرفة من لغة إلى أخرى لتمكين الاكتشاف بتكلفة منخفضة.

- المنهجية:

تتبع المنهجية آلية دورة حياة تزايدية تتكون من أربع مراحل رئيسية: التدريب المسبق، الكشف أحادي اللغة، الكشف العابر للغات، والصيانة.

في مرحلة التدريب المسبق، يتم بناء نموذج أولي باستخدام جزء صغير من الرسائل لبناء بيان مثل متعدد العلاقات وتدريب النموذج الأولي عليه. في مرحلتي الكشف (أحادي اللغة ومتعدد اللغات)، يتم تحديث البيان برسائل جديدة ويتم استخدام النموذج المدرب للكشف عن الأحداث. أما في مرحلة الصيانة، فيتم إزالة الرسائل القديمة وتحديث النموذج عبر التدريب التزايدى باستخدام البياني المحدث.

تعتمد المنهجية على خمس وحدات رئيسية: أولاً، يتم نمذجة رسائل التواصل الاجتماعي في بيان مثل متعدد العلاقات، حيث يتم تحويل شبكات المعلومات غير المتجانسة (HINs) التي تدمج عناصر الحدث المختلفة (مثل المواقع، الأشخاص، العلاقات) إلى بيان للرسائل. ثانياً، يتم استخدام إطار MarGNN (Multi-agent Reinforced weighted Multi-relational Graph Neural Network) لتعلم تضمينات رسائل أكثر تمييزاً. يستخدم MarGNN تعلم معزز متعدد العملاء لتعلم عتبات التجميع المثلى لكل علاقة، مما يساعد في اختيار الجيران الأكثر صلة (Top-p neighbor sampling) قبل تجميع المعلومات داخل العلاقة وبين العلاقات المختلفة. ثالثاً، لمعالجة مشكلة "الذيل الطويل" (Long Tail problem) في اكتشاف الأحداث، تم تصميم آلية BasCL (Balanced Sampling Strategy based Contrastive Learning Mechanism) التي تستخدم التعلم المقارن (Contrastive Learning) لتعلم ميزات مشتركة بين الرسائل المتشابهة وتمييز غير المتشابهة، مع الحفاظ على هيكل البيان. رابعاً، في مرحلة الكشف، يتم عنقدة الرسائل الممثلة إلى عناقيد باستخدام نموذج DRL-DBSCAN (Deep Reinforcement Learning guided DBSCAN) يستخدم هذا النموذج التعلم المعزز، لتعيين المعاملات المثلى لخوارزمية العنقدة DBSCAN بشكل تلقائي. أخيراً، لتمكين الكشف العابر للغات، تقدم المنهجية طريقة Cross-lingual social message embedding. تعتمد Crlme على نقل المعرفة عن طريق استخدام المتوسطات المدربة على لغة المصدر

(مثل الإنجليزية) لتحسين أداء تضمين الرسائل في اللغة الهدف، وتم استخدام نموذج مساعد مثل LNMAP لمطابقة الرسائل غير الإنجليزية مع الفضاء الدلالي للغة المصدر.

- مجموعة البيانات: Events 2012, Events 2018

- الملاحظات:

هذه الورقة تركز على جانب كشف الأحداث فقط (القناة مفلترة مسبقاً). والكود متوفر.

وتمت مقارنة النموذج مع: Word2vec, LDA, WMD, BERT, PP-CGN, EventX, KPGNN, QSGNN

ث. 2022 From Known to Unknown: Quality-aware Self-improving Graph Neural Network for Open Set Social Event Detection

- الملخص:

يقدم هذا العمل شبكة عصبونية بيانية ذاتية التحسين واعية بالجودة (QSGNN) لمواجهة التحدي المتمثل في اكتشاف الأحداث الاجتماعية في بيئة مفتوحة تتضمن أحداثاً جديدة وغير معروفة تظهر باستمرار. في حين أن الشبكات العصبونية البيانية (GNNs) حققت نجاحاً كبيراً في الكشف عن الأحداث ضمن مجموعة مغلقة من الأحداث، فإنها تتطلب كميات كبيرة من البيانات النمطة وتواجه صعوبة في التعامل مع البيانات غير المعروفة سابقاً، مما يجعل التوسع إلى سيناريوهات المجموعة المفتوحة أمراً صعباً. لمعالجة ذلك، تعمل QSGNN على توسيع المعرفة من الأحداث المعروفة إلى غير المعروفة من خلال الاستفادة القصوى من العينات المعروفة ونقل المعرفة الموثوق.

- المنهجية:

تعتمد منهجية هذا العمل على مرحلتين أساسيتين لمواجهة تحدي اكتشاف الأحداث الاجتماعية في بيئة مفتوحة: مرحلة التدريب المسبق الخاضع للإشراف ومرحلة الضبط الدقيق ذاتي التحسين. في مرحلة التدريب المسبق الخاضع للإشراف، يتم بناء مرمز GNN أساسي باستخدام بيانات نمطة (أحداث معروفة). يتم بناء بيان الرسائل على أساس العلاقات بين الرسائل، وتجمع ميزات الرسائل الأولية بين الدلالات اللغوية والمعلومات الزمنية. بعد ذلك، يتم تطبيق مرمز GNN لتعلم تمثيل شامل للرسائل. في مرحلة الضبط الدقيق ذاتي التحسين، يتم تعميم النموذج المدرب مسبقاً على بيانات غير نمطة (أحداث غير معروفة) ويتم تحسينه ذاتياً. تستخدم المنهجية هنا أشعة توزيع التشابه المرجعي (Reference Similarity (RSD Distribution، يتم قياس جودة التسميات Pseudo Labels بناءً على اتساقها (التشابه بين أشعة RSD)، حيث تكون الجودة أعلى كلما كانت قيمة الاتساق بعيدة عن عتبة التقسيم. لمقاومة الضجيج في التسميات الزائفة، يستخدم النموذج تحسيناً

موجهًا بالجودة وزن مساهمات التسميات بناءً على جودتها. يتم بعد ذلك ضبط مرمز GNN الأساسي بشكل مستمر باستخدام Pseudo Labels المنتقاة والموزونة للتكيف مع الأحداث الجديدة وغير المرئية من قبل.

- مجموعات البيانات: Events 2012, Events 2018

- الملاحظات

هذه الورقة تركز على جانب كشف الأحداث فقط (القناة مفلترة مسبقاً). والكود متوفر. وتمت مقارنة النموذج المطروح مع: Word2vec, LDA, BERT, PP-CGN, EventX, KPGNN

### 3. الطرائق المعتمدة على المستند Document Pivot

أ. Embed2Detect: Temporally Clustered Embedded Words for Event Detection in Social 2022 Medi

- الملخص:

يهدف هذا النظام المنجز في هذه الورقة إلى كشف الأحداث الواقعة في شريحة زمنية، والدخل هو قناة مفلترة Filtered Stream Data، ومن خلال ال Hyper Parameter يتم تحديد أهمية الحدث ومعدل التحديث للأحداث المهتم بها. تعتمد هذه الورقة على تضمين الكلمات باستخدام BERT والعنقدة الهرمية.

تم اختبار النموذج على مجموعتي بيانات من وسائل التواصل الاجتماعي واحدة رياضية والأخرى سياسية.

- المنهجية:

الأمر الجديد في هذه الورقة أنها تقوم بتعلم التضمين أثناء العمل (تدريب BERT في الزمن الحي). كما ويتم مراقبة التغيرات في العناقيد والمفردات المستخدمة، بين الشرائح الزمنية المتتالية لتحديد الحدث، وذلك لتجنب الاعتماد على العنقود لوحده مباشرةً. يحتوي النظام على أربع مكونات، وهي (1) Stream Chunker الذي يفصل دفع البيانات إلى نوافذ زمنية متتالية، (2) Word Embedding Learner الذي يتعلم التضمين في الزمن الحي، (3) Event Window Identifier الذي يحدد الشريحة الزمنية التي وقع فيها الحدث، (4) Event Word Extractor.

- مجموعة البيانات: Sports (English Premier League 19/20 on 20 October), Politics (Brexit Super Saturday 2019)

ب. 2024 An Incremental Clustering Baseline for Event Detection on Twitter

- الملخص:

كشف الأحداث في دفع نصي هو مهمة مهمة من أجل تحليل حي لوسائل التواصل الاجتماعي. وواحدة من التحديات الحالية ي هذا المجال هي وضع حد مرجعي قياسي. في هذه الورقة، تم استخدام خوارزمية عنقدة تزايدية بجمعها مع التقدمات الحالية في تضمين الجمل. والهدف من هذه هو مقارنتها مع الدراسات السابقة خصوصاً Mazoyer و Cao et al., 2024 et al., 2020.

- المنهجية:

تم اتباع المنهجية المستخدمة في Mazoyer et al. 2020 حيث تم استخدام خوارزمية عنقدة تزايدية. حيث يتم تمثيل الجمل كأشعة باستخدام SBERT، ومعالجة الدفع كدفعات صغيرة mini batches، حيث لكل مستند في الدفعة إذا كانت مسافته لأقرب جار أصغر من عتبة ما فتتم إضافته إلى العنقود، وإلا يتم إضافته لعنقود جديد. وتم استخدام تابع المسافة Cosine Similarity لقياس التشابه.

- مجموعة البيانات المستخدمة: Events 2012, Events 2018

ت. 2016 TwitterNews: Real Time Event Detection from the Twitter Data Stream

- الملخص:

في هذه الورقة تم اقتراح نظام كشف أحداث قابل للتوسع، تمت تسميته TwitterNews، لكشف الأحداث الجديدة بالذكر وتتبعها في الزمن الحقيقي، حيث تم اقتراح مقارنة جديدة، بدمج Random Indexing و LSH لتحسين أداء العنقدة المتزايدة ي تجميع التغريدات المتشابهة ضمن شريحة زمنية ثابتة. وتم التعامل مع مشكلة تجزئة العناقيد في العنقدة المتزايدة.

- المنهجية:

تم تقسيم النظام إلى وحدتين الأولى Search Module التي تحتوي على خوارزميات لإيجاد الكلمات التي يجب البحث عنها والاهتمام بها، الوحدة الثانية هي Event Cluster Module والتي تقوم بتنفيذ العنقدة المتزايدة ومعالجتها، ومن ثم يتم فترة العناقيد وأخذ العناقيد الجديدة بالذكر، وبعد ذلك لكل عنقود يتم أخذ تغريدة واحدة كمثلة للعنقود.

- مجموعة البيانات: Events 2012

ث. 2015 Arabic Event Detection in Social Media

- الملخص:

كشف الأحداث هو مفهوم مهم لضمان الأمان العام والإحاطة بأحداث العالم الحقيقي. في هذه الورقة تم تقديم إطار عمل جديد لكشف الأحداث، مع التركيز على الأحداث التخريبية باستخدام بيانات تويتر. المقارنة قائمة على خمس خطوات، أولاً جمع البيانات، ومن ثم تطبيق معالجة مسبقة، وبعد ذلك القيام بعملية تصنيف لفترة

الرسائل غير ذات الصلة، بعد ذلك يتم القيام بعملية عنقدة، وأخيراً تلخيص العقود. حيث تم استخدام نموذج Naïve Bayse للتصنيف، وطريقة عنقدة حيّة من أجل العنقدة. وكانت هذه الدراسة هي الأولى التي اهتمت بأحداث باللغة العربية. مجموعة البيانات: غير متاحة.

## 4. Graph Partitioning Approaches

أ. 2024 EDT BERT Event Detection and Tracking in Twitter using Graph Clustering

- الملخص:

كشف الأحداث من وسائل التواصل (مثل تويتر) هو مشكلة بحث ساخنة. وله تطبيقات في مجالات عديدة كالصحافة، التسويق، إدارة الأزمات، الاستجابة للأزمات. معظم المنهجيات الحالية تعتمد على تضخم الكلمات المفتاحية، ومشكلة هذه المنهجيات أنها تفشل في التعرف على الأحداث الجديدة بالذكر قبل أن تصبح ترند. وعلاوة على ذلك، فإن هذه المنهجيات تواجه صعوبة في متابعة تطور الحدث مع معلومات سياقية محدودة أو غير كافية. وفي هذه الورقة، اقترح الباحثون منهجية معالجة التغريدات قائمة على الشرائح الزمنية سموها EDT BERT لكشف الأحداث ومتابعة تطور الحدث مع الزمن.

- المنهجية:

تستخدم المنهجية المعلومات الهيكلية والدلالية الموجودة في التغريدات، حيث تبدأ بتوليد بيان للتغريدات، حيث كل تغريدة هي عقدة والوصلات بين العقد هي التشابه بين الرسائل (حيث تم استخدام الكيانات المسماة hashtags المتداخلة بينهما في قياس المسافة). وتم استخدام BERT لتمثيل التغريدات كأشعة، وبعد ذلك تم استخدام طريقة عنقدة للبيان لتحديد العناقيد. وبعد ذلك يتم ربط العناقيد الحالية مع عناقيد المرحلة السابقة لتشكيل سلسلة من العناقيد تعبر عن تطور الحدث.

مجموعة البيانات: Super Tuesday, US Election, FA Cup

ب. 2023 Efficient graph-based event detection scheme on social media

- الملخص:

يشارك المستخدمون آراءهم المختلفة على وسائل التواصل الاجتماعي، مما يوفر إمكانية لكشف العديد من الأحداث عبر تحليل بياناتها. في هذه الورقة، أقترح الباحثون آلية لكشف الأحداث معتمدة على البيان عبر تحليل بيانات وسائل التواصل الاجتماعي. الآلية المقترحة تعتمد على الظهور المتوازي للكلمات المذكورة في الحدث في شكل البيان. حيث يتم تشكيل بيان الكلمات المفتاحية ويتم عنقدة الكلمات المتعلقة بحدث من خلال تحليل البيانات. ومن خلال قياس اهتمام المستخدم وتغير نشاطه،



تستطيع الآلية المقترحة تحديد الأحداث التي تتلقى استجابة من المستخدمين وتسحين اعتمادية النتائج من خلال استبعاد المنشورات غير ذات الصلة.

- المنهجية:

تم جمع البيانات، وتحضيرها، ومن ثم استخراج الكلمات باستخدام محلل صرفي، وإزالة الكلمات التي يفترض أن تحدث فجأة. بعد ذلك تم بناء بيان الكلمات المفتاحية ومن ثم تطبيق إجراءات استخراج معلومات الحدث التي تستخلص البيانات المرشحة. تم اسناد أوزان للوصلات بناءً على قياس اهتمام المستخدم. ويتم استخراج البيانات المرشحة لتكون حدث من خلال فترة وعنقدة البيان. وأخيراً بعد إعطاء ترتيب لبيان كل حدث تم أخذ أفضل K حدث.

- مجموعة البيانات: غير متاحة.

## 5. الطرائق المعتمدة على السمة Feature Pivot

أ. 2019 SEDTWik: Segmentation-based Event Detection from Tweets using Wikipedia

- الملخص:

إنّ كشف الأحداث واحد من مجالات البحث في نطاق التقنيب في النصوص والذي جذب الانتباه خلال هذا العقد تبعاً للانتشار الواسع والإتاحة لوسائل التواصل الاجتماعي وخصوصاً تويتر. لقد أصبح تويتر المصدر الرئيسي للمعلومات حول أحداث العالم الحقيقي. الأعمال السابقة في مجال كشف الأحداث من تويتر تكون إما قابلة للتطبيق لكشف الأحداث المحليّة أو الأخبار العاجلة فقط أو تفقد العديد من الأحداث المهمة.

هذه الورقة تعرض المشاكل المرتبطة بكشف الأحداث من تويتر وتقدم نظام لكشف الأحداث قائم على التقطيع، تم تسميته SEDTWik، وهو قادر على تحديد الأحداث الجديدة بالذكر التي تظهر في أماكن متعددة من العالم ومن فئات متعددة. الفكرة الرئيسية لهذا العمل هي تجزئة التغريدة وال hashag إلى قطع segments، ومن ثم استخراج القطع المتضخمة في التغريدات bursty segments، وبعد ذلك عنقدهم، ومن ثم تلخيصهم.

- مجموعة البيانات: Events 2012