

Mendeteksi Otomatis Aspek Produk dari Ulasan Pelanggan

Fakry Adi Permana (1301164034)
Anggi Yuniar Putri (1301154492)

Program Studi Teknik Informatika
Telkom University, Bandung

Abstract—Makalah ini mengusulkan pendekatan baru untuk mendeteksi aspek secara otomatis. Tujuan dari algoritma yang kami gunakan untuk memperoleh ringkasan aspek dari produk tertentu, yang berasal dari ulasan pelanggan. Pendekatan kami dimulai dengan mencocokkan *dependency path* buatan dalam kalimat untuk menemukan aspek kandidat pada pendapat yang diungkapkan. Pendekatan kami tidak membutuhkan sepetah katapun ... atau pengetahuan khusus domain, karena hanya mempekerjakan leksikon sentiment. Yang kami bahas hanya hasil dari akurasi yang telah dilakukan berdasarkan data dari ulasan pelanggan.

Keywords—*aspek, dependency path, lexicon, sentiment.*

I. PENDAHULUAN

Pada saat ini minat seseorang untuk menganalisis telah bergeser ke arah analisis yang lebih kualitatif, yang bertujuan untuk mendeteksi aspek-aspek yang berbeda dari suatu topik ke arah mana suatu pendapat tersebut diungkapkan. Misalnya, dalam menganalisis ulasan sebuah produk. Dalam pengulasannya terhadap aspek-aspek seperti kondisi barang, kinerja produk, dan kualitas produk tersebut.

Dalam tulisan ini kami memperkenalkan pendekatan baru yaitu Aspektaktor, sebuah algoritma baru mendeteksi dan menilai aspek produk dari ulasan pelanggan. Saat ini dibatasi pengerjaan aspektaktor hanya sampai ekstraksi aspek. Aspektaktor bisa menemukan aspek kandidat dengan mencocokkan sintaks *dependency path*, sedangkan pendekatan lain membutuhkan inputan kata-kata awal dan menggunakan sintaks *dependency* atau untuk menemukan kata baru dan hubungannya. Akibatnya, sistem yang diusulkan dapat mendeteksi dan menilai aspek produk dalam domain apapun. Secara konkret Aspektaktor menggabungkan aspek kandidat yang telah diekstraksi dari kalimat individu melalui sintaks *dependency path*, dengan mencari akurasi, presisi dan recall dari hasil aspek-aspek yang telah dikeluarkan sistem berdasarkan data dari ulasan pelanggan.

II. TINJAUAN PUSTAKA

Pada tahun 2004 Hu dan Liu memperkenalkan masalah yang lebih menarik dari analisis sentiment berbasis aspek, dimana polaritas tidak ditugaskan untuk kalimat atau dokumen tetapi untuk aspek tunggal yang dibahas dalamnya. Dalam pendekatannya diberikan sejumlah ulasan besar pada produk tertentu, pertama-tama melakukan pengidentifikasian aspek-aspek produk yang menarik, menggunakan asosiasi penambangan, kemudian melampirkan nilai sentiment untuk setiap aspek yang telah dieksploitasi dari sekumpulan kata-

kata opini, bersamaan dengan kata sinonim dan antonim. Tahap selanjutnya mendeteksi kata pendapat yang kemudian di ekstraksi menjadi beberapa aspek-aspek.

Pekerjaan yang telah dibahas sejauh ini mengusulkan solusi dari domain independent yang diposisikan atau digunakan untuk menganalisis sentiment berbasis aspek. Dalam pengklasifikasiannya ulasan produk, membutuhkan inputan awal untuk membaca algoritma khusus mengenai *value* produk tersebut seperti kondisi produk, kualitas produk dan kinerja produk. Yang di ikuti dengan kata opini khusus sebagai domain untuk algorima mereka. Seperti yang diharapkan domain menghasilkan klasifikasi sentiment yang lebih akurat. Karena semua aspek domain yang relevan diberikan sebagai inputan, secara eksklusif pendeteksian mengenai opini masih berfokus terhadap aspek yang diberikan (1) Menangkap kata-kata opini baru melalui *dependency sintaks*, dan (2) Menilai aspek produk berdasarkan eksternal leksikon sentiment dan beberapa kata opini khusus yang telah diberikan domain.

Demikian pula, menurut Zhu, menggunakan aspek produk dan beberapa istilah terkait aspek sebagai input untuk algoritma mereka, tetapi kemudian berusaha untuk menemukan istilah terkait aspek baru dengan menerapkan algoritma bootstrap berdasarkan pada kejadian antara istilah kata utama dan istilah kandidat baru. Skor sentimen kembali diperoleh dengan mengakses leksikon sentimen eksternal. Sementara pendekatan kami mempertahankan dari karya-karya ini penggunaan leksikon eksternal, itu tidak memerlukan contoh berlabel atau pengetahuan khusus domain, sehingga memiliki penerapan yang lebih luas.

A. Mendeteksi Aspek Produk

Langkah pertama adalah untuk mengekstrak dari ulasan pelanggan menyebutkan aspek produk dan kata-kata yang mengekspresikan pendapat penulis terhadap aspek tersebut. Misalnya, diberi kalimat:

“i'd say that this camera is perfect.”

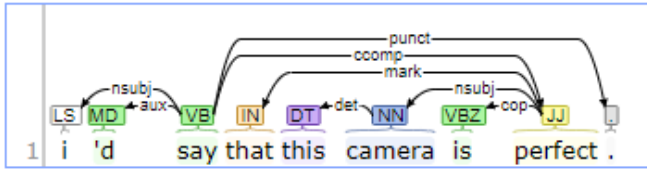
Aspektaktor akan mengekstrak kalimat menjadi berikut:

< Camera >
Aspect
Mention

Aspektaktor mengekstraksi pasangan pendapat dengan menggunakan enam *dependency path* sederhana buatan tangan secara manual, dalam tiga langkah:

1. Untuk setiap kalimat, aspektaktor mengekstrak sintak *dependency tree* menggunakan stanford *depedency parser*, seperti pada Gambar 1 yang menunjukan *dependencies* untuk contoh kalimat.

2. Menggunakan dependency tree, aspektator mencoba untuk mengeluarkan pasangan pendapat dasar yang dibentuk oleh pengubah sentimen kata tunggal.
3. Jika ada pasangan pendapat yang cocok dengan dependency path yang terdapat pada Table 1 dan Table 2, aspektator mencoba untuk memperluas kecocokan dengan kata terdekat dengan menerapkan dependency path tambahan.



Gambar 1. Contoh sintatik deteksi dependency oleh Stanford dependency parser

Tabel 1. Main dependency path digunakan Aspektator untuk mendeteksi sebuah aspek(A)

Type	Dependency path	Example
Adjectival modifier	$M \xleftarrow{amod} A$	$DT \quad NN \quad VBD \quad DT \quad JJ \xleftarrow{amod} NN$ The movie had an excellent storyline!
Direct object	$A \xleftarrow{nsubj} * \xrightarrow{dobj} M$	$DT \quad NN \xleftarrow{nsubj} VBD \xrightarrow{dobj} JJ$ The story got stale.
Adjectival complement	$A \xleftarrow{nsubj} * \xrightarrow{acomp} M$	$DT \quad NN \xleftarrow{nsubj} VBZ \xrightarrow{acomp} JJ$ The music does sound great.
Complement of a copular verb	$A \xleftarrow{nsubj} M \xrightarrow{cop} *$	$DT \quad NN \quad NN \quad VBN \quad IN \quad DT \quad NN \quad VBD \quad RB \quad RB \quad JJ$ The action music used in the movie wasn't too good.
Adverbial modifier to a passive verb	$A \xleftarrow{nsubjpass} * \xrightarrow{advmod} M$	$DT \quad NN \xleftarrow{nsubjpass} VBZ \xrightarrow{advmod} VBN$ The storyline is well written.

Tabel 2. Extension dependency path digunakan Aspektator untuk menangani multi-word aspek (A)

Type of extension	Dependency path	Example
Compound noun	$A' \xleftarrow{nn} A$	$DT \quad NN \xleftarrow{nn} NN \quad VBN \quad IN \quad DT \quad NN \quad VBD \quad RB \quad RB \quad JJ$ The action music used in the movie wasn't too good.

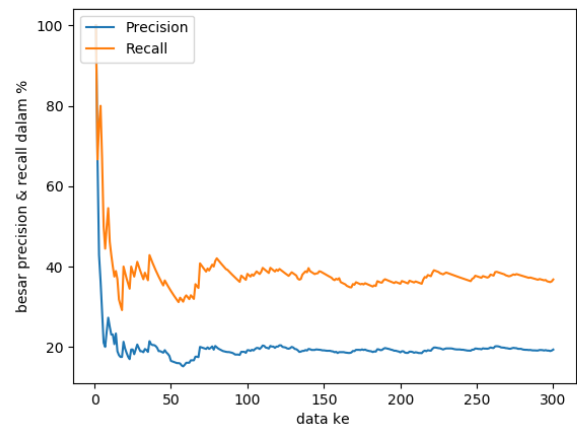
III. EXPERIMENT

Pada bagian ini, kami menyajikan evaluasi awal Aspektator. Tujuan percobaan kami adalah untuk menjawab “Bisakah pendekatan kami mendeteksi aspek produk yang menarik dan relevan?”.

A. Metodology

Telah tersedia dataset berupa kumpulan ulasan produk kamera, data tersebut dilakukan dependency parsing menggunakan Stanford dependency parser, hasilnya dilakukan *enumerate* untuk dianalisis dengan 6 aturan *dependency path* yang telah dibuat. Apabila termasuk ke dalam salah satu tipe *dependency path* maka aspek diekstraksi.

B. Hasil



Gambar 2. Grafik Precision dan Recall berdasarkan Data yang dioperasikan

Grafik diatas adalah hasil perhitungan dari dilakukannya ekstraksi aspek pada tiap kalimat ulasan dari dataset yang dimiliki, kemudian membandingkan hasilnya dengan aspek sebenarnya. Hasil precision didapatkan dari perhitungan :

$$Prec = \frac{Jml \text{ aspek terekstrak benar}}{Jml \text{ aspek terekstrak benar} + Jml \text{ aspek terekstrak salah}}$$

Hasil recall didapatkan dari perhitungan :

$$Recall = \frac{Jml \text{ aspek terekstrak benar}}{Jml \text{ aspek sebenarnya}}$$

Sehingga dari 300 data dihasilkan precision sebesar 19,335% dan recall sebesar 36,782%.

IV. CONCLUSION

“Aspektator”, algoritma baru untuk analisis sentimen berbasis aspek yang memasukan input ulasan pelanggan terhadap produk tertentu, dan secara otomatis akan mengekstraksi aspek. Pertama-tama aspektator mengumpulkan aspek kandidat dan opini terkait untuk dicocokkan dengan 6 aturan *dependency path*. Pendekatan kami adalah domain-independen dan tidak memerlukan contoh berlabel, sehingga dapat diadopsi untuk menganalisis ulasan pelanggan untuk produk dalam domain yang tidak terlihat.

Berdasarkan hasil perhitungan, jumlah data mempengaruhi precision dan recall, semakin banyak data yang dioperasikan maka precision dan recall semakin stabil diangka precision sekitar 18-19% dan recall sekitar 35-36%.

Untuk pekerjaan di masa mendatang, kami tertarik untuk memverifikasi, mulai dari beberapa contoh pasangan opini, kami dapat mempelajari *dependency path* yang telah dibuat, dan menemukan jalur tambahan yang digeneralisasi dengan baik di berbagai domain.