

Multimodalna analiza sentymentu memów z wykorzystaniem dużych modeli językowych oraz połączonych reprezentacji wizualno-tekstowych

Katarzyna Michalska
Matwej Novgorodtsev

Problem badawczy

Memy to specyficzna forma komunikacji, w której znaczenie wynika z jednoczesnej obecności obrazu, tekstu i często kontekstu kulturowego. Klasyczne metody analizy sentymentu skupiają się jednak głównie na tekście, co sprawia, że dla analizy sentymentu memów klasycznymi metodami tracimy połowę informacji. Memy bywają ironiczne, sarkastyczne, wizualnie przerysowane, a tekst i obraz potrafią ze sobą grać albo wręcz sobie przeczyć. To prowadzi do kilku realnych wyzwań badawczych:

- Trudność w uchwyceniu ironii i humoru opartego na kontraście między obrazem a tekstem.
- Problemy z uzyskaniem reprezentacji wizualnych, które faktycznie niosą ładunek emocjonalny, a nie tylko opisują treść obrazu.

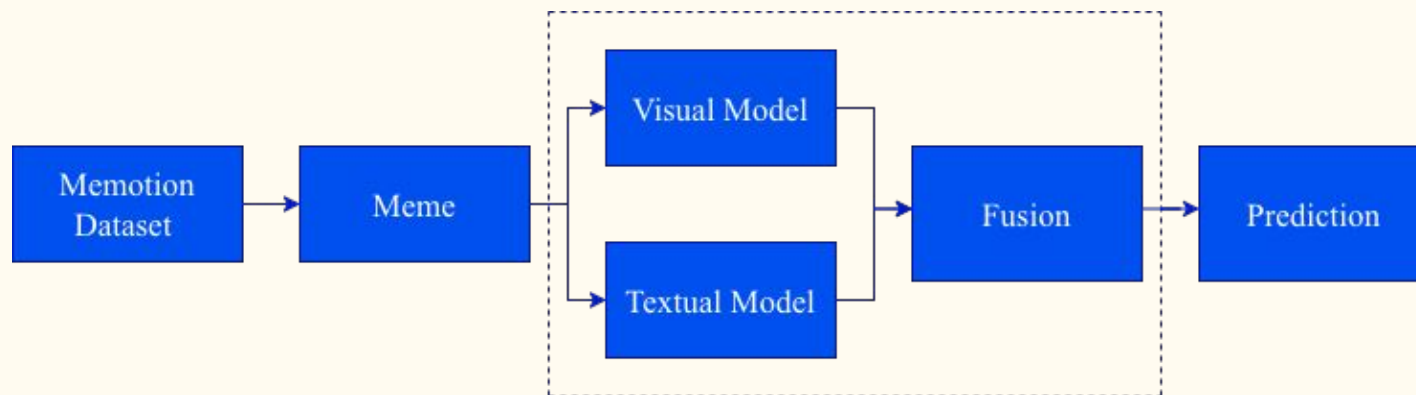
Badanie polega więc na tym, żeby sprawdzić, które modele najlepiej radzą sobie z tą trudną mieszaniną tekstu i obrazu, a także jak różne strategie fuzji wpływają na jakość klasyfikacji nastroju.

Cel pracy

Celem pracy jest zbadanie, jak różne podejścia do multimodalnej analizy sentymentu radzą sobie z interpretacją memów. Stworzony zostanie taki zestaw metod, który pozwoli nie tylko analizować tekst i obraz osobno, ale też zrozumieć ich wzajemne oddziaływanie. Praca ma więc kilka kluczowych zadań:

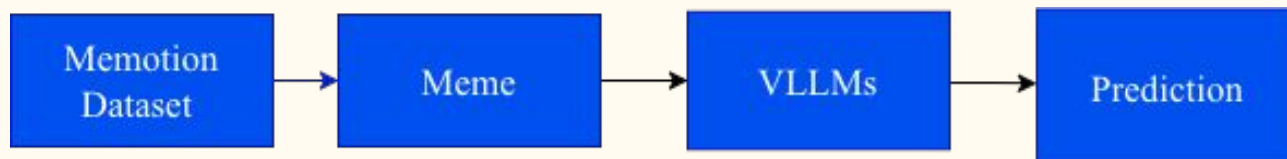
- Opracowanie pipeline'u, który potrafi wyciągnąć sensowne cechy z dwóch modalności: obrazu i tekstu.
- Zaprojektowanie i przetestowanie sposobów ich łączenia w jedną wspólną reprezentację, która nadaje się do klasyfikacji nastroju.
- Porównanie wyników modeli opartych na dużych modelach językowych z klasycznymi architekturami multimodalnymi, żeby sprawdzić, czy nowoczesne LLM-y mają przewagę w rozumieniu treści memicznych.
- Ocena jakości klasyfikacji w różnych wariantach, tak żeby finalnie móc wskazać, które podejście jest najbardziej skuteczne i dlaczego.

Dane wejściowe



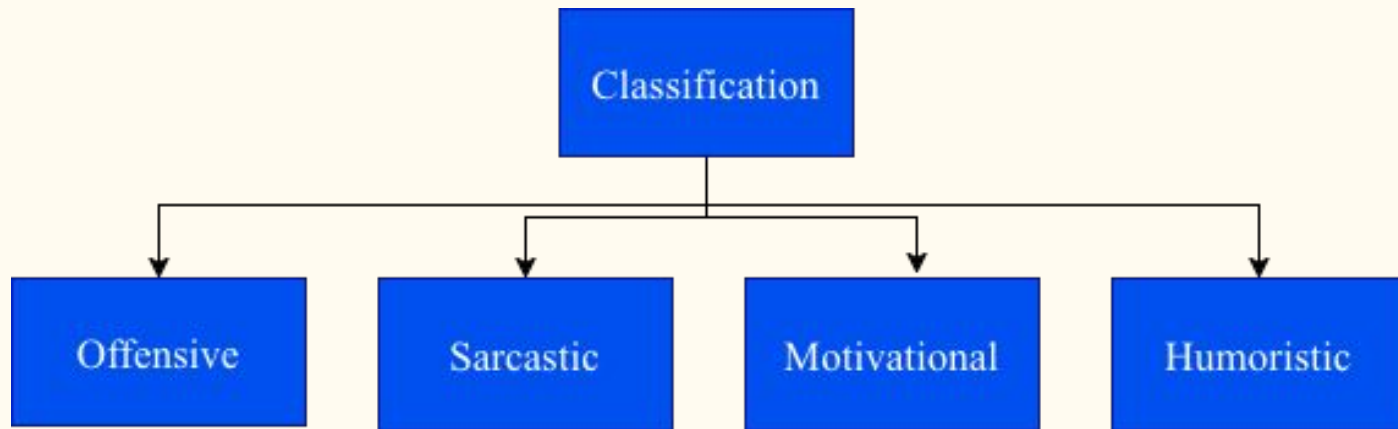
Rys. 1 Koncepcja architektury wizualno-tekstowej

Dane wejściowe



Rys. 2 Koncepcja przepływu architektury z VLLM

Output



Rys. 3 Możliwe warianty klasyfikacji

Wyniki badań

W ramach badań przewiduje się ocenę:

- jakości predykcji w zakresie rozpoznawania kategorii sentymentu
- wpływu zastosowanej metody fuzji modalności na uzyskiwane rezultaty
- roli reprezentacji wizualnych i tekstowych w procesie klasyfikacji
- zdolności poszczególnych architektur do uchwycenia niuansów treści charakterystycznych dla memów, takich jak ironia czy humor
- stabilności i powtarzalności wyników uzyskiwanych przez różne modele

Celem tej części pracy będzie ustalenie, które podejścia osiągają najwyższą skuteczność oraz określenie czynników decydujących o ich wydajności. Analiza wyników ma umożliwić wyciągnięcie wniosków dotyczących mocnych i słabych stron poszczególnych metod oraz wskazanie kierunków dalszego doskonalenia modeli przeznaczonych do multimodalnej analizy sentymentu.

Potencjalne zastosowania praktyczne

Z punktu widzenia zastosowań praktycznych opracowane rozwiązania mogłyby znaleźć wykorzystanie w kilku obszarach:

- W mediach społecznościowych analiza sentymentu memów mogłaby wspierać monitorowanie nastrojów użytkowników oraz wczesne wykrywanie negatywnych trendów.
- W marketingu narzędzia tego typu mogłyby służyć do oceny reakcji odbiorców na kampanie oparte na treściach wizualnych.
- W badaniach nad komunikacją internetową analizowanie memów mogłoby stanowić wsparcie dla socjologów i psychologów zajmujących się dynamiką dyskursu online.
- Dodatkowo modele multimodalne mogłyby być wykorzystywane w systemach moderacji treści, ułatwiając identyfikację memów o nacechowaniu obraźliwym lub szkodliwym.