Biz əvvəl sual cavablandırmanı qiymətləndirməkdən danışdıq. Əgər bizim faktiki suala

cavabımız varsa, biz həmin cavabı sistem tərəfindən qaytarılan faktla müqayisə edib düzgün

fakt olub-olmamağına baxaraq qiymətləndiririk. Qısa məzmunlar bu cür qiymətləndirilə bilməz,

çünki, bizim bir sənəd üçün yeganə və mükəmməl qısa məzmunumuz olmur. Belə ki, ROUGE

adlanan fərqli alqoritmi təqdim edəcəyik. ROUGE Lin və Hovy tərəfindən təklif edilmişdir və

Xülasənin Qiymətləndilirməsi Üçün Yada Salma Yönümlü Təkrarçı deməkdir. İdeyası da ondan

ibarətdir ki, o qısa məzmunları qiymətləndirmək üçün daxili ölçüdür və bizim qısa

məzmunumuzun yaxşı olub olmadığını təyin edəcək. Sadəcə qısa məzmun olmaqla

xarici tətbiqləri də vardır. Və o bleu və ya blue adlanan ölçüyə əsaslanır hansı ki, maşın

tərcüməsi (machine translation) üçün təyin edilmişdir. Təbii ki, ROUGE nəticə olaraq

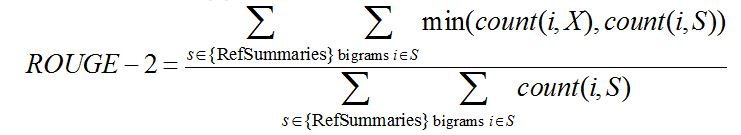
"Bu qısa məzmun istifadəçinin sualına cavab verirmi?" sualına düzgün cavab verməkdə

insanlar qədər yaxşı deyil. Əgər imkanımız olsa, biz istifadəçiləri işə götürərdik və onlara

cavabın istifadəçinin sualına cavab verdiyini yoxladardıq. Ancaq ROUGE biz sistemimizi

düzəltdiyimiz zaman test üçün çox uyğundur. Və o indi danışacağımız kimi işləyir.

* D sənədi, və avtomatik qısa məzmun X verilib:
  1. N sayda adam D sənədinin istinad qısa məzmunlar toplusunu yaradır.
  2. Sistemi işə salırıq və o bizə avtomatik qısa məzmun X-i verir.
  3. İstinad qısa məzmunlardan götürülmüş neçə faiz biqramlar X-da da mövcuddur?



Bizə D sənədi verilmişdir və deyək ki, bizim qısa məzmun çıxardan sistemimiz var və

həmin sistem avtomik qısa məzmun yaradır. Bu sorğu-əsaslı qısa məzmun ola bilər əgər

istidaçinin verdiyi sualı biliriksə və ya ümumi qısa məzmun ola bilər. N sayda adam

bu sənədin istinad qısa məzmunlar toplusunu yaradır. Onlar sorğu əsaslı qısa məzmun çıxartma

zamanı sorğu baxıb öz qısa məzmunlarını hazırlayırlar. Ümumi qısa məzmun çıxarmada isə

onlar sadəcə qısa məzmunlar hazırlayırlar. Məsələn D sənədinin əlimizdə olan bir, iki, üç və ya

dörd qısa məzmunu var. Və sistem yeni bir qısa məzmun yaradır və onu x adlandıraq. Artıq

bizim əlimizdə avtomatik qısa məzmun və dörd ədəd insan tərəfində hazırlanmış qısa məzmun

var. Beləliklə, biz sual verə bilərik ki, insanlar tərəfindən hazırlanmış qısa məzmunlardan

götürülmüş biqramlar neçə faizi X-da baş verir? Aydındır ki, biqramların hamısı X-da olmayacaq.

Yaxşı qısa məzmunda insanlar tərəfindən hazırlanmış qısa məzmunlarda baş verən çoxlu

biqramlar olacaq. Faizləri saymaqla sən yaxşı qısa məzmunun necə olacağı haqqında intuisiya

əldə edirsən. ROUGE'un müxtəlif növləri var: yuniqram ROUGE, biqram ROUGE və

uzunluqları müxtəlif variantlarda nəzərə alan digər versiyalar. Biz biqram ROUGE təqdim

edəcəyik, o sadəcə bütün istinad qısa məzmunlardakı bütün cümlələrdə olan biqramlara

baxacaq və onları sayını hesablayacaq. O sistem tərəfindən yaradılmış qısa

məzmundakı hər bir cümlədə biqramlara baxır və hər biqramın sisteminin qısa məzmununda

və insanların hazırladığı qısa məzmunda minimum sayının neçə olduğunu hesablayır.

Beləliklə, o sadəcə soruşur ki, nə qədər biqram həm bizim sistemimizin qısa məzmununda

həm də insan tərəfindən hazırlanmış qısa məzmunda baş verir. Nəticə olaraq biz baş vermələrin

sayını əldə edəcəyik. Gəlin, biz bir misala baxaq. Burda biz üç ədəd insan məhsulu qısa

məzmun və sistemin məhsulu olan qısa məzmun (sistemin suala cavabı) hazırlamışıq.

İndi, ROUGE hesablayaq. Bunun üçün biz surətdə insan məhsulu qısa məzmunlarda nə qədər

biqramın olduğunu və o biqramların nə qədərinin sistem cavabında da

əks olduğunu bilmək istəyirik. Biz başlaya bilərik, ilk qısa məzmunda birinci biqram

"water spinach"-dır. Bu sistemin cavabında da əks olunub. Sonrakı biqramlar

"spinach is" və "is a"-dır. İkinci qısa məzmunda da eyni biqramlar vardır. Və üçüncüdə isə

bu biqramlardan əlavə sistemin cavabı ilə ortaq olan " commonly eaten ",

" leaf vegetable ", " of Asia " biqramları da var. Əgər biz bunları cəmləsək, birinci qısa

məzmundan üç biqram, ikincidən də üç və üçüncüdən isə altı ortaq biqram əldə edəcəyik.

Bəs yaxşı insan məhsulu olan qısa məzmunlarda ümumi nə qədər biqram var? Siz onları

özünüz saya bilərsiniz. Birinci məzmunda 10, ikincidə 9 və üçüncüdə isə 9 biqram var.

Və biz (3+3+6)/(10+9+9) ifadəsini hesablayırıq və ROUGE nəticəsi olaraq .43 alırıq. Beləliklə biz

qısa məzmunların qiymətləndırmə alqoritmi olan ROUGE'u tanıtdıq. Qısa məzmunların ümumi

və ya sorğu-əsaslı olmasınından asılı olmayaraq biz insan məhsulu olan qısa məzmunlardan

neçə biqramın və ümumilikdə neçə N qramın maşın tərəfindən yaradılmış qısa məzmunda

baş verdiyinə baxdıq. Ən yaxşı qısa məzmun insan məhsulu olan qısa məzmunlarla daha

çox üst-üstə düşəndir. Ancaq ROUGE bu suala cavab verən zaman işləmir: "Bu cavab

həqiqətən istifadəçinin soruşduğu informasiyanı təmin edirmi?" Ancaq performans baxımından

bu çox uyğun olmaz, ona görə də biz ROUGE sürətli və daxili ölçü kimi sistemlərimizi test

etmək üçün istifadə edirik. Sonda nəticəni insanlara göstərməklə sistemin necə effektiv

olduğunu təyin edə bilərik.