Universitat Autònoma de Barcelona

escola d'enginyeria

5967: Classificació automàtica d'imatges de comptadors de gas

Memòria del projecte
d'Enginyeria en Informàtica
realitzat per
Júlia Giner Delgado
i dirigit per
Ernest Valveny Llobet
Bellaterra, 14 de setembre de 2015

Universitat Autònoma de Barcelona

escola d'enginyeria

El sotasignat, Ernest Valveny Llobet,

Professor de l'Escola d'Enginyeria de la Universitat Autònoma de Barcelona

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat sota la seva direcció per na Júlia Giner Delgado.

I per tal que consti firma la present

Signat: Ernest Valveny Llobet

Bellaterra, 14 de setembre de 2015

Índex

Índex de figures

Índex de taules

Agraïments

Primer de tot, m'agradaria donar les gràcies al meu director de projecte, Ernest Valveny Llobet, per guiar-me en aquest treball i per ajudar-me a aconseguir-ho.

També vull agrair el suport i dedicació per part de la meva família (en particular, la meva mare i les meves tres germanes, Carla, Maria i Helena), que m'han ajudat i m'han brindat l'energia necessària per completar la carrera en els moments difícils.

En especial, agrair l'esforç i confiança plena en mi que m'ha demostrat la meva parella, Carles, i que ha suposat l'última empenta per a assolir els meus objectius.

Finalment, mencionar a tots aquells que en diferent mesura han contribuït a que hagi arribat a on sóc ara. Gràcies.

1 Introducció al projecte

1.1 Motivació

Avui en dia, el tractament d'imatges per a diversos fins està molt estès. En la societat actual pràcticament tothom porta una càmera al damunt, ja que normalment ve incorporada per defecte en el telèfon mòbil i en la majoria de dispositius portàtils. Això facilita l'obtenció i l'intercanvi d'imatges en qualsevol moment.

Les imatges esdevenen una eina molt útil si s'aconsegueix extreure les dades que mostren perquè ens poden aportar molta informació. Si aquesta extracció dels continguts importants de la imatge es fa de forma automàtica es pot reduir el temps en l'adquisició de les dades i utilitzar-lo en el seu processament o en l'ús que s'hi vulgui fer d'elles.

Aquest és un àmbit que personalment trobo molt interessant i és en el que es desenvolupa el meu projecte de final de carrera. En el projecte es treballa directament extraient informació d'imatges de comptadors de gas, utilitzant un mètode basat en un descriptor de contorn i un classificador SVM per catalogar els diferents models de comptador, de manera que faciliti el reconeixement del valor del consum i de l'identificador del client.

El projecte de final de carrera m'ha servit per a posar en pràctica habilitats que he anat adquirint al llarg d'aquesta. A més a més, té un ús pràctic al món real, fora de la universitat, pel que m'ha permès veure altres punts de vista diferents als acadèmics.

En aquest cas, el tractament de les imatges està molt enfocat a una necessitat de l'empresa dels comptadors de gas per a agilitzar el seu procés de lectura, però també es podria usar en altres contextos. Un exemple seria l'adquisició d'imatges concretes que continguin certes paraules rellevants dintre d'un llarg nombre de fitxers.

1.2 Objectius

L'objectiu principal d'aquest projecte és avaluar la possibilitat d'identificar el model (marca i tipus) al que correspon un comptador de gas cercant certes paraules clau en una imatge del comptador.

Poder distingir entre models de comptadors de gas de forma automàtica és quelcom molt útil, ja que la ubicació del número de client i el valor del consum de gas realitzat pot variar depenent de quina marca i tipus de comptador es tracti.

En les imatges de la figura 1, podem veure un exemple on es pot observar les diverses maneres de mostrar la mateixa informació.

Figura 1: Mostra de diferents models de comptadors.
Es pot veure que la ubicació de la informació important varia segons els model del que es tracti. Emmarcat en verd tenim l'identificador del client que volem reconèixer i en vermell la lectura associada a aquest.

Inicialment es podria pensar en fer una classificació visual completa, comparant la totalitat de cada imatge per determinar a quin model de comptador pertany. Això però, pot comportar problemes a l'hora de fer la diferenciació, ja que com podem veure a la figura 2 hi ha models diferents que tenen una aparença similar i pot portar a confondre'ls fàcilment.

Figura 2: Models de la mateixa marca i diferent tipus.
Aquests dos models comparteixen la mateixa marca, però són de tipus diferent.

És per això que per a poder determinar a quin model pertany cada imatge, necessitem trobar exactament què els diferencia uns dels altres. L'alternativa triada en aquest cas és la identificació de seccions concretes de les imatges, determinant les paraules claus que indiquen de quin tipus de comptador es tracta. En algun model de comptador podem necessitar-ne només una, mentre que en d'altres facin falta més d'una per a distingir-les entre si. A més a més, la posició relativa d'aquestes paraules en la imatge del comptador també pot servir d'ajuda per a identificar-lo, i això fa que sigui important obtenir la seva localització.

Per a portar-lo a la pràctica es partirà d'un mètode per cercar text en imatges manuscrites que s'haurà de modificar i adaptar lleugerament per poder-lo aplicar a les imatges de comptadors. Aquest mètode es coneix com **EWS** (de l'anglès *Exemplar Word Spotting*) i permet localitzar paraules concretes a partir d'exemples positius i negatius del que volem trobar. A la secció 4 explicarem els fonaments del mètode.

Podem resumir els subobjectius intermedis més importants per a obtenir l'objectiu principal del projecte amb els següents tres punts:

1. Determinar les paraules clau que identifiquen cada model (per marca i tipus).
2. Modificar el codi proporcionat per a adaptar-lo al cas d'imatges de comptadors.
3. Realitzar diversos experiments modificant paràmetres del mètode per a veure l'adaptació feta i quins resultats aporta.

1.3 Organització de la memòria

L'estructura d'aquesta memòria es composa de sis capítols més annexos. El primer capítol (*Introducció al projecte*), on s'inclou aquesta secció, dóna una visió general dels aspectes relacionats amb el projecte. Segueixen els capítols *Estat de l'art*, *Viabilitat del projecte*, *Fonaments teòrics* i *Proves i resultats*, on es troba la descripció del treball realitzat. Finalment trobem una conclusió tècnica i una de personal al darrer capítol.

Amb una mica més de detall:

- *Introducció al projecte*
- *Estat de l'art*: revisió de l'estat des d'on comença l'àmbit de desenvolupament i d'aplicació d'aquest projecte.
- *Viabilitat del projecte*: on s'hi revisa els temps i els costos del projecte.
- *Fonaments teòrics*: s'explica el mètode original del que parteix aquest projecte i la descripció de l'adaptació que s'ha dut a terme.
- *Proves i resultats*: s'hi descriu tot el procés realitzat per provar l'adaptació i els resultats obtinguts.
- *Conclusions*
- *Annexos*: on es troba el diagrama de Gantt de la planificació del projecte i el conjunt de gràfics obtinguts de les diverses proves realitzades.

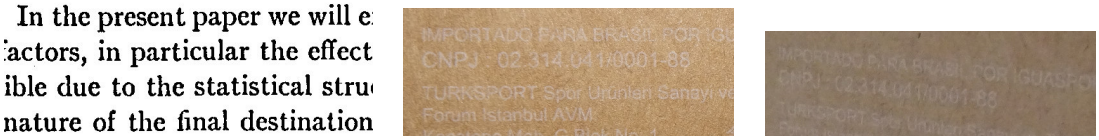
2 Estat de l'art

Tal i com hem comentat, el problema a tractar en aquest projecte es centra en l'extracció d'informació útil de determinades imatges. En concret, es vol localitzar certes paraules que apareguin en aquestes, pel que estem interessants en els mètodes de detecció de paraules (en anglès, *word spotting*).

Aquest problema és molt captivant, ja que tant en el món de la visió per computador com per a la resta de la societat hi ha molt d'interès en traduir la paraula escrita de forma manual o en alguna font computada, a dades en un ordinador. Les avantatges que pot proporcionar aquesta tecnologia són immenses, ja que es pot usar en innumerables aplicacions que poden anar des d'ajuts per al dia a dia de les persones amb discapacitats visuals fins a una millor organització empresarial.

Un procés dirigit a la digitalització de textos és l'anomenat **Reconeixement Òptic de Caràcters**, més comunament escurçat per l'acrònim **OCR** (de l'anglès *Optical Character Recognition*). Aquest procés fa referència a la conversió mecànica o electrònica d'imatges de text mecanografiat, escrit a mà o imprès, a text que pugui ser usat per programes d'edició de text o d'altres finalitats, com cercadors o compactadors de dades. Les tècniques usades habitualment en el camp de l'OCR han anat evolucionant al llarg de la seva història, i així com hi ha documents on els resultats obtinguts són extremadament satisfactoris, n'hi ha d'altres on la transcripció del text en imatge a text editable és més dificultosa.

Quan les paraules i el fons on estan escrites són ben diferenciats, com en el cas del document mecanoscrit digitalitzat en blanc i negre que es mostra a la figura 3a, l'extracció sol ser eficaç i acurada. En canvi, en les imatges reals (imatges preses en entorns sense restriccions), el ventall en que es mostra el text és molt variable. Depenent també de com han sigut fetes i amb quin dispositiu s'han capturat, s'ha de tenir en compte que atributs com la il·luminació de la imatge i la inclinació o enquadrament de les paraules pot fer diferir molt el valor dels píxels d'una imatge a una altra encara que l'element representat sigui originalment el mateix. En la figura 3b es pot observar un clar exemple d'aquest últim cas.



(a) Document escanejat on es diferencia fàcilment el text mecanografiat en negre en contraposició amb el fons blanc.

(b) Dues imatges on es pot apreciar la dificultat afegida en les imatges provinents del món real, on s'ha de tenir en compte la forma en que s'ha obtingut la imatge, a més a més de la similitud entre el text i el fons on està escrit.

Figura 3: Mostra de diferents dificultats en l'extracció de text en imatges.
Es pot observar com clarament les imatges provinents del món real comporten una major dificultat en el seu processament i transcripció digital.

En els casos en què no es pot aplicar un OCR per la dificultat de les imatges, una alternativa és el *word spotting*. Consisteix en trobar paraules concretes en una imatge usant la aparença visual, sense reconèixer explícitament quina paraula és. Aquest és el cas del nostre projecte, on només volem localitzar certes paraules clau en imatges que tenen un fons complex.

Entre els diversos articles referents a la detecció de paraules, en podem trobar un gran nombre especialitzats en la transcripció de text manuscrit, sobretot en el cas de documents d'interès històric o científic. Aquests textos solen ser difícils de tractar amb les tècniques utilitzades normalment en l'OCR, ja que la seva digitalització acostuma a aportar imatges amb un gran alt nivell de soroll. A més a més, els documents poden provenir de diferents escriptors i llengües, pel que la forma de de la tipografia variarà inclús en un mateix document. La extracció del text inclòs en aquests documents és interessant ja que permet un millor accés a la informació que hi apareix. Així com la preservació d'aquests documents a llarg a plaç pot ser solucionada digitalitzant-los i emmagatzemant les imatges resultants en grans llibreries de dades, la seva indexació comporta un gast de recursos humans considerable i les tècniques de visió per computador poden servir de gran ajuda.

Per a buscar paraules en textos escrits a mà apareixen les tècniques de reconeixement com l'HMM (que correspon al model ocult de Márkov, de l'anglès *Hidden Markov Model*) o el DTW (de l'anglès *Dynamic Time Warping*). Aquestes ajuden a trobar seqüències similars en els textos i són usades en varietat d'articles com [10], [4] o [11]. La majoria d'articles que es poden trobar fàcilment fan referència a textos llatins, però també es pot observar

aplicacions en llengües que utilitzen altres alfabets com la japonesa en l'article [5] que es basa en un classificador DTW, o l'àrab en l'article [8] que es decanta per fer una extracció de les característiques del text usant GHT (de l'anglès *Generalized Hough Transform*) que s'adapta millor a la forma de les paraules d'aquest alfabet. En contraposició amb l'extens ús de HMM i DTW podem trobar altres articles com [7] o [9], que utilitzen altres tècniques com l'emmagatzemament de les característiques en estructures *hash* usant *Loci Features* o l'ús de codi de forma dels caràcters modificats (en anglès *Modified Character Shape Code*) per al cas específic del text en cursiva.

Un altre mètode que s'usa habitualment en els centres de recerca de diferents departaments de visió per computador arreu del món, és l'ús de descriptors de contorn (**HOG**, de l'anglès *Histogram of Oriented Gradients*) per a retratar les característiques de les porcions que formen una imatge. D'aquesta manera es divideix la imatge en cel·les de la mateixa mida, on es guarda el càlcul dels valors dels gradients d'intensitat. Això permet processar-les amb la finalitat de detectar objectes, o en aquest cas paraules, presents en elles.

Altres articles d'interès es centren també en l'extracció de paraules en imatges reals, com l'article [3] que usa descriptors HOG i un classificador NN (de l'anglès *Nearest Neighbor*) per a enfocar el problema de navegació i del dia a dia de les persones cegues o amb altres discapacitats visuals. També podem veure una altra aproximació, tant al problema del text manuscrit així com en imatges reals, en l'article [6], usant HMM, PHOC (de l'anglès *pyramidal histogram of characters*) i un classificador NN.

Per al projecte que ens incumbeix ens centrem en el mètode emprat en l'article realitzat al CVC a Barcelona (Centre de Visió per Computador) [1], adaptant-lo al nostre cas particular per a poder-lo aplicar a les imatges dels comptadors de gas. Aquest mètode utilitza descriptors HOG per a representar els documents, finestres lliscants per a localitzar les regions similars a les paraules buscades i un classificador de tipus SVM per a representar-les de forma no supervisada.

A banda de tenir accés al codi usat a l'article [1] per a poder modificar-lo i adaptar-lo al problema dels comptadors de gas, hi ha altres avantatges que ens encoratja a utilitzar-lo. Els resultats a l'hora de cercar paraules en textos manuscrits van ser bons, i s'adapta bé a les necessitats d'aquest projecte, pel que ofereix un bon inici de partença. A més a més, l'entrenament es pot dur a terme amb una única instància de la paraula a cercar, pel que

es pot fer de manera totalment no supervisada en el cas de no disposar d'un nombre més gran d'imatges d'exemple. De totes maneres es pot millorar el resultat usant-ne més d'una en l'entrenament per a poder diferenciar millor a l'hora de l'avaluació. També cal dir, que en aquest mètode s'usa una finestra lliscant per a recorre tota la imatge de manera exhaustiva, pel que ens dóna una manera eficient de buscar paraules i alhora ens permet obtenir la localització de la paraula dintre de la imatge.

A la secció 4 fem una revisió a l'article on es descriu el mètode i n'aprofundim en els detalls del seu funcionament.

3 Viabilitat del projecte

3.1 Estudi de la viabilitat

En aquesta secció s'estudia la viabilitat del projecte des del punt de vista tècnic, econòmic i legal.

Viabilitat tècnica

Per a realitzar aquest projecte cal modificar el codi en C++ i Matlab proporcionat, que va ser desenvolupat en el seu moment per a ser executat en una màquina on es feia córrer Linux. Per tant, es necessari disposar d'un ordinador amb aquest sistema operatiu o com a mínim de la família Unix, ja que part del codi està compilat per a aquest.

Matlab[12] és un llenguatge d'alt nivell i un entorn interactiu que va ser creat per la companyia MathWorks i que és usat per milions d'enginyers i científics d'arreu del món. Permet una manipulació fàcil de matrius i vectors, i dibuixar-ne gràfics amb les seves dades. Està disponible tant per Linux, com Mac OS X i Windows, pel que s'adapta perfectament a les necessitats de cada tipus d'usuari.

Per a aquest projecte, els ordinadors als quals podem accedir són els següents:

- Ordinador de taula:
  - Sistema Operatiu: Ubuntu 14.04, 64 bit.
  - Processador: Intel(R) Core(TM) i5-650 CPU @ 3.20GHz (2 cores).
  - Memòria: 4 GB 1333 MHz DDR.
- Portàtil A:
  - Sistema Operatiu: OS X Yosemite 10.10.5, 64 bit.
  - Processador: Intel(R) Core(TM) i5-3427U CPU @ 1.80GHz (2 cores).
  - Memòria: 4 GB 1600 MHz DDR3.

- Portàtil B:
  - Sistema Operatiu: Ubuntu 14.04, 64 bit.
  - Processador: Intel(R) Core(TM) i7-4702MQ CPU @ 2.20GHz (4 cores).
  - Memòria: 8 GB 1600 MHz DDR3.

Per altra banda, cal disposar d'imatges dels comptadors de gas per a poder fer l'aprenentatge i les proves corresponents. Aquestes imatges són cedides per l'Ernest Valveny, tutor del projecte.

Viabilitat econòmica

Matlab té una llicència de programari propietari, pel que per a poder usar-lo s'han de comprar els drets a fer-ho. L'empresa Mathworks facilita diverses llicències segons l'ús que se li vulgui donar. Com el projecte està en l'àmbit acadèmic d'estudiant, es pot adquirir la llicència personal d'estudiant (*Individual Student License*), que té un cost de 35€. En el cas de que fos usat per a per a una altra finalitat, el cost varia d'entre 105€ a 2000€ (depenent de si la llicència és d'ús acadèmic, familiar o empresarial).

Finalment, s'ha de tenir en compte que es tracta d'un projecte de final de carrera, pel que no hi ha un cost monetari real de les hores dedicades. Tot i així es pot fer una aproximació considerant el sou d'un enginyer informàtic recentment titulat. Segons el Ministeri de Treball i Seguretat Social[2], el sou mínim per a un Enginyer o Llicenciat és de 6.37€/hora. Tenint en compte que el projecte de final de carrera equival a unes 375 hores (corresponents als 15 crèdits assignats a l'assignatura), això comporta un cost de 2388.75€.

Viabilitat legal

En la part legal, hem de tenir en compte que l'ús de la llicència d'estudiant per al Matlab permet usar-lo únicament per l'estudiant que l'ha comprat en un àmbit acadèmic, sense cap fi governamental, comercial o de qualsevol ús diferent en altres organitzacions.

La llicència està disponible per a usar Matlab en qualsevol plataforma, ja sigui Windows, Linux, o Macintosh.

3.2 Planificació del projecte

En aquesta planificació es veuen reflectides les tasques que al inici del projecte es van estimar necessàries per a assolir-ne els objectius 1.2. En l'apartat anterior, hem considerat que els 15 crèdits del projecte de final de carrera equivalen a 375 hores, que s'han de repartir per les diverses tasques a realitzar. Per a això, considerant una mitjana de treball d'unes 6 hores el dia, la totalitat d'aquest es podria fer en uns 63 dies laborables.

Les tasques van ser dividides en els següents punts:

- Tasca 1 (60 hores): Estudi dels comptadors, codi usat i altres
  - Tasca 1.1 (10h): Estudi dels diferents tipus de comptadors i com es diferencien
  - Tasca 1.2 (30h): Estudi del mètode i codi usat a l'article [1]
  - Tasca 1.3 (20h): Altres estudis necessaris que apareguin
- Tasca 2 (187 hores): Implementació dels canvis necessaris per a modificar el mètode
  - Tasca 2.1 (80h): Modificació del codi en els llocs pertinents
  - Tasca 2.2 (70h): Proves necessàries per a validar el funcionament
  - Tasca 2.3 (37h): Altres implementacions necessàries que puguin sorgir
- Tasca 3 (128 hores): Redacció dels documents associats al projecte
  - Tasca 3.1 (8h): Informe previ
  - Tasca 3.2 (95h): Memòria del projecte
  - Tasca 3.3 (25h): Preparació de la presentació

Aquestes hores estaven distribuïdes segons la planificació que s'observa en el diagrama de Gantt de l'annex A.

4 Fonaments teòrics

En la secció 2 hem introduït el mètode que s'usa a l'article del qual partim, però en aquesta secció explicarem de forma breu les parts més destacades.

4.1 Mètode usat a l'article

Aquest mètode està pensat per a que donat un conjunt d'imatges de manuscrits digitalitzats i una imatge exemple de la paraula a buscar, identifiqui possibles seccions en les imatges d'aquest conjunt on es puguin trobar instàncies d'aquesta mateixa paraula.

Està focalitzat en l'aprenentatge no supervisat, on no hi ha etiquetatge de dades disponible per a realitzar l'entrenament. Tradicionalment es segmenta el document de text segons les paraules candidates, que després es compararan amb la representació de la paraula buscada (usant mètodes com DTW i HMM), però en l'article es va optar per a usar una representació on la llargada està fixada. Això fa que la comparació sigui més ràpida i que es pugui utilitzar una finestra lliscant per a recórrer tot el text de forma exhaustiva.

Hi ha tres punts a destacar en el mètode utilitzat:

- **Representació de la informació de les imatges**

  En aquest cas es tracta amb text manuscrit, pel que hi ha molta variabilitat en una mateixa paraula depenent de la forma o estil de l'escriptor. Això fa que els descriptors que s'usin hagin de ser ràpids de calcular i comparar per a poder usar la cerca amb finestra lliscant, sobretot en el cas de grans conjunts de dades. Per això aquí s'opta per usar els descriptors **HOG**, ja que encara que no arribi a tenir una representació de les característiques de la imatge tan acurada com amb d'altres mètodes, té un bon equilibri entre precisió i velocitat a l'hora de realitzar comparacions.

  Per tant, per a representar les imatges (tant dels documents com de la paraula a cercar) s'usa una graella d'histogrames HOG, on cada element d'aquesta conté el mateix nombre de píxels. Per a cada cel·la s'usen histogrames de HOG de 31 dimensions, on es codifica el gradient local de la imatge usant diferents mesures de les característiques. La graella es recorre usant una finestra lliscant que inclou tantes

cel·les com el nombre usat per a representar la paraula clau. La finestra es mourà per la imatge desplaçant-se una o vàries cel·les en cada pas, segons la configuració inicial del programa, fins que hagi recorregut tota la imatge.

Com que la mida de la finestra lliscant ve donada per la representació de la paraula buscada, tenim que el nombre de cel·les que s'usi, i per tant la dimensió total del descriptor, depèn de la mida de la imatge de la paraula que es cerca. Si anomenem H al nombre de files i W al nombre de columnes del descriptor HOG de la paraula clau, aleshores la finestra constarà de $H \times W$ cel·les.

- **Entrenament i avaluació del mètode**

  Per a calcular la puntuació d'una regió x de $H \times W$ cel·les de la imatge del document respecte la paraula cercada q, es calcula la seva convolució `sim(q,x)`. Això es pot entendre també com el producte escalar del vector normalitzat de HOGs de x amb el vector normalitzat de q. Movent la finestra per tota la imatge es pot calcular la semblança de cada regió amb la paraula clau, per després ordenar aquests valors i poder quedar-se amb un conjunt definit de les millors solucions.

  Els resultats poden donar finestres solapades, pel que per arreglar-ho i no obtenir dos o més regions per a una mateixa possible solució s'aplica el **NMS** (de l'anglès *Non-Maximum Suppression*). Aquest és emprat per eliminar els solapaments de les finestres resultants amb unions més grans al 20%.

  Per a maximitzar la puntuació de les regions significatives per a la cerca de la paraula clau, s'utilitza el classificador **Exemplar SVM**. Amb això s'aconsegueix donar diferent valors a les regions que resulten positives o negatives per a la cerca de la paraula. Per a fer-ho, en el moment de crear el descriptor de la imatge exemple de la paraula clau, aquesta es modifica creant diverses imatges de la paraula a partir de la imatge d'on prové l'exemple de manera que la paraula quedi descentrada. Així s'obté un conjunt d'imatges que serviran com a exemples positius i a més a més es dóna més flexibilitat a l'hora de comparar regions d'una imatge amb la finestra lliscant, ja que la majoria de cops no es podrà tenir la paraula clau centrada en els textos a processar. Per al conjunt d'exemples negatius que s'usaran en el classificador, s'agafen regions a l'atzar de la mida dels exemples positius de la imatge original d'on prové l'exemple

inicial. S'ha de tenir en compte que amb el retall aleatori aquests exemples negatius poden contindre o no paraules, que poden coincidir amb part de la paraula a cercar.

Al aprendre l'SVM, es dóna diferents pesos a les dimensions de les regions dels descriptors segons la rellevància respecte la paraula buscada. Amb això aconseguim una nova representació de q, que donarà puntuacions molt altes a les zones rellevants al fer el producte escalar amb el vector normalitzat de x i puntuacions molt baixes a les que no ho siguin. També tenim l'avantatge que d'aquesta manera al cercar la imatge amb la finestra lliscant no ens cal que la paraula estigui centrada per a obtenir bones puntuacions, ja que els exemples positius que hem creat no són tots centrats.

- **Compressió de les dades**

  Inicialment, els descriptors de HOGs són reduïts en dimensió al aplicar-los l'anàlisi de components principals (**PCA**, de l'anglès *Principal Component Analysis*). Aquest anàlisi converteix un conjunt de variables possiblement correlacionades a un conjunt de valors de variables no correlacionades linealment, pel que els descriptors passen de 31 a 24 dimensions per histograma.

  A més a més d'això, en els experiments realitzats a l'article, es van comprimir els descriptors HOGs usant **PQ** (de l'anglès, *Product Quantization*). Aquesta compressió permet reduir el cost en memòria i agilitzar el càlcul de les comparacions usant la finestra lliscant. Per altra banda, això comporta fer una aproximació a l'hora de realitzar certs càlculs pel que hi ha una lleugera pèrdua d'informació. Malgrat això, els experiments realitzats mostren que la precisió en els resultats es manté alta, mentre que hi ha un significatiu guany en la reducció de l'espai usat.

Tot i que al principi hem mencionat que es requereix d'una única instància d'exemple de la paraula a buscar, es pot usar múltiples imatges d'aquesta. Amb més exemples es pot ampliar el nombre del conjunt d'exemples positius i negatius, donant-li d'aquesta manera més informació al mètode per a fer l'entrenament.

4.2 Adaptació per al cas particular del projecte

Les imatges de comptadors de gas contenen text imprès en diverses fonts i poden contindre diferents graus de nitidesa, inclinació i contrast depenent de com s'hagin realitzat la fotografies. La variabilitat que aporten les imatges del projecte és similar al cas del que es parteix a l'article, encara que aquesta variabilitat tingui diferents orígens. És per això que l'ús de descriptors de HOG i finestra lliscant de l'article és adient. La representació que explicada en la secció 4.1 anterior és la mateixa que usem en aquest projecte. Les imatges estan dividides en cel·les i la finestra lliscant ve determinada pel nombre de cel·les que ocupa la paraula clau.

En aquest cas, per a l'aprenentatge usant el classificador *Exemplar SVM*, partim directament de múltiples exemples de la paraula clau que volem trobar, per a poder incloure la possible versatilitat de les imatges en el moment de l'aprenentatge. De cada instància de la paraula a cercar creem els conjunts d'exemples positius i negatius, que després s'escalaran a partir de la mitja aritmètica de les amplades i les alçades de les imatges d'exemple positives. Així creem un sol conjunt d'exemples positius i un d'exemples negatius amb el mateix nombre de cel·les per a cada element d'aquests .

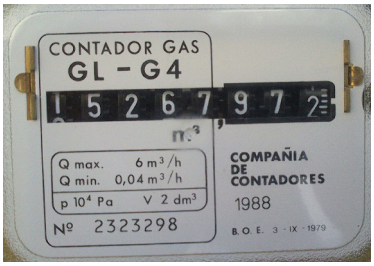
Quan es treballa amb textos manuscrits com els de l'article, depenent de quina paraula es busqui segurament es repetirà més d'un cop en un mateix document. Això comporta una diferència amb aquest projecte, ja que en una imatge d'un comptador de gas com a molt hi haurà una instància de la paraula a cercar si la imatge correspon al comptador cercat o a un comptador de la mateixa marca, i cap si es tracta d'un comptador completament diferent. Per tant, quan obtenim el conjunt de finestres trobades per a una imatge només ens fa falta quedar-nos amb la regió amb millor puntuació per a aquell model.

Finalment, cal mencionar que encara que als descriptors es redueixin quan apliquem l'anàlisi de components principals, no afegim en aquest cas la compressió amb PQ. D'aquesta manera encara que la rapidesa a l'hora de fer els càlculs pogués augmentar, ens assegurem de mantenir tota la informació possible de la que disposem. A més, donat que no treballem amb grans volums de dades, l'ocupació de memòria i la velocitat no ens generen problemes.

5 Proves i resultats

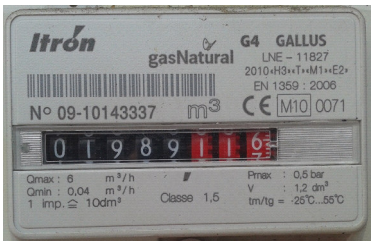
En aquesta secció explicarem les diverses proves realitzades aplicant el mètode explicat a la secció 4.

Per a fer els experiments hem tingut en compte 9 models diferents de comptadors de gas, que corresponen a aquells models dels quals disposàvem de més imatges. En la figura 4 podem veure l'aparença d'aquests models.



(a) Comptador 58

(b) Comptador 68

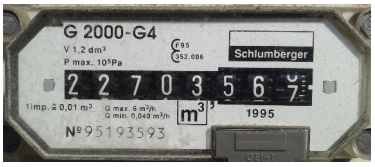


(c) Comptador 70

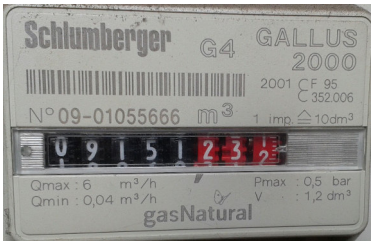
(d) Comptador 72

(e) Comptador 75

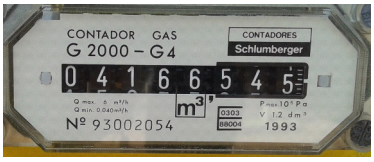
(f) Comptador 86



(g) Comptador 89



(h) Comptador 90



(i) Comptador 94

Figura 4: Mostra dels models usats en les proves realitzades.

De cada un d'aquests models de comptador de gas disposem d'entre 30 i 40 imatges per a realitzar l'entrenament del mètode, i unes altres 30 o 40 imatges diferents per a fer l'avaluació.

5.1 Proves realitzades

Els experiments que hem dut a terme per a comprovar l'efectivitat del mètode aplicant-lo a les imatges dels comptadors de gas, s'han basat en modificar diversos paràmetres que s'usen dintre del programa per veure quins canvis provoca en els resultats.

Els paràmetres a modificar es poden dividir en tres grups: tipus d'imatge, número d'imatges d'exemple usades a l'entrenament del mètode i número de paraules claus associades a cada model de comptador de gas.

Veiem en què consisteixen cada un d'aquests canvis:

- Tipus d'imatge

  Les imatges dels comptadors de gas de les que disposem estan fetes a color, pel que cada píxel d'una imatge està definit per tres valors que ens indiquen el grau de vermell, verd i blau que conté. En un principi el programa original (usat a l'article [1]) es va utilitzar per a tractar imatges de documents escanejats en escala de grisos, on per cada píxel d'una imatge només hi ha un valor associat que representa el nivell on troba en el rang entre blanc i negre. No obstant, això no implica que no estigui adaptat a poder usar imatges a color, ja que en aquest cas el programa agafa un dels tres valors associats a cada píxel com a representant d'aquest.

  Per a provar com s'adapta el mètode als diferents tipus d'imatge, hem realitzat proves partint de les imatges a color i altres partint de les mateixes imatges però en escala de grisos. A més a més, també hem generat el complement de les imatges en escala de grisos per a veure si afecta d'alguna manera als resultats. Podem veure un exemple dels tres tipus d'imatge dels que partim en la figura 5.

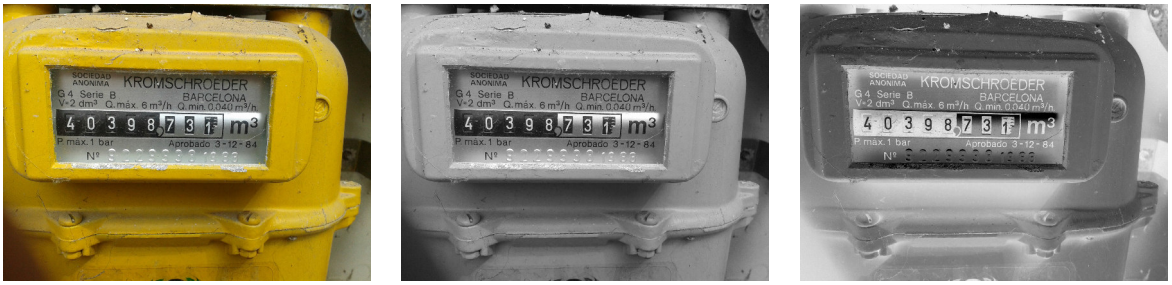
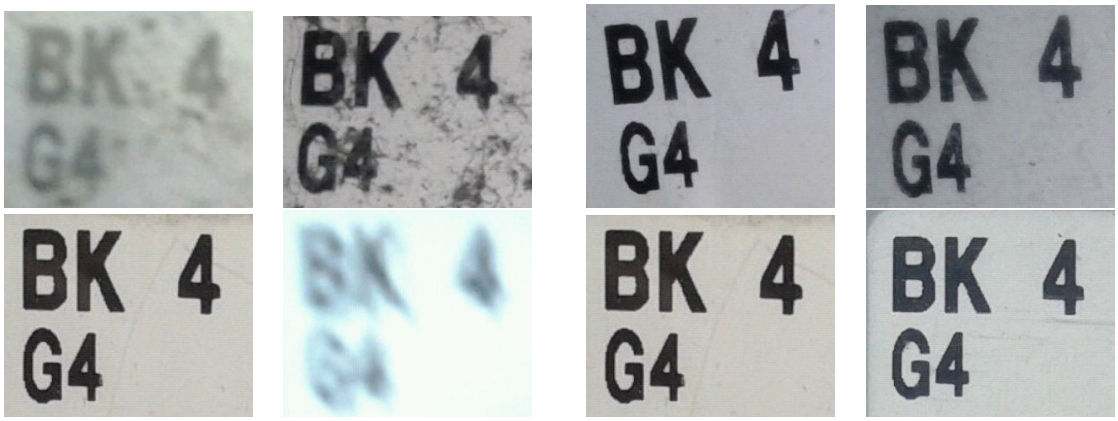


Figura 5: Mostra dels diferents tipus d'imatge en les proves.
Podem veure la imatge original a color, la seva representació en escala de grisos i la imatge del complement d'aquesta última.

- Número d'imatges d'exemple

  Per a fer l'entrenament del mètode, necessitem les imatges d'exemple positives i negatives, tal com hem explicat a la secció 4.1. Variant el nombre d'elements d'aquests dos conjunts d'imatges podem veure la influència que tenen sobre els resultats del mètode, pel que hem fet diverses proves modificant-los.

  Les imatges dels comptadors de gas de les que disposem són imatges que provenen del món real, pel que no estan fetes per a que estiguin totes a la mateixa escala, igual d'enfocades o amb el mateix impacte de llum. És per això que pel conjunt d'imatges d'exemple positives tenim dos casos possibles (figura 6). La primera opció és extreure els exemples de les paraules clau de totes les imatges que disposem per a fer l'entrenament, per a cada un dels models de comptador. Per altra banda, podem centrar-nos només en aquelles imatges que estiguin més nítides, excloent totes les que es vegin borroses o continguin massa soroll. En aquest últim cas, el conjunt d'exemples positius és veu minvat i amb menys variabilitat, però resulta en imatges ben definides, amb canvis de contorn més marcats. Hem de recordar que en ambdós casos, de cada imatge d'exemple aconseguim 9 imatges positives al generar-ne de noves descentrant la paraula clau, pel que el nombre total d'exemples positius hauria de ser sempre suficient.



(a) Imatges positives extretes de tot el conjunt d'imatges d'entrenament

(b) Imatges positives extretes només d'aquelles imatges d'entrenament ben definides.

Figura 6: Mostra dels dos casos d'imatges d'exemple positives.

En el cas del conjunt d'exemples negatius, podem variar el nombre de retalls aleatoris que realitzem i d'on provenen. En les proves, hem començat agafant 10 imatges negatives per cada positiva, i hem anat variant aquest nombre fins a arribar a 50 de negatives per cada una de positiva. Pel que fa a l'origen d'aquests retalls, també realitzem dos tipus de prova. Per un costat agafem els exemples negatius aleatòriament sobre les imatges d'un mateix model de comptador i, per altra banda, les extraiem a partir de les imatges de la resta de models comptadors. En el primer cas hem de dir que a més a més excloem la paraula clau de la zona de retalls aleatoris, mentre que en el segon cas podem agafar qualsevol zona de la imatge.

- Número de paraules clau

  Finalment l'últim paràmetre que hem modificat és el número de paraules que usem per a identificar un model de comptador. Per un costat, reduïm a una sola paraula per comptador per a diferenciar-los, i per l'altre, ho comparem a agafar-ne com a mínim dos o més paraules.

  En el cas en que tenim més d'una paraula per identificar el model de comptador també ens sorgeixen més maneres de determinar com fem aquesta identificació. En aquest cas fem servir quatre opcions diferents per a declarar una imatge com a pertanyent a un model determinat de comptador:

- fent que com a mínim es necessiti que contingui una paraula clau de les que defineixen el model,
- que en contingui com a mínim dues,
- que hagi de contindre totes les paraules que el defineixen, o
- que l'acceptem mentre només li falti una paraula clau.

De vegades, quan tinguem un model definit per 2 o 3 paraules clau, algunes d'aquestes opcions que avaluem es veuran repetides.

Per a cada prova hem anat avaluant els resultats del mètode a mesura que augmentàvem el valor de *cutoff*, és a dir, el valor de puntuació mínima en els resultats d'una imatge per declarar que una paraula clau està continguda en aquella imatge.

5.2 Resultats obtinguts

Com que als experiments realitzats hem usat imatges de les quals sabem a quin model de comptador de gas pertanyen, podem usar els resultats de les diverses proves per a obtenir quatre valors que ens poden proporcionar molta informació sobre el mètode empleat.

		Ground Truth	
		Positives	Negatives
Results	Positives	**TP** True Positives	**FP** False Positives
	Negatives	**FN** False Negatives	**TN** True Negatives

Taula 1: Relació entre els resultats obtinguts aplicant el mètode i els valors de *Ground Truth*, que corresponen a l'origen real de cada imatge de comptador

Els quatre valors que es mostren a la taula 1 fan referència a la classificació obtinguda en els resultats després d'aplicar el mètode. Per un costat tenim les imatges ben classificades, que serien les veritables positives (**TP**, de l'anglès *True Positives*) i les veritables negatives (**TN**, de l'anglès *True Negatives*), i per l'altre costat les classificades errròniament, que serien

les falses positives (**FP**, de l'anglès *False Positives*) i les falses negatives (**FN**, de l'anglès *False Negatives*). El primer conjunt està format per les imatges identificades correctament com a pertanyents o no pertanyents a un model de comptador determinat. En canvi, el segon està format per les imatges identificades, de forma incorrecta, com a pertanyents a un model quan no hi pertanyen o com a no pertanyents quan sí que hi pertanyen.

Aquests valors ens poden ajudar a entendre l'efectivitat del mètode en les diverses proves que hem realitzat, i combinant-los podem obtenir quatre proporcions que ens informen d'una manera més clara com són aquests resultats.

Una d'elles és l'anomenat ***recall*** en anglès (que també es coneix com a **sensitivitat** i **taxa de veritables positius**, o *sentivity* i *true positive rate* en anglès). Aquest fa referència al percentatge d'imatges identificades correctament com a procedents del model al que pertanyen dintre de totes les imatges que disposem d'aquell model en concret:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Per a valors molt baixos de *cutoff*, tindrem que el valor de *recall* és del 100% o molt proper, ja que tots els resultats seran positius i no tindrem cap imatge amb un fals negatiu (el conjunt de FN estarà buit). A partir d'un punt, començarà a baixar, fins que ja no s'identifiqui cap imatge positivament per a cap model, moment en el qual el valor de *recall* esdevindrà 0.

Un altre valor que es considera sovint conjuntament amb el *recall* és la **precisió** (anomenada *precision* en anglès i sovint també com *positive predictive value*, en ambients com la medicina). Aquest valor fa referència al percentatge d'imatges identificades correctament com a procedents del model al que pertanyen dintre de totes les imatges identificades per a aquell model:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Amb un valor de *cutoff* baix, el valor de precisió és també baix però no arriba al 0%, ja que el valor de TP serà màxim encara que en el quocient de la precisió apareguin totes les imatges de les que disposem.

Amb la precisió només es tenen en compte els resultats que donen positiu per a algun

model de comptador, pel que quan el *cutoff* sigui massa alt per a acceptar-ne cap com vàlid, el valor de precisió deixarà de tenir sentit perquè no tindrem cap resultat per a poder fer els càlculs. És per això que a partir d'un valor de *cutoff* la corba de la precisió en les gràfiques desapareixerà (que coincideix en el moment en que el *recall* passa a ser 0, és a dir, quan ja no hi ha cap TP, i encara és menys esperable cap FP).

Un altre valor que també ens aporta visió general de com està resultant el mètode és l'***accuracity***, que ens dóna la proporció d'imatges ben classificades dintre de totes les imatges que s'han processat:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

El valor de l'*accuracity* és més gran amb valors de *cutoff* alts, ja que donada una paraula clau per a un model de comptador, tenim més imatges a les que no hi apareix que a les que hagi d'aparèixer. Un bon resultat seria que tant la precisió com el *recall* fossin alts en el moment en que les seves corbes es tallen a la gràfica, ja que aquest és el punt on el mètode ens ha aportat més imatges classificades correctament en el model que els hi pertoca alhora que s'equivocava menys cops en aquesta classificació. Idealment aquest serà el punt on hi hagi menys imatges classificades erròniament, que fa referència al pic més alt del valor de l'*accuracity*. Per tant, ens interessa que la corba generada pel *recall* tingui una baixada lenta, i la corba de la precisió pugui ràpidament.

Finalment l'últim valor que també és interessant d'obtenir és la **taxa de veritables negatius** (coneguda també per **especificitat**, o *specificity* i *true negative rate* en anglès). Mentre que el *recall* ens dóna informació sobre el percentatge d'imatges que s'han classificat correctament en el model d'on pertanyen, la taxa de veritables negatius té en compte la part negativa de la classificació. És a dir, aquesta taxa ens informa del nombre d'imatges classificades com a no pertanyents a un model de comptador quan realment no hi pertanyen, dintre de totes les imatges de les que disposem que no formen part d'aquell model:

$$\text{True Negative Rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Així com la corba del *recall* partia de valors propers al 100% quan el *cutoff* és baix, i anava disminuint a mesura que augmentaven aquest valor de tall, la taxa de veritables

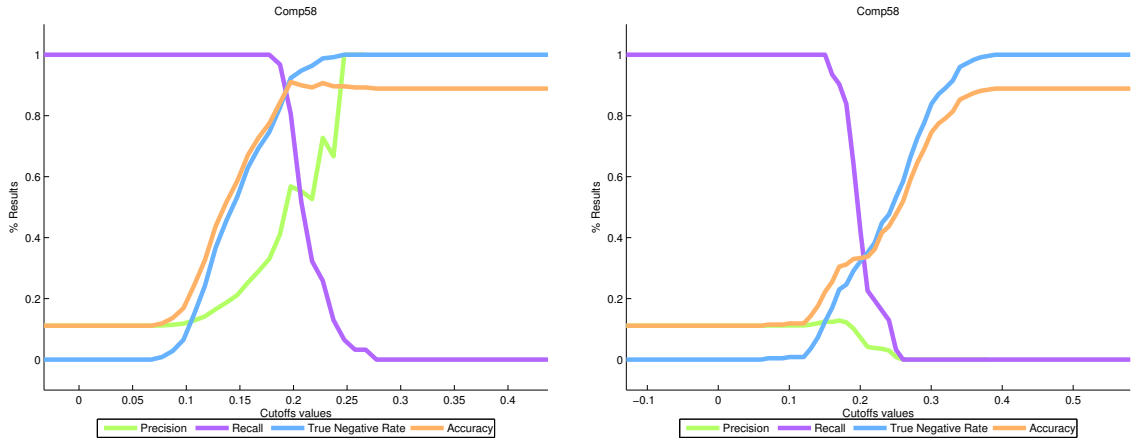
negatius té el comportament contrari. De manera que la seva corba serà propera al 100% quan el *cutoff* sigui alt, punt on el nombre d'imatges catalogades com a negatives de forma correcta és màxim.

Podem veure els resultats de totes les proves que hem realitzat en l'annex B. S'hi mostren totes les gràfiques generades com a resultat d'avaluar cada comptador amb el mètode, usant diferents valors de *cutoff*. En elles apareixen les corbes resultants de fer la mitja de cada un dels quatre valors que hem esmentat: *recall*, *precision*, *accuracity* i *true negative rate*.

Amb els resultats que hem obtingut de les proves realitzades no podem concloure que hi hagi una configuració concreta en els paràmetres del mètode, en front a una configuració diferent, que faci que el mètode funcioni clarament millor a l'hora d'identificar qualsevol model de comptador. Tot i així, sí que hem pogut constatar que tant el rendiment com els paràmetres òptims depenen del model de comptador, pel que a continuació ens centrarem en tres dels resultats més significatius que hem trobat: els resultats per als models de comptador 58, 70 i 94.

Començant pels comptadors 58 i 70, podem veure que són models de comptador que podem diferenciar bé de la resta si ens fixem en les paraules claus que els defineixen (paraules com *COMPAÑÍA DE CONTADORES* en el cas del comptador 58, i *Itrón* en el cas del comptador 70, figura 4).

En el cas del model de comptador 58 veiem que els resultats són molt millors quan usem imatges en escala de grisos que a color, i quan el conjunt d'imatges usat per a generar les imatges d'exemples positius per a l'entrenament està format només per imatges ben definides (figura 7).

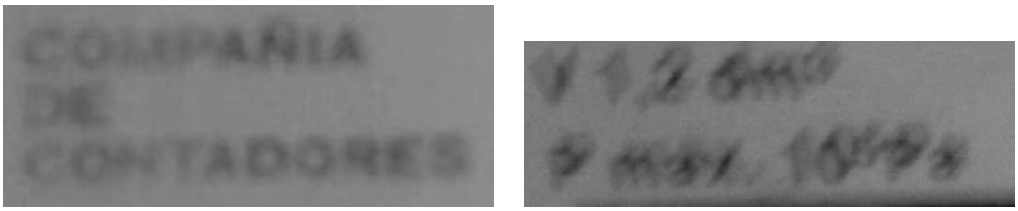


(a) Amb imatges en escala de grisos i usant només les imatges ben definides per a crear els exemples positius.

(b) Amb imatges a color i amb tot el conjunt d'imatges d'exemples positius.

Figura 7: Resultats amb el model de comptador 58.

Podem veure que en el cas d'incloure totes les imatges per a fer els exemples positius, serà més difícil diferenciar aquesta paraula que el distingeix dels altres al semblar-se a altres paraules clau de diferents models quan no les imatges no estan ben definides (figura 8).



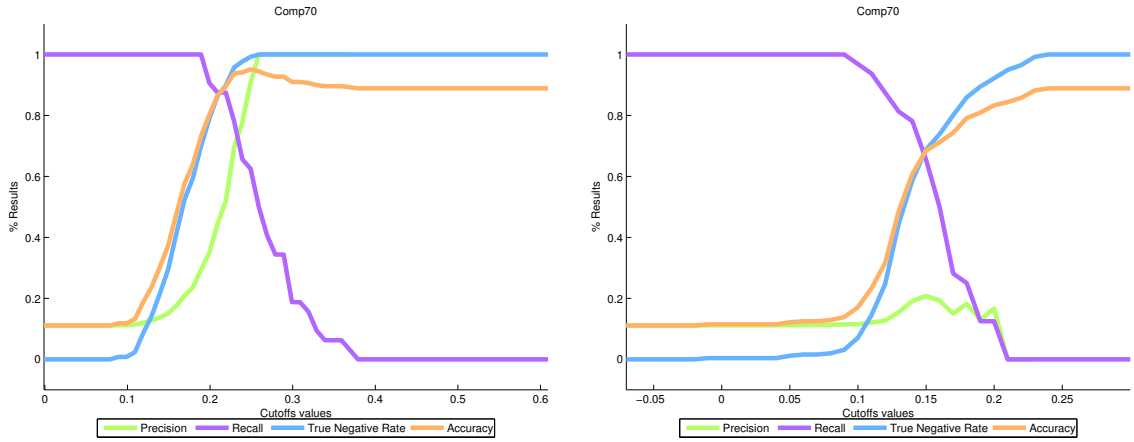
(a) Paraula clau del model de comptador 58 d'una imatge no nítida.

(b) Paraula procedent d'una imatge del model de comptador 89.

Figura 8: Utilitzant el mètode no podem diferenciar fàcilment entre les paraules dels dos models de comptador.

Pel model de comptador 70 també veiem que es detecta millor la paraula que defineix el model quan es tracta amb imatges del complement de les imatges en escala de grisos que en el cas d'usar imatges a color. En canvi en aquest cas, hi ha millor resultat quan les imatges d'exemples positius són extretes de tot el conjunt d'imatges en compte de tenir en compte només aquelles més nítides (figura 9). Això pot ser fàcilment degut a que el contorn de la paraula clau *Itrón* és poc típic i al afegir més imatges d'exemple millorem la

varietat de l'entrenament.

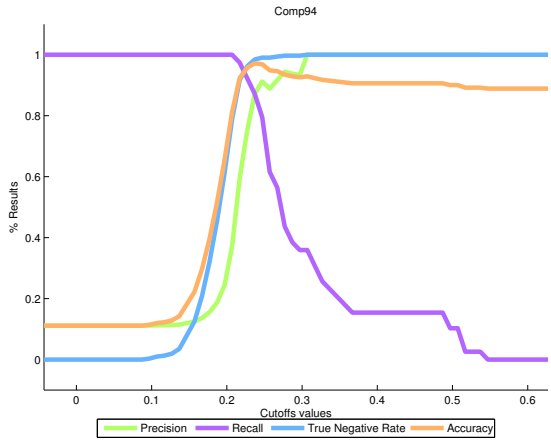


(a) Amb imatges en complement de les imatges en escala de grisos i tenint en compte totes les imatges d'exemple positives.

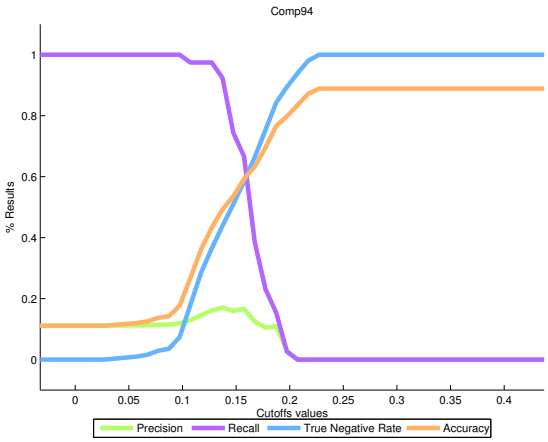
(b) Amb imatges a color i només usant imatges d'exemple positives nítides.

Figura 9: Resultats amb el model de comptador 70.

Finalment, el cas del model de comptador 94 és un bon exemple per mostrar que amb alguns comptadors necessitem més d'una paraula clau per a definir-los de forma que el mètode els pugui identificar correctament. L'aparença d'aquest model és molt semblant a la del model 89, pel que al usar com a mínim dos paraules clau millorem molt la seva identificació (figura 10).



(a) Fent que com a mínim es necessitin dos paraules en una imatge per a classificar-la com pertanyent a aquest comptador.



(b) Necessitant només que contingui una paraula per a classificar-la com a pertanyent a aquest model.

Figura 10: Resultats amb el model de comptador 94.

5.3 Modificacions a la planificació inicial

En aquesta secció veiem com s'ha vist modificat el temps emprat per a cada tasca durant el transcurs del projecte respecte a la planificació inicial que es va fer.

En els punts següents veim els canvis que s'han produït, comparant en gris les hores que es van estimar que s'usarien per a la tasca amb les que s'han destinat realment a aquesta:

- Tasca 1 (60 hores → 67 hores): Estudi dels comptadors, codi usat i altres

  Podem veure que el temps necessari per a estudiar els diferents models de comptadors, el mètode i codi usat a l'article ha sigut menys de l'estimat, mentre que s'ha necessitat més temps per a altres estudis. En aquests estudis s'inclou la recerca per a solucionar els diversos problemes d'execució del programa en les diferents màquines usades i per a la recerca de certes funcions que s'usen dintre de Matlab.

  - Tasca 1.1 (10h → 4h): Estudi dels diferents tipus de comptadors i com es diferencien
  - Tasca 1.2 (30h → 20h): Estudi del mètode i codi usat a l'article [1]
  - Tasca 1.3 (20h → 43h): Altres estudis necessaris que apareguin

- Tasca 2 (187 hores → 215 hores): Implementació dels canvis necessaris per a modificar el mètode

  El canvi significatiu en el temps dedicat a aquestes tasques es troba en la tasca 2.3. Això és degut a que durant un període de temps només es va poder usar l'ordinador amb el sistema operatiu OS X Yosemite, i això va fer que certes parts del programa fallessin al executar-se, al estar compilades específicament per a Linux. Això va requerir diverses instal·lacions de paquets i configuracions noves per a poder seguir amb el treball, pel que el temps dedicat a aquesta tasca es va veure incrementat.

  - Tasca 2.1 (80h → 60h): Modificació del codi en els llocs pertinents
  - Tasca 2.2 (70h → 80h): Proves necessàries per a validar el funcionament
  - Tasca 2.3 (37h → 75h): Altres implementacions necessàries que puguin sorgir

- Tasca 3 (128 hores → 113 hores): Redacció dels documents associats al projecte
  - Tasca 3.1 (8h → 8h): Informe previ
  - Tasca 3.2 (95h → 90h): Memòria del projecte
  - Tasca 3.3 (25h → 15h): Preparació de la presentació

Aquestes hores es poden comparar amb distribuïdes segons la planificació inicial que s'observa en el diagrama de Gantt de l'annex A.

6 Conclusions

Després de realitzar aquest projecte de final de carrera, podem afirmar que s'ha acomplert amb l'objectiu i subobjectius esmentats a la secció 1.2.

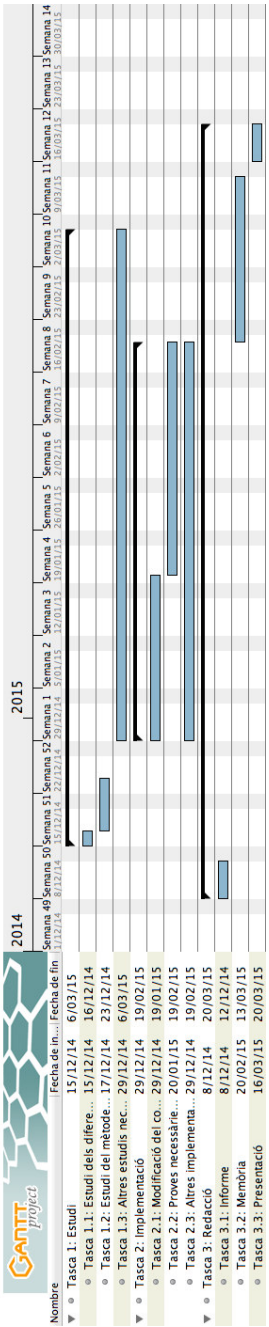
Hem sigut capaços de diferenciar les diverses característiques que identifiquen cada tipus de model de comptador, per així poder adaptar de forma satisfactòria el mètode usat a l'article [1] a la cerca de paraules clau en imatges de comptador de gas. Amb els experiments que hem realitzat un cop feta l'adaptació del codi, hem vist que depenent del model de comptador és millor usar uns paràmetres que d'altres i per així poder-lo identificar millor.

Ja que hem sigut capaços d'identificar diversos models de comptadors en les proves realitzades, podem dir que l'objectiu principal del projecte s'ha aconseguit. El que s'ha de tenir en compte és que no hi ha una configuració general que funcioni per a tots els tipus de models de comptador. El més adient és definir els models segons els paràmetres utilitzats en les proves amb millor resultats per a cada un, sense intentar definir uns paràmetres estàndards per a tots alhora.

Personalment he gaudit al realitzar aquest projecte, ja que el processament d'imatges i els mètodes d'aprenentatge són temes que m'agraden molt. A més, he pogut posar en pràctica la teoria apresa durant la carrera i m'ha donat una visió d'una aplicació real dels coneixements adquirits a la universitat. De totes maneres, em sembla oportú mencionar que la part que més m'ha frustrat ha sigut haver de tractar amb els problemes d'adaptar el codi d'un sistema operatiu a un altre. Malgrat això he pogut solucionar els problemes que han anat apareixent.

Annexos

Annex A Diagrama de Gantt



Annex B Resultats obtinguts

En aquest annex mostrem tots els resultats obtinguts segons el tipus d'imatge que s'ha usat per a generar els càlculs i resultats: imatges a color, imatges en escala de grisos o imatges del complement de les imatges en escala de grisos.

Per cada un d'aquests tipus d'imatge, podem dividir els resultats segons si s'ha usa una paraula o vàries per al definir cada model de comptador de gas (que denominarem com a opció *One word* i opció *Many words*), i segons el conjunt triat per a generar els exemples positius de les paraules claus, és a dir, agafant totes les imatges disponibles o només aquelles que estan ben definides (que denominarem com a conjunt *All images* i conjunt *Clear images*).

Hem de tenir en compte que en les proves que s'usa l'opció de *One word*, només tenim una manera d'acceptar una imatge com a procedent d'un model concret: quan conté la paraula clau que el defineix. En canvi, quan tenim més d'una paraula per a definir un model (usant l'opció *Many words*), les opcions per a acceptar una imatge en un model concret són les següents:

- Opció *One word minimum*: fent que com a mínim es necessiti que contingui una paraula clau de les que defineixen el model,
- Opció *Two words minimum*: que en contingui com a mínim dues,
- Opció *All words*: que hagi de contindre totes les paraules que el defineixen, o
- Opció *All but one words*: que l'acceptem mentre només li falti una paraula clau.

A més a més, per cada podem usar 10 imatges d'exemple negatives per cada imatge d'exemple positiva, o 50 imatges d'exemple negatives per cada imatge d'exemple positiva. I aquestes imatges poden provenir de retalls aleatoris dintre del mateix model de comptador (que anomenarem com a opció Same model), o provenir de retalls aleatoris dintre de les imatges de la resta de models (que anomenarem com a opció Different model).

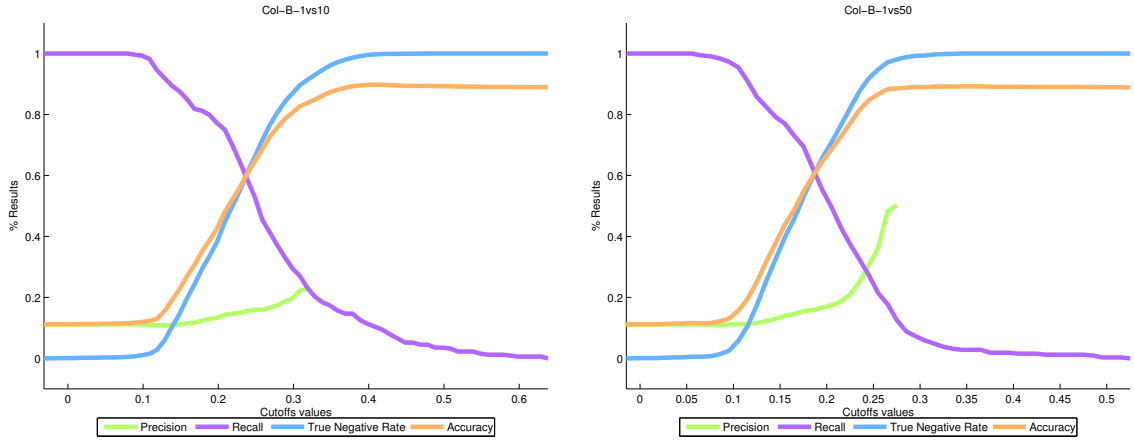
En les figures dels següents annexos podem veure els resultats per a cada una de les proves realitzades usant cada tipus d'imatge. En cada gràfica es mostres les corbes resultants de fer la mitja de cada un dels quatre valors *recall*, *precision*, *accuracity* i *true negative rate*,

al avaluar cada comptador amb el mètode usant diferents valors de *cutoff*.

Resultats obtinguts amb les imatges a color

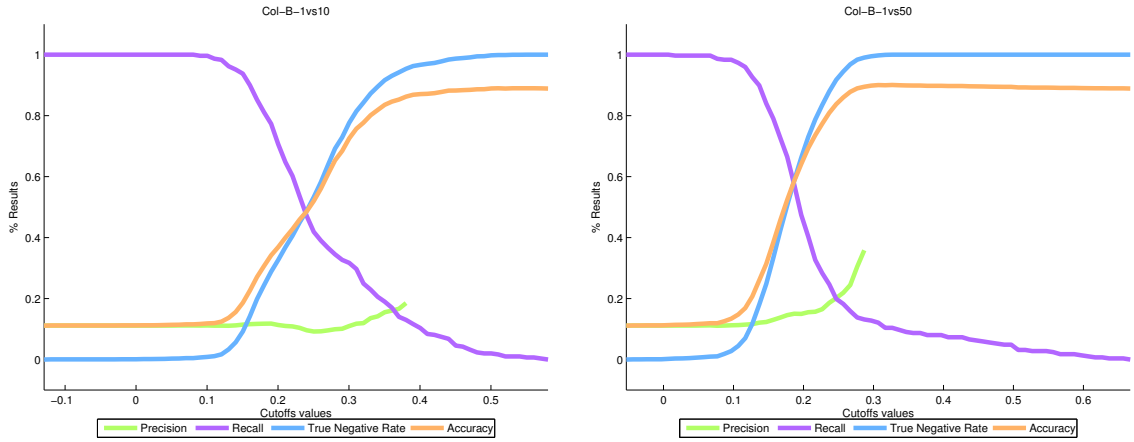
Aquí mostrem tots els resultats obtinguts utilitzant les imatges a color:

One word + All images



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

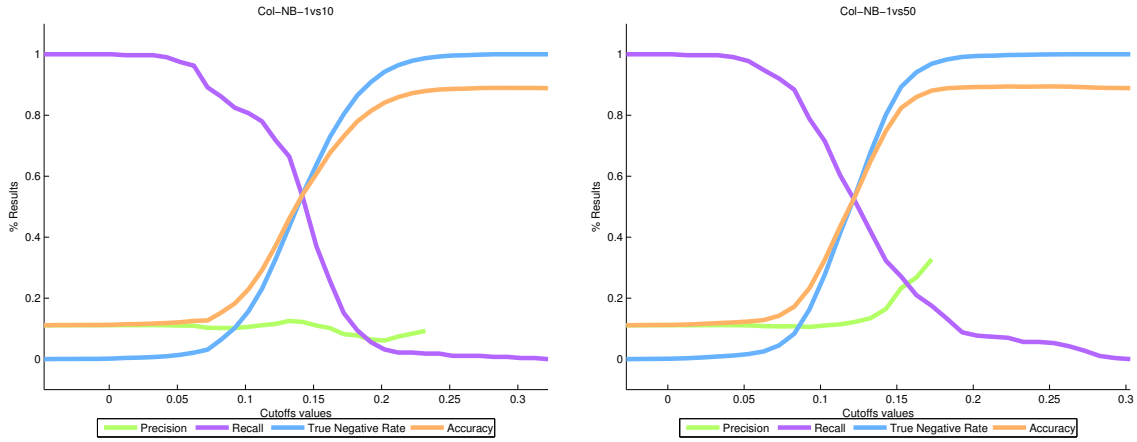
Figura 11: One word + All images: Opció Same model



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

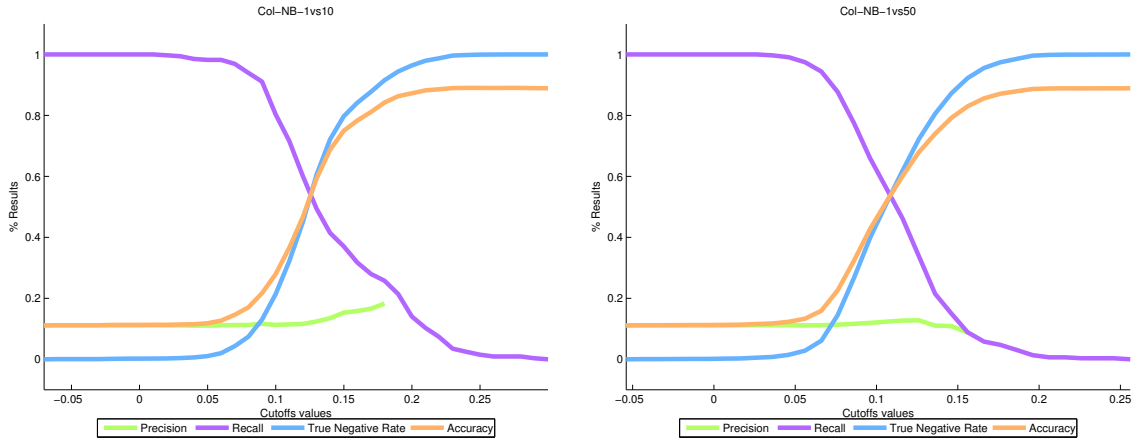
Figura 12: One word + All images: Opció Different model

One word + Clear images



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

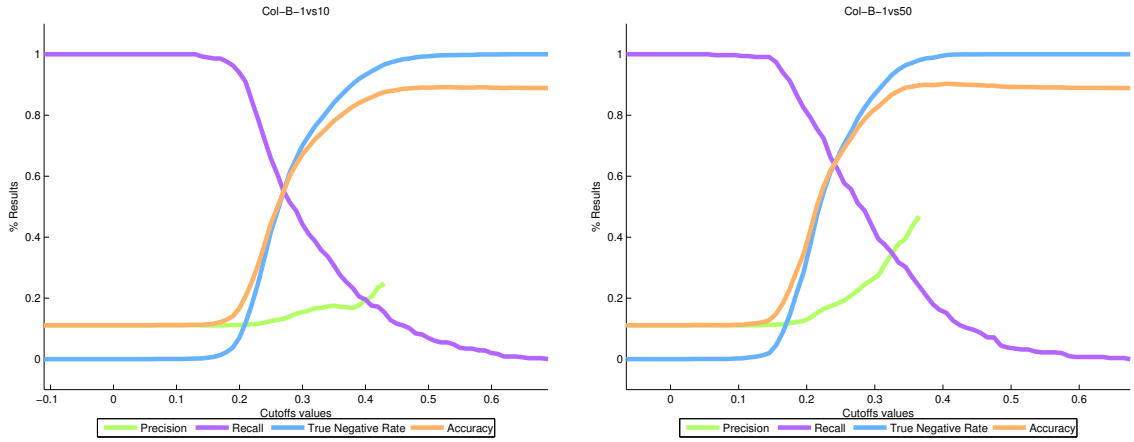
Figura 13: One word + Clear images: Opció Same model



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

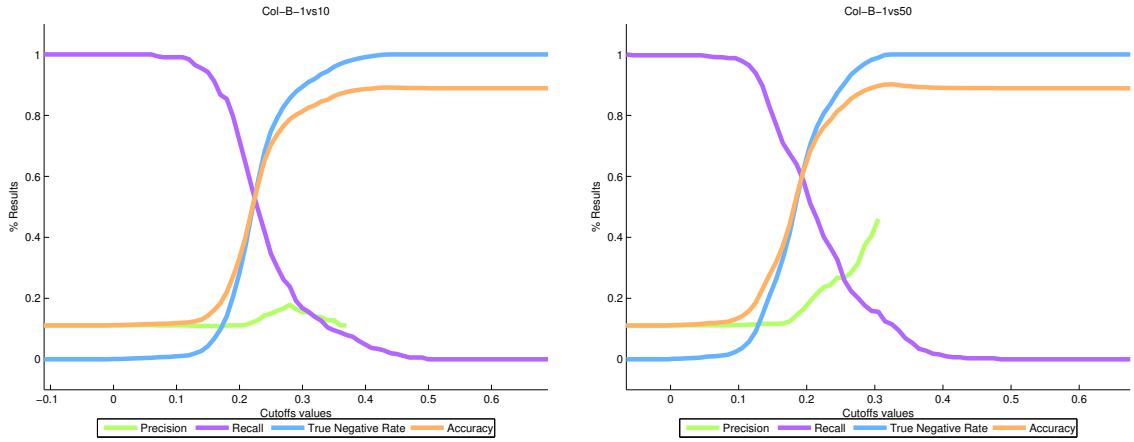
Figura 14: One word + Clear images: Opció Different model

Many words + All images



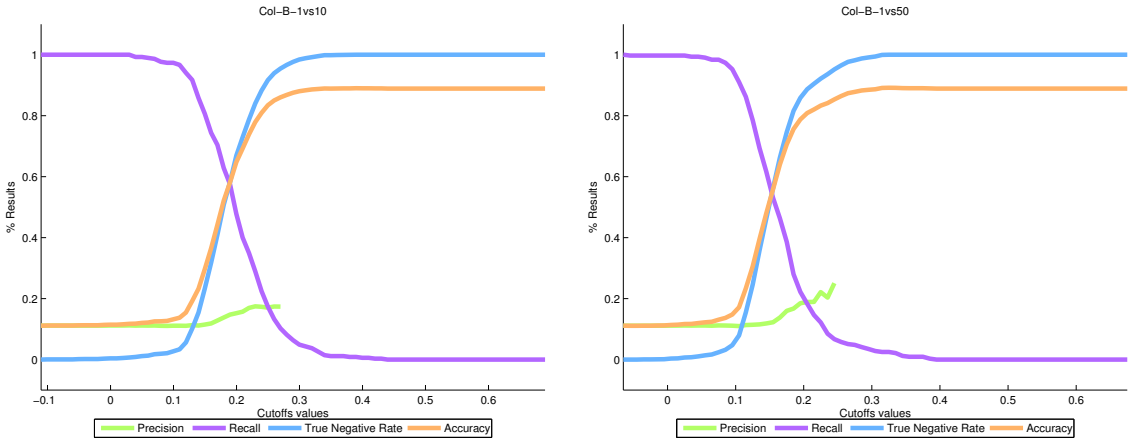
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 15: Many words + All images: Opcions Same model i One word minimum



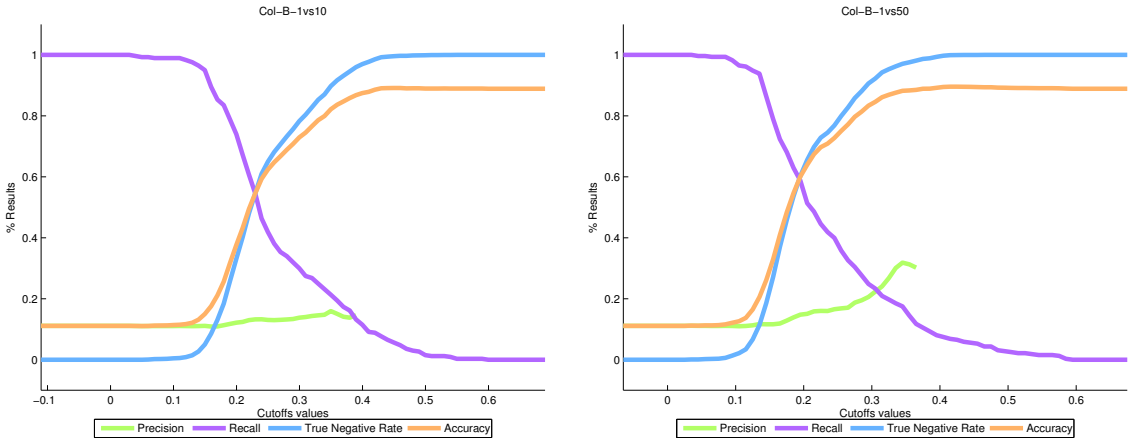
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 16: Many words + All images: Opcions Same model i Two words minimum



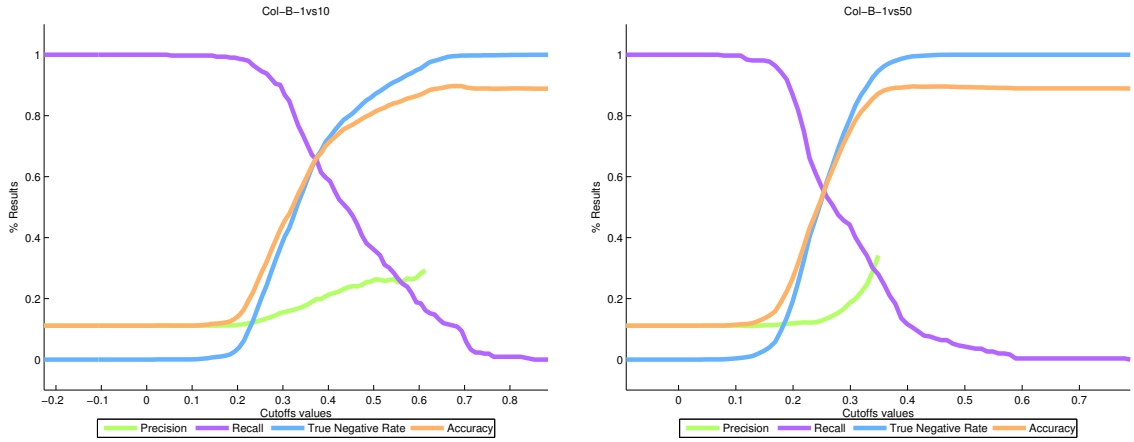
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 17: Many words + All images: Opcions Same model i All words



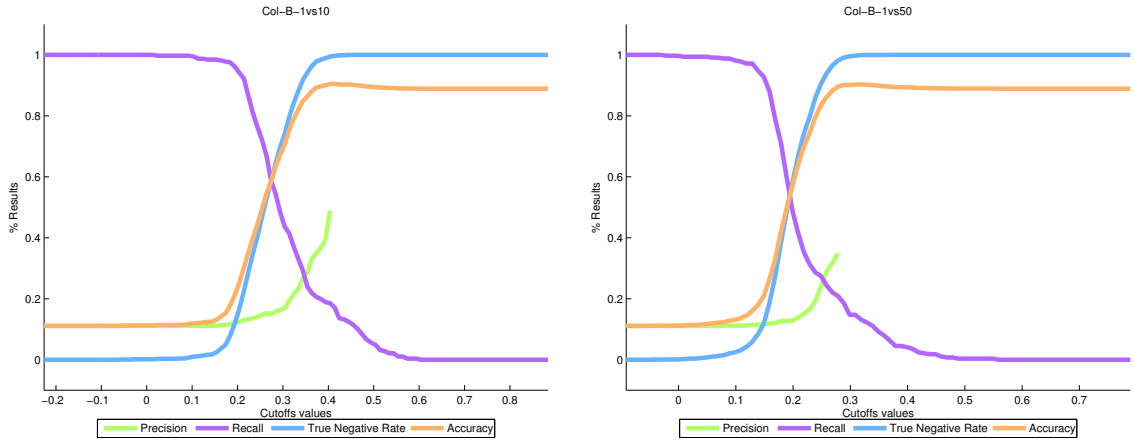
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 18: Many words + All images: Opcions Same model i All but one words



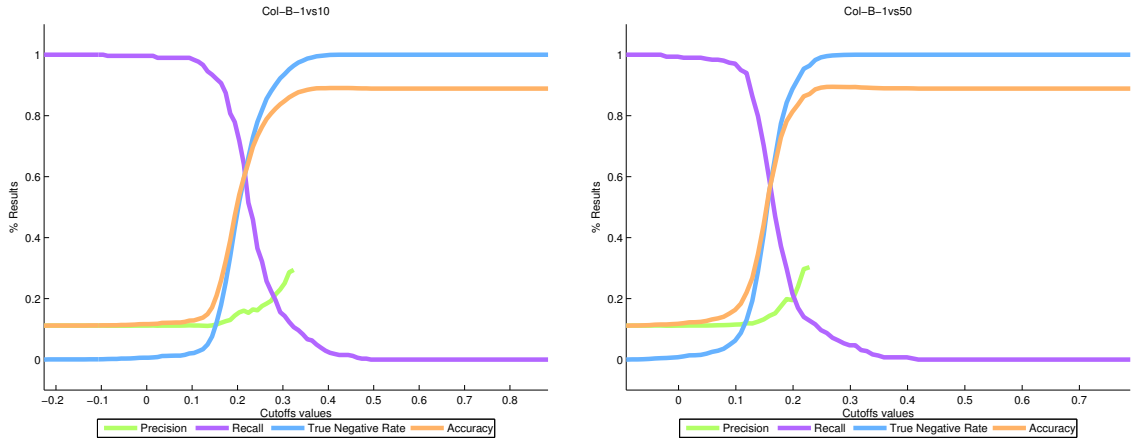
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 19: Many words + All images: Opcions Different model i One word minimum



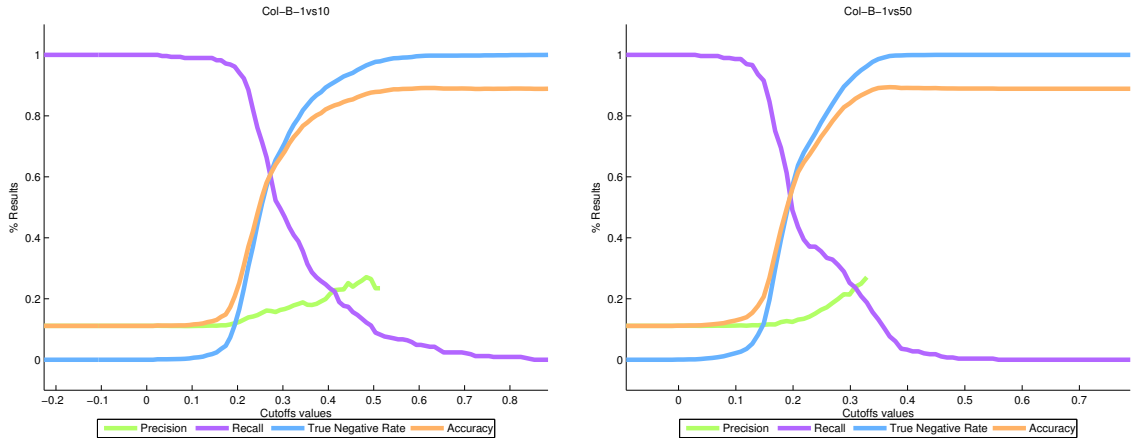
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 20: Many words + All images: Opcions Different model i Two words minimum



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

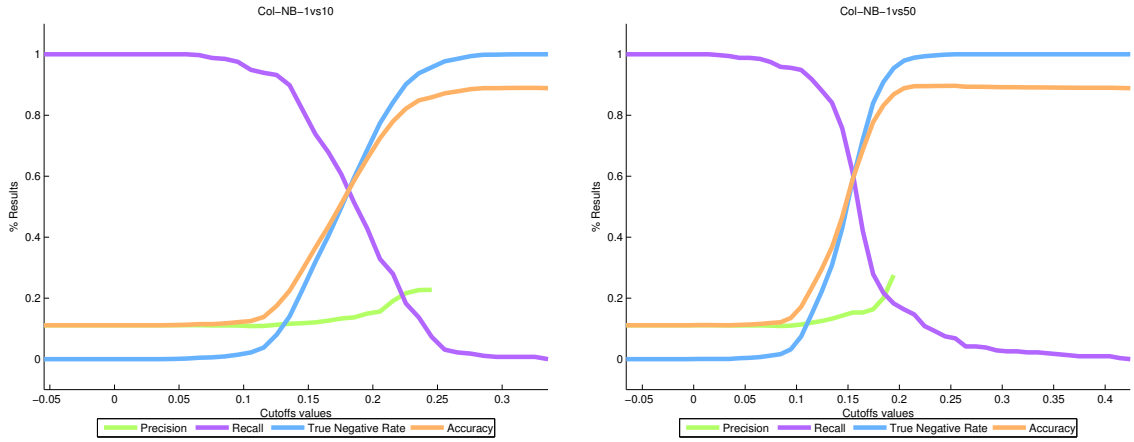
Figura 21: Many words + All images: Opcions Different model i All words



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

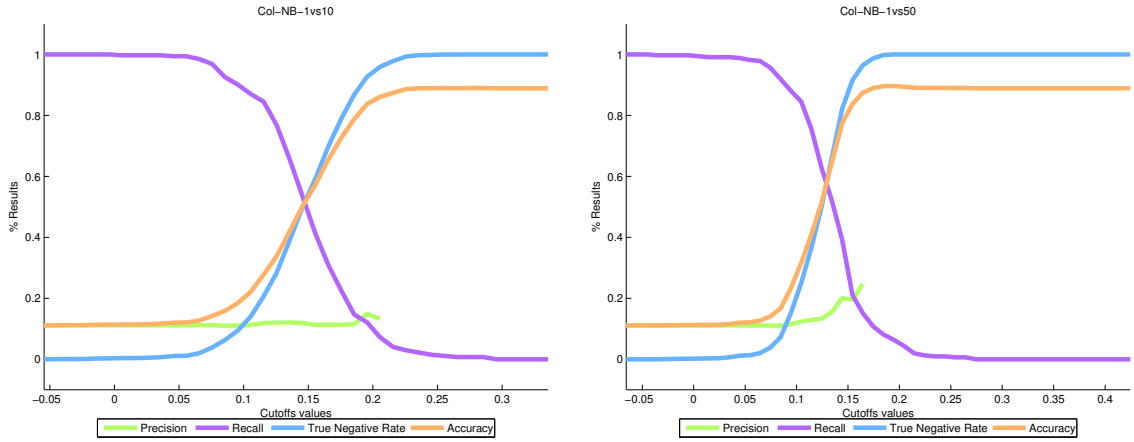
Figura 22: Many words + All images: Opcions Different model i All but one words

Many words + Clear images



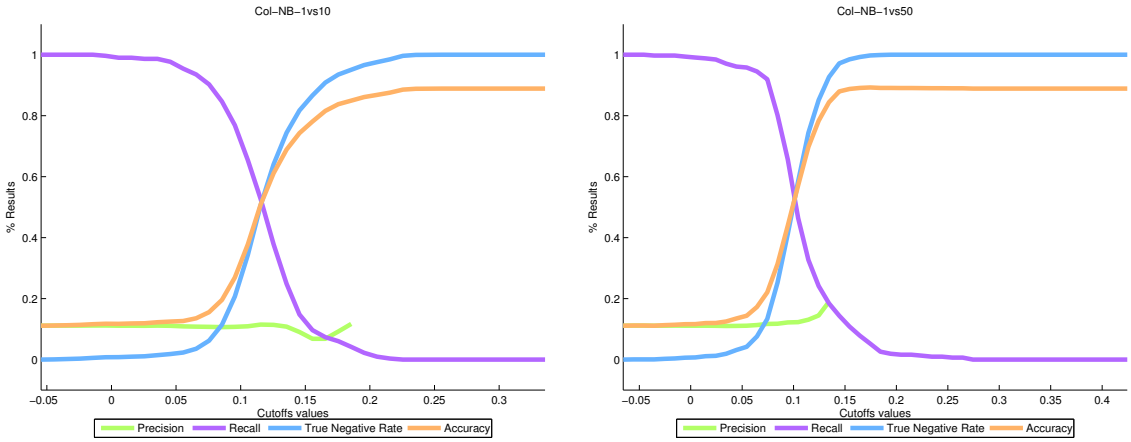
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 23: Many words + Clear images: Opcions Same model i One word minimum



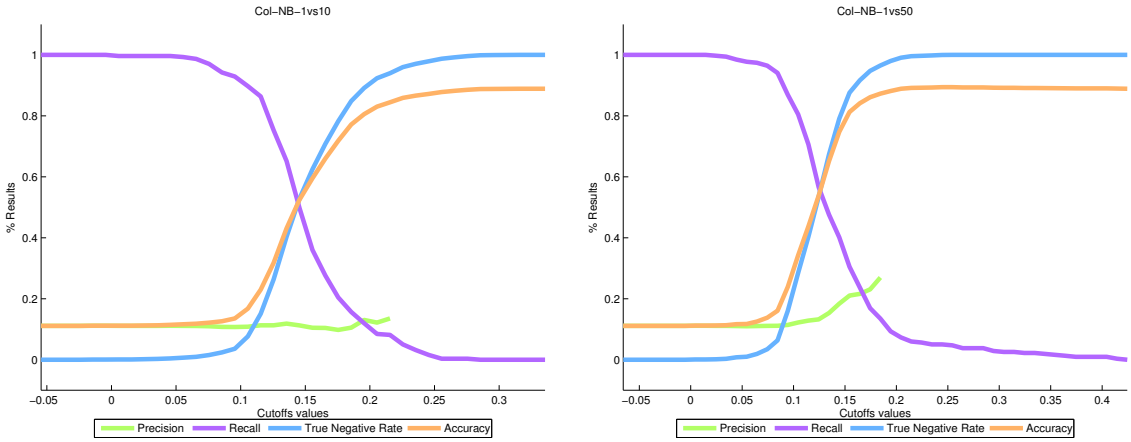
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 24: Many words + Clear images: Opcions Same model i Two words minimum



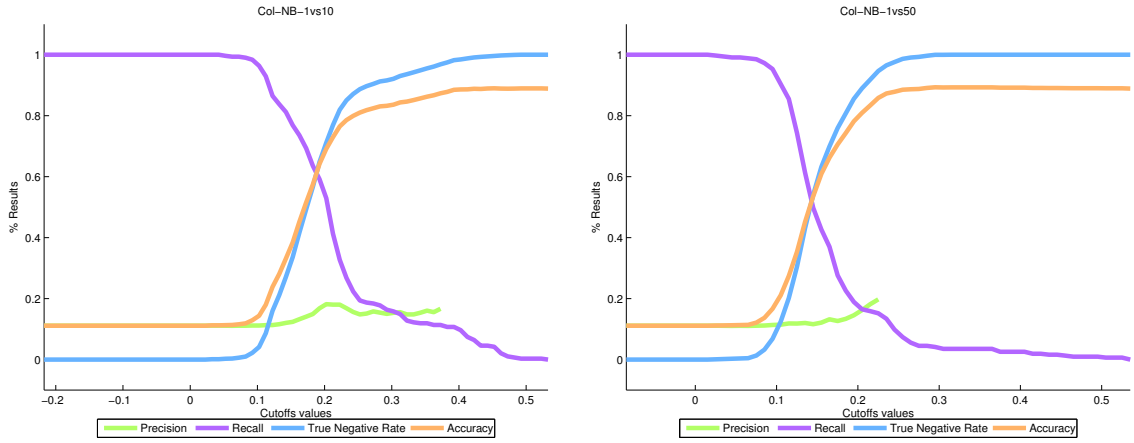
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 25: Many words + Clear images: Opcions Same model i All words



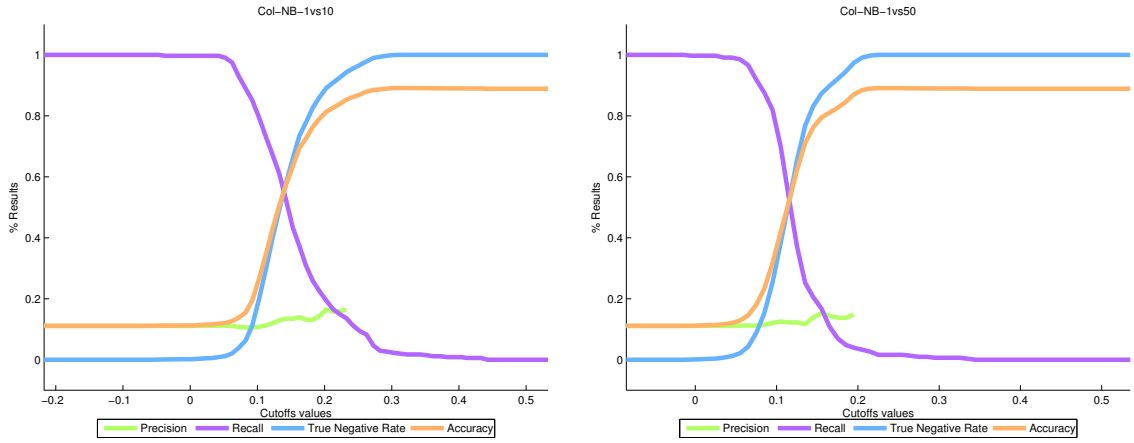
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 26: Many words + Clear images: Opcions Same model i All but one words



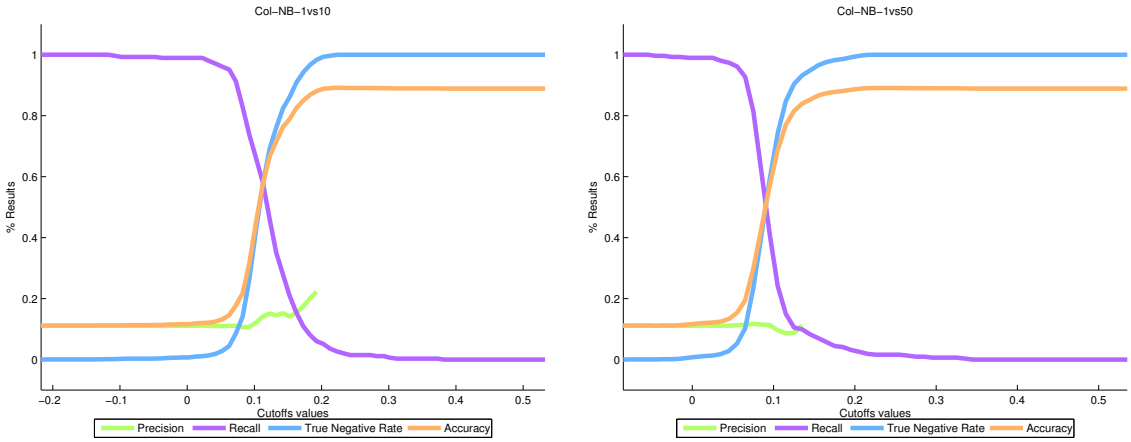
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 27: Many words + Clear images: Opcions Different model i One word minimum



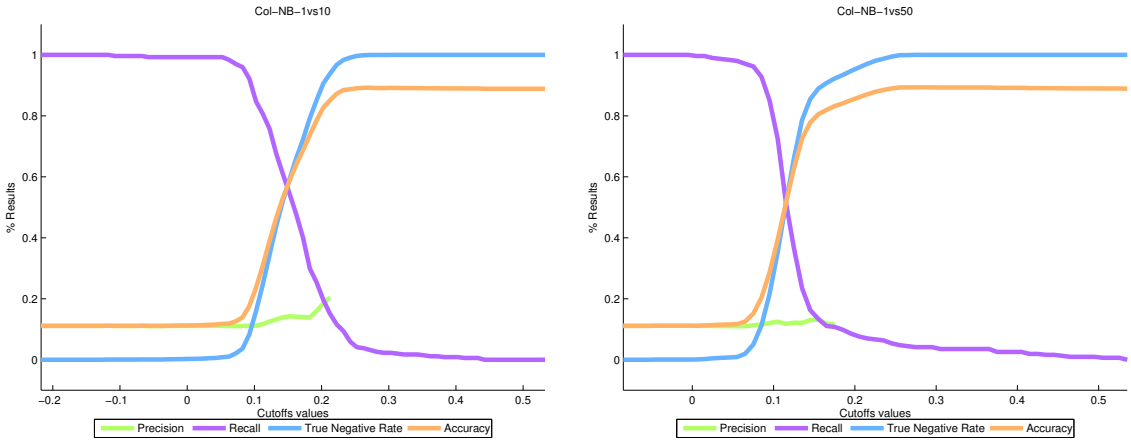
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 28: Many words + Clear images: Opcions Different model i Two words minimum



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 29: Many words + Clear images: Opcions Different model i All words



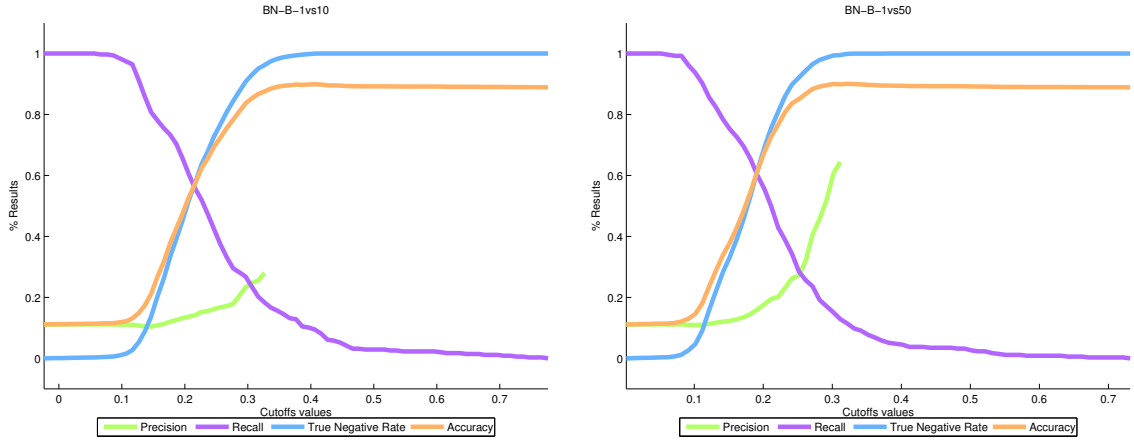
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 30: Many words + Clear images: Opcions Different model i All but one words

Resultats obtinguts amb les imatges en escala de grisos

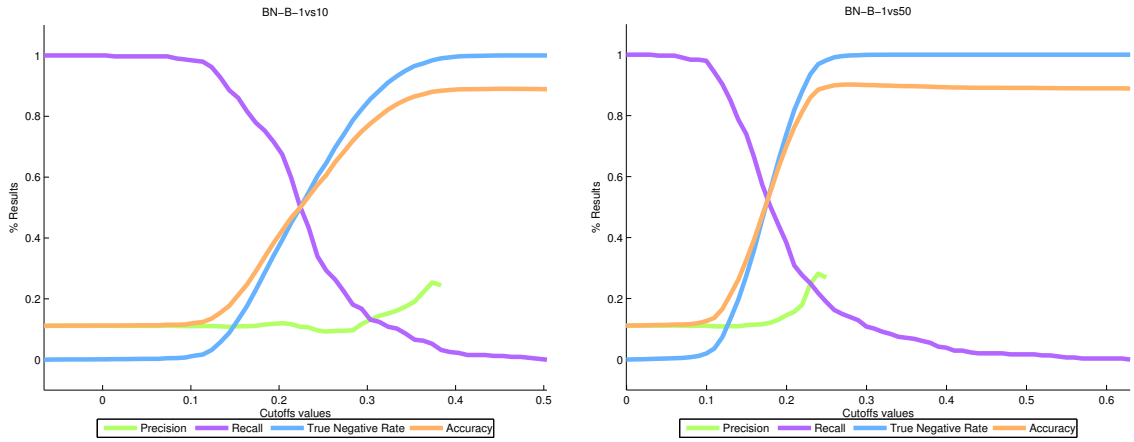
Aquí mostrem tots els resultats obtinguts utilitzant les imatges en escala de grisos:

One word + All images



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

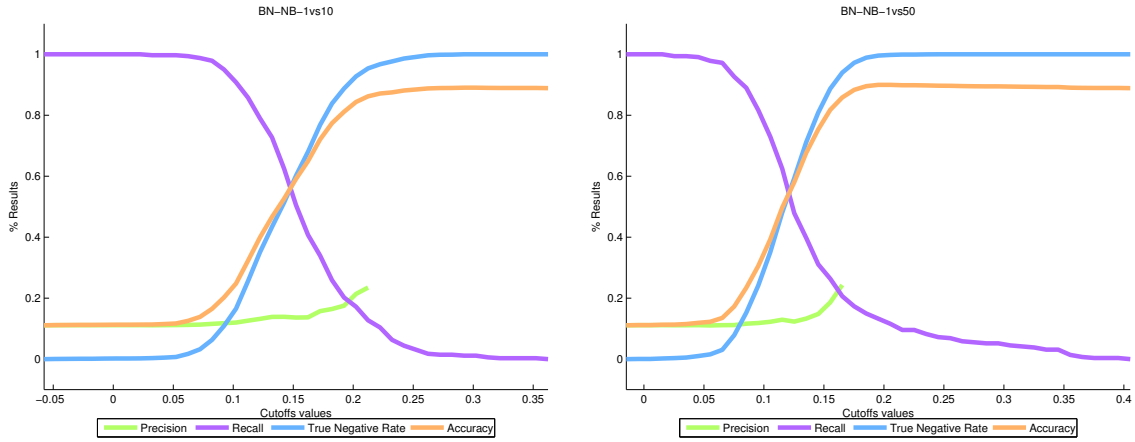
Figura 31: One word + All images: Opció Same model



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

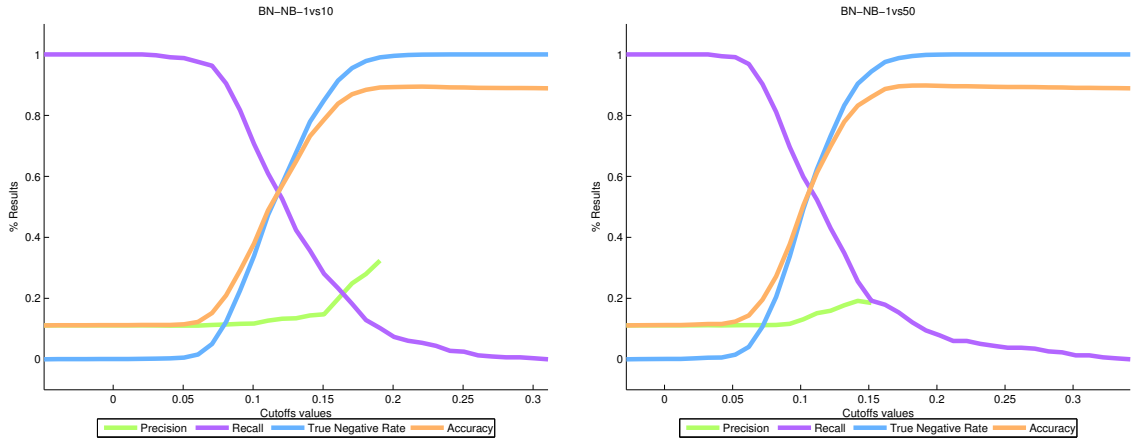
Figura 32: One word + All images: Opció Different model

One word + Clear images



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

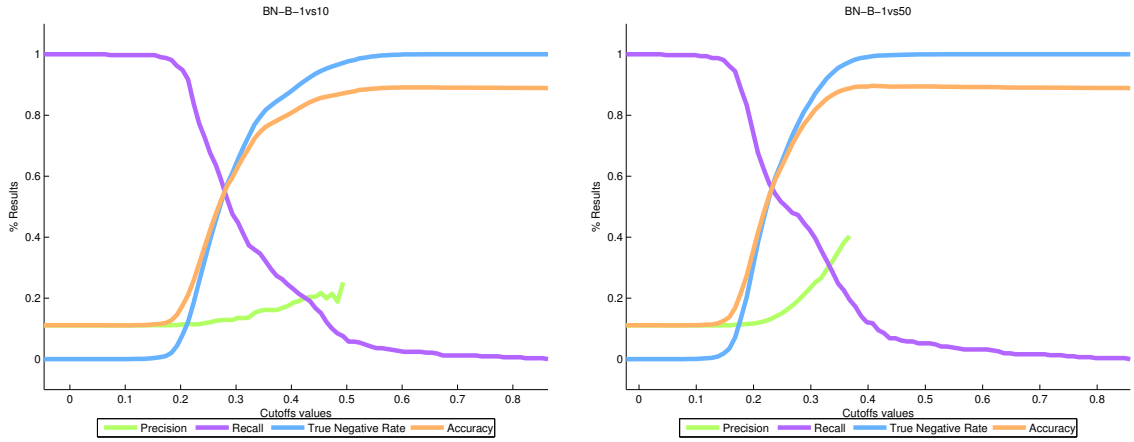
Figura 33: One word + Clear images: Opció Same model



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

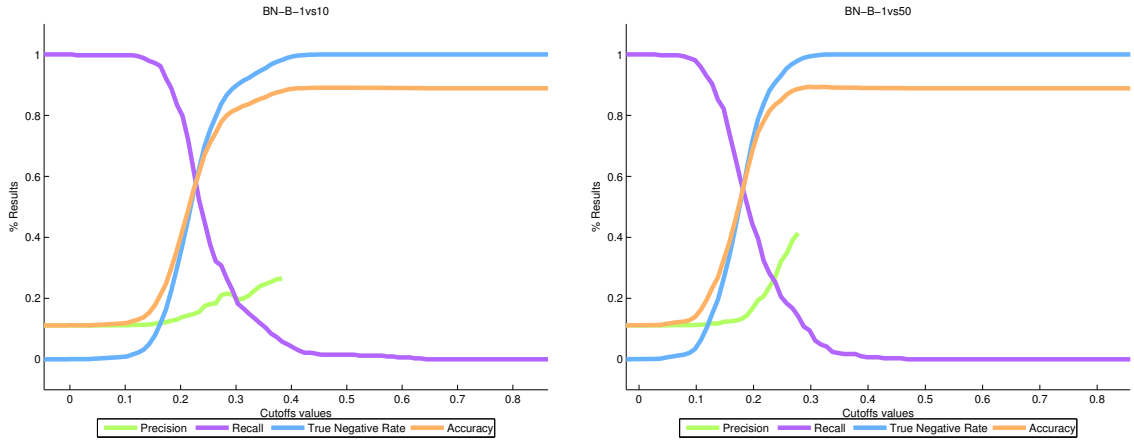
Figura 34: One word + Clear images: Opció Different model

Many words + All images



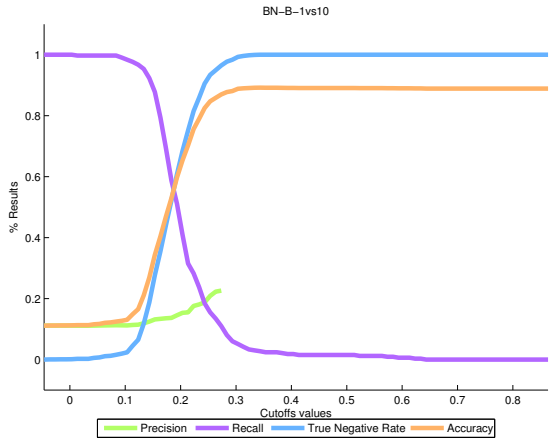
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 35: Many words + All images: Opcions Same model i One word minimum

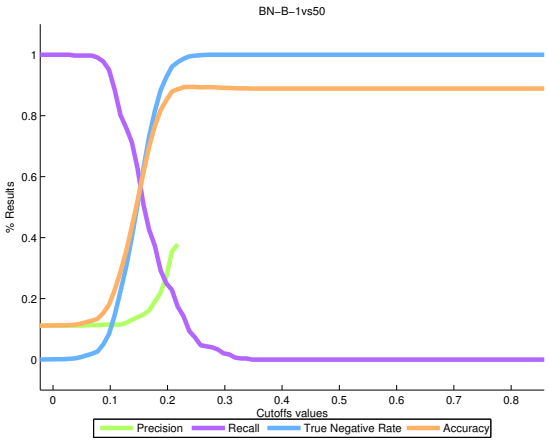


(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 36: Many words + All images: Opcions Same model i Two words minimum

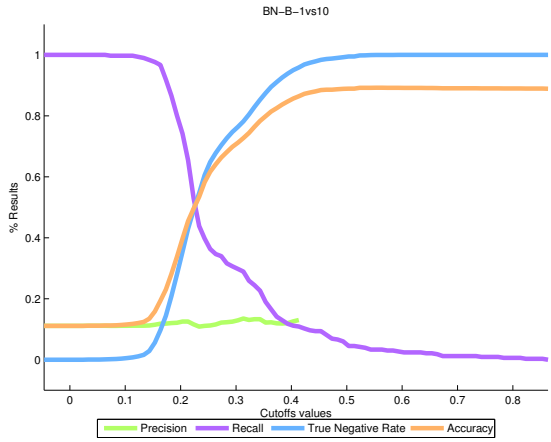


(a) 10 negatives per cada positiva

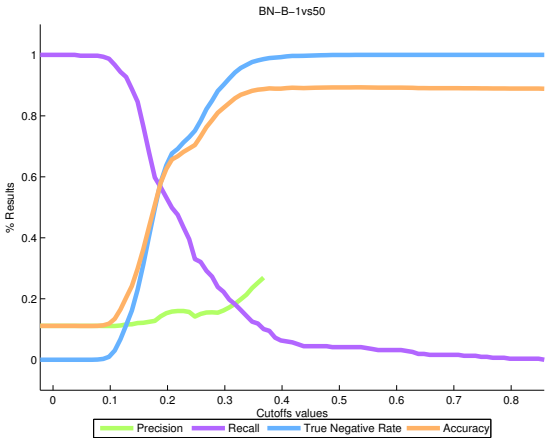


(b) 50 negatives per cada positiva

Figura 37: Many words + All images: Opcions Same model i All words

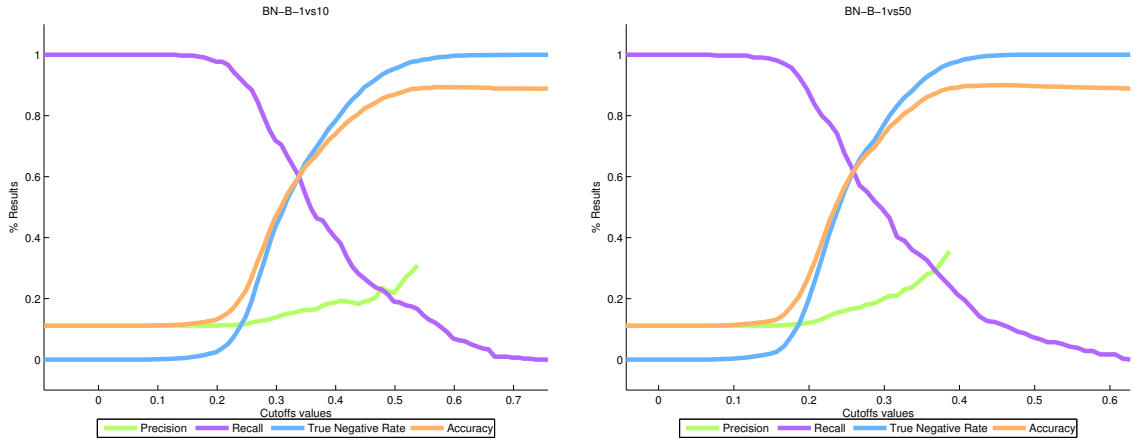


(a) 10 negatives per cada positiva



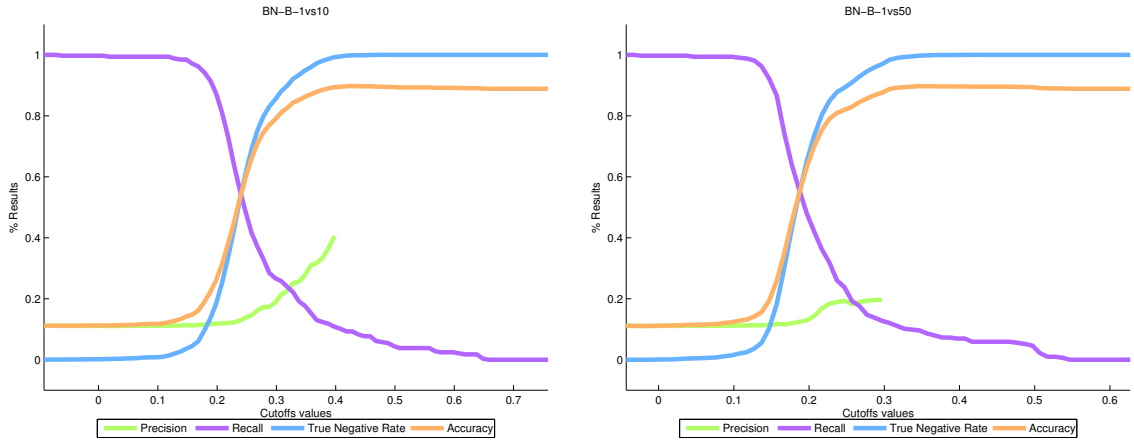
(b) 50 negatives per cada positiva

Figura 38: Many words + All images: Opcions Same model i All but one words



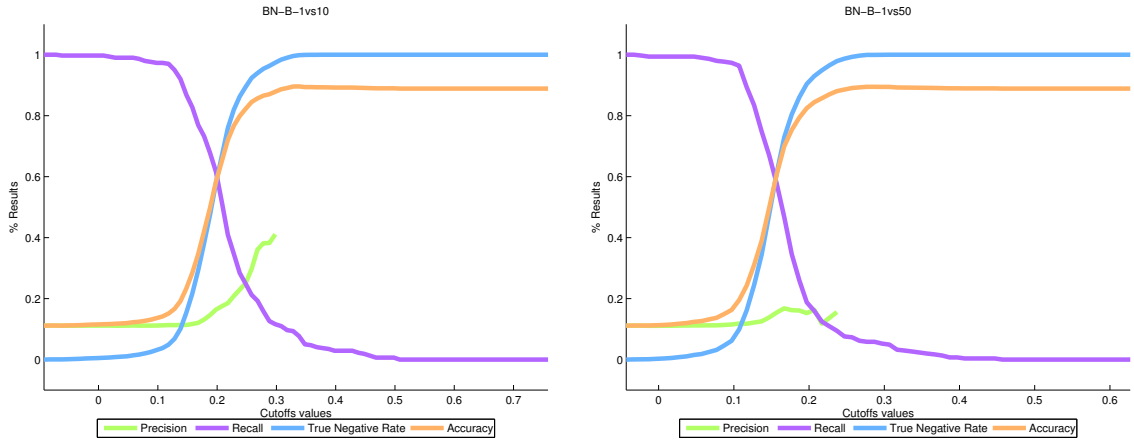
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 39: Many words + All images: Opcions Different model i One word minimum



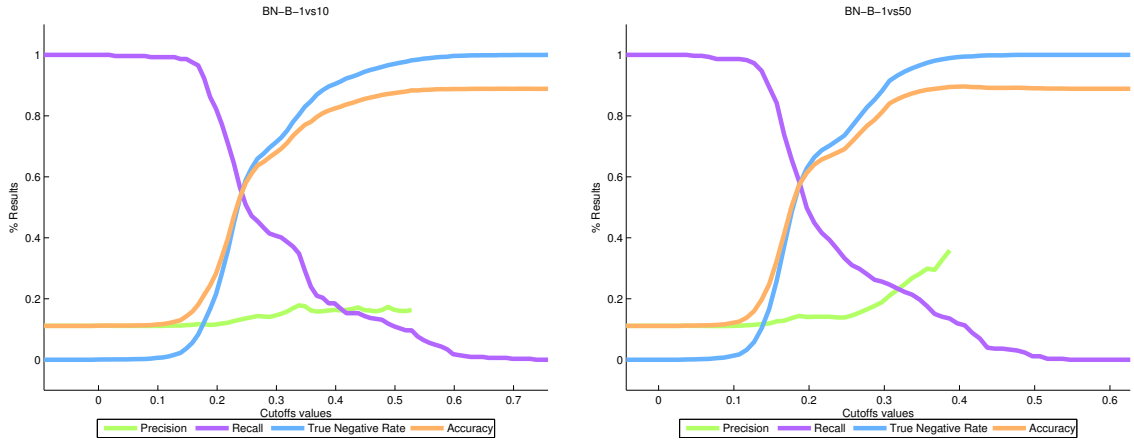
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 40: Many words + All images: Opcions Different model i Two words minimum



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

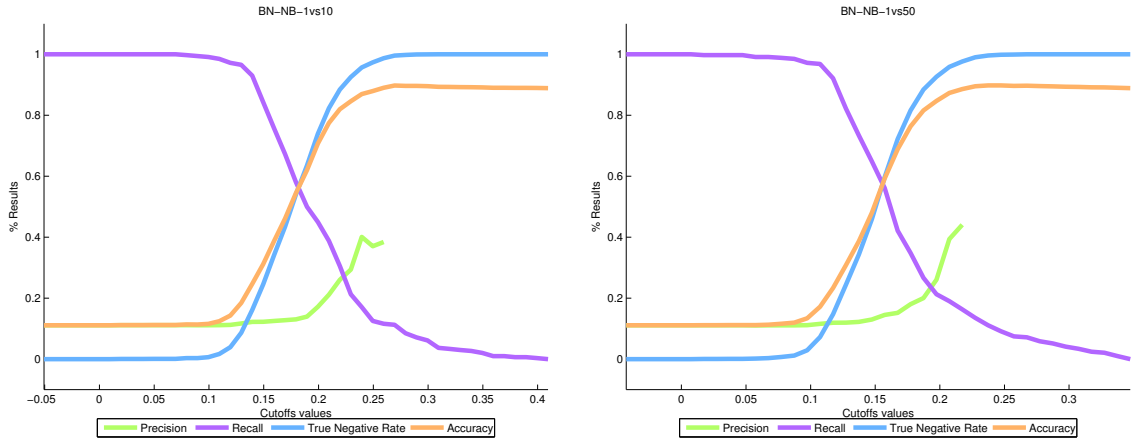
Figura 41: Many words + All images: Opcions Different model i All words



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

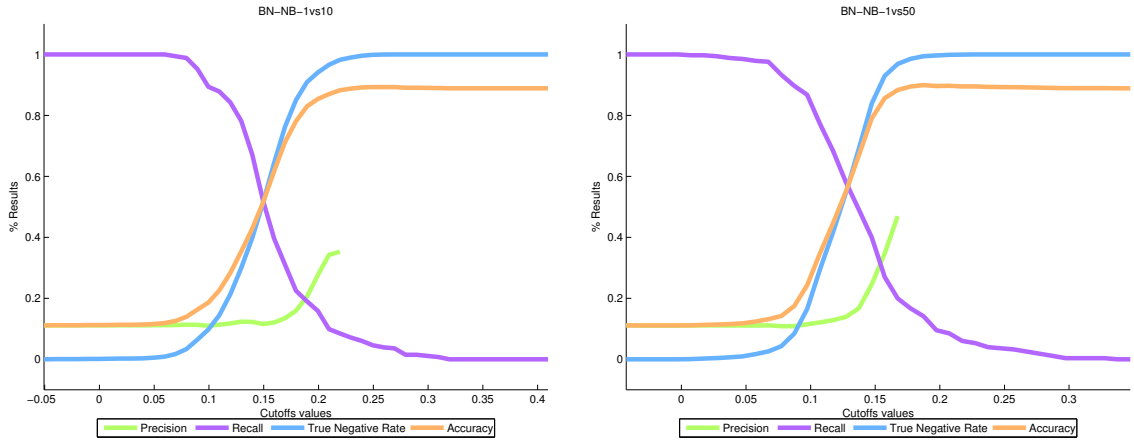
Figura 42: Many words + All images: Opcions Different model i All but one words

Many words + Clear images



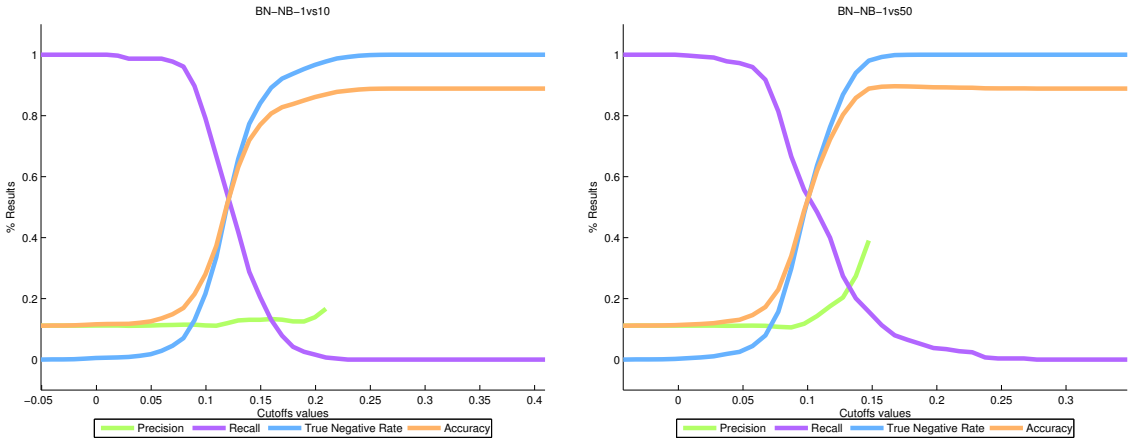
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 43: Many words + Clear images: Opcions Same model i One word minimum



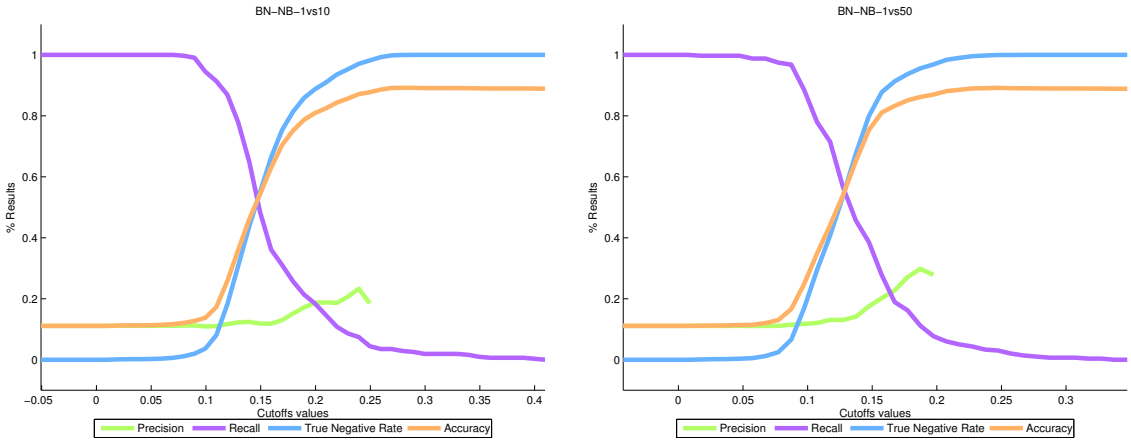
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 44: Many words + Clear images: Opcions Same model i Two words minimum



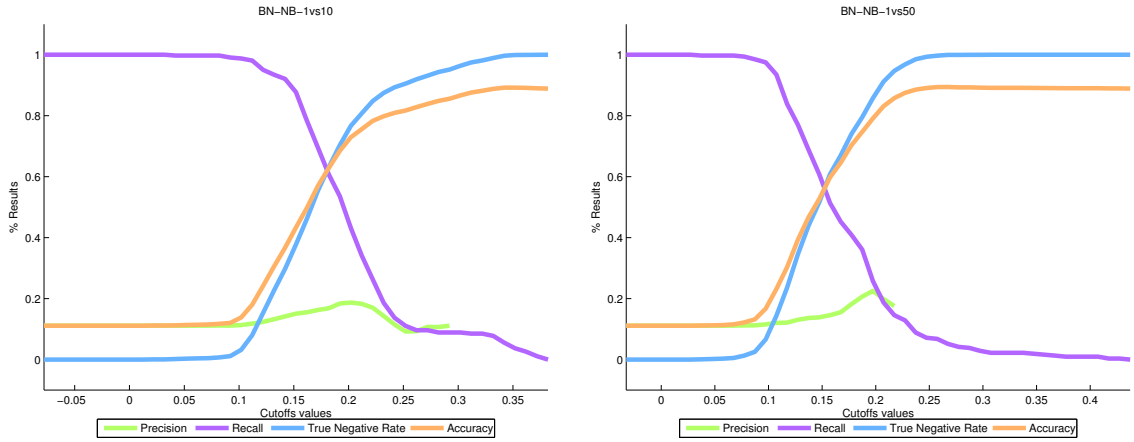
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 45: Many words + Clear images: Opcions Same model i All words



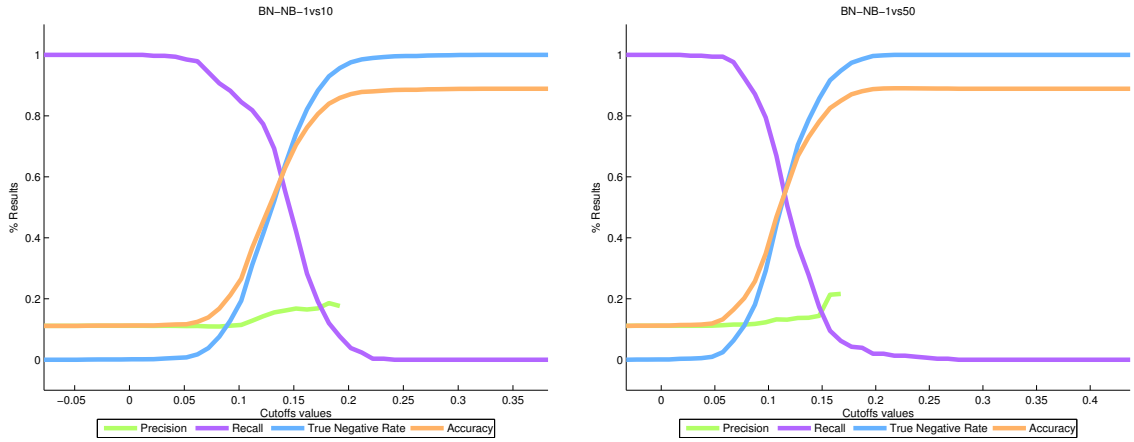
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 46: Many words + Clear images: Opcions Same model i All but one words



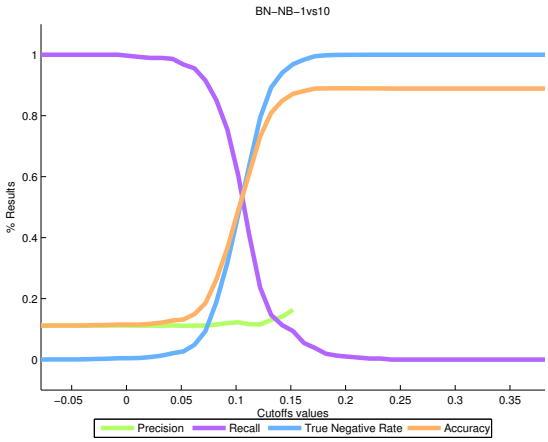
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 47: Many words + Clear images: Opcions Different model i One word minimum

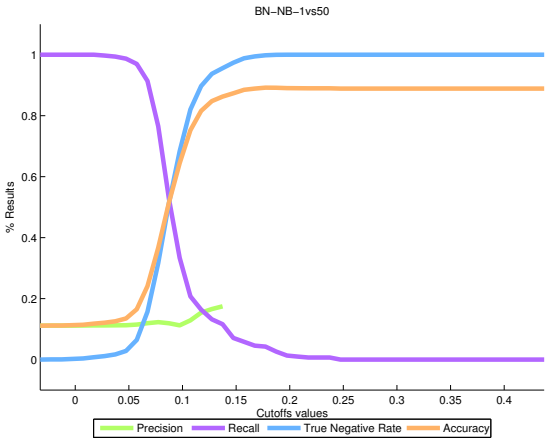


(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 48: Many words + Clear images: Opcions Different model i Two words minimum

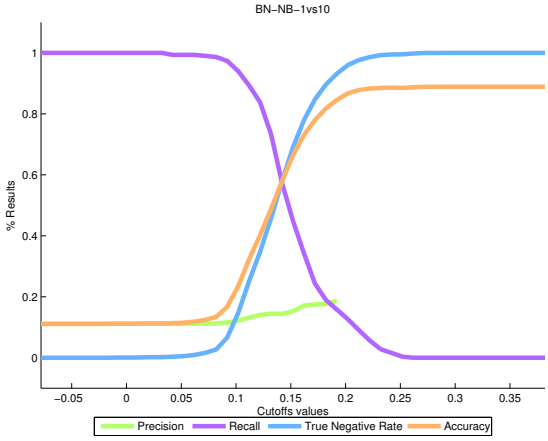


(a) 10 negatives per cada positiva

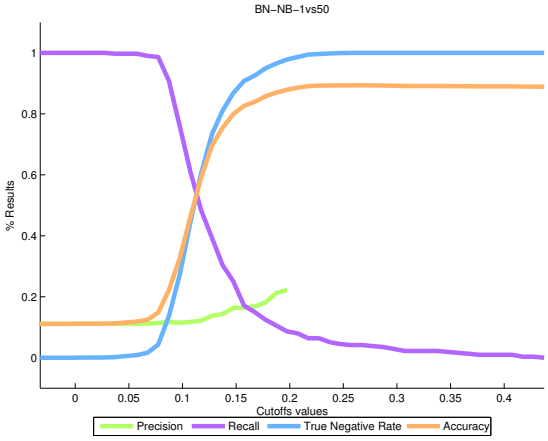


(b) 50 negatives per cada positiva

Figura 49: Many words + Clear images: Opcions Different model i All words



(a) 10 negatives per cada positiva



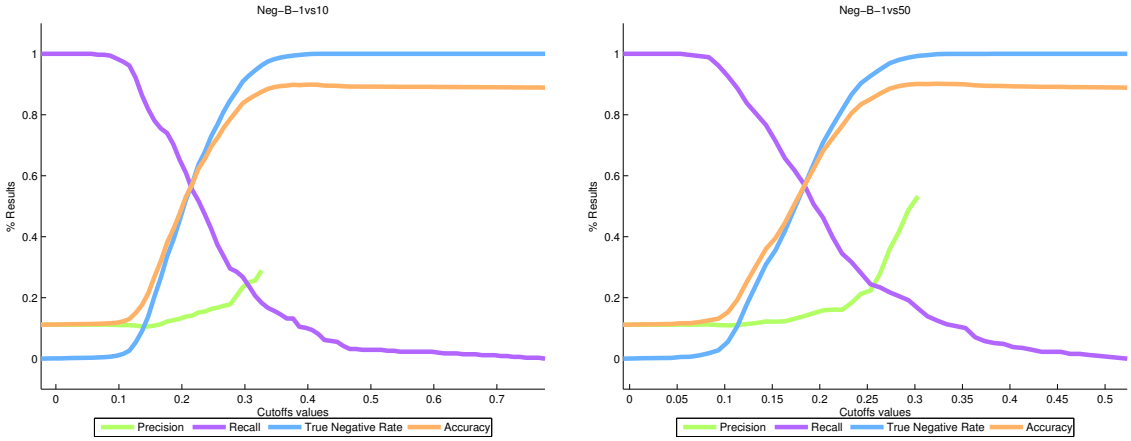
(b) 50 negatives per cada positiva

Figura 50: Many words + Clear images: Opcions Different model i All but one words

Resultats obtinguts amb les imatges del complement de les imatges en escala de grisos

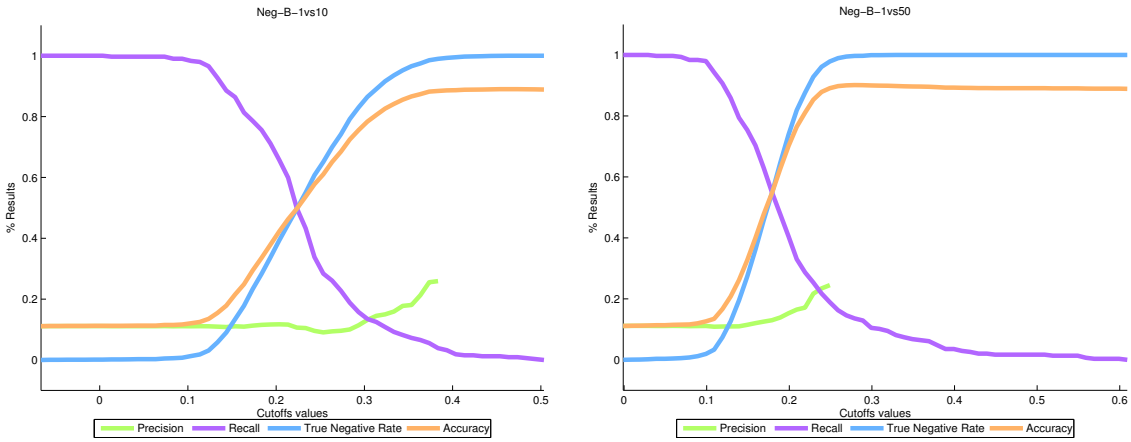
Aquí mostrem tots els resultats obtinguts utilitzant les imatges del complement de les imatges en escala de grisos:

One word + All images



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

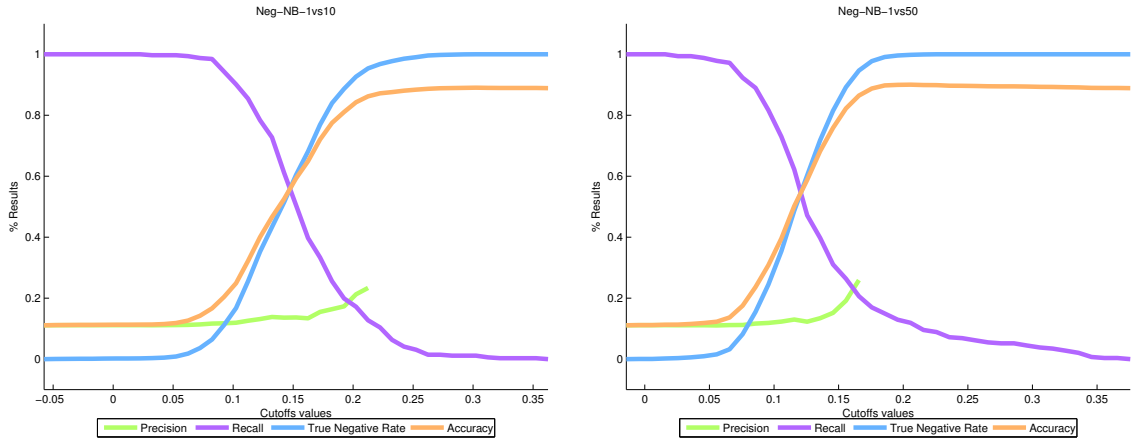
Figura 51: One word + All images: Opció Same model



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

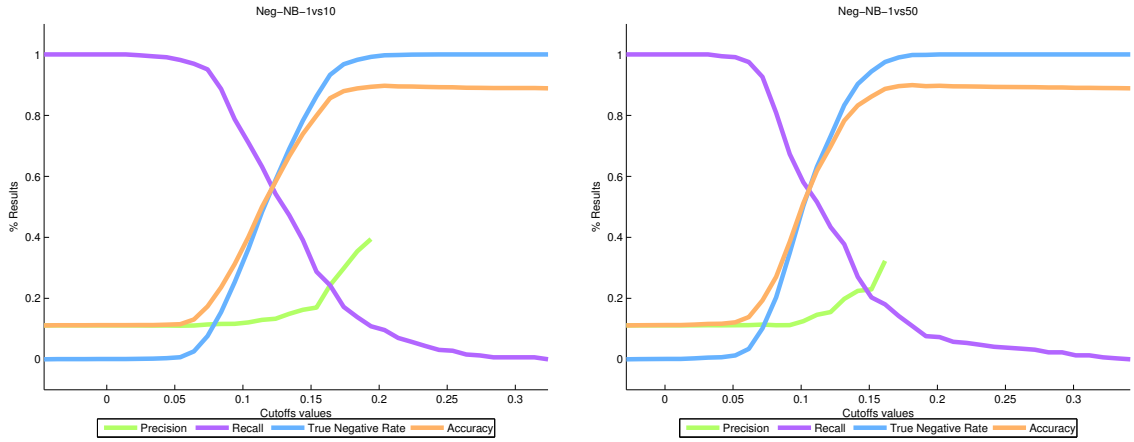
Figura 52: One word + All images: Opció Different model

One word + Clear images



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

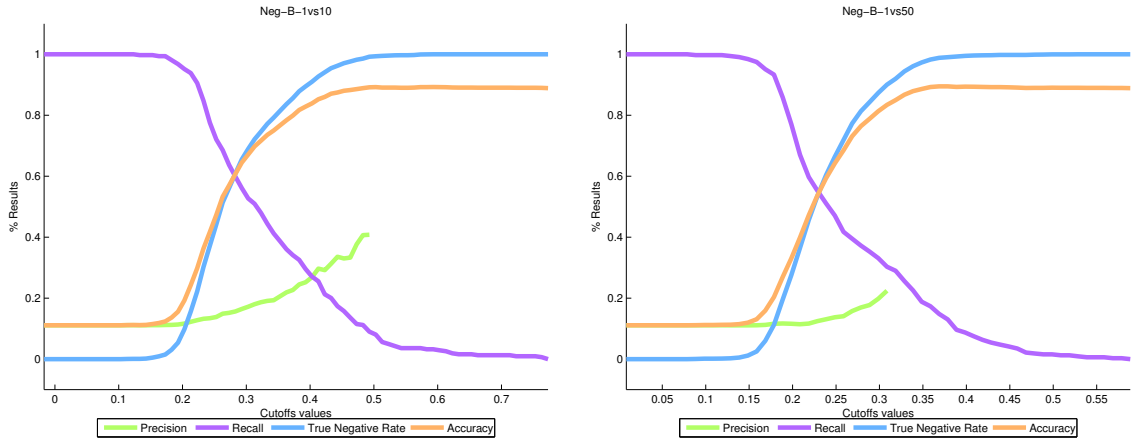
Figura 53: One word + Clear images: Opció Same model



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

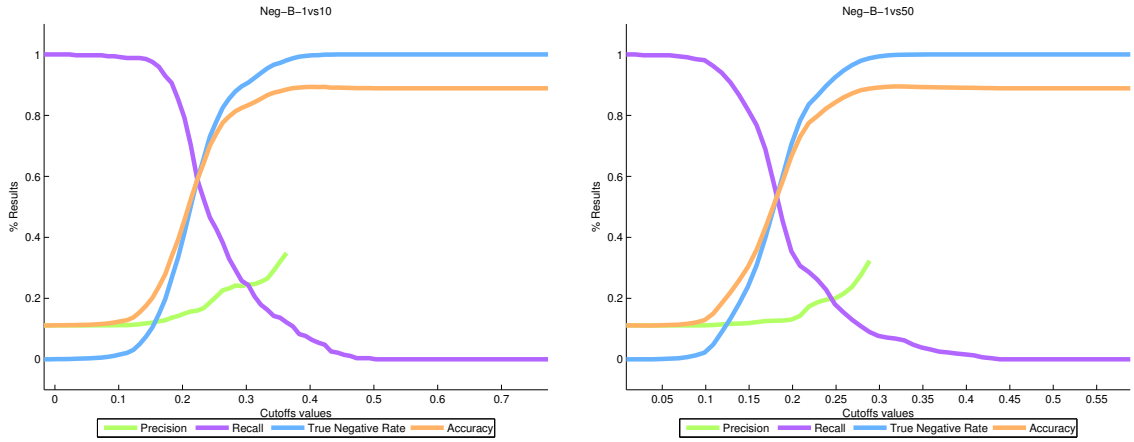
Figura 54: One word + Clear images: Opció Different model

Many words + All images



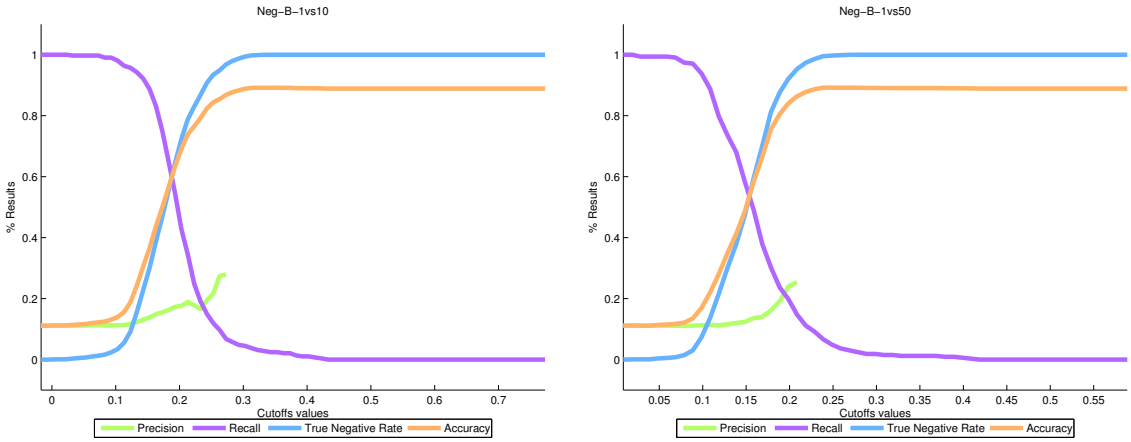
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 55: Many words + All images: Opcions Same model i One word minimum



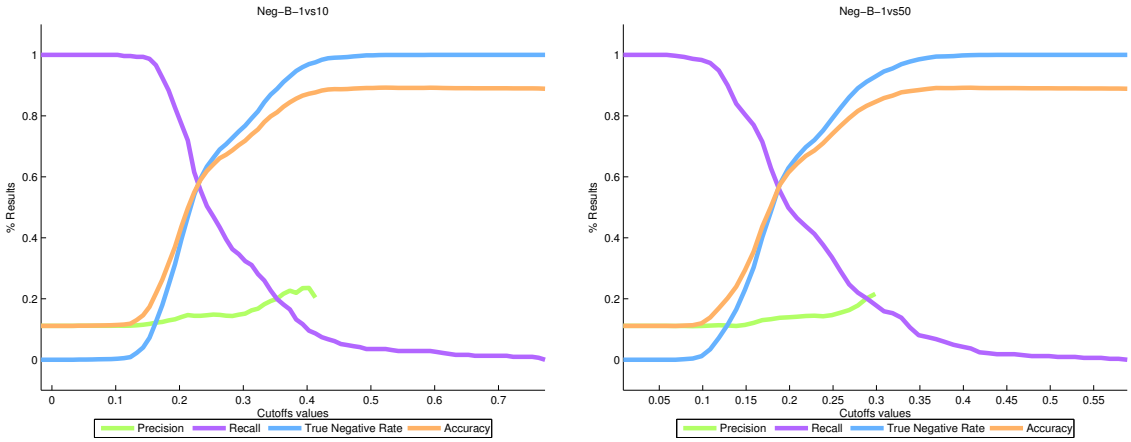
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 56: Many words + All images: Opcions Same model i Two words minimum



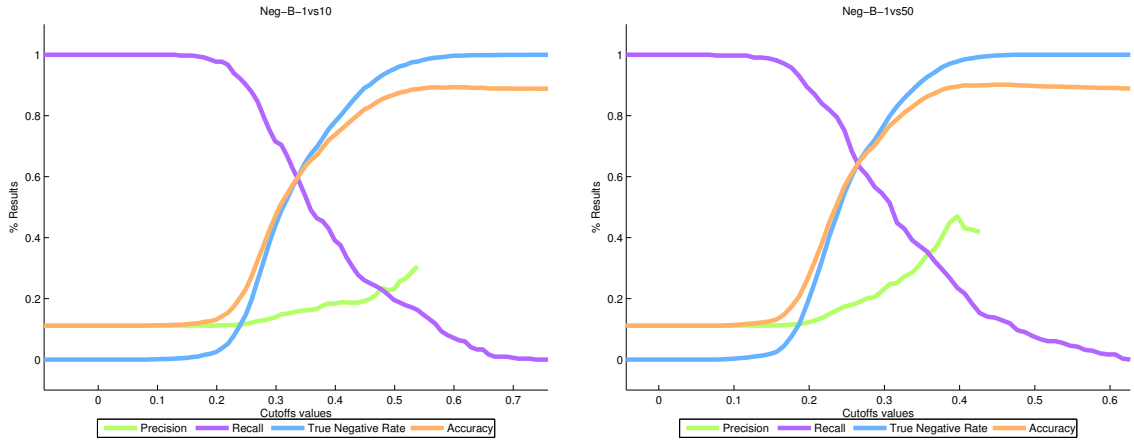
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 57: Many words + All images: Opcions Same model i All words



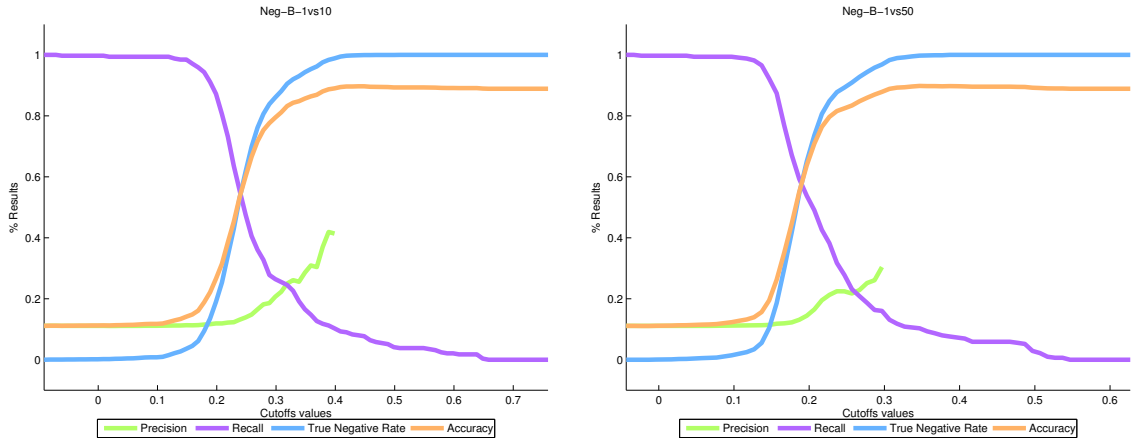
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 58: Many words + All images: Opcions Same model i All but one words



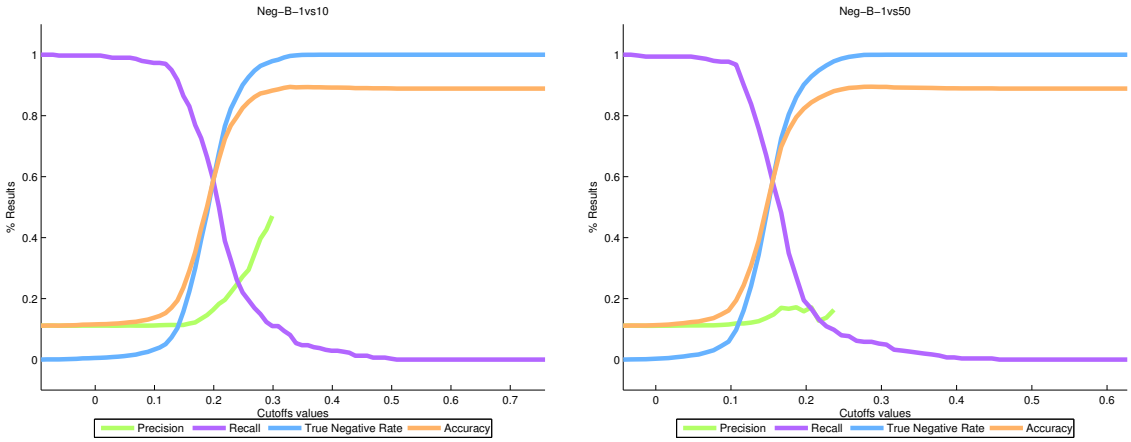
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 59: Many words + All images: Opcions Different model i One word minimum



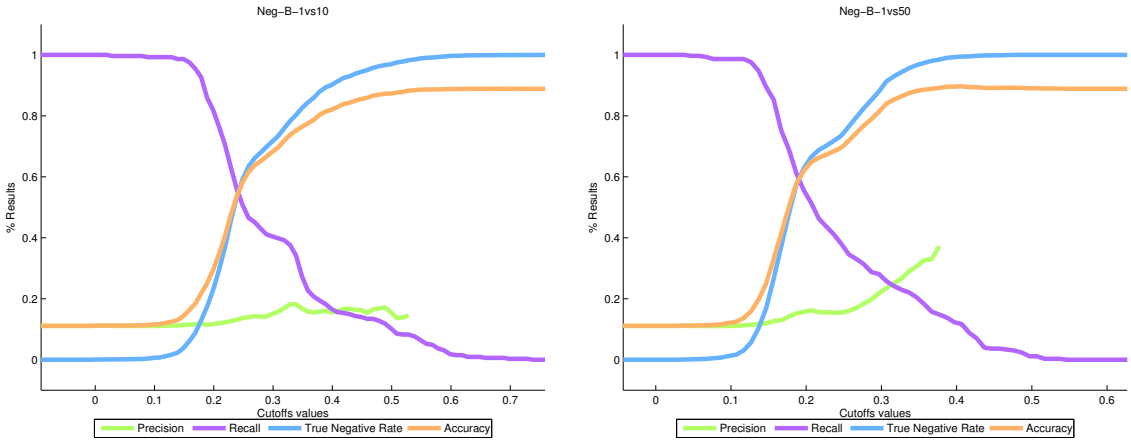
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 60: Many words + All images: Opcions Different model i Two words minimum



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

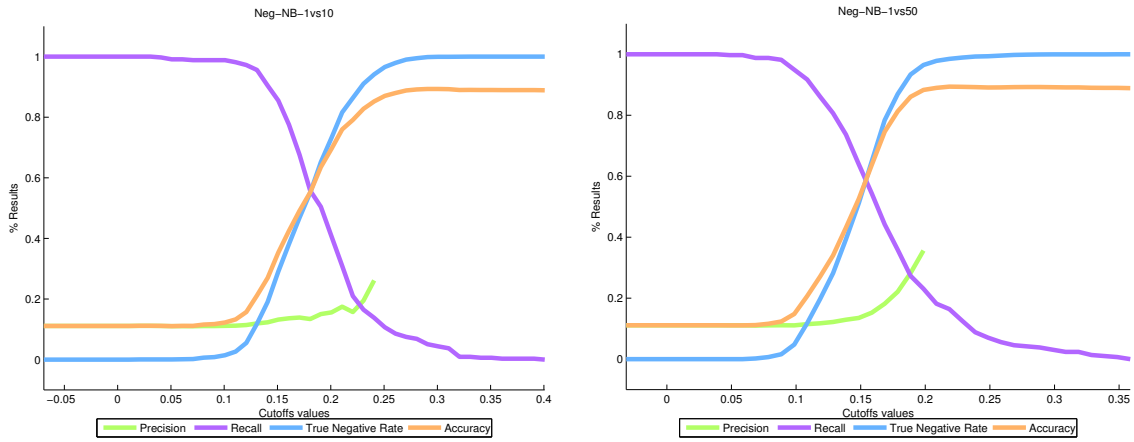
Figura 61: Many words + All images: Opcions Different model i All words



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

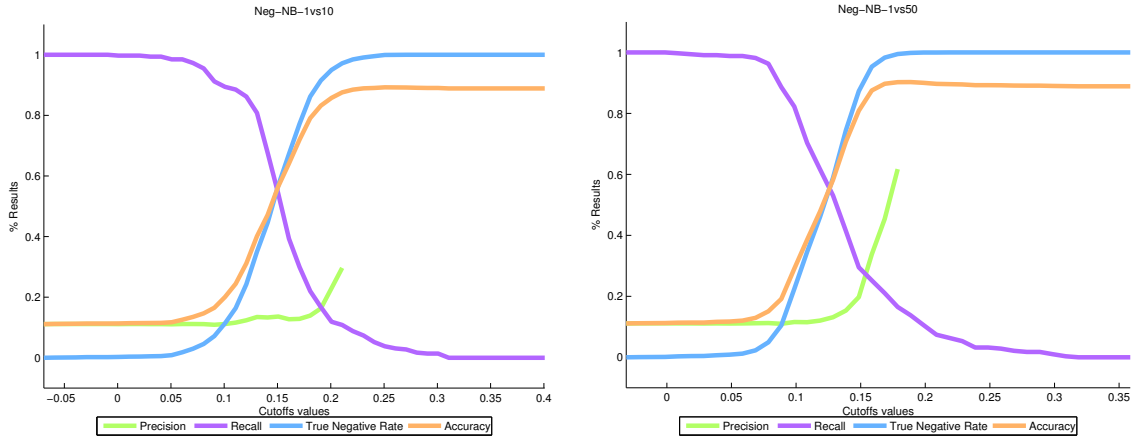
Figura 62: Many words + All images: Opcions Different model i All but one words

Many words + Clear images



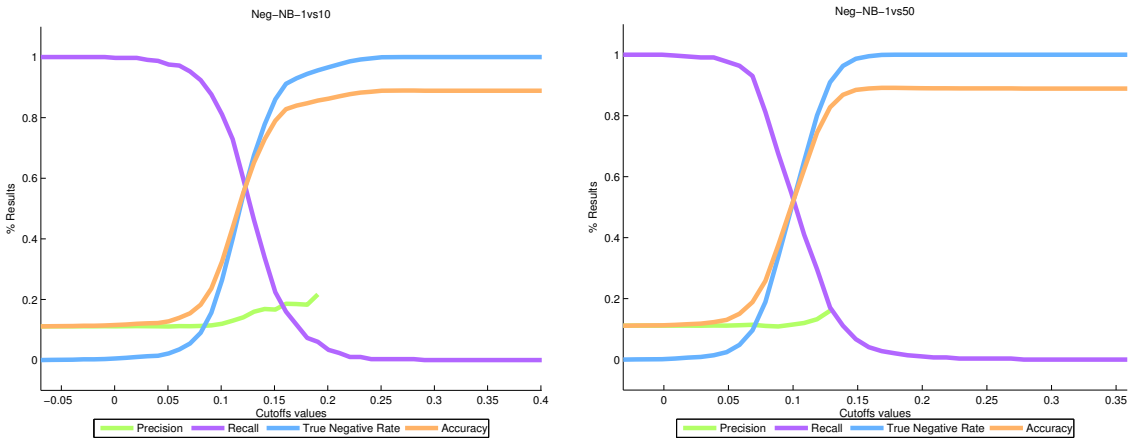
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 63: Many words + Clear images: Opcions Same model i One word minimum



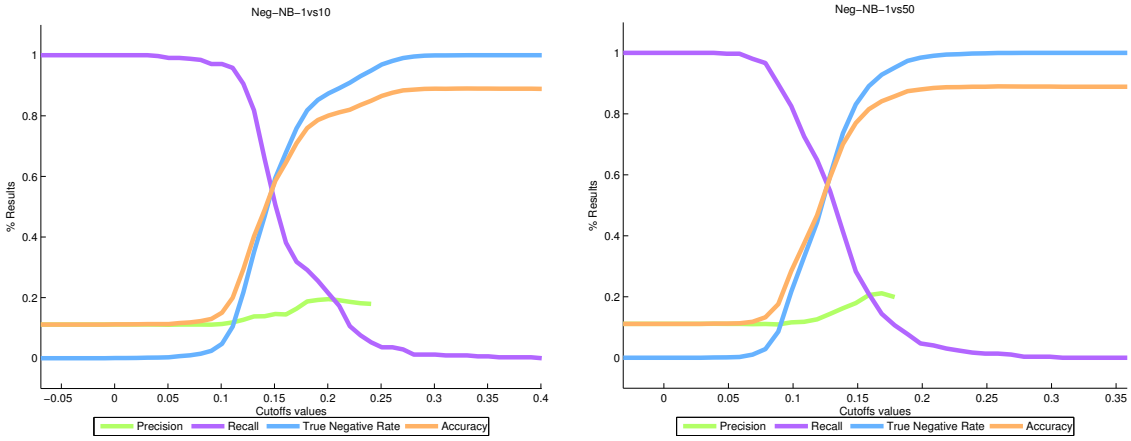
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 64: Many words + Clear images: Opcions Same model i Two words minimum



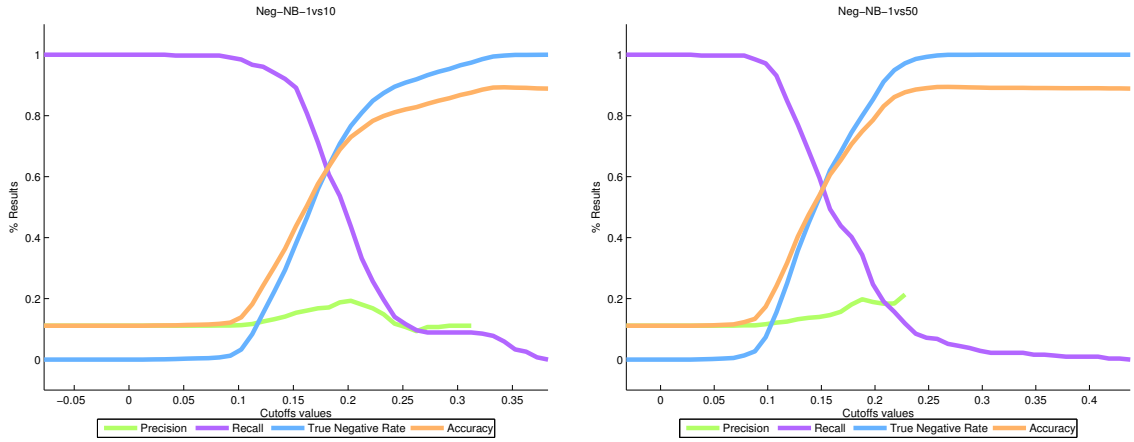
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 65: Many words + Clear images: Opcions Same model i All words



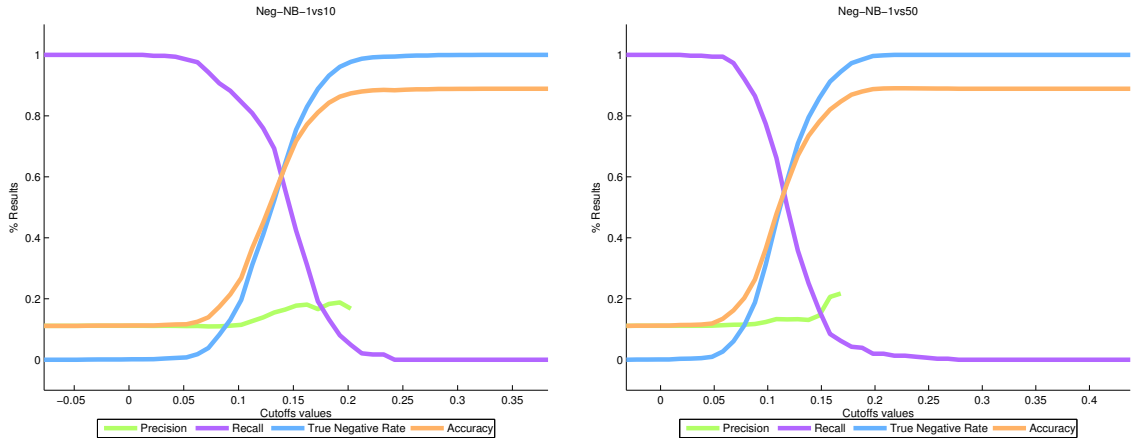
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 66: Many words + Clear images: Opcions Same model i All but one words



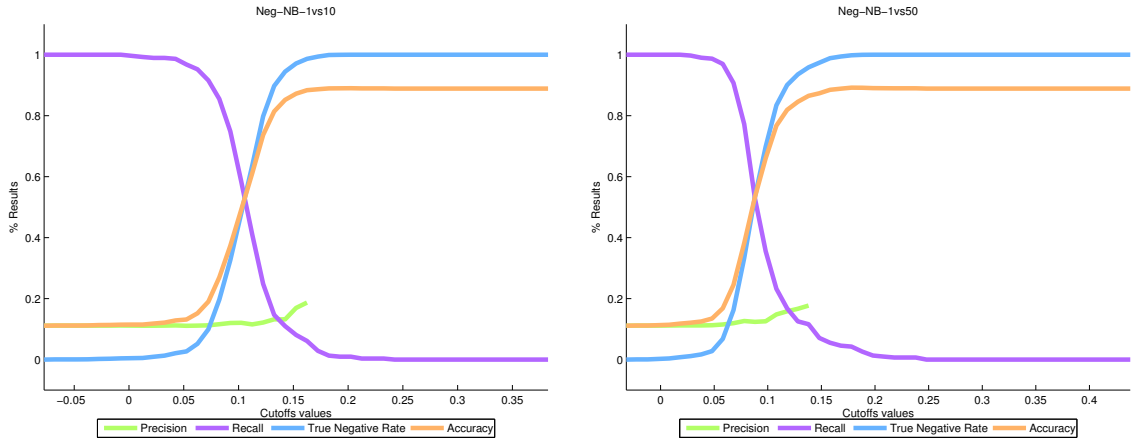
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 67: Many words + Clear images: Opcions Different model i One word minimum



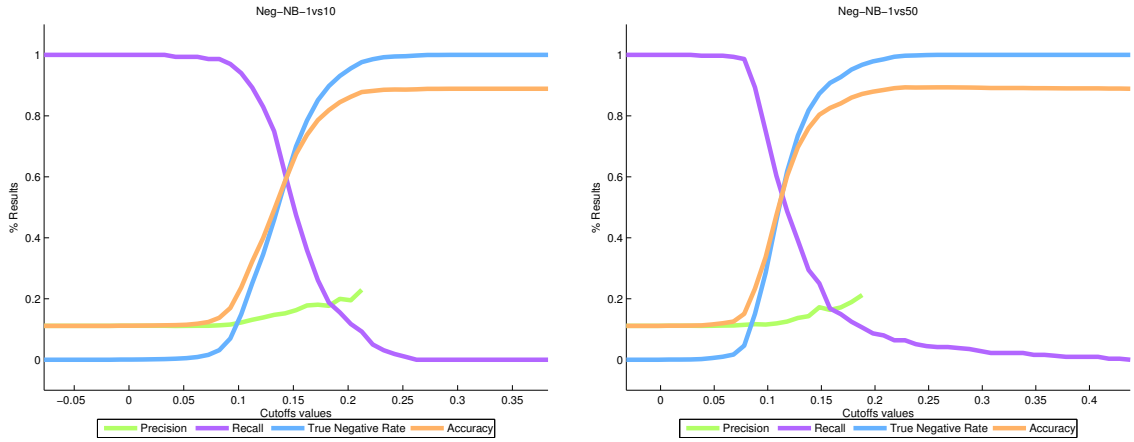
(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 68: Many words + Clear images: Opcions Different model i Two words minimum



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 69: Many words + Clear images: Opcions Different model i All words



(a) 10 negatives per cada positiva (b) 50 negatives per cada positiva

Figura 70: Many words + Clear images: Opcions Different model i All but one words

Referències

[1] J. Almazán, A. Gordo, A. Fornés, E. Valveny. "Segmentation-free Word Spotting with Exemplar SVMs", *Pattern Recognition, volume 47, issue 12, pages 3967-3978,* 2014.

[2] "Ordre ESS/86/2015, de 30 de gener, per la qual es despleguen les normes legals de cotització a la Seguretat Social, atur, protecció per cessament d'activitat, Fons de Garantia Salarial i formació professional, que conté la Llei 36/2014, de 26 de desembre, de pressupostos generals de l'Estat per a l'any 2015", *Boletín Oficial del Estado, Suplement en llengua catalana al núm. 27, Sec I, pàg. 25,* Gener 2015.

[3] K. Wang, S. Belongie. "Word Spotting in the Wild", *European Conference on Computer Vision (ECCV), Heraklion, Crete, Sept.,* 2010.

[4] T. Rath, R. Manmatha. "Word Spotting for Historical Documents", *International Journal on Document Analysis and Recognition (IJDAR), pp. 139 - 152,* 2005.

[5] K. Terasawa, Y. Tamaka. "Slit Style HOG Feature for Document Image Word Spotting", *10th International Conference on Document Analysis and Recognition,* July 2009.

[6] J. Almazán, A. Gordo, A. Fornés, E. Valveny. "Word Spotting and Recognition with Embedded Attributes", *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), volume 36, issue 12, pages 2552-2566,* 2014.

[7] D. Fernández, J. Lladós, A. Fornés. "Handwritten Word Spotting in Old Manuscript Images Using a Pseudo-Structural Descriptor Organized in a Hash Structure", *Pattern Recognition And Image Analysis, Lecture Notes in Computer Science, vol. 6669. pp.628-635.* Editorial: Springer-Berlag, Berlin. ISBN 978-3-642-21256-7, 2011.

[8] N. Aouadi, A. Kacem. "Word Spotting for Arabic Handwritten Historical Document Retrieval using Generalized Hough Transform", *The Third International Conferences on Pervasive Patterns and Applications*, 2011.

[9] S. Sarkar. "Word Spotting in Cursive Handwritten Documents using Modified Character Shape Codes", *Proceedings Of The Second International Conference On Advances In Computing And Information Technology,* July, 2012.

[10] A. Kovalchuk, L. Wolf, N. Dershowitz. "A Simple and Fast Word Spotting Method", *Proceedings of the Fourteenth International Conference on Frontiers in Handwriting Recognition (ICFHR), Crete, Greece, pp. 3-8,* September 2014.

[11] V. Frinken, A. Fischer, R. Manmatha, H. Bunke, "A Novel Word Spotting Method Based on Recurrent Neural Networks", *IEEE Pami. 34(2): 211-224,* 2012.

[12] Matlab: http://es.mathworks.com/products/matlab/

7 Resum

L'àmbit d'aquest projecte és la visió per computador i el seu objectiu principal és avaluar la possibilitat d'identificar el model (marca i tipus) al que correspon un comptador de gas cercant certes paraules clau en una imatge del comptador. Per a assolir aquest objectiu hem partit d'un mètode ja existent i hem realitzat una adaptació per a aquest fi específic.

Resumen

El ámbito de este proyecto es la visión por computador y su principal objetivo es evaluar la posibilidad de identificar el modelo (marca y clase) al que corresponde un contador de gas buscando ciertas palabras clave en una imagen del contador. Para alcanzar este objetivo se ha partido de un método ya existente y se ha realizado una adaptación a este fin concreto.

Summary

The field where this project was developed is "computer vision", and its objective is to evaluate the possibility of identifying the model (brand and class) to which corresponds a gas meter by searching keywords in a picture of the gas meter. To achieve this goal we started from an existing method and we performed an adaptation for this specific purpose.